2018

Enactment of Ongoing Formative Assessment: Challenges and Opportunities for Professional Development and Practice

Dante Cisterna

University of Nebraska-Lincoln

Amelia Wenk Gotwals

Michigan State University

Follow this and additional works at: <https://digitalcommons.unl.edu/natrespapers>

Part of the [Natural Resources and Conservation Commons](#), [Natural Resources Management and Policy Commons](#), and the [Other Environmental Sciences Commons](#)

Cisterna, Dante and Gotwals, Amelia Wenk, "Enactment of Ongoing Formative Assessment: Challenges and Opportunities for Professional Development and Practice" (2018). *Papers in Natural Resources*. 985.

<https://digitalcommons.unl.edu/natrespapers/985>

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Enactment of Ongoing Formative Assessment: Challenges and Opportunities for Professional Development and Practice

Dante Cisterna & Amelia Wenk Gotwals

To cite this article: Dante Cisterna & Amelia Wenk Gotwals (2018) Enactment of Ongoing Formative Assessment: Challenges and Opportunities for Professional Development and Practice, Journal of Science Teacher Education, 29:3, 200-222, DOI: [10.1080/1046560X.2018.1432227](https://doi.org/10.1080/1046560X.2018.1432227)

To link to this article: <https://doi.org/10.1080/1046560X.2018.1432227>

Published online: 26 Feb 2018.

Submit your article to this journal [↗](#)

Article views: 423

View Crossmark data [↗](#)

Enactment of Ongoing Formative Assessment: Challenges and Opportunities for Professional Development and Practice

Dante Cisterna ^a and Amelia Wenk Gotwals^b

^aSchool of Natural Resources, University of Nebraska–Lincoln, Lincoln, Nebraska, USA; ^bDepartment of Teacher Education, Michigan State University, East Lansing, Michigan, USA

ABSTRACT

Formative assessment practices have been shown to improve science instruction. However, examples of teachers' moment-to-moment informal formative assessment practices are still sparse. This multiple-case study explores the ongoing formative assessment practices of 4 in-service science teachers who participated in a statewide professional development program. We examined what teachers (can) do when enacting formative assessment with specific science content. Our findings suggest that these teachers were able to enact some components of formative assessment in a piecemeal fashion, but they tended to struggle with integrating formative assessment practices to enact seamless science instruction that was both rigorous and responsive to their students' ideas. This was especially true as they attempted to explore and use students' scientific ideas to push for deeper understanding. The analysis of formative assessment critical moments, however, suggests that these teachers could effectively enact aspects of formative assessment in ways that may have supported student learning of science and may have acted as stepping-stones to improve teachers' formative assessment practices. We discuss how teachers' current practices currently align (or do not align) with the Next Generation Science Standards and the implications of this nonalignment for the design of professional development in formative assessment in science.

KEYWORDS

classroom observation;
classroom practice;
formative assessment;
professional development

Formative assessment (FA) has been promoted as a central classroom practice that impacts student learning. Several studies posit the importance of teachers and students collecting ongoing evidence of student understanding and using this evidence to improve educational outcomes (e.g., Black & Wiliam, 1998; Shepard, 2000). FA is well researched as a domain-general practice (e.g., Heritage, 2007), and several aspects of FA practices are applicable across domains. For example, regardless of discipline, expert teachers provide opportunities for students to share their thinking and use that evidence to guide instruction. Coffey, Hammer, Levin, and Grant (2011), however, questioned the benefits of using these discipline-neutral strategies. From their perspective, FA practices are highly embedded in disciplinary structures that are rooted in cultural and discursive practices (Hodgen & Marshall, 2005; Pryor & Crossouard, 2010). For instance, attending to students' ideas requires different expertise depending

CONTACT Dante Cisterna dcisterna@unl.edu School of Natural Resources, University of Nebraska–Lincoln, 317 Hardin Hall, Lincoln, NE 68583, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uste.

© 2018 Association for Science Teacher Education

on the content being taught as well as beliefs about teaching and learning in a particular discipline. Moreover, the ways in which feedback is provided to students are affected by the nature of the discipline and the contexts of teaching (Havnes, Smith, Dysthe, & Ludvigsen, 2012), and teachers' instructional decisions are often framed by their understanding of the discipline (Bennett, 2011). Despite this discipline specificity to FA, there is scarce research on FA within specific disciplines (Coffey et al., 2011).

In particular, more research is needed to illustrate what FA looks like in science classrooms. The implementation of the Next Generation Science Standards (NGSS; NGSS Lead States, 2013) poses additional challenges for science teachers' assessment practices. Teachers need to plan and enact a variety of assessment tasks that are embedded in instruction and assess the new three-dimensional framework for science teaching and learning that is described in the NGSS (National Research Council, 2014).

Given that enacting FA practices effectively tends to be challenging for teachers (e.g., Heritage, Kim, Vendlinski, & Herman, 2009), professional development (PD) may be a strategy for supporting teachers' learning around FA. A recent report from the National Academies of Science, Engineering, and Medicine (2015) described recent research findings on PD in science. This report emphasized the importance of identifying core instructional practices that teachers need to develop to meet the new requirements stated in the NGSS, specifically for teachers supporting their students in their progress toward the NGSS performance expectations and guiding students to more complex levels of science understanding. Among these practices, the report stated that "given the central role of ongoing assessments in informing instruction, teachers need to master a range of formative and summative assessment strategies" (p. 103). Science-related PD also needs to support teachers in gathering and interpreting information from instructional tasks that provide assessment evidence, an aspect that is directly related to ongoing FA practices.

Although the use of FA in science classrooms has been linked to increased student learning and engagement (e.g., Furtak et al., 2016; Herman, Osmundson, Dai, Ringstaff, & Timms, 2015; Ruiz-Primo & Furtak, 2006), some studies have shown that the implementation of FA in science classrooms is neither simple nor straightforward and does not always translate into measurable student learning outcomes (e.g., Buck & Trauth-Nare, 2009; Gómez & Jakobsson, 2014). Moreover, few research studies describe how teachers and students interact around specific science content (Coffey et al., 2011). Thus, in the context of the new challenges posed by the implementation of the NGSS, the field needs a better understanding of teachers' current FA practices and the nuances of FA within specific science content. This starting point may be helpful (a) to identify what teachers need and are able to do and (b) to design effective PD opportunities for teachers that address the ambitious demands for science FA. Accordingly, we designed a multiple-case study to explore four science teachers' enactment of ongoing FA practices that have the potential of supporting student learning. The context of this study was a statewide PD program focused on FA knowledge and classroom enactment. Our research question for this study was as follows: What is the nature of a group of teachers' engagement of ongoing FA practices in the context of a PD program?

