Formative Assessment for Next Generation Science Standards: A Proposed Model

Joan Herman

September 2013

K-12 Center at ETS

Invitational Research Symposium on Science Assessment

Formative Assessment for Next Generation Science Standards: A Proposed Model

Joan Herman
National Center for Research on Evaluation, Standards and Student Testing (CRESST)
UCLA Graduate School of Education and Information Studies

Introduction

Historically, educational policymakers have focused on and invested heavily in accountability testing *of learning* to leverage improvement in student learning. Through accountability testing, policy makers aim to communicate standards, establish performance goals, provide data through which educators can analyze and improve school programs and student performance, and establish incentives and sanctions to motivate action. Today, however, there is growing recognition of the limitations of accountability testing *of learning* and wide acknowledgment and accumulating evidence of the crucial role that formative assessment—*assessment for learning*—can play in helping all students achieve rigorous standards. Rather than looking back to judge what has been learned, formative assessment projects forward. It involves the ongoing collection and use of assessment during instruction to understand where students are relative to intended goals, as well as the use of that data to take immediate action—to adapt teaching and learning—to help students get to where they need to go. Attesting to the popularity of formative assessment in current educational policy and practice, the two Race to the Top Common Core State Standards assessment consortia are charged with developing formative and interim tools and practices, in addition to end-of-year accountability tests. Formative assessment must also be an essential—if not *the* key—component of any assessment system for the Next Generation Science Standards (NGSS Lead States, 2013).

Yet, despite its popularity, formative assessment remains an elusive concept, its definition muddied in research, policy, and practice. The predominant view sees formative assessment as an evidence-based process that uses ongoing assessment to "form" subsequent instruction (Black & Wiliam, 2004; Formative Assessment for Students and Teachers, 2008). Teachers are continually assessing where students are relative to learning goals and taking immediate action to adapt teaching and learning to students' needs.

At the same time, however, it seems clear that the process of formative assessment requires sound evidence to achieve its goals. Teachers need to employ assessment strategies and tools that will yield valid inferences about student learning. Without such validity, formative assessment may yield faulty decisions that could delay rather than promote student progress. In other words, if formative assessment is a process—a verb, it needs effective tools and strategies—nouns—to reach its promise. Measurement-oriented researchers and assessment developers tend to have the latter as a focus: the design and validation of tools and strategies that may be necessary to enact effective formative assessment practice. And indeed, commercial vendors try to make the case for their interim or benchmark testing serving formative purposes. Still other researchers bring to bear diverse theoretical perspectives in their definition and consideration of essential mechanisms and outcomes of the formative assessment process, highlighting, for example, advances in cognitive, motivation, and sociocultural theory (see reviews by Assessment Reform Group, 2006; Herman, Osmundson, & Silver, 2010; Shepard, 2005).

This paper attempts a synthesis of these various perspectives to propose a conceptual model underlying the design and implementation of formative assessment to support the learning goals of the Next Generation Science Standards (NGSS Lead States, 2013). The paper starts with a selective review of diverse literatures bearing on effective formative assessment practice and uses this literature to describe and justify a proposed model. The paper then highlights special considerations for and provides examples of the model's application in Next Generation Science Standards, and ends with implications for the design of coherent assessment systems that support student learning.

Literature Review

The sections below summarize formative assessment theory and research from three distinct perspectives. These include the role of formative assessment in improving teaching and learning; the nature and effects of assessments that support learning; and the perspectives of cognitive, motivation, and sociocultural theories.

The Role of Formative Assessment in Improving Learning

The role of assessment in improving student learning has been a motivating concern throughout the history of educational measurement (see, for example Baker & Popham, 1973; Glaser, 1963; Thorndike, 1918; Tyler, 1949). Black and Wiliam's (1998) landmark metareview of studies related to the use of classroom formative assessment ignited the worldwide interest in its use. Their review incorporated a wide variety of studies related to the use of assessment to improve teaching and learning—ranging, for example, from studies of mastery learning to those involving teachers' classroom assessment practices and use of feedback, and student engagement in self-assessment. The researchers concluded that formative assessment had an effect size of between .4 and .7 on standardized tests, making it demonstrably one of the most effective educational interventions in practice, particularly for low achieving students. They argued that the feedback to close the gap between where students were and a given benchmark or goal was the critical feature of formative assessment and posed social context and student empowerment as key considerations in the formative assessment process. The Black and

Wiliam analysis echoed key elements that in Sadler (1989) proposed as essential to the formative assessment process: (a) clear learning goals, (b) information about the present state of the learner, and (c) action to close to gap between the learner's present state and the learning goals.

Definitions of formative assessment as a process. Following on both Black and Wiliam's and Sadler's work, Great Britain's Assessment Reform Group (1999, 2002) defined formative assessment—or *assessment for learning* as the group called it—as the "process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (Assessment Reform Group, 2002, p. 2). More than a decade later in the United States, the Council of Chief State School Officers' (CCSSO's) Formative Assessment for Students and Teachers (2008) definition echoes these same themes: "Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes" (Assessment Reform Group, 2002, p. 3)

Both definitions emphasize formative assessment as a *process* during *ongoing instruction*, where *both teachers and students* use evidence of current learning to improve subsequent learning. The questions that formative assessment—or assessment *for* learning—answers essentially involve diagnosing students' learning progress and taking appropriate action based on it. Among the questions:

- Where are students relative to my immediate learning goals? What stands in their way of accomplishing the goals?
- Have students progressed as I expected? Has their thinking advanced as I had planned? If not, what misconceptions or learning obstacles do they evidence?
- How can I help students to bridge the gap between where they currently are and where I want them to be, that is, accomplishing immediate and subsequent learning goals, progressing toward mastery?
- Based on the data, what are next steps for teaching and learning? What kinds of instructional activities will best respond to individual students' learning needs?

Figure 1 graphically displays this general process of assessment for learning, courtesy of my colleague, Margaret Heritage (2010).

Figure 1. The formative assessment process. From **Formative Assessment: Improving Teaching and Learning*****, by M. Heritage, 2007. Paper presented at the CRESST 2007 Assessment Conference, Los Angeles, CA.***

From a teacher-centric perspective, teachers start by making their learning goals clear and knowing how that learning is expected to develop. Teachers collect evidence of where students are relative to that progression by asking students questions, observing student activity, and analyzing student work. Teachers' analysis of student responses enables them to interpret the status of student learning and to identify the gap between where students are and where they need to be. Teachers then use these interpretations to provide feedback to students and take instructional action to help students clarify their misconceptions and bridge identified gaps…and the process starts all over again.

