

GATHERING EVIDENCE OF STUDENT UNDERSTANDING¹

MARGARET HERITAGE

A central practice in formative assessment is teachers' generation and collection of information about how learning is developing while instruction is underway. Teachers engage in continually taking stock of learning by paying close, firsthand attention to specific aspects of students' developing understanding and skills as teaching and learning is taking place in real time (Erickson, 2007).

Consider the following illustration of one teacher's evidence collecting strategies:

In Ms. Alonzo's fifth-grade class, the students are working in a *writers' workshop* setting. They are writing a persuasive argument to encourage their readers to take more care of the natural environment. In previous workshops, they have learned about the nature and purpose of arguments and counterarguments and evidence to support the argument as well as how to organize their argument effectively.

While the students are involved in the independent writing part of the workshop, Ms. Alonzo sits with Edgar to discuss his writing progress. She has a three-ring binder open to a page with these headings at the top: Child's Name/Date, Research Compliment, Teaching Point, and What's Next for This Child? Further

down the page is a self-adhesive note that lists five students' names. These are the other children she wants to meet with during the session.

Ms. Alonzo's initial purpose with Edgar is to follow up from 2 days ago when she provided him with feedback based on the evidence she had elicited from her interaction with him; in that interaction, she determined that Edgar needed to provide stronger sources of evidence to support his argument. On this occasion, she wants to see how he has used her prior feedback.

Ms. Alonzo begins her interaction with Edgar: "You're working on evidence? What was your source? Where did you find it?"

Edgar responds, "In the book of the Environmental Protection Agency and on the Internet."

Ms. Alonzo continues, "And what do you think about what you found so far? Do you think that it supports your argument?"

Edgar is unsure and responds, "I guess . . ."

At this stage, Ms. Alonzo reminds Edgar that the purpose of the evidence is to support his argument and asks him to read his argument aloud. Having established that the focus of his

¹My thanks to Fritz Mosher for his insightful comments on an earlier draft of this chapter.

argument is to “stop dumping in the ocean because all the beautiful animals we see are going to start vanishing,” Ms. Alonzo then asks, “So what evidence did you find to support that? What evidence did you find that will help you to strengthen that argument?”

In the ensuing interaction, Ms. Alonzo helps Edgar to recognize which of the information he has located is from a reliable source and will be effective in supporting his argument. Satisfied that Edgar can move forward on his own to incorporate his evidence, she then asks him to go over the organization of his persuasive argument and to let her know where he will place the evidence. When Edgar does this, it is apparent to Ms. Alonzo that he has some confusion about the overall structure and that his writing needs to be reorganized. She goes over the organization with him and writes the organizational elements on a self-adhesive note saying, “So make sure that you put them in order, but when you do that, you know, focus on the organization because that’s gonna help it to flow so that once we read it to our audience or somebody else reads it, it makes sense.”

And she adds, “You might need some transitional sentences. Remember that we talked about those?”

Edgar nods at this point and she leaves him, saying, “So go ahead and work on those.”

Throughout this interaction, Ms. Alonzo has made notes in her three-ring binder. Under Research Compliment she writes that he had recognized the reliability of his source, in the section labeled Teaching Point she wrote that she had discussed how evidence supported his argument, and under the heading What’s Next for This Child? she wrote “organization and transitional sentences,” noting that Edgar was still unsure about how to organize his writing to effectively convey his argument to the reader. What do we see in this example?

Black and Wiliam (1998) referred to formative assessment as “encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (pp. 7–8). In a further elaboration, Torrance and Pryor (2001) considered the practice to be routine classroom assessment (CA) “integrated with pedagogy to maximize its formative potential in

promoting learning” (p. 616). In the example of Ms. Alonzo and Edgar, we see a teacher who is gathering evidence in the context of a student’s developing learning. The evidence is not gathered in the form of a “test” but rather in a carefully executed investigation through which Edgar’s teacher is able to determine the next steps to move his learning forward through pedagogical action, including feedback to him.

In short, we see evidence gathering in action: evidence gathering that is integrated with pedagogy and that provides information that the teacher uses as feedback to further learning.

The focus of this chapter is how teachers and their students can engage in the process of generating tractable information to be used as feedback in support of learning. First is a consideration of the purpose of evidence in formative assessment practice. Then follows a discussion of the range of sources of evidence. Next is a section devoted to assessment quality in relation to evidence sources in formative assessment, which is followed by a consideration of gathering evidence in the context of learning progressions. The chapter concludes with a discussion of the role of the student in gathering evidence.

PURPOSE OF EVIDENCE IN FORMATIVE ASSESSMENT PRACTICE

Assessment has two fundamental purposes: (1) to provide information on students’ current levels of achievement to the present time and (2) to inform the future steps that teachers need to take in classrooms to ensure that students make progress toward desired outcomes. In broad terms, assessments that provide information on students’ current levels of achievement represent a *past-to-present* perspective of learning in the sense that they indicate what has been learned to date. Many goals of assessment require this past-to-present view of learning, for example, accountability, placement, and certification. By contrast, the goals of assessment in support of prospective learning while it is developing require a *present-to-future* perspective, in which the concern is not solely with the actual level of performance but with anticipating future possibilities (Heritage, 2010). The contrast between these two perspectives is nicely captured by Frederick Erickson, when he observed that in addition to looking “upstream at what has been learned, assessment needs to look

downstream at what *can be learned*⁹ (F. Erickson, personal communication, October 28, 2009). In a related discussion, Torrance and Pryor (2001) distinguished between two kinds of assessment. The first centers on trying to find out *if* the learner knows, understands, or can do a predetermined thing and is characterized by closed or pseudo-open questioning and tasks. The second is intended to discover *what* the learner knows, understands, and can do. The latter is characterized by open questioning and tasks and oriented more to future development rather than measurement of past or current achievement.

Anticipating future possibilities for learning inherent in a present-to-future assessment perspective accords with Vygotsky's stance on instruction (Heritage, 2010; Heritage & Heritage, 2011; Torrance & Pryor, 2001). He stated that instruction "must be aimed not so much at the ripe as at the ripening functions" (Vygotsky, 1986, p. 188). To aim instruction at the "ripening functions," teachers need an indication about a student's *zone of nearest development* (also termed *the zone of proximal development [ZPD]*), described by Vygotsky as "those processes in the development of the same functions, which, as they are not mature today, still are already on their way, are already growing through, and already tomorrow will bear fruit" (Vygotsky, 1935, p. 120).