Framing FA

FA has the purpose of providing evidence that can support teachers' practices and influence student learning (Bell & Cowie, 2001; McMillan, 2013). Although the definition of FA varies across researchers, there is some shared understanding that FA is embedded in instruction and involves more than the mere use of a particular assessment tool or strategy (e.g., Bennett, 2011; Council of Chief State Schools Officers, 2008; Furtak et al., 2015). Likewise, FA can be framed around three guiding questions related to setting goals, gathering evidence from students, and making adjustments in instruction to guide and promote student learning (Hattie & Timperley, 2007; Sadler, 1989; Wiliam & Thompson, 2007). These guiding questions are (a) "Where are we going?" which includes developing and sharing instructional goals or learning targets; (b) "Where are we now?" which implies creating opportunities to elicit evidence of students' thinking as well as noticing and recognizing students' ideas (e.g., Hiebert, Morris, Berk, & Jansen, 2007; Van Es & Sherin, 2002); and (c) "How do we close the gap?" which refers to providing feedback and adjusting learning and instruction to reach learning targets. These three guiding questions involve a continuum of classroom practice that is organized in different levels of length, formality, and planning (Shavelson et al., 2008). These questions can be framed as FA loops with different levels of temporal and structural organization. In this article, we focus more on the ongoing side of the FA continuum—mainly informal, moment-to-moment, and on-the-fly interactions between teachers and students (Bell & Cowie, 2001; Ruiz-Primo, 2011; Ruiz-Primo & Furtak, 2006; Shavelson et al., 2008).

FA and science instruction

Framing quality teaching should include the discipline-specific aspects of pedagogy rather than solely domain-general pedagogical techniques (Knight & Duschl, 2015). Coffey et al. (2011), however, posited that research on FA has focused more on teachers' use of relatively domain-general assessment strategies than an examination of students' thinking in science. They argued that many studies on FA tend to conceive science content as information or vocabulary but have overlooked the examination of students' scientific ideas as well as the organization of the science discipline. Moving forward, we adopt the argument that FA practices are rooted in disciplinary structures framed by cultural and discursive practices (Hodgen & Marshall, 2005; Pryor & Crossouard, 2010).

Given the complexities inherent in working with students' ideas in the moment, the enactment of many of these FA practices is challenging for teachers (Furtak et al., 2016). Many teachers are able to use questioning strategies to elicit student ideas, but they tend to struggle with responding and using those ideas as productive stepping-stones to move students' learning forward. Ateh (2015) examined the classroom practices of two experienced science teachers who reported using strategies to notice, elicit, and respond to students' ideas. Although both teachers used questioning strategies to elicit student responses to examine student thinking, the actual focus of their practices was on the verification of factual knowledge. The adjustments in the lesson based on their students' ideas were minimal—mainly to help students say the right response. Similarly, a case study described by Izci and Siegel (2015) showed a teacher who held understanding of reform-based assessment principles but who ultimately enacted traditional assessment practices in

the classroom. Furtak (2012) found that some teachers struggled to use students' current ideas to support ongoing learning. Gómez and Jakobsson (2014) explored classroom assessment practices in secondary biology and chemistry in Sweden. They described lessons in which the predominant assessment pattern was initiation–response–evaluation (Cadzen, 2001). This pattern was so anchored in teachers' practice that it persisted even when students posed questions that demanded higher engagement from teachers.

In sum, these studies suggest that although teachers may elicit student ideas for FA, they may struggle to use these ideas to support deeper science understanding. Our study begins to examine the characteristics of FA enactment and the implications for teaching science and student learning in science.

Materials and methods

This study used mixed-methods strategies within a multiple-case study approach (Yin, 2013) in which we consider each participant teacher as one case. The study was conducted in the context of a statewide PD program targeted at promoting teachers' learning about FA theory and practice in a Midwestern U.S. state. Teachers participated in a common kickoff workshop in which they learned the components of the FA process (Council of Chief State Schools Officers, 2008), and after this workshop the PD activities occurred in local learning teams. In these teams, teachers learned about and discussed the nuances of FA and supported one another in implementing these practices in their classrooms. See Kintz et al. (2015) and Cisterna et al. (2016) for more information about the PD model and features.

Participants and data sources

We worked with four science teachers from two different focal learning teams that were selected purposely for more intensive study by our larger research program (Kintz et al., 2015). We were able to observe the PD meetings of these two focal learning teams for at least 1 year, and we had evidence of the teachers' engagement in the program. Based on this, we asked two science teachers from each learning team to volunteer to allow us to conduct classroom observations and videotaping. These teachers showed a willingness to share their lessons with us. Nevertheless, we do not aim to make connections between these teachers' practices and the characteristics of the PD program. Likewise, we cannot make claims about the effectiveness of the PD model for teachers' FA practices or generalize the findings of the study to all participant teachers in the program. We also note that the four teachers' lessons were videotaped between 2012 and 2014, which means that they planned their science instruction according to the state standards and not the NGSS (the state adopted the NGSS thereafter).

The four female teachers each had at least 5 years of classroom experience. The pairs of teachers on the same learning team taught the same grade level and content. Similar to Gómez and Jakobsson (2014), we conducted consecutive lesson observations to gather evidence of teachers' practices and abilities to connect lessons, actions, students' understandings, and events (Danielson & McGreal, 2000). Each teacher was videotaped for at least one round of 3 consecutive days of teaching. Although there is some debate about how many observations are necessary to gain stable measures of teacher practice, research

Table 1. Characteristics of participants and lessons collected in the study.

Teacher	Learning team	Grade level	Content area	Lessons observed ^a	Videotaping schedule
Karen	Blue	10th HS	Chemistry	13	2012: Two rounds of consecutive lessons First round: Three lessons Second round: Three lessons 2013: Two rounds of consecutive lessons First round: Four lessons Second round: Three lessons 2012: Two rounds of consecutive lessons
Heather	Blue	10th HS	Chemistry	10	 First round: One lesson Second round: Three lessons 2013: Two rounds of consecutive lessons First round: Four lessons Second round: Two lessons 2012: One lesson
Julie	Green	6th MS	General science	4	2014: Three consecutive lessons in one round 2014: Two rounds of three consecutive lessons each
Michelle	Green	6th MS	General science	6	

Note. Teacher and team names are pseudonyms. HS = high school; MS = middle school.

^aFor the four teachers, a single lesson lasted 55–60 min.

suggests that observing teachers at least three times allows for a valid picture of their practice (Gargani & Strong, 2014; Praetorius, Pauli, Reusser, Rakocky, & Klieme, 2014). In total, we collected 33 1-hr lessons (see details in Table 1). Before videotaping, teachers collected consent and assent forms from students and their parents. In a few cases, students or their parents did not give consent to participate in the videotaping, and these students sat in an area of the classroom where they could not be captured by the video camera. Members of the research team, including both of us, videotaped the lessons. Our role was as nonparticipant observers.

Data analysis

Data reduction

We used a research-based FA practice progression protocol to code lessons (Gotwals et al., 2015). The progression was framed by current research on FA (Gotwals et al., 2017) and structured around three main dimensions that connect the guiding questions of FA described previously: (a) Where are we going? (b) Where are we now? (c) How do we close the gap? We defined indicators for specific teacher practices in each dimension and developed a four-level novice-to-expert progression to capture the nuances in teachers’ FA practices. The FA dimensions and their specific indicators for the protocol are in Table A1 in Appendix A. In order to illustrate the specific levels of FA practice, Table B1 in Appendix B presents three examples of indicators (one example for each dimension) that are included in the observation protocol.