Because student involvement and responsibility also is paramount in definitions of formative assessment, Figure 1 also can characterize student processes. Students, too, may be responsible for setting learning goals for themselves, monitoring and self-assessing their progress, providing feedback to their peers, and proactively acting on feedback to move their learning forward. Technology, too, may play a role in the enactment of formative assessment practices. Simulations and games, for example, can be designed and used to support each and all stages of the formative assessment process (e.g, by eliciting and analyzing evidence of student understanding, providing individualized and focused feedback and reports of results, and linking to resources for next steps). Whether from the perspective of teachers and/or students, enhanced by technology or not, formative assessment is an ongoing and dynamic process that is integrated with ongoing instruction.

More recent research on learning effects. In addition to articulating key components of the formative assessment process, more recent theory and empirical research has subjected Black and Wiliam's claims to more rigorous quantitative analysis and has investigated key components of the formative assessment process. A few selected studies are reviewed below:

Unlike Black and Wiliam's metareview, which incorporated studies of uneven methodological quality, Kingston & Nash's (2011) recent meta-analysis focused on studies of formative assessment meeting rigorous standards. Their study examined not only the average effect size documented in studies of formative assessment, but also the extent to which effect sizes are moderated by grade, content area, or specific formative assessment interventions. Based on 13 studies with 42 independent effect sizes, the authors found a weighted effect size for formative assessment of .20, substantially lower than that reported in Black and Wiliam's seminal review. Further, analysis of moderator variables indicated stronger effect sizes for English language arts than for math or science (.32 compared to .17 and .09, respectively) and for treatments based on professional development or on use of technology-based formative systems (.30 and .28, respectively), rather than for curriculum-based interventions. However, interventions included in the study tended to focus on formal assessment activity at key juncture points, rather than ongoing formative assessment practices. The Kingston and Nash (2011) findings, in short, reinforce the potential of formative assessment but also show the wide variation in observed effects; the researchers concluded with the need for high quality research that considers critical variables in practice.

Feedback, a prominent component in the Black and Wiliam review, also has been the subject of recent research. Hattie and Timperley's (2007) reanalysis of existing meta-analyses confirmed that feedback can be one of the most powerful influences on student learning, but that its effects are conditional based on the nature of the feedback; the learning goals; and learner knowledge, self-beliefs, and self-regulation. Defining feedback as "information provided by an agent (e.g., teacher, peer, book, parent, self experience) regarding aspects of one's performance or understanding," (Hattie & Timperley, 2007, p. 81), the researchers contended that the feedback must provide answers to at least one of three overlapping questions, reminiscent of Sadler's (1989) formulation: "Where am I going? How am I going? Where to next?" Feedback is most effective, according to the researchers' analysis, when learning goals are clear and specific and students are committed to attaining them; when feedback provides effective cues for better accomplishing a task; and when the feedback is sensitive and/or adapted to students' prior knowledge and understanding (i.e., students must have the prior knowledge to understand and act on the feedback).

The researchers differentiate three levels of potentially effective feedback, in ascending order of both effectiveness and difficulty in implementing well: task or product feedback, which provides substantive cues specific to accomplishing the task; process feedback, which cues previously learned processes that can be used to accomplish the task; and self-regulation feedback, which prompts the use of metacognition. A fourth level of feedback, personal feedback or reinforcement—such as "Good job!

Try harder!"—they find ineffective. The researchers' formulation conceives of both teachers and students as learners—actively seeking, responding, and learning from feedback and using it to improve teaching and learning—and conceptualizes assessment as activities that provide teachers and/or students with evidence to answer at least one of the three key questions, providing cues to support learning at one or more levels (task, process, self-regulation). As Hattie (2012, p. 22) pointed out, "the biggest effects on student learning occur when teachers become learners of their own teaching and when students become their own teachers." (Later, I summarize some of the cognitive theory that underlies these views, e.g., Vygotsky's [1978] zone of proximal development, motivation theory, and research on self-regulation and metacognition.) Teachers and students must take responsibility for continually seeking feedback on whether students are on track to reach intended goals and on what problems they may be encountering along the way and for taking action to support continuous progress toward goals (Corcoran, Mosher, & Ragat, 2009).

At the same time, however, research reveals challenges in bringing such practices to fruition. Prime among these are teachers' knowledge and skill in implementing formative assessment. For example, Heritage, Kim, Vendlinski, and Herman (2009) documented middle school mathematics teachers' limited capacity to accurately analyze student work or to generate next steps for instruction; Herman and colleagues (Herman, Osmundson, Ayala, Schneider, & Timms, 2006; Herman, et al., 2010) found similar limitations in science teachers' content and pedagogical knowledge and in their formative assessment practices. Other studies also have found the challenges teachers face in eliciting students' understanding, providing productive feedback, and bridging the gap to take students to deeper levels of understanding (Furtak et al., 2008; Herman et al., 2006; Shavelson et al., 2009).

Centrality of learning progressions. The current state of knowledge about how students' science knowledge develops over time presents another key underlying challenge. That is, the whole formative notion of looking forward and assessing where students are relative to desired learning goals and taking action to support and/or accelerate their goal attainment implies knowledge of the pathway through which students are expected to develop to achieve mastery. Otherwise, how does one know where students are relative to a goal and how to take action to achieve it? Learning progressions serve this function (see also Corcoran, et al., 2009; Heritage, 2008). In science, they are empirically grounded and testable hypotheses about how students' understanding and ability to apply scientific concepts and related practices develop and grow more sophisticated over time, in the context of appropriate instruction (National Research Council [NRC], 2007). They are grounded in research on how students' learning actually develops, rather than in traditional curriculum sequences or logical analysis of how learning components may fit together.

The Next Generation Science Standards (NGSS Lead States, 2013) and underlying Framework (NRC, 2012a) show this same commitment to learning progressions as a key theme in science learning and in coherent science education. The Framework conceptualizes learning as a trajectory through which students progress over the course of a unit, year, or K–12, particularly as they engage in practices that involve them in active analysis, argument, inquiry, and so forth. To develop disciplinary core ideas, for example, the Framework maintains that "…students need sustained opportunities to work with and develop the underlying ideas and to appreciate those ideas' interconnections over a period of years rather than weeks or months" (NRC, 2012a, p. 26). At the same time, however, research on "learning progression in science is at an early stage [and] many aspects of the core ideas and their progressions over time with instruction remain unexplored territory" (NRC, 2012a, pp. 13–14).