Kozulin, Gindis, Ageyev, and Miller (2003) suggested Vygotsky used the concept of the ZPD in three different contexts. First, in the developmental context, the ZPD refers to the child's emerging psychological functions. Second, in the context of assessment and classroom learning, the ZPD refers to the differences between the unaided and the assisted performance of the child. Third, the ZPD is conceived of as a metaphoric space in which the everyday concepts of the child meet *scientific* concepts introduced by teachers or by other mediators of learning. Elaborating further, Chaiklin (2003) observed that new psychological functions develop in conditions where there is a fundamental contradiction between the child's current capabilities—the developed psychological functions—the child's needs and desires, and the challenges and possibilities afforded by the learning context.

In this vein, one can make a distinction between two types of learning: (1) learning that occurs within an extant paradigm of knowledge (cf., Kuhn, 1962) through the deployment of

existing competencies in order to develop and extend them within a common paradigm and (2) the discontinuous moments in learning involving the reorganization of knowledge through the application of new competencies that permit a significantly greater depth of understanding in a particular domain. A learner will bump up against the limits of a paradigm or learning framework before being able to move on to a higher, more sophisticated level. For example, returning to the chapter's opening scenario, Edgar is engaged within a new paradigm of writing that takes him beyond his current competence. Now he has to explicitly recognize and address the positions of others (counterarguments), develop the skills to counter them, and organize his writing so that the evidence he has found effectively bolsters his arguments and counterarguments. Ms. Alonzo's interaction reveals that these are *maturing* functions, intermittently grasped and not yet consolidated.

In the course of teaching and learning, then, teachers have to recognize and act on two orders of learning. One is a progressive steplike form of learning in which the accumulation of observations and actions consolidates and deepens a given understanding. In the other, there is learning that occurs in the midst of change, during which the learner shows a fragmentary and inconsistent grip of new concepts: An accumulation of observations or data points yields confusing or incompatible inferences and conclusions, requiring movement to a different order of generalization or abstraction. From a teacher's perspective, the key element in acting upon both orders of learning is feedback. Feedback *from* evidence of student learning helps the teacher establish if the student is moving forward without the need for tailored intervention or to determine which cognitive structures are emerging so that a subsequent pedagogical move can be made to assist these cognitive structures to mature. Feedback *to* the student is the essential means through which appropriate pedagogical moves can be made.

In his seminal article, Sadler (1989) identified feedback to the student as the essential component in formative assessment. Adopting a cybernetic perspective, he began with Ramaprasad's (1983) definition of feedback as "information about the gap between the actual level and the reference level of a system parameter that is used to alter the gap in some way" (p. 4). Thus, Sadler

conceptualized formative assessment as a feedback loop designed to close the gap between the learner's current status and desired goals. For Sadler (1989), a critical component in closing the gap is the "judgments teachers make about student responses that can be used to shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial-and-error learning" (p. 120). This echoes the work of Pellegrino and Glaser (1982), who noted that of prime importance in formative assessment is teachers' careful probing and analysis of student learning, which can lead to sensitive adjustments to individual students' learning and to the pathways along which they will progress. Therefore, the overall purpose of evidence gathering in formative assessment is to enable teachers to respond to student learning in order to enhance that learning while the student is in the process of learning (Bell & Cowie, 2000).

SOURCES OF EVIDENCE

As noted earlier, the term *formative assessment* does not apply to a specific tool or measurement instrument. Over 80 years ago, Dewey (1928) pointed the way to evidence collection in support of learning with his comments that what is required is "a much more highly skilled kind of observation than is needed to note the results of mechanically applied tests" (p. 204). In more current literature, "highly skilled observation" can occur in the context of teacher–student interaction, student–student interaction, tasks, and observations of actions. Griffin (2007) argued that humans can only provide evidence of cognitive and affective learning through four observable actions: (1) what they say, (2) write, (3) make, or (4) do. These behaviors act as indicators of an underlying learning construct and are the ways in which learning can be inferred by the observer. Whatever the source of the evidence, the role of the teacher is to construct or devise ways to elicit responses from students that are revealing of their current learning status (Sadler, 1989).

Interactions

Interaction between teacher and students has been characterized as a principal source of evidence in formative assessment (Allal, 2010;

Black & Wiliam, 2009; Harlen, 2007; Heritage & Heritage, 2011; Jordan & Putz, 2004; Ruiz-Primo & Furtak, 2006, 2007; Torrance & Pryor, 1998). More specifically, Black and Wiliam (2005) noted that productive strategies in formative assessment include questions designed by the teacher to explore students' learning and generate teachable moments when they can intervene and further learning. Shavelson et al. (2008) suggested that to find the gap between what students know and what they need to know that teachers need to develop a set of central questions that get at the heart of what is to be learned in a specific lesson. Additionally, teachers have to know the right moment to ask these questions so that they can enable students to reveal what they understand and what evidence they can provide to back up their knowledge.

Chin (2007) illustrated the process of questioning and response in a science context: The teacher asks conceptual questions to elicit students' ideas and assists them to engage in productive thinking, invites and encourages multiple responses and questions, and offers ongoing comments to their responses. Finally, in a more in-depth analysis focused on acquiring evidence of children's cognitive processes in mathematics, Ginsburg (2009) advocated a three-pronged approach centered on observations, task performance, and the clinical interview. The latter is the most significant: In it, the teacher follows up task performance "with questions designed to elicit thinking, and in general follows the child's thought process to where it leads" (p. 113).

The essential point about teacher–student interaction as a source of evidence is that it enables teachers to have access to student thinking so that they can advance from the current state. In this context, the type of questioning reflected in the Initiation-Response-Evaluation (I-R-E) or *recitation* paradigm (Cazden, 1988; Mehan, 1979; Sinclair & Coulthard, 1975) does not lead to productive evidence. In this kind of model, the outcome is not to make student thinking visible but rather to let the students know if their responses are right or wrong, which tends to end the exchange and prevents further dialogue (Webb & Jones, 2009).

In a related discussion, Harlen (2007) pointed out that it is not just the framing of the question that matters but also the timing, particularly the time allowed for answering the question. Rowe (1974) found that teachers generally allow very

few seconds for students to answer questions, which as Harlen (2007) noted is very short even for questions that ask for recall—let alone for questions that demand students to provide explanations or express their ideas.

In a project designed to improve teachers' use of formative assessment, researchers worked with 24 science and mathematics teachers and brought together the ideas of improved questioning techniques and wait time—the length of the silence between when a teacher asks a question before speaking again if no student has responded (Black, Harrison, Lee, Marshall, & Wiliam, 2003). Teachers altered their questioning practices to include more wait time—for example, asking students to discuss their thinking in pairs before being randomly called on to respond (a no hands-up policy was instituted). Additionally, teachers did not refer to the answers as correct or incorrect but instead asked students to provide reasons for their answers, which gave them an opportunity to explore student thinking. The net result of this practice was that teachers asked fewer questions and spent more time on each (Black et al., 2003).