Coders trained in the protocol and with expertise in science segmented each video by classroom activity. Within each segment, they registered the level of FA practices by indicator (Levels 1–4, if evidence was present). Coders also took notes of key FA moments that they noticed or went beyond the descriptions in the observation protocol. Using a

social moderation process (e.g., Frederiksen, Sipusic, Sherin, & Wolfe, 1998), two researchers coded each classroom video and, if there were disagreements, worked to an agreement in the coding. Some of these disagreements were also discussed with the research team in order to receive feedback and decide on coding rules.

Based on the video coding and the coders' notes, we selected FA episodes or critical moments (Myhill & Warren, 2005; Wiliam & Leahy, 2007) in the videos that represented episodes of FA practices that illustrated rich information in terms of varied FA practices. For example, we focused on moments that included multiple codes simultaneously or reflected richness in FA evidence. We also selected critical moments that were unique in problematizing FA in science classrooms and that had the potential to illustrate the interplay among science content, instructional practice, and classroom contexts. For instance, we selected segments in which teachers made an instructional decision, which might have included a public announcement to the class about something that the teacher noticed about student learning or a feedback episode that was connected with the lesson learning target. We note that these critical moments do not necessarily reflect teachers' expertise in FA, describe teachers' overall levels of performance, or relate to particular frequencies of practice.

Within-case and cross-case analysis

We first examined our codes within each teacher to examine patterns in FA practices. We focused on the frequency of observations of specific FA practices and computed weighted averages and descriptive statistics. We also created graphs by teacher and practice in order to create preliminary teacher profiles. The information from each teacher profile was complemented with a narrative description of the FA critical moments that we selected. These critical moments were transcribed and described in detail. We integrated evidence from the lesson codes, the video segments selected, and the coders' notes. Each teacher profile was iteratively refined to describe and illustrate teachers' FA enactment. We then compared the profiles of each teacher in order to understand similarities and differences in the process of FA implementation and to examine the nature of FA practices.

Results

In this section, we succinctly describe aspects of each teacher's profile by including evidence from the lessons observed in terms of activity typologies and FA practices. We also present one example of a critical moment for each teacher that attempts to describe how FA may be seen in science classrooms.

Karen's profile

Karen, a 10th-grade chemistry teacher, exhibited an interactional style with her students characterized by a continuous monitoring of student work through observation and questioning. While students worked in small groups, Karen walked around to monitor their work, respond to their questions, and make suggestions. Karen also asked numerous questions of the students, although in the majority of the episodes they tended to be in the lowest levels of practice (see more details in [Appendix B](#)). [Figure 1](#) shows Karen's use and level of FA practices. For example, she frequently provided feedback to the students based

Figure 1. Karen's use of different formative assessment practices (13 lessons). Level 1 refers to the lowest level of formative assessment practice and Level 4 to the highest.

on what she noticed on students' work. The types of feedback varied: In some cases, her feedback just directed students to change their wrong answer to the correct one; in others, her feedback helped students develop conceptual understanding and identify steps for the completion of a particular task.

Karen's FA critical moment: Referring back to core concepts

Karen responded to one student (Amanda) who struggled to solve a problem about oxidation-reduction chemical reactions. Amanda struggled to find the oxidation states for manganese in the redox equation for potassium permanganate combined with hydrochloric acid. Amanda called for the teacher and told her teacher, "I don't know if the oxidation number is right ...". After that, the following conversation occurred between Karen and Amanda:

- Karen: Here you do! So you're saying this is a +5? [The state of oxidation for Mn in MnO_4^-]
- Amanda: Yes ...
- Karen: No, that's seven ...
- Amanda: I don't understand, how do I know this?
- Karen: They all together [the oxygen atoms] are going to be -8 that is going to be +7.
- Amanda: Oh, okay! I kind of get it ...

Based on Amanda's responses, Karen noticed that she did not understand why the oxidation state for manganese in MnO_4^- was +7. Then, a few minutes after this interaction, Karen drew two diagrams on the whiteboard that represented two atoms with 14 protons in their nuclei (she used the + symbol to represent the protons) in order to illustrate a change in the state of oxidation from +7 to +2. In the first drawing, she drew seven lines (-) that represented electrons on the model orbits. The teacher then called Amanda over to the board:

- Karen: Amanda, ready? This is an atom. I never ever, ever lose my protons. They are in the middle, so if I have 14 protons here, there are 14 protons here [The teacher points to each atom]. If I have a +7 charge, that means ... [The teacher counts from 1 to 7], I have 14 protons but I have 7 electrons, because my overall charge is +7. Does it make sense?
- Amanda: Yes.
- Karen: ... If I'm going to +2, that means ...?
- Amanda: Hmm, there's ..., to 5?
- Karen: So, how many more electrons do I have to reduce my charge?
- Amanda: Five.
- Karen: This one has more electrons now. It gained, so the charge is reduced. Does that help? [In the model she counts the electrons]

Initially Karen's guidance was focused on solving problems and completing mechanical redox procedures, but Amanda's responses did not necessarily mean that she understood what an oxidation or a reduction meant in terms of changing the number of electrons of an atom (i.e., "Oh, okay! I kind of get it ..."). Karen then used her understanding of Amanda's struggles to guide her instructional moves. Karen noticed this problem and she decided to adjust her instruction by using a representation of an atom to illustrate that the state of oxidation of an atom depends on the balance between its protons and electrons. Karen retaught in a different way, and the new explanation pointed to the concepts that were essential and core for Amanda in that moment: understanding(a) that the charge of an atom relates to the number of protons and electrons and (b) that a change in an atom's charge implies a change in the number of electrons.

Heather's profile

Heather was a 10th-grade chemistry teacher who taught next door to Karen. Usually Heather lectured at the beginning of a lesson, and then students worked in their instructional guides. At the beginning of a lesson Heather tended to paraphrase the lesson main idea (instead of using formal learning targets) and usually made connections with the unit instructional sequence. Heather's lectures tended to be slow paced; she carefully explained the nuances of the content and the steps for completing the chemistry problems. However, these practices, seen from the FA observation protocol, tended to be at the lowest levels. Almost all of the questions Heather asked students focused on connecting the lecture to prior chemistry knowledge and verifying understanding of the main ideas that were explained. Later, while students worked in their student guides, Heather was attentive in responding to students' questions and comments. In a few opportunities, she interrupted students' work to make a public announcement or emphasize a difficult idea or exercise. Figure 2 shows Heather's use and level of FA practices .

Heather's FA critical moment: Promoting opportunities for self-assessment

In her assessment episode, Heather reviewed and discussed the results of a lab activity completed the previous day. In that lab, which aimed to verify Hess's law, students created three related chemical reactions to measure the amount of heat using a Styrofoam calorimeter.

Figure 2. Heather's use of different formative assessment practices (10 lessons). Level 1 refers to the lowest level of formative assessment practice and Level 4 to the highest.