Assessments that Support Learning

Many of the same themes that permeate discussions of the formative assessment process recur in theory and research on the nature of assessments—tests, assessment tools—that benefit teaching and learning.

Quality in learning-based assessment. More than a decade ago, *Knowing What Students Know* (KWSK; Pellegrino, Chudowsky, & Glaser, 2001) synthesized advances in cognitive and psychometric theory to provide the foundation for the design and development of new kinds of assessments that will help students learn and succeed in school by clarifying for students, their teachers, and other education stakeholders the nature of student accomplishments and the progress of their learning. KWSK established that "Every assessment….rests on three pillars: a model of how students represent knowledge and develop competence in a subject matter domain; tasks of situations that allow one to observe students' performance; and an interpretation method for drawing inferences from the performance evidence thus obtained" (Pellegrino et al., 2001, p. 2).

Mirroring core components in Heritage's model formative assessment process, the oft-cited KWSK assessment triangle (see Figure 2) starts with cognition, which can be taken as specification of learning goals and how learning is expected to develop (i.e., a learning progression) and highlights the need for both appropriate observations or tasks through which student understanding will be elicited and an explicit interpretative framework for analyzing student performance and making inferences about student progress relative to focal learning goal(s). Assessment development closely connects and coordinates all three elements, evoking the ideas of evidence-centered design (ECD) and assessment as a process of reasoning from evidence to make inferences about student learning (see, for example, Mislevy, Almond, & Lukas, 2003).

Inherent in KWSK's formulation is another important recognition: assessment validity cannot be an afterthought but rather must designed in, to both assure the close coordination of the three pillars at the levels and assure that the assessment will provide appropriate evidence to serve the intended assessment purposes. For purposes of formative assessment, students' responses must yield inferences about the gaps and/or misconceptions that obstruct students' pathways to immediate learning goals, that is, the assessment must yield diagnostic information to inform subsequent teaching and learning. To do so, it seems axiomatic that the assessment must be based on a robust cognitive model of how learning is expected to develop, a robust learning progression that provides the foundation for diagnosis, and gap analysis. These models have been variously conceptualized as ontologies (Baker, 2012), conceptual flows (DiRanna et al, 2008; assessment-centered teaching), and learning progressions (Forster & Masters, 2004; Heritage, 2008; Smith, Wiser, Anderson, & Krajcik, 2006; Wilson & Sloane, 2001).

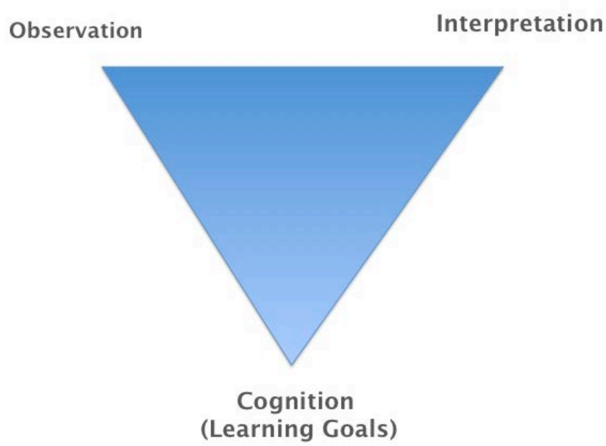

Figure 2. Knowing what students know: Assessment triangle, Adapted from Knowing What Students Know, by J. Pellegrino, N. Chudowsky, and R. Glaser, 2001, Copyright 2001, by the National Academies Press.

Typologies of formative assessment tools and strategies. A complete ontology or learning progression documents how learning is expected to develop, but can do so at levels of detail that may or may not directly serve a specific level or type of formative assessment. For example, Wiliam and Thompson (2007) proposed a typology of formative assessment based on the duration of instruction assessed that may well have strong implications for the extent of the progression and nature of the gap an assessment is intended to uncover:

- Short cycle—minute-by-minute and/or day-by-day assessments that focus on student learning within and between lessons.
- Medium cycle—focusing within and between instructional units, typically covering a 1- to 4-week sequence of instruction.
- Long cycle—focusing on quarterly, semester, annual, or other time period progress, typically traversing instruction over 4 weeks to 1 year.

Richard Shavelson and colleagues (2008), in turn, defined a continuum of formative assessment that ranges from informal to formal and varies relative to specific preplanning. The continuum essentially focuses on short and medium cycle tools, running the gamut from unplanned “on the fly” assessments that particularly capitalize on “teachable moments,” through planned interactions during the course of instruction that are specifically designed to both support learning and elicit evidence of it, to formal embedded assessments at key juncture points to evaluate what all students have learned. The informal activities are directly embedded in instruction and indistinguishable from it. Teachers (and students) may draw inferences and provide feedback on student learning by observing student behavior and interactions during classroom activities; analyzing students' work, such class assignments, lab work, science notebook entries, and/ or homework; or analyzing and responding to whole class and/or small group discussion questions (see also Bell & Cowie, 2001).

The constant, regardless of cycle time or formality, is that the interaction or task starts with a clear sense of the learning goal and how it is likely to develop, leverages a specifically designed activity or spontaneous interaction to reveal student understanding relative to the trajectory, and applies an interpretative framework to analyze student responses and provide feedback relative to the goal—and may indeed call upon subsequent interaction to fine-tune the interpretation and feedback. Such design and interpretation calls on a detailed sense of progression: where students are likely to start relative to a given disciplinary idea(s), cross cutting concept(s) and/or practice(s); how they are expected to develop over the course of a lesson, week, unit, year, or across years; and how to probe and interpret student responses along the way—for example, what does typical progress look like, what are intermediate points along the way, what are common misconceptions or obstacles along the way, and how are they detected.

As noted above, research-verified progressions, particularly at this level of detail, do not yet exist, so teachers' or developers' designs will be largely dependent on their experience-based sense of progression, which can then be revised over time (see, as an example, Lehrer, Wilson, Ayers, & Kim [2011] for a developmental process combining substantive and psychometric analysis to identify and validate a specific learning progression). Similarly, the strength of teachers' interpretative frameworks may be expected grow over time.

It's important to note that while we expect formative assessment to follow the same general KWSK design process, teachers' approaches clearly will be more qualitative and cannot be expected to meet the psychometric and other validity demands of large scale, high stakes tests. Moreover, because teachers can draw on and triangulate a variety of evidence sources and can probe further as necessary, standards for strict reliability and validity can be relaxed. The reliability or validity of a single instance of formative assessment is not crucial, as important consequences are not attached.