Ruiz-Primo (2011) expanded on questions as a source of evidence to extended interactional sequences referred to as assessment conversations. Developed from the notion of instructional dialogues as a pedagogical strategy linked to embedded assessment (Duschl & Gitomer, 1997), assessment conversations are conceived of as dialogues that embed assessment into an activity already occurring in the classroom, which enable teachers to gain insights into the nature of student thinking and act pedagogically on those insights (Ruiz-Primo & Furtak, 2004, 2006, 2007). Ruiz-Primo also noted that research suggests that in classrooms where teachers frequently engage in assessment conversations, students achieve at higher levels (Applebee, Langer, Nystrand, & Gamoran, 2003; Nystrand & Gamoran, 1991; Ruiz-Primo & Furtak, 2006, 2007).

Still in the context of student talk, Harlen (2007) contended that teachers can gain insights about student thinking when they set up a situation in which students converse with each other while the teacher “listens in” without participating in the discussion. Of course, it will be essential that the situation is well structured and promotes student exchanges of ideas so that their thinking is revealed.

Other Sources of Evidence

Beyond questions, interactions, discussions, and assessment conversations, there is a range of other sources of evidence. In mathematics, Lesh, Hoover, Hole, Kelly, & Post (2003) proposed model-eliciting activities that are useful for both assessment and instruction. Lesh and colleagues contrasted their model-based eliciting activities with traditional problem-solving activities found in textbooks. The latter requires students to produce an answer to a question that was formulated by someone else, whereas model-based activities require students to recognize the need to develop a model for interpreting the goals and potential solutions of an authentic, relevant problem. While students are working on the activities, they reveal how they are interpreting and mathematizing the problem.

The approach of Cognitively Guided Instruction in mathematics (Carpenter, Fennema, & Franke, 1996; Carpenter, Fennema, Peterson, & Carey, 1988) provides students with learning tasks created from a model of student thinking in arithmetic that permits teachers to interpret and respond to ongoing events in real time as they unfold during instruction. Also in mathematics, Heritage & Niemi (2006) proposed a framework for considering students' mathematical representations as evidence in formative assessment and, more specifically, Kouba and Franklin (1995) proposed that student representations, for instance, of division and multiplication situations and their explanations of the relationship among those representations, can function as evidence.

Harlen (2007) suggested that students' work, for example writing, drawings, and other artifacts resulting from well-designed tasks, can be a rich source of evidence about their ideas and skills. However, she cautioned that the tasks must be constructed so that students are able to express their ideas and that the teacher must have the knowledge and skills to notice the significant features of the work. Harlen (2007) also added that while student products can provide evidence, they rarely provide sufficient details about how certain skills have been used, noting that observation of how students are working can provide insights into their skills and attitudes. She suggested that observation can, for example, provide detail about how pupils make changes in their investigation of

variables in science and in mathematics about how they read scales, draw graphs, and use number grids.

Reporting on a program entitled *Every Child a Reader and Writer*, which was designed to improve writing instruction, Poppers (2011) echoed Harlen's (2007) caution about teacher knowledge and skills with respect to student work products as sources of evidence. Researchers in this program found that initially teachers did not have sufficient background knowledge to notice the significant features of the written work, but with professional development and analysis tools, such as rubrics, they were able to gain skills in interpretation and the use of the evidence to improve student writing.

With respect to other forms of gathering evidence, in the area of reading, Bailey and Heritage (2008) have offered a range of strategies to gain information about students' reading, including student read-alouds, strategic questions focused on the text, and prompted written responses about text as ways to elicit evidence about students' learning. Specifically in the context of science curricula, but equally relevant to other areas, Shavelson et al. (2008) referred to "embedded-in-the-curriculum" formative assessment. These are assessments placed in the ongoing curriculum by teachers or curriculum developers at key junctures in a unit and designed to create goal-directed teachable moments. Finally, Wiliam (2011) provided 50 *techniques* for eliciting evidence ranging from learning logs to ranking exemplars to students generating test items with correct answers.

Technology

Some promising avenues in the ways in which technology can support gathering evidence are emerging. One such example is ASSISTments, which makes use of digital teaching platforms to blend assessment and assistance in a tool that can be adapted and used in a variety of ways with different cognitive models and different content. ASSISTments is designed to augment, replicate, and promote effective assessment practices, including uncovering detailed diagnosis of misconceptions, providing immediate, specific feedback, and monitoring student practice (Heffernan, Militello, Heffernan, & Decoteau, in press). ASSISTments also has a feature that allows teachers to create their questions

(on the fly or prepared in advance), ask the students to respond to them, anonymously post the answers using a projector or interactive white board—one source of evidence—and then generate discussions about the questions that can be an additional source of evidence. Similarly, the online program Agile Assessment enables secondary school teachers to construct a range of cognitively demanding assessments to assess higher-order thinking in mathematics (Cook, Seeley, & Chaput, 2011). A Web-based tool, Strategic Reader, designed for use with struggling middle-school readers, provides a flexible assessment and instruction environment so that teachers can gather evidence of student performance during the instructional episodes and employ interventions as needed for individual students (Cohen, Hall, Vue, & Ganley, 2011).

EVIDENCE QUALITY

Validity and reliability are central to all assessment, yet to date, the application of these concepts to formative assessment is an underdeveloped area of study (Brookhart, 2003; Ploegh, Tillema, & Segers, 2009; Smith, 2003). Erickson (2007) described formative assessment as different from professional psychometrics because it involves the clinical judgment of teachers about students' actions in the classroom. Similarly, Dierick and Dochy (2001) argued that when the notions of validity and reliability are applied to a new assessment context, the predominant view of them needs to be widened and other more appropriate criteria developed. This section addresses some of the ways in which the view of validity and reliability can be considered in the context of formative assessment.

Validity

In line with the idea of an argument-based approach to validity (Kane, 1992), there are a number of arguments that can underpin validity in formative assessment (see also the extended discussion of validity by Sarah M. Bonner in Chapter 6). Let us first consider two fundamental principles concerned with validity: (1) the assessment measures that it is intended to measure and (2) it provides sound evidence for specific decision-making purposes (Herman, Heritage, & Goldschmidt, 2011). The evidence generated by

the variety of means discussed earlier is intended to provide information about the students' learning status in relation to the specific learning goals and to be used to inform decisions about next steps in teaching and learning. Ideas of content relevance, construct underrepresentation, and construct-irrelevant variance can be applied here in the sense that the evidence-gathering strategy should be aligned to the learning goal (the construct being addressed), it should be an appropriate representation of the construct, and should include the important dimensions of the construct. In other words, the strategy should not be so broad that it contains dimensions that are irrelevant to the construct nor too narrow that it fails to include the important dimensions of the construct. From the perspective of content relevance, the assessment strategy should be meaningful to students and situated in an authentic context (Frederiksen & Collins, 1989; Newman, 1990). Selecting the appropriate evidence-gathering strategy to meet the conditions that were previously discussed will require teachers to be very clear about what is to be learned and what evidence is needed to determine their students' current learning status.