Heather started the review by having Kate (one student) share the results for q —a measure for heat change—that she experimentally got in the lab for each reaction. Heather explained to the class that she would use Kate's results to illustrate the calculations on the board and encouraged the rest of the class to use their own data to review the problems. Based on Kate's results, Heather completed the procedures to obtain, for each reaction, the ΔH value (i.e., the change in enthalpy). Then Heather used those ΔH values to compare Reactions 1 and 2 (added together) to Reaction 3. She asked the class the following:

Heather: [In this case] added 1 and 2 together, does it equal 3? Yes or no?

Students: No ...

Heather: No ..., if yes ... groups are going to be different. [The value for ΔH in reactions] 1 plus 2 is not going to equal 3. The difference between the two was 11 KJ [She calculated the difference with a calculator machine]. That's pretty good! She [Kate] only had 11 KJ of difference ...

Then Heather redirected the conversation to the entire class and emphasized the following:

So, there must have been some sort of errors in the lab, I want you go back and say ..., did I add exactly 2 grams each time? Did I measure the highest temperature each time? Did I wait? Did I stir with my stir rod and get the temperature, or did I just and get the temperature? Go back and look at your lab and how you did your procedures and the lab itself, and the different steps, and make sure you account for anything that may have gone wrong if the answer is no. If the answer is yes, it is done.

Heather's episode illustrates how her explanation supported, scaffolded, and set up students' self-assessment. For example, she used Kate's experimental results to explain the set of procedures to solve the calculations and make sure that the students had a common understanding of how to solve the problem. Instead of just giving the right answer, she encouraged the students to review their own results in light of the procedures

that were explained. Furthermore, Heather provided criteria for self-assessment—although not in the most explicit way—as instructions for how students could review their own results, considering the procedures to solve the problems or the possible sources of error in their measurements. However, from our perspective as external observers, we do not know how students actually self-assessed their work. We know that they were engaged in the activity, but we do not have evidence of how these students used their scientific ideas for self-assessment.

Julie's profile

Julie, a sixth-grade general science teacher, often organized her lessons by combining interactive lecture and whole-class discussions. Unlike Karen and Heather, Julie mentioned the learning targets at the beginning of each lesson and made sure that students understood how the lesson activities were connected to a specific component of a learning target. She also made clear to the students the ways in which the activities were embedded in an instructional sequence and how they would help students to reach the learning targets.

Julie's questioning practices were included in her lectures and served to guide the discussion. Cycles of questions and responses embedded in the lectures and whole-class discussions were frequent (see the observation protocol details in [Appendix B](#)). Likewise, her questioning strategies combined questions to verify knowledge acquisition and questions that provided opportunities for students to share the depth of their understandings. In some lessons, Julie displayed episodes in which questioning was organized to scaffold student ideas from a low to higher depth of knowledge (Kennedy, 2005). Similarly, Julie provided different types of feedback to her students, focusing on descriptive feedback and feedback loops to deepen student learning (see [Figure 3](#)).

Julie's FA critical moment: Promoting the use of assessment criteria

In Julie's episode, students worked on a study guide about geologic time scale in preparation for the summative test to be administered the following day. Once students completed the study guides, Julie handed them out to other students in order to facilitate the review in a peer assessment modality. When reviewing Question 6 of the study guide, Julie mentioned that the question was directly connected to one of the learning targets posted on the wall. Then Julie described possible ways of responding to that question and preparing for the unit test. She listened to students who were not sure about the quality of their peer's responses, as illustrated below. Two students, Tashari and Ashley, shared the written responses to the question:

Tashari: It is useful because the geologic times have eras, periods, and... they also show organisms in the earth ...

Ashley: They use it to show a geological form of history ...

Julie responded to Tashari and Ashley with a probing question—"Does it [the response] talk about earth history being so massive or so large?"—to guide students' peer assessment and help them revise their responses. A few minutes later, another student (Austin) shared

Figure 3. Julie's use of different formative assessment practices (four lessons). Level 1 refers to the lowest level of formative assessment practice and Level 4 to the highest.

a response that was very similar to the previous ones and wanted to know if his response “got it”:

Austin: The geological time scale is to show how earth history went ...

Julie: Yeah, it [the response] doesn't get it [the point of the question], but it is asking why we need it ...

Julie noticed that the majority of students' questions referred to the same issue, and then she brought the whole class together to clarify the statement:

The question doesn't say *what* does the time geological scale do, the question is *why* do we need it and ... your number one reason right now why you get these things wrong is because you don't read this question, you do not answer the question that is being asked. You give information that is absolutely correct but it is not answering the question that is being asked ...

This episode shows how Julie used this FA instance before the summative assessment. Thus, students had the opportunity to review responses of their classmates, and, when they were unsure, they used the assessment criteria given by Julie. When Julie noticed that many students were struggling with the same type of response, she made an instructional decision to clarify the difference between expected and actual student answers. Therefore, Julie decided to provide timely and descriptive feedback by pointing out how students were responding and how they should shift their thinking to better match the learning target and structure of the question. Other important points of this episode—and frequent components of Julie's practice—are how she referred back to the learning targets and her efforts to make the target comprehensible for the students.

This episode, however, poses some questions regarding how Julie resolved students' doubts. Although she outlined the assessment criteria to the students, Julie ultimately resolved whether or not the response was correct instead of allowing students to see where

their answers did not match the question. Furthermore, we do not know whether students actually used that evidence to refine their ideas.

Michelle's profile

Michelle, a sixth-grade teacher who taught down the hall from Julie, often used an instructional pattern type called “More is better” (Hanuscin et al., 2016) in which many activities were loosely related to a scientific topic and not organized in a coherent manner. Michelle's activities included the use of different resources and hands-on activities that partially addressed the content associated with the learning targets. For example, Michelle's typical lesson included lectures supported with the use of multimedia (video, websites, among others) and the use of representations combined with hands-on activities.

Regarding FA practices (see Figure 4), Michelle tended to state the learning targets at the beginning of the lesson. In the lessons observed, all of the questions Michelle asked, such as guessing games and initiate-respond-evaluate, were at the lowest levels of practice, which means that the goal was to verify what students knew (Cadzen, 2001). Similarly, the feedback that Michelle provided tended to be more evaluative and focused on getting the right response. Opportunities for self- and peer assessment in Michelle's lessons were rare.