Systems of assessment. The assessment of students relative to a learning progression and with assessments of various cycle lengths implies a system of assessments, all of which are coherent with the intended learning goals and providing continuous streams of information, at appropriate grain sizes, to fuel teaching and learning. Immediate goals and short cycle assessments build to intermediate and medium cycle assessments at key juncture points, and onward to longer term, end-of-year, and over-time goals presumably assessed by long cycle tools (see also the 3Cs from Pellegrino et al., 2001; Herman, 2010; and NRC, 2005). Figure 3 shows this general progression.

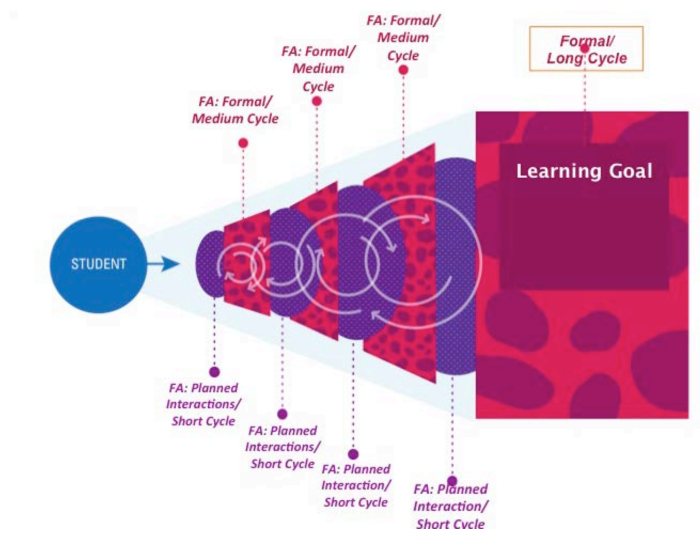

Figure 3. Coherent progression supporting learning.

A recent study of the effectiveness of adding a systematic, coherent assessment system to Full Option Science System (FOSS), a hands-on K-8 science program, provides a case in point (see Ringstaff, Timms, & Tiu, 2013). FOSS features kit-based modules that engage students in a series of investigations to develop their understanding of a given module topic. With funding from the National Science Foundation, researchers and developers from the Lawrence Hall of Science, the Berkeley Evaluation and Assessment Research (BEAR) Center from the University of California Berkeley, and SRI International revised selected units to incorporate a series of embedded and benchmark assessments. The embedded assessments, with associated rubrics keyed to specific lesson goals, generally involved teacher observation of students' inquiry practices during investigation activities, analysis of written work in science notebooks and response sheets, individual and whole-class feedback, and students' self-assessment of their science learning. The embedded assessments were designed to provide teachers and students with continuous information about students' learning so that ongoing instruction could be adjusted as need to help all students acquire essential scientific concepts and processes. The benchmark assessments, completed after each investigation, provided a more summative view of whether all students had mastered the key intended concepts and were also used formatively when students self-assessed and reflected on their learning. An end-of-unit assessment was also included. So the basic design mirrors that of Figure 2:

- Students engage in Part 1 of investigation, including embedded assessments (notebook entries, response sheets, etc.); teacher reviews embedded assessments, provides feedback, and plans next steps.

- Students engage in Part 2 of investigation, including embedded assessment (notebook entries, response sheets, etc.); teacher reviews embedded assessments, provides feedback, and plans next steps.
- Students engage in Part 3 of investigation, including embedded assessment (notebook entries, response sheets, etc.); teacher reviews embedded assessments, provides feedback, and plans next steps.
- Class completes end-of-investigation benchmark and engages in self-assessment and reflection on results.
- Class engages in further instruction on investigation concepts, as needed, and moves on to the next investigation, which continues to process.

The WestEd-CRESST efficacy study, using a randomized control design, found that revised curriculum featuring a comprehensive assessment system, has statistically significant effects on teacher knowledge and student learning (see Timms et al., 2013).

The effects of assessment on learning: A view from cognitive literature. In addition to studies documenting the learning benefits of adding formative assessment to instruction, longstanding research in cognitive psychology also provides relevant evidence supporting the effects of teaching. Research has long demonstrated that giving students a recall test after they have learned something increases the likelihood that they will retain that newly acquired knowledge when tested later (e.g., Bartlett & Tulving, 1974; Donaldson, 1971; Izawa, 1970;). This so-called testing effect was hypothesized as a function of students' opportunity to study and practice prior to the final retention test and might also be a related to the signaling function that we know assessment serves in large scale contexts (see, for example, Hamilton, Stecher & Yuan, 2012; Herman, 2008).

Recent studies, however, have revealed a more complex relationship. Repeated testing boosts the effect (Roediger & Karpicke, 2006), and constructed response tests produce higher learning gains than selected response ones (McDaniel, Roediger, & McDermott, 2007). For example, students who were required to construct answers rather than just select the correct response performed better on the final test, regardless of the response format of the final test (Kang, McDermott, & Roediger, 2007).

Other Related Theory and Research

Other concepts from cognitive psychology provide an additional theory base for the effective design of formative assessment tasks that in themselves can directly benefit student learning. As Shepard (2005) observed in sharing the implications of learning theory for formative assessment, "Contemporary learning theories—including constructivism, cognitive theory and sociocultural theory—share several core principles. Most important are two core ideas: that we construct knowledge, and that learning and development are culturally embedded, socially supported processes" (p. 66).

Design principles. Theory and research, for example, suggest that transfer is supported when teaching and learning actively engage students in constructing and organizing their knowledge, for example, by connecting to students' prior knowledge; by engaging students with multiple and varied representations; by asking students to explain, elaborate, and question their thinking; by using modeling; by teaching with examples and cases; by encouraging metacognition and self-reflection on learning; and by engaging students in challenging tasks while also providing guidance and feedback (see Mayer, 2010; NRC , 2012; Shepard, 2005).

The latter principle highlights the relevance to formative assessment of Vygotsky's theories on the zone of proximal development (ZPD; 1978) and Bruner's early work on scaffolding (Wood, Bruner & Ross, 1976). The ZPD is the distance between what a student can do independently and what that student potentially could do with appropriate guidance from an adult or in collaboration with more capable peers, so that if a student is at the ZPD for a particular task, providing appropriate support will enable the students acquire the learning to be successful on the task. Scaffolding is the support provided during the teaching and learning process, tailored to the individual's needs (and ZPD) and may take the form of such things as modeling, coaching, prompting, key questions, and other forms of feedback (see also Heritage, 2010; Shepard, 2005).