In the same way that issues of fairness and bias are applicable in traditional psychometric approaches, they are also relevant to formative assessment. Because students' maturing functions do not develop in lockstep, formative assessment is inevitably personalized and teachers will need to employ strategies that tap into the individual knowledge that students manifest. Whatever strategies a teacher selects, they should account for the range of students present in the class so that all students have the opportunity to show where they are in their learning and have the prospect of moving forward from their current status. Similarly, formative assessment strategies should not include any elements that would prevent some students from showing where they are relative to goals, such as the use of language they cannot understand or images that could be offensive to certain subgroups of students.

Erickson (2007) introduced the notion that there can be threats to the "formativity" of formative assessment (p. 189). He argued that for assessment to be formative it must be both timely and produce information that can inform teaching practice during its ongoing course. For this reason, the immediate or proximate timing of evidence is a key component of formative assessment

validity. Moreover, and in addition, for formative assessment to be valid it must also yield tractable insights: insights into students' current learning status that are sufficiently tractable to be used in subsequent pedagogical moves (Heritage, 2010).

Messick (1989) viewed validity as an "integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of the inferences and actions based on test scores and other modes of assessment" (p. 13). The preceding discussion centered on the rationale for the inferences drawn from evidence gathered in formative assessment. In what follows, the focus is on the actions that are taken based on evidence. Cronbach (1988) developed a functional perspective on the validity argument in terms of whether the actions result in appropriate consequences for individuals and institutions. This issue was taken up in Messick's (1994) discussion of the interplay between evidence and social consequences in validity argument: Because action resulting from the use of formative assessment evidence is intended to result in benefits to student learning, consequences represent an important component of the validity argument. Indeed, Stobart (2006) referred to the arguments made by Wiliam and Black (1996) that formative assessment is validated primarily in terms of its consequences; Black and Wiliam (2005) suggested that even if assessments are formative in intention they may not be so in practice if they do not generate further learning. In a later paper, Stobart (2008) went further, suggesting that if learning does not improve as a result of formative assessment, then the issue of validity should be addressed in terms of an investigation of why the assessment and its use were not successful.

Frederiksen and Collins (1989) argued that assessment has systematic validity if it encourages behaviors on the part of teachers and students that promote the learning of valuable skills and knowledge and conclude that encouraging deep approaches to learning is one aspect that can be explored in considering consequences of assessment use. In a related conceptualization of validity that placed particular emphasis on the importance of social consequences, Crooks, Kane, and Cohen (1996) identified "pedagogical decisions" as an important factor in the validity of formative assessment,

noting that two students who had performed similarly on a task might benefit from differential pedagogical responses and encouragement based on their personal preferences and needs.

More recently, Crooks (2011), taking a broader conception of validity in formative assessment, identified a number of considerations that arise from the inclusion of student agents in the gathering and use of evidence (which is discussed later in this chapter). These considerations range from committed, motivated participants (teachers and students) on the one hand through the issue of trust so that students feel able to admit the difficulties and uncertainties they are having and includes the notion of teacher insights into the difficulties students are having.

Reliability

Shepard (2001) has argued that reliability is less critical for CA because errors in instructional decisions can be rectified quickly through gathering more evidence of learning (see also Chapter 7 of this volume for more on reliability in CA). Reliability in relation to instructional decisions has been conceived as “sufficiency of information” (Smith, 2003, p. 30). Teachers have to be confident that they have enough information about the student’s learning to make a reasonable judgment about the current status of that learning. In the classroom context, a crucial point with regard to reliability has to do with whether the teacher will get the same result again—whether it is characteristic of the student’s level of performance or is alternatively a chance outcome or “fluke” and, correlatively, whether the performance seems to change in response to trivial alterations in the classroom situation or only in response to big/important changes? From a teacher’s perspective, the sufficiency issue could be reframed as follows: “How do I know this isn’t a fluke or artifact of what is going on today—including whether the student may have gotten up on the wrong side of the bed. Do I know enough to rule that out, or do I have to try again another, brighter day?” (Fritz Mosher, personal communication, October 2011). This conception of reliability argues for multiple sources of evidence before a teacher makes an instructional decision. The wider the range of information and the more frequently the information is collected, the more accurately both specific and

generalized learning can be inferred (Griffin, Murray, Care, Thomas, & Perri, 2010).

In practical terms, this might mean that before making a judgment about learning a teacher has evidence from a student representation, from observations of the student constructing the representation, and from probing questions about the nature of the representation—why the student constructed it in a particular way and what it means. The more this kind of evidence can be gathered in the context of learning tasks, and so not to take time away from instruction, the more the number of learning events as assessment tasks can be increased to improve the reliability of the information gathered (Linn & Baker, 1996).

Anastasi (1990) contributed a further dimension to the consideration of reliability: the users’ responsibilities in interpreting evidence. She suggested that information needs to be considered in relation to a “backward and forward reference” to past and prospective aspects of the students’ experience and performance (p. 482). Thus, assessment evidence is not considered as a single instance but rather within the contextualized knowledge a teacher has of all the dimensions of student learning, both antecedent and anticipated.

A final consideration with regard to the judgments that teachers make based on the evidence is the accuracy of their judgment. In other words, how precise is their interpretation of the evidence they have gathered? To this end, Harlen and James (1997) suggested that teachers need to be given opportunities to develop both the expertise and the confidence to make and use reliable judgments about student learning that they can use as the basis for instructional decisions. It may be that, as Erickson (2007) argued, teachers’ clinical judgment is undervalued because of the dominance of summative testing as a tool of evaluation and the associated disprivileging of teachers’ discretionary authority relative to professional psychometrics. This is clearly a situation that will need to be addressed if teachers are to be supported in developing the skills needed for reliable judgments related to students’ ongoing learning.