Michelle's FA critical moment: (Not) recognizing the centrality of students' questions

Michelle's sixth-grade class was learning about the different organisms that showed up and lived along the geologic time scale. At the beginning of the lesson, Michelle reminded students of the specific learning targets that were posted on the wall that related to this lesson and previous ones. The learning targets were “I can explain what the geologic time scale is” and “I can explain why scientists use geologic time scales.” She started the lesson

Figure 4. Michelle's use of different formative assessment practices (six lessons). Level 1 refers to the lowest level of formative assessment practice and Level 4 to the highest.

by having students provide a definition in their own words of the geologic time scale based on what they had previously learned. Michelle mentioned the specific concepts that students needed to know for this lesson and the activities that students would complete in this and the subsequent lessons. Then Michelle projected a few slides on the board and showed a video with additional information on the changes that had occurred through the earth's history. Seven minutes later, the lesson continued with a discussion of the definition of a vertebrate and the order of appearance of the different classes of vertebrates in the geologic time scale. She had students read the information on board about the Precambrian and said that during this time, "There is some evidence of bacteria, algae, very very few invertebrates, but not many..." Then Mark and Lee, two students, asked the following:

- Mark: How do they find it? [evidence of living things]
 Michelle: Fossils ..., inference ...
 Mark: But they are bacteria ...!
 Michelle: On the rocks there may be small circles where they found ...
 Mark: But they are bacteria ...!
 Michelle: On microscope, microscope ...
 Mark: Bacteria can't ...
 Michelle: You can see it on the microscope
 Lee: Bacteria can fossilize, can't it?
 Michelle: It may get impressed ... I know ... Somehow they found it. I can't tell you exactly how, but somehow they found it ...
 Mark: So a bacteria is smaller than a grain of dirt or whatever? So ... you can't, not many, small features, the dirt can't cover ... It just ...
 Michelle: Why don't you look that up? They do ... they know that bacteria were found in the Precambrian ... Figure out how they know that bacteria were found in the Precambrian ... look that up and bring it in 'cause I don't want to give you the wrong answer.

Then Michelle redirected the conversation and continued lecturing, projecting a new slide on the board and telling the class, "So this is the Paleozoic, this is the early age of life, these are the periods ..."

This episode shows that, on the one hand, Michelle was aware at the beginning of the lesson of the importance of explicitly connecting different learning targets and activities for the students. On the other hand, the ways in which Michelle dealt with student questioning are troublesome, especially from the perspective of using students' ideas purposefully. The question that Mark posed was directly connected with the second lesson learning target and had the potential to be used as a scaffold for promoting conversations with a higher depth of knowledge. However, Michelle did not fully engage Mark's question and did not use the question as an opportunity to explore and build on students' ideas. Moreover, Mark's question was directly related to one crucial aspect for science learning, that is, how scientists collect and validate scientific evidence. Nevertheless, Michelle was not able to recognize the centrality of this question and gave responses that, although not incorrect ("It may get impressed"), were not adequate to respond to Mark's inquiry. Finally, Mark's question was not addressed.

Overall, this episode suggests that although that Michelle showed some desirable practices related to learning target use, she was unable to notice how the evidence from student questioning accounted for building pieces of the learning targets. Based on their questions, Mark and Lee demonstrated understanding that the evidence from fossilized bacteria is indirect, given that they are microscopic organisms, but the way that Michelle responded emphasizes factual information (“They know that bacteria were found in the Precambrian”). Furthermore, this episode suggests difficulties Michelle had in noticing the importance of students’ questioning as both a science practice as well as evidence of their current levels of understanding.

Discussion

In this study, we conceptualized FA as a process organized by three guiding questions that involve setting learning goals, gathering evidence of student understanding, and making instructional decisions based on student evidence. We described the classroom practices of four science teachers and focused on particular FA moments. Here we synthesize our findings from the four cases, discuss challenges the teachers faced, and suggest some implications for the enactment of FA and PD design.

Where are we going?

Setting learning targets is important for teachers and students because they provide direction and guidance during a lesson and orient students toward the focal concepts and practices that students need to master. In our observation protocol, the highest levels of FA practices for learning targets implied that teachers and students had clarity of the types of knowledge, skills, and abilities students would be expected to accomplish at the end of a lesson or instructional unit as well as the ways by which students could progress toward accomplishing the learning targets. Three out of the four teachers in this study scarcely referenced the lesson learning targets, and when they did, it tended to be at the beginning of the lesson and at the lowest levels of practice (e.g., “Remember the ‘I can’ on the board”). We observed that learning targets tended to be enacted in terms of informational pieces about the content that students would be learning in the particular lesson or the activity that students would be completing. In all cases, the learning targets were introduced by the teacher, and student participation was only included to verify that students understood the context of the learning targets or knew where to look to find them.

The NGSS (NGSS Lead States, 2013) will involve moving away from learning targets that focus mainly on content and moving toward three-dimensional learning goals that include science practices and crosscutting concepts. These learning targets should not give away what students will be investigating or learning; however, they should help to guide both teachers and students as they move through a unit (Penuel, Novak, McGill, Van Horne, & Reiser, 2017). Simply referencing these new types of learning targets at the beginning of a lesson will not be sufficient. Learning targets aligned with the NGSS will need to include (a) goals for students and success criteria that outline what this type of learning will look like and (b) examples that scaffold students’ self- and peer assessment practices (Moss & Brookhart, 2012).

Teachers may face several challenges as they shift to align their learning targets to the NGSS. One is crafting learning targets that involve more than just content. For example, although Michelle’s learning targets included explaining, when we look more carefully, we see that they really focused on content definitions, not on explaining phenomena, as

suggested by the NGSS. If she had crafted these learning targets more as scientific explanations (e.g., see McNeill, 2011) that prioritized using evidence and scientific reasoning, she might have been more able to recognize that the students' questions about what counted as fossil evidence were directly related to the learning target. In addition, if she had planned based on this learning target, she might have been better prepared to address the student question about how scientists used fossil evidence.

Another challenge related to learning targets comes when one is attempting to use them with students. For example, Karen rarely related her instruction to learning targets. Although she was able to provide feedback to her struggling student on a given problem, perhaps if she had had a learning target associated with a driving question (e.g., Krajcik & Czerniak, 2013) or a puzzling phenomenon, students might have known why they were learning what they were learning. In addition, she drew a model for her student to better understand what was happening. Perhaps if her learning targets had included a modeling component, students would have been able to visualize the process better.

Where are we now?

In our observation protocol, the highest levels of practice for this dimension were related to the collection of multiple types of evidence that allowed teachers to explore students' ideas and scaffold students' thinking in relation to the learning targets and to help students collect evidence of their own thinking. We saw that the four teachers in this study used strategies for eliciting evidence from their students relatively frequently. With the exception of Julie (at times), teachers' use of elicitation strategies and questioning tended to be at the lowest levels, which means that they mostly focused on the verification of factual knowledge and used questioning strategies that were merely looking for students' right responses. These challenges are not different from the findings presented in other studies (e.g., Ateh, 2015; Gómez & Jakobsson, 2014) and illustrate the difficulties teachers face when enacting strategies to elicit evidence of students' nuanced thinking.

The NGSS will involve gathering new types of evidence of student understanding. Just asking questions to determine whether students know content will not suffice. If teachers prioritize the science practices and crosscutting concepts rather than just the content (i.e., disciplinary core ideas), then they will likely be better able to elicit richer evidence of student understanding, which (with practice) may allow them to better work with students' ideas in a productive way (National Research Council, 2014). However, this new three-dimensional approach implies that teachers need to reconsider what (and how) they assess.