CRESST's POWERSOURCE project, which provided formative assessment tools for middle school mathematics teachers to use in teaching major concepts thought essential for success in Algebra 1, specifically incorporated a number of these features and was specifically designed to foster transfer (see Phelan, Kang, Niemi, Vendlinski, & Choi, 2009). Based in a comprehensive ontology detailing knowledge and skill underlying each major concept, the short, 20-minute assessments used multiple-item formats to diagnose student misconceptions while involving students in partially worked examples, explanations, and applications across a variety of context—practices which are associated with transfer (see, for example, Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Mayer, 2010; Sweller, vanMerrienboer, & Paas, 1998; VanLehn, 1996). Accompanying resources helped teachers to implement subsequent activities—linked to patterns of test performance—to scaffold students' learning. A randomized control study of POWERSOURCE revealed statistically significant, positive effects on student performance on a transfer test (see Phelan et al., 2012). At the same time, however, like the FOSS study reported above, the POWERSOURCE study also revealed limitations in teachers' content and assessment knowledge.

Roles of motivation and metacognition. An in-depth treatment of theories of motivation, metacognition, and self-regulation and their implications for formative assessment is beyond the scope of this paper. However, motivation and self-regulation play a prominent role in the conceptions of prominent formative assessment theorists (see, for example, Black & Wiliam, 2004; Harlen, 2006; Sadler, 1989). The process of learning, as has been continually noted, involves taking students along a pathway from where they currently are to ever greater knowledge and capability and more sophisticated scientific thinking. Inevitably, however, there are challenges and obstacles along the way. If students are to be successful, they must act to remain on course and marshal the intellectual resources and their own behavior to pursue and achieve the goal (Hoyle, 2011). Hattie's feedback model, discussed earlier, highlights this same point. As a number of prominent and theorists have noted, motivation and self-regulation are thoroughly intertwined in students' learning (Pintrich, 1999; Zimmerman, 2002), and these predispositions are at least equally important as cognitive/academic achievement in predicting educational and socioeconomic success (Heckman, 2006, as cited by Hoyle, 2011).

Adaptive problem solving provides an important case in point. It has been identified as a key capacity in most formulations of the skills students need for success in the 21st century (Partnership for 21st Century Skills; NRC, 2012b), yet clearly the ability to solve challenging, nonroutine problems requires self-motivation and self-management in establishing goals and monitoring progress, in addition to transferable knowledge. Similarly, motivation and self-regulation are essential to lifelong learning, another ultimate capacity for student development. Many, in fact, might say that the ultimate goal of education is to produce lifelong learners who are able to adapt successfully to changing requirements and circumstances.

Fostering motivation and self-regulation thus are important goals of formative assessment and require consideration of the sociocultural context in which assessment and learning are conducted. One leading edge of research in this arena is Carol Dweck's conception of *mindsets*, based on her decades of research on motivation (Dweck, 2006). Dweck maintains that how students *think* about themselves and their abilities—their mindsets—as much as their ability and talent—are critical to their success. Students' mindsets reflect their views of themselves as learners, and particularly their theories of the nature of intelligence, and exist on a continuum. At one end of the continuum are those who have an entity or fixed theory of intelligences and view their intelligence as fixed and unchangeable. These students are motivated to prove themselves to others, to look smart, and to avoid performing poorly relative to others. Because they view that circumstances are beyond their control, they may avoid challenging situations and may ultimately stop trying altogether.

In contrast, at the other end of the continuum are those who have an incremental theory of intelligence, who believe that their ability is malleable and can be increased through hard work and effort. Students holding an incremental theory of intelligence—also termed a *growth mindset*—believe that their success is determined by their effort; they embrace challenge and are motivated to seek feedback to overcome obstacles—they are resilient because they have confidence in their ability to improve and are oriented toward improvement. Dweck's theories have strong implications for how goals are framed, high expectations communicated, and feedback provided relative to effort in the learning process, including taking on challenges, learning from mistakes, accepting feedback and criticism, taking on risks, perseverance, and so forth (see for example, Mindset Works Inc., 2012). Dweck's work also shows the impact of environmental cues and thus has implications for the culture of the classroom, its learning orientation, and trust atmosphere. For example, creating a "risk-tolerant learning zone" is one important principle in helping students develop a growth mindset

A Model for Moving Formative Assessment Forward

Research thus shows the value of formative assessment as a process and as tool and points to common design features and elements of quality for both. Further, studies on cognition, motivation, and self-regulation have strong implications for how a formative assessment process is best implemented and for the nature of tools whose use may themselves foster learning. I maintain that effective models for pushing formative assessment forward will integrate these various literatures and will particularly incorporate the elements of both effective processes and effective assessment tools. The model provides an underlying validity argument supporting the effective use of formative assessment for improving learning.

Model: A Single Instance of Assessment

Figure 4 shows, based on the literature reviewed earlier in this paper, the basic structure of an effective learning-based assessment and an effective formative process of assessment for learning. For simplicity's sake, this initial model represents a single instance of assessment, which we build out below to incorporate how assessment operates over time in a system of assessment.

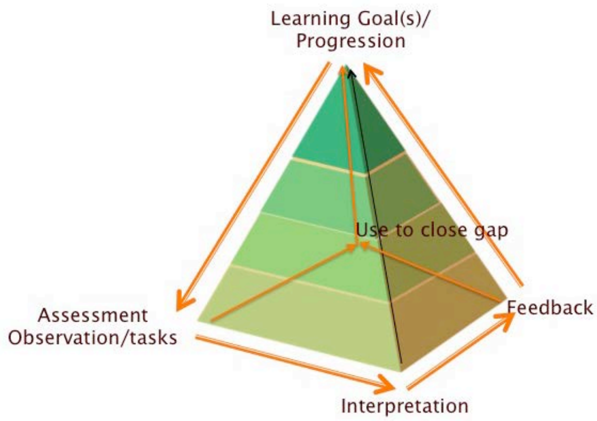

Figure 4. Critical formative assessment components.

Figure 4 essentially starts with the KWSK assessment triangle (Pellegrino et al., 2001)[1] and adds to it the additional elements that are central to assessment tools and processes that can benefit learning. Because we are dealing with a single instance of assessment, Figure 4 depicts a single intended goal or set of goals, with the assumption that the goal(s) lies within a learning progression. The goal is the starting point for considering how to observe student learning relative to the intended goal(s) and the design of tasks and/or activities that simultaneously engage students in and can be used to elicit evidence of their learning relative to the goals. The tasks or activities simultaneously can be designed to themselves promote learning, based on theory and research from cognitive science. Further, an interpretative framework must be designed/applied to student responses or interactions to ascertain where students are relative to the goal and to provide diagnostic information for moving forward. Moving beyond the assessment triangle, if the assessment is to be formative, the relevant interpretation must be transformed by and for teachers and/or students into informative feedback and subsequently acted upon to help students reach the intended goal. And the entire process of assessment and use is embedded in a classroom culture and structures that support—or do not support—learning goals, collaboration, and accountability for learning.