If we return to the chapter’s opening scenario, we can see the validity and reliability arguments manifest in the interaction between Ms. Alonzo and her student Edgar. Ms. Alonzo’s assessment of Edgar takes place in the context of

an authentic task—his ongoing writing of a persuasive argument—and attends to deepening his learning about argumentation and extending his writing skills. She is assessing his current status within a developmental trajectory: She wishes to see how he has incorporated her feedback about use of evidence sources, and she has an eye on his future learning, a more developed piece of writing. Her evidence-gathering strategy is aligned to the learning goal, is an appropriate representation of the construct—well-organized persuasive writing that included arguments, counterarguments, and evidence—and permits Edgar to show how he has used evidence sources as well as to reveal some problems in the organization of his writing. The strategy has relevance to Edgar because it centers directly on the development of his persuasive argument, of which he already has ownership because it is on a topic of importance to him: saving the environment. Ms. Alonzo uses several evidence-gathering strategies: Edgar’s writing, his documented sources of evidence, and his responses to questions she asks. From this range of evidence, she draws inferences about his learning and determines his next step is to improve his organization and provides feedback accordingly. The value of her interventions is only enhanced by the fact that she has seen what he has done before and will see what he does next—and again and again.

USING THE EVIDENCE

Teachers’ use of evidence, particularly in terms of feedback and instructional correctives, is dealt with in greater length by Wiliam (see Chapter 12 of this volume) and the chapter by Black (see Chapter 10 of this volume), which discusses some of the issues involved in integrative summative and formative functions of assessment. Nonetheless, some essential points about gathering and using evidence are in order. The first point is that observations and evidence are in fact distinct. Essentially *observations* do not become *evidence* unless there is a structure, such as teachers’ understandings of learning progressions, into which the observations can be fitted in order to support interpretation. Second, even before the evidence is generated, teachers will need to have the knowledge and skills to formulate or select evidence-gathering tasks that reveal the nature of student understanding or

skills. Third, evidence gathering is a planned process, and assessment tasks “must have a place in the ‘rhythm’ of the instruction, built-in as part of the constant interaction that is essential to ensure that the teacher and the learner are mutually and closely involved to a common purpose” (Black, Wilson, & Yao 2011, p. 98). This means teachers should determine in advance at what points in the lesson they will need evidence to maintain the forward momentum of learning. Of course, this does not preclude actionable assessment opportunities arising spontaneously in the lesson but rather that evidence gathering should not be left to chance.

In contrast to standardized assessment, formative assessment practice rests mainly with teachers. In addition to determining how and when to gather evidence, they also determine whom to assess. For example, in Ms. Alonzo’s lesson, she had decided on the specific students she wanted to assess in individual one-on-one interactions. In other instances, teachers may use an evidence-gathering strategy that engages the whole class or groups simultaneously.

Once the evidence is gathered, teachers will have to interpret the student response against their knowledge of what a fully formed understanding or skill would look like in the context of the task. Finally, teachers will need to translate that interpretation into an appropriate pedagogical move to take the student’s understanding to a more advanced state. Sometimes pedagogical decisions need to be made on the fly, so the interpretation of evidence and the subsequent pedagogical move may need to be made in situ rather than at a later point of reflection. These kinds of interventions require flexibility and the nimble use of evidence and insight in real time. Without the attempt to support or influence new learning from the evidence, the label *formative assessment* cannot be applied to the process of evidence gathering. Instead, the term *dangling data* (Sadler 1989, p. 121) is more apt.

GATHERING EVIDENCE IN THE CONTEXT OF LEARNING PROGRESSIONS

The evolution of the standards movement in the United States has led to the development of academic standards that define what students should know and be able to do in the core academic subjects at each grade level. These standards are

often complemented by performance standards that describe what level of performance is needed for students to be classified as advanced, proficient, basic, below basic, or by some other performance level (Phelan, 2003). While these standards may provide better descriptions of what students needed to learn than teachers had before such standards existed—and can serve as guideposts for assessment for summative purposes (e.g., annual tests that provide information about student performance in relation to them)—they fall short of what is needed to engage successfully in the practice of formative assessment. Instead, Black and Wiliam (1998) suggested the following:

[We need to] develop methods to interpret and respond to the results in a formative way. One requirement for such an approach is a sound model of students' progression in the learning of the subject matter, so that the criteria that guide the formative strategy can be matched to students' trajectories of learning. (p. 37)

To gather evidence that can be used to keep learning moving forward, teachers need to understand the pathways leading to increasing expertise in a domain. Learning progressions can provide such a pathway by specifying the steps that students are likely to traverse with adequate instruction along the way to achieving landmark goals in the development of expertise. In the section that follows, we will consider current views of what learning progressions are, how they can be developed and validated, and how they can be used to support evidence gathering (see also Chapters 2 and 6 of this volume).

Defining Learning Progressions

Alonzo and Steedle (2008) noted that the *learning progression* label has operationalized in a variety of ways. These include the following:

- “Learning progressions are tied to big ideas, the central concepts and principles of a discipline. At the core of learning progressions is enactment or use of big ideas in practices, namely, the learning performances.” (Duschl, 2006, p. 116)
- “Learning progressions . . . describe successively more sophisticated ways of

reasoning within a content domain that follow one another as students learn: They lay out in words and examples what it means to move toward more expert understanding.” (Smith, Wiser, Anderson, & Krajcik, 2006, p. 2)

- “Learning progressions are successively more sophisticated ways of thinking about a topic that can be used as templates for the development of curricular and assessment products.” (Songer, Kelcey, & Gotwals, 2009, pp. 2–3)
- Vertical maps that provide “a description of skills understanding and knowledge in the sequence in which they typically develop: a picture of what it means to ‘improve’ in an area of learning.” (Masters & Forster, 1996, p. 1)
- Learning progressions “assume a progression of cognitive states that move from simple to complex; while not linear, the progression is not random and can be sequenced and ordered as ‘expected tendencies’ or ‘likely probabilities’” (Confrey & Maloney, 2010).

Although there are variations in these descriptions of progressions, they share two common views: (1) progressions lay out in successive steps, increasingly more sophisticated understandings of core concepts and principles in a domain and (2) progressions describe typical development over an extended period of time. They reflect the idea that early in their schooling students develop concepts and skills at a rudimentary level, and over time, their understanding of the concepts and their acquisition of skills are developed in progressively more sophisticated forms. Progressions provide a connected map of the steps along the way to increasing expertise. These steps, which are components of a connected and coherent landscape of learning in a domain, can serve as guides for instruction and act as a touchstone for gathering evidence for formative purposes.

Developing and Validating Learning Progressions

Elsewhere, I have suggested that approaches to developing progressions can be loosely characterized in two ways: (1) top down and (2) bottom up (Heritage, 2008). In a top-down approach to progressions, experts in a domain (e.g., physicists,

mathematicians, or historians) construct a progression based on their domain knowledge and research on how children's learning unfolds in the domain. They conduct empirical research to determine the extent to which their hypothesis holds up in reality and make refinements based on the resulting data (e.g., Black et al., 2011; Confrey & Maloney, 2010; Songer et al., 2009). Ideally, learning progressions should be developed from a strong research base about the structure of knowledge in a discipline and the kind of thinking that is involved and be subject to a rigorous validation process. However, given that a strong research base does not exist in many domains, a bottom-up approach has to fill the gap until it such a research base materializes.