Teachers may face many challenges when transitioning to NGSS-aligned instruction. One challenge can be seen by looking at Michelle's critical moment, which illustrates a trend we observed in other teachers—they seemed to struggle to make sense of the evidence that they collected with relation to the learning targets for the lesson. Because Michelle was so focused on providing her students with a list of facts, she did not realize that her students were asking relevant and interesting questions about how scientists find evidence. Similarly, although Heather did provide students with opportunities to self-assess around their calculations, she never pressed students to see whether they knew what these calculations meant. However, we did see Karen take some time to process her students' struggles with a chemistry problem and then return to work with her student to better support her understanding of atomic structure, which was underlying her

struggles. Thus, considering the types of evidence of student understanding that teachers will need to gather as they align their instruction with the NGSS will be very important.

How do we close the gap?

Research tells us that even when teachers are able to interpret the nuances in student thinking, they may struggle to identify specific learning needs (e.g., Ruiz-Primo & Furtak, 2007) and ways to address those learning needs (e.g., Heritage et al., 2009). The highest levels of the observation protocol for this dimension emphasized the importance of providing descriptive and evaluative feedback to students (from teachers or peers) so that teachers and students would be able to make decisions and adjustments in their learning process with a clear rationale. Similar to other studies (e.g., Kloser, Borko, Martinez, Stecher, & Luskin, 2017), we found more variability across teachers in the practices associated with this dimension of FA compared to the previous ones. Karen and Julie tended to show higher levels of providing feedback to students—considering descriptive and evaluative feedback in some opportunities. In contrast, the feedback practices of Heather and Michelle tended to focus on merely providing the right answer to students.

Aligning instruction to the NGSS will require feedback that focuses on the nuances of students' abilities to engage in sense making using science practices, crosscutting concepts, and disciplinary core ideas. Teachers will need to support students as they make sense of phenomena. Rubrics and other supports (e.g., McNeill, 2011) can help teachers and students to make progress in this type of learning.

The challenges in providing that richer feedback can be evidenced when one contrasts Karen's and Michelle's critical moments. Karen was able to recognize her student's struggle, which was connected to a main scientific idea related to atomic structure. Karen was able to select, *in the moment*, a representation that was adequate for providing feedback and closing her student's gap. By contrast, Michelle was not able to recognize the nature of her students' questioning, which could have promoted rich science discussion around what counts as evidence of geological time periods and was directly related to the lesson learning targets.

Summary of teacher practices

FA is not a single teaching practice; in fact, it can be understood as a complex interplay of planning, developing and communicating clear learning goals, eliciting students' ideas, and using these ideas to provide feedback and make instructional and learning decisions (e.g., Black & Wiliam, 2009; Popham, 2008). The use of FA involves teaching practices that are interconnected and can be observed multiple ways. Although the four teachers presented different levels of FA practice, across dimensions, their practices tended to be at the two lowest levels of FA. This is not surprising. Ruiz-Primo and Furtak (2006) found that even though many teachers enacted questioning strategies to elicit students' ideas as part of their FA practices, they did not tend to complete assessment cycles to promote student learning. Although we did observe isolated episodes of FA practices that aimed to support students' understanding of scientific ideas, these practices were not organized in a seamless manner such that evidence from students was continuously elicited and addressed in relation to learning targets.

Implications for PD in FA

Despite participating in PD for at least 1 year in their school learning teams, Karen, Heather, Julie, and Michelle tended to struggle with the enactment of ongoing FA practices in different ways. This suggests that experts will need to consider specific design principles as they develop PD to support teachers in enacting a more NGSS-aligned view of FA. Therefore, professional developers who aim to support science teachers in developing FA knowledge and skills need to consider two important components. First, developers need to make sure that FA is organized as a connected and coherent process instead of piecemeal practices that are enacted in isolation. In our previous research (Cisterna et al., 2016) we saw that many teachers tend to focus their initial learning about FA on learning target use but then struggle to connect other aspects of FA in a seamless manner. Therefore, it is important that PD includes activities and representations that help teachers understand how and when the elicitation of students' ideas is needed and how to make instructional adjustments based on evidence of student learning and learning targets. This may include an examination of classroom video so that teachers can pause and decompose ongoing instruction (e.g., Koellner & Jacobs, 2015; Van Es & Sherin, 2010). The use of these strategies would help teachers such as Michelle and Heather to recognize how they are recognizing and using students' ideas. Second, PD needs to emphasize FA in regard to (core) scientific ideas. We observed that the emphases of teachers' instruction tended to focus on supporting students' learning of scientific information and procedures. Thus, the use of learning targets and strategies to elicit and respond to students needs to be aligned to three-dimensional science framework, not simply a decomposition of disciplinary core ideas presented as facts. Similarly, feedback provided to students needs to be guided to support students in understanding and being able to explain scientific phenomena instead of just providing the right (factual) response.

Given the challenges presented by the NGSS, PD in FA needs to address the complexities of the three-dimensional framework. Many teachers enact FA as the elicitation of informational pieces with a low depth of knowledge, and a shift is needed. Professional developers may use different tools to support teachers in understanding what needs to be assessed (and how). Learning progressions may be an important tool to help teachers understand how students' ideas may develop (e.g., Furtak, 2012). Teachers often consider students' ideas as either right or wrong (Otero, 2006), which leads them to want to just correct the wrong ideas. Learning progressions can help teachers conceptualize students' ideas on a continuum with certain ideas being productive stepping-stones to more sophisticated understandings (Alonzo, 2011). In addition, when learning progressions are embedded in a system of supports such as curriculum or frameworks for teaching (e.g., Furtak, Thompson, Braaten, & Windschitl, 2012), teachers have the support for working with students' ideas in more productive ways. Likewise, the use of lesson plan strategies that unpack the concepts and practices enacted in a lesson learning target may be useful for teachers, for example, by using conceptual storylines (Ramsey, 1993) in lesson planning.

Finally, a sound PD focused on FA needs to be created along with a deep understanding of teachers' own instructional practices. For learners, ongoing FA is difficult to differentiate from other classroom practices (Wylie & Lyon, 2012). The collection of samples of teachers enacting high-quality FA may be helpful to promote teachers' understanding of analysis of this practice. Video evidence, teachers' reflections, analysis of instructional lesson plans and artifacts, and the selection of critical FA episodes might

be organized in cases to help teachers understand FA in depth. With teachers transitioning their instruction to align better with the NGSS, it is important that FA practices are considered a crucial component of the changes being made, and PD (not just adopting new curricula) should be a key factor in supporting teachers along the way.

Acknowledgments

An earlier version of this article was presented at the April 2016 annual meeting of the National Association for Research in Science Teaching.

Disclosure Statement

The authors report no potential conflict of interest.