The design principles for the process and for assessment tools and strategies are roughly parallel, although the process orientation clearly emphasizes and leverages interactions and reciprocity that are thought critical to the success of formative assessment. For example, in establishing and clearly communicating goals for student learning and assessment, theory and research would suggest the importance of goals that reflect challenging and high expectations, and in designing assessment tasks or instructional activities to support these goals, research would suggest the value of linking them to students' prior knowledge, incorporating appropriate levels of challenge, engaging students in multiple representations, and asking students to explain their thinking (Dweck, 2006; Mayer, 2010; NRC, 2012b). Further, research would suggest that initial interpretative frameworks be developed in concert and not separate from task or activity design (Baker, 2012; Herman, Aschbacher, & Winters, 1992), coordinated with understanding of how students' knowledge and skill is expected to develop in the goal area (Heritage, 2010). Feedback for a formal test might be a descriptive score report, potentially involving students in self-reflection and providing them with concrete task or process suggestions for next steps (cf. Hattie & Timperley, 2007). In the context of a formative process, however, the feedback process is likely to more interactive, involving teachers, students, and their peers, and more seamlessly connected to next steps in instruction and a continuing formative assessment process. For example, in responding to a student's misconception during a classroom discussion, a teacher might call on other students to propose alternative representations and/or conceptions, might probe or reframe the question to more closely connect it to students' current level of understanding, and/or otherwise engage students in

[1] Note that Figure 4 also is a further adaption of the Center for the Assessment and Evaluation of Student Learning (CAESL) model, developed collaboratively with colleagues at WestEd, UC Berkeley, and Stanford.

instructional conversations to coach them through cycles of scaffolding-response-feedback to enable students to reach a higher level of understanding. Figure 4 is an oversimplification, not only in focusing on a single instance of assessment, but also in its failure to convey the complex nature of the Next Generation Science Standards' (NGSS Lead States, 2013) learning goals. We attempt to deal with these complications next.

A System of Formative Assessment

The learning goals—or performance expectations—of the Next Generation Science Standards (NGSS Lead States, 2013), of course, purposively fuse the three dimensions of science learning: science and engineering practices, cross-cutting concepts, and disciplinary core ideas. Rather than a single learning progression, then, these goals involve the fusion of three developmental trajectories, each of which is expected to yield generalizable and transferable knowledge and skill, both alone and in combination. The three must be artfully interwoven in instruction and ongoing formative assessment to achieve these intended goals.

Figure 5 attempts to portray this complex interplay. It shows three learning progressions, one for each targeted science learning dimension, that start from where students currently are in their understanding and skills and project an effective pathway through which students can progress from simpler to more sophisticated conceptions and applications to reach intended science learning goals—by the end of a learning sequence, be it a unit, a semester, a year, or a period ranging across years. For example, in the context of a given unit of science instruction, where are students expected to start in their understandings of a given topic or idea, where are they expected to be at the end of the unit, and descriptively what will they know and be able to do? What are typical intermediate levels of understanding that likely will mark students' learning progress? What common errors and misconceptions are likely to provide barriers and obstacles for students along the pathway? In what practices should students be engaged to support their deeper content learning, and where are students likely to start and end (for this unit) in their ability to apply to apply the identified practice(s)? What are intermediate points and typical errors that students are likely to face along the way? And finally, how might the unit coordinate with specific crosscutting ideas and where do students start and end on that continuum?

The figure, in short, does not adequately display the sophisticated intertwining of the three dimensions that mark productive pathways, or the complex ways in which various elements within a dimension may interact within or across units. The Framework (NRC, 2012a) and the Next Generation Science Standards (NGSS Lead States, 2013), for example, make clear their intent that students be engaged in science and engineering practices as they gain knowledge and understanding of disciplinary core ideas and in that process make use of crosscutting ideas to make connections across topics. Students should be using and expanding their understanding of core disciplinary ideas and crosscutting concepts as they engage in science and engineering practices. At the same time, there may not be a one-to-one correspondence among the three and over time, the potential combinations within and across units and time may be daunting. It also seems obvious that over a sequence of instruction, at times each of targeted dimensions is best treated individually and in strategic doublets. Indeed, getting a fix on where students are and supporting their learning on each progression may require some targeted teaching, assessment, and diagnosis. The intent, however, is to design progressions on which teaching, learning, and assessment are based that move students purposefully toward increasingly more sophisticated conceptual understanding, evidence-based analysis and argumentation, inquiry skill, and so forth—and to success on the focal science standards.

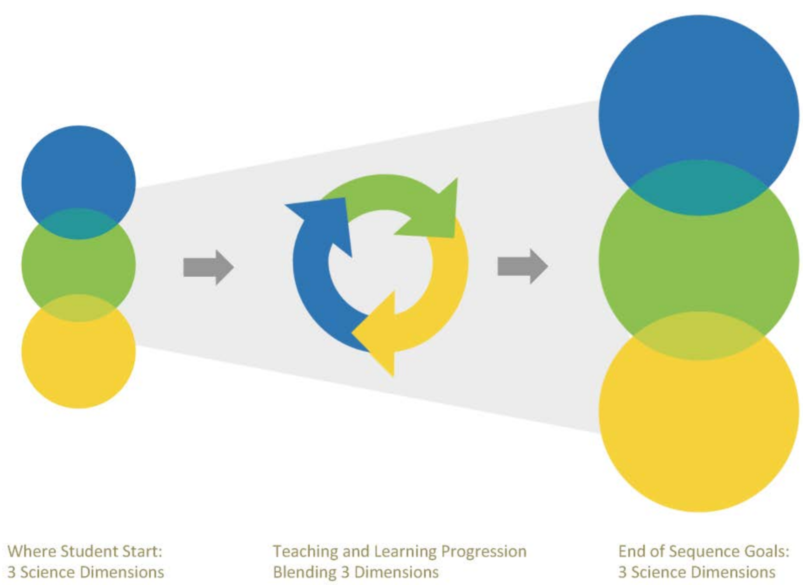

Figure 5. Progressions of fused science knowledge.