A bottom-up approach involves teachers and curriculum content experts in developing a progression that is based on their experience of teaching children. Their sources for developing the progression are curricula, their views of what is best taught when, and their knowledge of children's learning. In this context, validation involves working together, testing each other's hypotheses against their professional knowledge, making refinements accordingly, trying out the progression to see if their model actually predicts what happens in terms of student learning, and then making further refinements from this experience (e.g., Riley, 2009). A by-product of teacher-developed progressions is an associated deepening of teacher knowledge about learning in a domain, which can have considerable payoff for evidence gathering and use. In the end, top-down and bottom-up approaches to learning progressions are not, and should not be, mutually exclusive. Indeed, creating institutional contexts for self-conscious and systematic integration of the two approaches will be an important development in the next phase of formative assessment practice.

Learning Progressions and Evidence Gathering

Whatever the source of progressions, if they are to be used to support effective evidence gathering they must reflect clear goal steps to index how learning progresses in a domain. Lest these goal level steps be interpreted as a laundry list of objectives, it should be stressed that progressions document connected steps on the way to increasing expertise in relation to

core principles, ideas, or concepts and skills in a domain. They are not discrete objectives. However, even the most well-developed and rigorously validated progressions will not be able to support evidence gathering and formative pedagogical responses if the steps are specified at too gross a level. Herein lies the problem with *summative* standards—even those that suggest elements of progression (e.g., the Common Core State Standards show more of a connected pathway than other standards) are still much too gross-grained for formative assessment. In relation to progressions, the Goldilocks maxim applies: Progressions to support evidence gathering should be at the *just right* level of detail so that they enable teachers to map evidence-gathering strategies onto the progression, to interpret the student responses against the background of the progression, and to make appropriate pedagogical moves that will advance learning.

Black et al. (2011) provided a useful description of how progressions work in the ongoing gathering and use of evidence:

Success overall then depends *first* on the power of the opening questions or activities to provoke rich discussion, but then *secondly* on the capacity of the teacher to listen, to interpret the responses, and to steer the discussion with a light but firm touch, by summarizing, or by highlighting contradictions, or by asking further questions. To do this skillfully and productively, one essential ingredient that the teacher needs is to have in mind an underlying scheme of *progression* in the topic; such a scheme will guide the ways in which students' contributions are summarized and highlighted in the teacher's interventions and the orientation which the teacher may provide by further suggestions, summaries, questions and other activities. (p. 74)

In the absence of clear progressions, despite their best intentions, teachers may not have the necessary resources to guide their evidence-gathering strategies in planned and systematic ways at the level required for formative assessment, nor be able to recognize the import of spontaneous evidence when it arises. In a worst-case scenario, the evidence gathering will be random, ad hoc, and unconnected to a picture of the progressive development of understanding and skills. The

best-case scenario of systematic, planned approaches to gathering evidence that can be used to consistently move learning forward is enabled by progressions.

As an endnote to this section on progressions, it is important to recognize that while progressions offer an important resource to teachers to gather evidence and to use it, teachers' experience and knowledge about responses to the evidence that will benefit students' learning are an essential adjunct to the value of progressions. Without due attention to the skills needed by teachers to formulate the next step (Heritage, Kim, Vendlinski, & Herman, 2009), progressions, while necessary, may not be sufficient to facilitate emergent learning.

THE ROLE OF THE STUDENT IN GATHERING EVIDENCE

The practice of formative assessment is a joint enterprise in which teachers and students play distinctive, yet complementary, roles in the common purpose of furthering learning. With respect to evidence gathering, students also play an active role in generating internal feedback on their learning through self-assessment (Black & William, 1998; see also Chapter 21 of this volume). Self-assessment has two aspects: (1) self-appraisal and (2) self-management (Paris & Winograd, 1990). Self-appraisal refers to students' ability to evaluate their learning status and learning strategies through a range of self-monitoring processes (Hattie & Timperley, 2007). Self-management is the students' capacity to take appropriate action to sustain ongoing learning. Engaging in self-assessment contributes to the development of learners' self-regulation processes when they activate and focus their cognitions and behaviors on attaining their specific, personal goals (Zimmerman & Schunk, 2011). The employment of these processes has been associated with achievement differences among students (Schunk, 1981, 1984; Zimmerman & Martinez-Pons, 1986, 1988). Of most relevance to this chapter is that students who are proactive self-regulators monitor and assess their progress toward goals, using the internal feedback they generate to determine when to seek assistance, when to persist with an approach, and when to adjust

their learning strategies (Zimmerman & Schunk, 2011).

That students are involved in self-monitoring does not mean that teachers abrogate their responsibilities with regard to evidence gathering. The point here is that both teachers and students generate evidence that they use in reciprocally supportive ways to progress learning. For this to be an effective process, teachers create the conditions in which they and their students develop a shared understanding of their respective roles. For example, students must come to understand that it is the teachers' role to elicit evidence about their learning status and that it is their responsibility to provide responses that help the teacher gain insights they can use to advance learning. In turn, the students understand that it is their role to generate internal evidence during learning and that the teachers' responsibility is to help them reflect on the evidence and build a repertoire of strategies through feedback that supports them to become self-regulated learners. Without this kind of partnership in relation to gathering evidence, students will remain overwhelmingly dependent on the teacher as the primary resource for learning and lack the capacity to become self-sustained lifelong learners (Heritage, 2010).

Sadler (1989) stressed that for students to be able to monitor their own learning they must come to hold a conception of quality similar to the teacher's and that developing this conception depends on the following:

- (i) possessing a concept of the *standard* (or goal, or reference level) being aimed for; (ii) comparing the *actual* (or current) *level of performance* with the standard; and (iii) engaging in appropriate *action* which leads to some closure of the gap. (p. 121)

He also made clear that self-monitoring does not happen automatically but has to be learned. To enable successful self-assessment then, the teacher needs to help students understand the goal being aimed for, understand the criteria for meeting the goal, assist them to develop the skills to make judgments about their learning in relation to the goal, and establish a repertoire of operational strategies to direct their own learning. Teachers must also ensure that there is time

for students to systematically engage in reflection. Teacher Sharon Pernisi—when summing up the changes she had made to preserve time for this—noted, “I used to do more but now I do less. Now I work hard to save time for student reflection rather than filling every minute with activity.” (Heritage, 2010, p. 4) This should be the goal of all teachers who are concerned with self-assessment.