Funding

The research reported here was supported in part by the Michigan Department of Education and Michigan State University. The opinions expressed are our own and do not represent the views of the Michigan Department of Education or Michigan State University. The funding sources had no involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

ORCID

Dante Cisterna <http://orcid.org/0000-0002-3304-0813>

References

- Alonzo, A. C. (2011). Learning progressions that support formative assessment practices. *Measurement: Interdisciplinary Research and Perspectives*, 9(2–3), 124–129.
- Ateh, C. M. (2015). Science teachers' elicitation practices: Insights for formative assessment. *Educational Assessment*, 20(2), 112–131. doi:10.1080/10627197.2015.1028619
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85, 536–553. doi:10.1002/(ISSN)1098-237X
- Bennett, R. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy, & Practice*, 18(1), 5–25. doi:10.1080/0969594X.2010.513678
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, & Practice*, 5(1), 7–73. doi:10.1080/0969595980050102
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability*, 21, 5–31. doi:10.1007/s11092-008-9068-5
- 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. doi:10.1007/s10972-009-9142-y
- Cadzen, C. (2001). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.
- Cisterna, D., Gotwals, A. W., Kintz, T., Lane, J., & Roeber, E. (2016). Potentials and challenges of a situated professional development model. In T. Petty, A. Good, M. Putnam (Eds.), *Handbook of research on professional development for quality teaching and learning* (pp. 151–180). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0204-3.ch008

- Coffey, J., Hammer, D., Levin, D., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, 48(9), 1109–1136. doi:[10.1002/tea.20440](https://doi.org/10.1002/tea.20440)
- Council of Chief State Schools Officers. (2008). *Attributes of effective formative assessment*. Washington, DC: Author.
- Danielson, C., & McGreal, T. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: ASCD.
- del Carmen Gómez, M., & Jakobsson, A. (2014). Everyday classroom assessment practices in science classrooms in Sweden. *Cultural Studies of Science Education*, 9(4), 825–853. doi:[10.1007/s11422-014-9595-y](https://doi.org/10.1007/s11422-014-9595-y)
- 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. doi:[10.1207/s15326977ea0504_1](https://doi.org/10.1207/s15326977ea0504_1)
- Furtak, E. M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. *Journal of Research in Science Teaching*, 49(9), 1181–1210. doi:[10.1002/tea.v49.9](https://doi.org/10.1002/tea.v49.9)
- Furtak, E. M., Cartun, A., Chrzanowski, A., Circi, R., Grover, R., Heredia, S. C., . . . White, K. H. O. (2015, April). *Toward a participation metaphor for formative assessment*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., De Leon, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science*, 44(3), 267–291. doi:[10.1007/s11251-016-9371-3](https://doi.org/10.1007/s11251-016-9371-3)
- Furtak, E. M., Thompson, J., Braaten, M., & Windschitl, M. (2012). Learning progressions to support ambitious teaching practices. In A. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 461–472). Rotterdam, The Netherlands: Sense.
- Gargani, J., & Strong, M. (2014). Can we identify a successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education*, 65(5), 389–401. doi:[10.1177/0022487114542519](https://doi.org/10.1177/0022487114542519)
- Gómez, M. C., & Jakobsson, A. (2014). Everyday classroom assessment practices in science classrooms in Sweden. *Cultural Studies of Science Education*, 9(4), 825–853.
- Gotwals, A. W., Cisterna, D., Lane, J., Kintz, T., & Ezzo, D. (2017). Characterizing observable formative assessment practices: A literature-based framework. Manuscript submitted for publication.
- Gotwals, A. W., Philhower, J., Cisterna, D., & Bennett, S. (2015). Using video to examine formative assessment practices as a measure of expertise for mathematics and science teachers. [Special Issue] *International Journal of Science and Mathematics Education*, 13(2), 405–423.
- Hanuscin, D., Lipsitz, K., Cisterna, D., Arnone, K. A., van Garderen, D., & de Araujo, Z. (2016). Developing coherent conceptual storylines: Two elementary challenges. *Journal of Science Teacher Education*, 27(4), 393–414.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. doi:[10.3102/003465430298487](https://doi.org/10.3102/003465430298487)
- Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38(1), 21–27. doi:[10.1016/j.stueduc.2012.04.001](https://doi.org/10.1016/j.stueduc.2012.04.001)
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140–145. doi:[10.1177/003172170708900210](https://doi.org/10.1177/003172170708900210)
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice*, 28(3), 24–31. doi:[10.1111/j.1745-3992.2009.00151.x](https://doi.org/10.1111/j.1745-3992.2009.00151.x)
- Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., & Timms, M. (2015). Investigating the dynamics of formative assessment: Relationships between teacher knowledge, assessment practice and learning. *Assessment in Education: Principles, Policy, & Practice*, 22(3), 344–367. doi:[10.1080/096594X.2015.1006521](https://doi.org/10.1080/096594X.2015.1006521)
- Hiebert, J., Morris, A. K., Berk, D., & Jansen, A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education*, 58(1), 47–61. doi:[10.1177/0022487106295726](https://doi.org/10.1177/0022487106295726)

- Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: A comparison. *Curriculum Journal*, 16(2), 153–176. doi:[10.1080/09585170500135954](https://doi.org/10.1080/09585170500135954)
- Izci, K., & Siegel, M. A. (2015, April). *Making a new high school chemistry teacher's struggles with classroom assessment visible*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Chicago, IL.
- Kennedy, M. M. (2005). *Inside teaching*. Cambridge, MA: Harvard University Press.
- Kintz, T., Lane, J., Gotwals, A. W., & Cisterna, D. (2015). Professional development at the local level: Necessary and sufficient conditions for critical collegueship. *Teaching and Teacher Education*, 51, 121–136.
- Kloser, M., Borko, H., Martinez, J. F., Stecher, B., & Luskin, R. (2017). Evidence of middle school science assessment practice from classroom-based portfolios. *Science Education*, 101(2), 209–231. doi:[10.1002/sce.21256](https://doi.org/10.1002/sce.21256)
- Knight, S. L., & Duschl, R. A. (2015). Current trends in the framing of teacher qualities. In G. K. LeTendre & A. W. Wiseman (Eds.), *Promoting and sustaining a quality teacher workforce: International perspectives on education and society* (Vol. 27, pp. 443–465). Bingley, UK: Emerald Group.
- Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *Journal of Teacher Education*, 66(1), 51–67. doi:[10.1177/0022487114549599](https://doi.org/10.1177/0022487114549599)
- Krajcik, J. S., & Czerniak, C. (2013). *Teaching science in elementary and middle school classrooms: A project-based approach* (4th ed.). London, UK: Routledge.
- McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 2–17). Thousand Oaks, CA: Sage.
- McNeill, K. L. (2011). Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching*, 48(7), 793–823. doi:[10.1002/tea.v48.7](https://doi.org/10.1002/tea.v48.7)
- Moss, C. M., & Brookhart, S. M. (2012). *Learning targets: Helping students aim for understanding in today's lesson*. Alexandria, VA: ASCD.
- Myhill, D., & Warren, P. (2005). Scaffolds or straitjackets? Critical moments in classroom discourse. *Educational Review*, 57(1), 55–69. doi:[10.1080/0013191042000274187](https://doi.org/10.1080/0013191042000274187)
- National Academies of Sciences, Engineering, and Medicine. (2015). *Science teachers learning: Enhancing opportunities, creating supportive contexts*. Washington, DC: National Academies Press.
- National Research Council. (2014). *Developing assessments for the Next Generation Science Standards*. Washington, DC: National Academies Press.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.
- Otero, V. K. (2006). Moving beyond the “get it or don't” conception of formative assessment. *Journal of Teacher Education*, 57(3), 247–255. doi:[10.1177/0022487105285963](https://doi.org/10.1177/0022487105285963)
- Penuel, W. R., Novak, M., McGill, T., Van Horne, K., & Reiser, B. J. (2017). *How to define meaningful daily learning objectives for scientific investigations?* Retrieved from the STEM Teaching Tools website: <http://stemteachingtools.org/brief/46>
- Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: ASCD.
- Praetorius, A., Pauli, C., Reusser, K., Rakocky, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12. doi:[10.1016/j.learninstruc.2013.12.002](https://doi.org/10.1016/j.learninstruc.2013.12.002)
- Pryor, J. B., & Crossouard, B. (2010). Challenging formative assessment: Disciplinary spaces and identities. *Assessment & Evaluation in Higher Education*, 35(3), 265–276. doi:[10.1080/02602930903512891](https://doi.org/10.1080/02602930903512891)
- Ramsey, J. (1993). Developing conceptual storylines with the learning cycle. *Journal of Elementary Science Education*, 5(2), 1–20. doi:[10.1007/BF03173540](https://doi.org/10.1007/BF03173540)
- Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37(1), 15–24. doi:[10.1016/j.stueduc.2011.04.003](https://doi.org/10.1016/j.stueduc.2011.04.003)

- Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3), 237–263. doi:[10.1080/10627197.2006.9652991](https://doi.org/10.1080/10627197.2006.9652991)
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84. doi:[10.1002/\(ISSN\)1098-2736](https://doi.org/10.1002/(ISSN)1098-2736)
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. doi:[10.1007/BF00117714](https://doi.org/10.1007/BF00117714)
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., ... Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21(4), 295–314. doi:[10.1080/08957340802347647](https://doi.org/10.1080/08957340802347647)
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. doi:[10.3102/0013189X029007004](https://doi.org/10.3102/0013189X029007004)
- Van Es, E., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571–596.
- Van Es, E. A., & Sherin, M. G. (2010). The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education*, 13(2), 155–176. doi:[10.1007/s10857-009-9130-3](https://doi.org/10.1007/s10857-009-9130-3)
- Wiliam, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In H. McMillan (Ed.), *Formative assessment classroom: Theory into practice* (pp. 29–42). New York, NY: Teachers College Press.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wylie, E. C., & Lyon, C. J. (2012, April). *Quality instruction and quality formative assessment: The same or different?* Paper presented at the annual meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.
- Yin, R. K. (2013). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

Appendix A

Table A1. Dimensions of practice progressions.

Dimension	Practice	Definition
1. Using learning targets and goal setting: Where are we going?	Learning target use	How teachers introduce the learning target to students and connect it to instructional activities
2. Eliciting evidence of student understanding: Where are we now?	Type of question/information elicited	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 examine their own work and regulate their own learning
3. Closing the gap/responding to students: How do we close the gap?	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 and alterations

Appendix B

Table B1. Examples of three observation protocols for formative assessment dimensions.

Level 1	Level 2	Level 3	Level 4
Practice progression: Dimension 1. Learning target use			
The teacher makes passing mention of the learning target (e.g., “You all know the learning target and so we will keep moving on …”) but does not go over the learning target or relate it explicitly to the class activities.	The teacher reads the learning target aloud but does not explicitly relate the learning target to class activities (e.g., “You know that our learning target is ‘I can…’ So now we are moving on to …”).	The teacher explains and/or reminds students of the learning target and relates it to their activities but does not thoroughly describe the types of knowledge, skills, and abilities students will be expected to master (e.g., “You know our learning target is ‘I can …,’ and today we will do ____ activity to get at this learning target …”).	The teacher thoroughly explains and/or reminds of the learning target and what will be taught and learned and the types of knowledge, skills, and abilities students will be expected to master over the course of the lesson and/or unit (e.g., “You know our learning target is ‘I can …,’ and today we will do ____ activity to get at this learning target. Specifically, you should be able to do ____ by the end of this activity …”). The teacher asks students what the learning targets are for the day (i.e., student involvement in creating their own learning goals).
Practice progression: Dimension 2. Type of questions/information elicited			
The teacher asks only low-DoK questions that only have one correct answer (e.g., recalling and reproducing others’ ideas) or are skill/concept questions.	The teacher mainly asks lower DoK questions but has some higher DoK questions included (or asks students to explain their responses sometimes).	The teacher asks a mix of questions (perhaps building from lower DoK questions to higher DoK questions as a scaffolding tool). The teacher asks students why they arrived at an answer (e.g., a reflective toss) but may not push students to be metacognitive about their reasoning process.	The teacher asks a mix of questions (perhaps building from lower DoK questions to higher DoK questions as a scaffolding tool). The teacher asks the students to explain their ideas (asking “Why”) and pushes students to be metacognitive in how they arrived at their answer. The teacher may push students to advance their thinking (e.g., “Will your method always work?”).
The teacher does not ask students to explain their ideas but may ask students “Do you understand?” or “Does this make sense?”	The teacher may still ask “Does this make sense?” as evidence for student understanding.		
OR Tasks require only low-DoK responses (e.g., multiple-choice worksheets, presentations that are like reading from a textbook).	OR Tasks require mainly low DoK but some explanation (e.g., worksheets with short answers).	OR Tasks require a mix of low and high cognitive demand (e.g., worksheets that scaffold students toward higher level understanding, presentations that include student opinion or synthesis).	OR Tasks require a mix of low and high cognitive demand (e.g., worksheets that scaffold students toward higher level understanding, presentations that include student opinion or synthesis or next steps).

(Continued)

Table B1. (Continued).

Level 1	Level 2	Level 3	Level 4
Practice progression: Dimension 3. Feedback loops			
Feedback is mainly evaluative (e.g., correct/incorrect). The teacher may provide the right answer if the response is incorrect.	Feedback is mainly evaluative and focused on the outcome (the answer) of the task and not the process for getting the answer (problem-solving skills), but the teacher may provide ideas for moving forward (but the feedback may be too directive (essentially gives the answer) or too vague (it is not actionable). There may be not enough or too much feedback throughout the class.	Feedback is descriptive and focuses on both the outcome of the task (the answer) and how the student arrived at the answer (the process). The teacher may provide feedback on how to move forward, but the feedback may be either too directive (essentially gives the answer) or too vague (it is not actionable).	The teacher provides descriptive feedback about the <i>process</i> for completing the tasks (strategies used) and on the <i>task</i> itself. Feedback is specific enough for students to know what to do next but not so specific that the task is not done for them (giving students the answer).
AND/OR			
Feedback focuses on student(s) rather than on the learning or the tasks (e.g., "Great, you are so smart").		The teacher uses students' questions as a means of providing feedback.	The teacher provides feedback that moves learning forward—working with students to provide them the information they need to better understand problems and solutions (including promoting metacognition).
The teacher does not use students' questions as a means of providing feedback (either answers them or does not engage with them).	The teacher uses students' questions as a means of providing feedback.		The teacher uses students' questions as a means of providing rich feedback.

Note. DoK = depth of knowledge.