Figure 6 displays how the single instance of assessment portrayed in the pyramid can be spiraled into a coherent system of assessment supporting students' success in targeted Next Generation Science Standards and designated performance expectations. As with any good formative assessment process, the system starts from wherever students are relative to the designated three-dimensional science learning goal(s) and focuses teachers and students on the learning pathway and sets of progressively more sophisticated concepts and applications that will support student success.

Figure 6. Coherent progression-based systems.

Consider learning progressions as sets of progressively more sophisticated learning goals (and potential misconceptions and obstacles that may arise) that provide guideposts for instruction and assessment. They enable teachers to be clear on goals and expected learning, to clearly communicate these expectations to students, and to connect what is to be learned to prior learning and knowledge. Based on these goals, as teachers and students engage in instruction, they use short cycle assessments that probe and diagnose student learning on immediate or daily goals, which build to the more sophisticated understandings intended in intermediate goals, which may be the targets of both short and medium cycle assessments, which in turn build to ultimate goals and formal long cycle assessments.

The critical feature is that all of the assessments (and instruction) are aligned with the projected learning progressions so that they can integrally support teaching and learning. The short cycle assessments, for example, may tend to be highly interactive and thoroughly embedded in instruction. Interactions are purposively designed so that the same instructional activities that are intended to support students' learning also yield evidence of students' progress that can be used to fuel students' immediate learning—through feedback, probing, scaffolding, and other immediate instructional moves. Teaching and assessment are merged in a reflective process that helps to move students forward.

Early in a learning cycle, these embedded assessments may ask students to grapple (through oral discourse or written work) with investigations, observations, data, and/or applications that deal with discrete concepts or components underlying a major disciplinary idea and later in the sequence may deal with a more unified or sophisticated view. Teachers may observe, probe, analyze student work to identify gaps or obstacles, provide feedback on and otherwise scaffold student thinking as students discuss applications of a given concept, analyze data, or provide evidence for an emerging theory. At more intermediate goal points, addressing intermediate levels of knowledge and practice, teachers and students may observe, question, probe to understand students' conceptions and use of evidence as students build and justify a larger model or theory, conduct and justify investigations of a larger concept, and/or provide a more nuanced explanation of a given phenomenon. And at key intermediate or advanced juncture points along the progression, more formal assessments, including lab reports, projects, demonstrations, presentations, research syntheses, or inquiries, can provide a more uniform view of what all students have accomplished and what gaps and obstacles may exist for some individuals or subgroups within a class. These too can be occasions for self-reflection, feedback, and peer collaboration and/or adaptive learning activities to move learning forward. And so on and so on in a process of continuous improvement, buoyed by a learning-oriented classroom culture and growth-oriented learners (Dweck, 2006) that support students as they move from beginning through intermediate to the sophisticated levels of knowledge and practice represented by the standards and targeted learning goals—and ultimately to lifelong science learning. The system is coherent; each cycle of assessment, including communicating and linking goals, eliciting evidence, analysis, feedback and action, builds to the next and toward ultimate accomplishment. The goal throughout the progression is to help students and teachers to understand what learning has been accomplished, to identify misunderstandings or misconceptions that may exist, to diagnose areas that require further effort, and to use this evidence to provide informative learning-based feedback and action to close identified gaps.

Concluding Comments

This paper has laid out a model for the design and implementation of formative assessment. As with all models, it is a simplification of the complex processes and elements that need to be orchestrated in effective systems of formative assessment that will support all students attaining Next Generation Science Standards (NGSS Lead States, 2013). The model calls on elements that are yet to be developed on a broad scale:

- Learning progressions to undergird the entire process: Progressions for core ideas and crosscutting concepts are in their infancy, as noted earlier. Those for science and engineering practices are even more immature. And how to weave the three together in effective sequences for teaching and assessment is yet to be discovered. In the short term, most teaching and assessment will need to be built on best guesses of expert teachers, curriculum developers, and subject matter researchers and theorists, and then refined as evidence is accumulated.

- Measurement capacity to assess fused knowledge: The fused knowledge expectation of the science standards also presents a challenge for measurement as well. Certainly, there are examples of assessments that purposefully fuse disciplinary ideas in science or engineering content with practices (see, for example, advanced placement biology redesign; CRESST integrated learning assessment models), and cognitive diagnostic models have been applied to try to disaggregate the relative contributions of each dimension. Clearly separating the two in order to get separate measures of students' understanding of content ideas versus practice has proved challenging, particularly since scientific and practices cannot be enacted in the absence of content knowledge (see, for example, NRC 2012b). Integrating the three dimensions of disciplinary ideas, practices, and crosscutting concepts brings takes the field into even more challenging territory. Needed are tasks that can accurately locate students on multiple learning trajectories and provide diagnostic feedback that can support their forward movement across the three.
- Curriculum embedded formative assessment tools and strategies: Assessment typically has been an afterthought in curriculum development. The Next Generation Science Standards require innovative curriculum materials that are built on thoughtfully interwoven learning progressions that provide a strong foundation for both teaching and assessment. Short, medium, and long cycle formative tools and strategies must be systematically designed and built into new materials.
- Supplementary tools and strategies: As new curriculum materials are systematically developed, supplementary tools and strategies can help schools and the educators and students within them to understand the new expectations of the Next Generation Science Standards. Supplementary materials also can help districts and schools who are lacking resources for a total retooling/purchasing of new curriculum to augment their existing materials. Effective supplementary materials will be educative for teachers and schools and provide models for future development.
- Teacher capacity to engage in formative assessment: Also, as noted earlier, study after study shows limitations in teachers' ability to elicit student thinking and reasoning and to diagnose student understanding (Heritage et al., 2009; Herman et al., 2006). Studies of teachers' feedback show mostly right-wrong responses and little task or process cueing (Bransford, Brown, & Cocking, 1999; Clare, Valdes, & Patthey-Chavez, 2000). Moreover, students suggest that teachers are least agile in prescribing next steps when their initial strategies have left gaps for students (Heritage et al., 2009; Herman, Osmundson, & Silver, 2010). The formative assessment model requires expert teaching and deep science content pedagogical

knowledge. Quality tools may help to both bridge the gap and support teachers' capacity (see Dai, Herman, Osmundson, & Chai, 2013).

- Classroom culture: The pressures of current accountability mandates encourage a performance- rather than learning-oriented culture (Assessment Reform Group, 2006). Moreover, the relentless focus on academic achievement gives short shrift to student motivation and self-regulation processes that ultimately are necessary to support learning and resilience.