In terms of supports for student self-assessment, Allal (2010) suggested that self-assessment tools can be either embedded in the curriculum materials or devised by the teacher. Allal (2010) also advised that self-regulation can be enhanced when teachers assist students to analyze the purposes and uses of the tools and to consider the results of their use. Going further, she proposed teachers can also assist students in developing their own self-assessment tools, for example, checklists, and internal questions to ask oneself when engaging in a task. Hattie and Timperley (2007) suggested that teacher feedback can be focused at the self-regulation level to assist students’ evaluative skills. Such feedback draws students’ attention to the criteria needed to achieve the goal, puts the responsibility squarely on the student to evaluate the evidence in relation to the criteria, and provides a model for how students can assess their learning.

Another approach to supporting students in the development of self-assessment skills is providing opportunities for peer-assessment and feedback (see Chapter 22 of this volume). These kinds of peer processes help students develop the skills to make judgments about evidence in relation to specific goals, which can then be transferred when students engage in and regulate their own work (Boud, Cohen, & Sampson, 1999; Gibbs, 1999). A by-product of student peer assessment for teachers can be the opportunity to gain insights into how well students understand the learning goal and the quality criteria for meeting the goal.

Ultimately, self-assessment by students is neither an optional extra nor a luxury but has to be seen as essential to the practice of formative assessment (Black & Wiliam, 1998). Students have to be active in their own learning, since no one else can learn for them, and unless they are able to evaluate their own strengths and weaknesses and how they might deal with them, they are unlikely to make progress (Harlen & James, 1996).

CONCLUSION

This chapter has presented some basic dimensions of evidence gathering in the context of formative assessment practice. It has suggested that evidence gathering is a fundamental means to provide information on students’ current levels of learning and to guide what teachers and students do to ensure progress. There is no single way to collect evidence for formative purposes. Evidence-gathering strategies can range from planned interactions between teacher and student, to examinations of student work products, and to technology-assisted affordances. However, evidence gathering will only be effective if a number of criteria obtain. First, teachers must have the knowledge to formulate tasks and occasions that provide insights into student thinking. Second, evidence in and of itself is of little use if it is not used to “form” new learning. To this end, the evidence elicited needs to provide a present-to-future perspective so that teachers and students are able to use it for the purpose of extending current learning within and through the students’ ZPD. Third, current standards do not provide teachers or students with a clear pathway of learning that can guide both instruction and assessment. While research-based and empirically validated learning progressions that could fulfill this role are emerging, they are presently insufficient to guide evidence gathering. Teachers will have to find ways to develop these progressions using available expertise. Fourth, while concepts of validity and reliability have not been extensively applied to evidence in the context of formative assessment, whatever strategies are employed must be of sufficient quality to provide actionable evidence that will have a strong probability of improving learning. Finally, evidence gathering is not solely within the purview of teachers. In formative assessment practice, students and teachers are collaborators in the common purpose of progressing learning. Without student involvement in the process, students remain passive recipients of teacher judgment and action. Students, too, must be engaged in reflective practices through the process of self-assessment, which in turn, supports the development of self-regulation, a characteristic of effective learning. Only in this way will they come to understand what it means to learn how to learn and be equipped with the skills they need for success beyond school.

REFERENCES

- Allal, L. (2010). Assessment and the regulation of learning. In P. Peterson, E. Baker, & B. McGraw (Eds.), *International encyclopedia of education* (Vol. 3, pp. 348–352). Oxford, UK: Elsevier.
- Alonzo, A. C., & Steedle, J. T. (2008). Developing and assessing a force and motion learning progression. *Science Education*, 93, 389–421.
- Anastasi, S. (1990). Ability testing in the 1980's and beyond: Some major trends. *Public Personnel Management*, 18, 471–485.
- Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, 40, 685–730.
- Bailey, A., & Heritage, M. (2008). *Formative assessment for literacy, grades K–6*. Thousand Oaks, CA: Corwin Press.
- Bell, B., & Cowie, B. (2000). The characteristics of formative assessment in science education. *Science Education*, 85, 536–553.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Buckingham, UK: Open University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice*, 5, 7–73.
- Black, P., & Wiliam, D. (2005). Changing teaching through formative assessment: Research and practice. In *Formative assessment: Improving learning in secondary classrooms* (pp. 223–240). Centre for Educational Research and Innovation (CERI). Paris: Organisation for Economic Cooperation and Development.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability*, 21, 5–31.
- Black, P., Wilson, M., & Yao, S. Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives*, 9(2–3), 71–122.
- Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education*, 24, 413–425.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12.
- Carpenter, T. P., Fennema, E. T., & Franke, M. L. (1996). Cognitively Guided Instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal*, 97, 3–20.
- Carpenter, T. P., Fennema, E., Peterson, P. L., & Carey, D. A. (1988). Teachers' pedagogical content knowledge in mathematics. *Journal for Research in Mathematics Education*, 19, 385–401.
- Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.
- Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 39–64). Cambridge, UK: Cambridge University Press.
- Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *Journal of Research in Science Teaching*, 44, 815–843.
- Cohen, N., Hall, T. E., Vue, G., & Ganley, P. (2011). Becoming strategic readers: Three cases using formative assessment, UDL, and technology to support struggling middle school readers. In P. Noyce & D. T. Hickey (Eds.), *New frontiers in formative assessment* (pp. 129–140). Cambridge, MA: Harvard Education Press.
- Confrey, J., & Maloney, A. (2010, October). *Building formative assessments around learning trajectories as situated in the CCSS*. Paper presented at fall meeting of the State Collaborative on Assessment and Student Standards: Formative Assessment for Students and Teachers, Savannah, GA.
- Cook, K., Seeley, C., & Chaput, L. (2011). Customizing and capture: Online assessment tools for secondary mathematics. In P. Noyce & D. T. Hickey (Eds.), *New frontiers in formative assessment* (pp. 69–85). Cambridge, MA: Harvard Education Press.
- Cronbach, L. (1988). Internal consistency of tests: Analyses old and new. *Psychometrika*, 53, 63–70.
- Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation*, 37, 71–77.
- Crooks, T., Kane, M., & Cohen, A. (1996). Threats to the valid use of assessments. *Assessment in Education*, 3, 265–285.
- Dewey, J. (1928). Progressive education and the science of education. *Progressive Education*, V, 197–201.
- Dierick, S., & Dochy, F. (2001). New lines in edometrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27, 307–329.
- Duschl, R. A. (2006). Learning progressions as babushkas. *Measurement*, 14(1 & 2), 116–119.
- Duschl, R. D., & Gitomer, D. H. (1997). Strategies and challenges to change the focus of assessment and instruction in science

- classrooms. *Educational Assessment*, 4(1), 37–73.
- Erickson, F. (2007). Some thoughts on “proximal” formative assessment of student learning. *Yearbook of the National Society for the Study of Education*, 106, 186–216.
- Frederiksen, J., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27–32.
- Gibbs, G. (1999). Using assessment strategically to change the way students learn. In S. Brown & A. Glasner (Eds.), *Assessment matters in higher education: Choosing and using diverse approaches* (pp. 41–53). Buckingham, UK: Open University Press.
- Ginsburg, H. P. (2009). The challenge of formative assessment in mathematics education: Children’s minds, teachers’ minds. *Human Development*, 52(2), 109–128.
- Griffin, P. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation*, 33, 87–99.
- Griffin, P., Murray, L., Care, E., Thomas, A., & Perri, P. (2010). Developmental assessment: Lifting literacy through professional learning teams. *Assessment in Education: Principles, Policy & Practice*, 17, 383–397.
- Harlen, W. (2007). Formative classroom assessment in science and mathematics. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 116–135). New York: Teachers College Press.
- Harlen, W., & James, M. (1996). *Creating a positive impact of assessment on learning*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365–379.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Heffernan, N., Militello, M., Heffernan, C., & Decoteu, M. (in press). Effective and meaningful use of educational technology: Three cases from the classroom. In C. Dede & J. Richards (Eds.), *Digital teaching platforms*. New York: Teachers College Press.
- Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment*. Washington, DC: Council of Chief State School Officers. Retrieved from www.ccsso.org/content/PDFs/FAST_Learning_Progressions.pdf
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin Press.
- Heritage, M., & Heritage, J. (2011). *Teacher questioning: the epicenter of instruction and assessment*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- 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.
- Heritage, M. & Niemi, D. (2006). Toward a framework for using student mathematical representations as formative assessments [Special issue]. *Educational Assessment*, 11(3 & 4), 265–284. Mahwah, NJ: Lawrence Erlbaum.
- Herman, J. L., Heritage, M., & Goldschmidt, P. (2011). *Developing and selecting assessments of student growth for use in teacher evaluation systems*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Jordan, B., & Putz, P. (2004). Assessment as practice: Notes on measurement, tests, and targets. *Human Organization*, 63, 346–358.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527–535.
- Kouba, V., & Franklin, K. (1995). Multiplication and division: Sense making and meaning. *Teaching Children Mathematics*, 1, 574–577.
- Kozulin, A., Gindis, B., Ageyev, V. S., & Miller, S. M. (2003). Introduction: Sociocultural theory and education: Students, teachers, and knowledge. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky’s educational theory in cultural context* (pp. 1–11). Cambridge, UK: Cambridge University Press.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Cambridge: University of Chicago Press.
- Lesh, R., Hoover, M., Hole, B., Kelly, A. E., & Post, T. (2003). Principles for developing thought revealing activities for students and teachers. In R. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Mahwah, NJ: Lawrence Erlbaum.
- Linn, R. L., & Baker, E. L. (1996). Can performance-based student assessment be psychometrically sound? In J. B. Baron & D. P. Wolf (Eds.), *Performance-based assessment—challenges and possibilities: 95th yearbook of the National Society for the Study of Education part 1* (Vol. 95(1), pp. 84–103). Chicago: National Society for the Study of Education.
- Masters, G., & Forster, M. (1996). *Progress maps: Assessment resource kit*. Melbourne, Victoria, Australia: The Australian Council for Educational Research.
- Mehan, H. (1979). *Learning lessons*. Cambridge, MA: Harvard University Press.

- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education and Macmillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
- Newman, L. S. (1990). International and unintentional memory in young children: Remembering vs. playing. *Journal of Experimental Child Psychology*, 50, 243–258.
- Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English*, 25, 261–290.
- Paris, S. G., & Winograd, P. W. (1990). How metacognition can promote academic learning and instruction. In B. J. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15–51). Hillsdale, NJ: Lawrence Erlbaum.
- Pellegrino, J. W., & Glaser, R. (1982). Analyzing aptitudes for learning: Inductive reasoning. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 269–345). Hillsdale, NJ: Lawrence Erlbaum.
- Phelan, R. (2003). *Assessment, curriculum, and instruction: Overview of the standards movement*. Retrieved from www.sonoma.edu/users/p/phelan/423/standards.html
- Ploegh, K., Tillema, H., & Segers, M. R. S. (2009). In search of quality criteria in peer assessment practices. *Studies in Educational Evaluation*, 35, 102–109.
- Poppers, A. E. (2011). Identifying craft moves: Close observation of elementary students writing. In P. Noyce & D. T. Hickey (Eds.), *New frontiers in formative assessment* (pp. 89–107). Cambridge, MA: Harvard Education Press.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13.
- Riley, J. (2009, June). *Developing learning progressions*. Paper presented at the Council of Chief State School Officers National Conference on Student Assessment, Detroit, MI.
- Rowe, M. B. (1974). Wait time and rewards as instructional variables, their influence in language, logic, and fate control: Part II, rewards. *Journal of Research in Science Teaching*, 11, 291–308.
- 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. (2004, April). *Informal assessment of students' understanding of scientific inquiry*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- 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), 237–263.
- 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 Educational Research in Science Teaching*, 44(1), 57–84.
- Sadler, D. R. (1989). Formative assessment and the design of instructional strategies. *Instructional Science*, 18, 119–144.
- Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology*, 73, 93–105.
- Schunk, D. H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist*, 19, 48–58.
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A. et al. (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.
- Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1066–1101). Washington, DC: American Educational Research Association.
- Sinclair, J. M., & Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. London: Oxford University Press.
- Smith, C., Wiser, M., Anderson, C., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and atomic-molecular theory. *Measurement*, 14, 1–98.
- Smith, J. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, 22(4), 26–33.
- Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal for Research in Science Teaching*, 46, 610–631.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 133–146). Thousand Oaks, CA: Sage.

- Stobart, G. (2008, November). *Validity in formative assessment*. Paper presented at the Ninth Annual AEA-Europe Conference, Hisar, Bulgaria.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment*. Buckingham, UK: Open University Press.
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27, 615–631.
- Vygotsky, L. S. (1935). *Umstvennoie razvitie detei v protsesse obuchenia*. Moscow, Russia: Gosudarstvennoie Uchebno-pedagogicheskoe Izdatel'stvo.
- Vygotsky, L. S. (1986). *Thought and language* (Rev. ed.) (A. Kozulin, Ed.). Cambridge, MA: MIT Press.
- Webb, M. E., & Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 16, 165–184.
- Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- Wiliam, D., & Black, P. J. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22(5), 537–548.
- Zimmerman, B., & Schunk, D. (2011). *Handbook of self-regulation of learning and performance*. New York: Routledge.
- Zimmerman, B. J., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, 23, 614–628.
- Zimmerman, B. J., & Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, 80, 284–290.