These challenges suggest an ambitious research and development agenda to support the implementation of the Next Generation Science Standards. Luckily, productive prototypes already exist. These prototypes provide powerful exemplars for building, refining, and using learning progressions to support rich science learning; for creating curriculum materials that intertwine multiple content and process progressions and coordinate multiple layers of assessment; and for using technology to incorporate the power of simulation and games to support coherent formative assessment systems. The presentations that follow provide such exemplars, and each as well provides a case study of its own continuous use of assessment for improvement.

References

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box.* Cambridge, England: University of Cambridge School of Education. Retrieved from http://www.nuffieldfoundation.org/sites/default/files/files/beyond_blackbox.pdf

Assessment Reform Group. (2002). *Assessment for learning: 10 Principles.* Cambridge, England: University of Cambridge School of Education. Retrieved from http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf

Assessment Reform Group. (2006). *Testing, motivation and learning.* Cambridge, England: University of Cambridge School of Education. Retrieved from http://arrts.gtcni.org.uk/gtcni/bitstream/2428/4624/1/Testing%2c%20Motivation%20and%20Learning.pdf

Baker, E. L. (2012). *Ontology-based educational design: Seeing is believing* (Resource Paper No. 13). Los Angeles, CA: CRESST.

Baker, E. L., & Popham, W. J. (1973). *Expanding dimensions of instructional objectives*. Upper Saddle River, NJ: Prentice-Hall.

Bartlett, J. C., & Tulving, E. (1974). Effects of temporal and semantic encoding in immediate recall upon subsequent retrieval. *Journal of Verbal Learning and Verbal Behavior, 13*, 297–309.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, *8*5, 536–553.

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*(1), 7–74.

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability* (pp. 20–50). Chicago, IL: National Society for the Study of Education.

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

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

Clare, L., Valdes, R., & Patthey-Chavez, L. (2000). *Learning to write in urban elementary and middle schools: An investigation of teachers' written feedback on student compositions* (CSE Technical Report No. 526). Los Angeles, CA: CRESST.

Corcoran, T., Mosher, F. A., & Ragat, A. (2009). *Learning progressions in science: An evidence-based approach to reform*. Philadelphia, PA: CPRE.

Dai, Y., Herman, J., Osmundson, E., & Chai, Y. (2013, April). *Multi-level analysis of an embedded assessment system: Methodological issues.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

DiRanna, K., Osmundson, E., Topps, J., Barakos, L., Gearhart, M., Cerwin, K.… Strang, C. (2008). *Assessment-centered teaching*. Thousand Oaks, CA: Corwin Press.

Donaldson, W. (1971). Output effects in multitrial free recall. *Journal of Verbal Learning & Verbal Behavior, 10*, 577–585.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in Grades K-8*. Washington, DC: National Academies Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.

Formative Assessment for Students and Teachers. (2008). *Attributes of effective formative assessment.* Washington, DC: Council of Chief State School Officers.

Forster, M., & Masters, G. (2004). Bridging the conceptual gap between classroom assessment and system accountability. In M. Wilson (Ed.), *Toward coherence between classroom assessment and accountability*: *103 Yearbook of the National Society for the Study of Education, Part 2* (pp. 51–73). Chicago, IL: NSSE.

Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P., Shavelson, R. J. & Yin, Y. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education, 21*(4), 360–389.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, *18*, 519–521.

Hamilton, L., Stecher, B., & Yuan, K. (2012). Standards-based accountability in the United States: Lessons learned and future directions. *Education Inquiry*, *3*, 149–170.

Harlen, W. (2006). The role of assessment in developing motivation for learning. In J. Gardner (Ed.), *Assessment and learning*, (pp. 61–80). London, England: Sage Publications.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. London, England: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment.* Washington, DC: CCSSO.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin Press.

Heritage, M., Kim, J., Vendlinksi, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28*(3), 24–31.

Herman, J. L. (2008). Accountability and assessment in the service of learning: Is public interest in K-12 education being served? In L. Shepard & K. Ryan (Eds.), *The future of test-based accountability*. New York, NY: Taylor and Francis.

Herman, J. L. (2010) *Coherence: Key to next generation assessment success* (CRESST Policy Brief). Los Angeles, CA: CRESST.

Herman, J., Aschbacher, P., & Winters, L. (1991). *A practical guide to alternative assessment*. Thousand Oaks, CA: Corwin Press

Herman, J. L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). *The nature and impact of teachers' formative assessment practices* (CSE Technical Report No. 703). Los Angeles, CA: CRESST.

Herman, J., Osmundson, E., & Silver, D. (2010). *Capturing quality in formative assessment practice: Measurement challenges* (CSE Report No. 770). Los Angeles, CA: CRESST.

Hoyle, R. (2011, January). *Assessment of self-regulation and related constructs.* Paper presented at NRC Workshop on Assessing 21st Century Skills, Irvine, CA.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, *83*, 340–344.

Kang, S. H., McDermott, K., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long term retention. *European Journal of Cognitive Psychology, 19*, 528–558.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*, 28–37.

Lehrer, R., Wilson, M., Ayers, E., & Kim, M. J. (2011, September). *Assessing data modeling and statistical reasoning.* Paper presented at the fall conference of the Society for Research on Educational Effectiveness (SREE), Washington, DC.

Mayer, R. E. (2010). *Applying the science of learning*. Upper Saddle River, NJ: Pearson

McDaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14,* 200–206.

Mindset Works Inc. (2012). *Mindset Works*. Retrieved from http://www.mindsetworks.com

Mislevy, R. J., Almond, R., & Lukas, J. (2003). *A brief introduction to evidence-centered design* (ETS Research Report RR-03-16). Princeton, NJ: Educational Testing Service.

Pellegrino, J, Chudowsky, N., & Glaser, R.(Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

National Research Council. (2005). *Systems for state science assessment*. Washington, DC: National Academies Press.

National Research Council. (2007). *Taking science to school*. Washington, DC: National Academies Press.

National Research Council. (2012a). *A framework for K-12 science education*. Washington, DC: National Academies Press.

National Research Council. (2012b). *Education for life and work*. Washington, DC: National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academies Press.

Pellegrino, J, Chudowsky, N., & Glaser, R. (2001). *Knowing what students know*. Washington, DC: National Academies Press.

Pellegrino, J. W., & Hilton, M. L. (2012a). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: National Academies Press.

Phelan, J., Choi, K., Niemi, D., Vendlinski, T., Baker, E., & Herman, J. (2012). The effects of POWERSOURCE assessments on middle-school students' math performance. *Assessment in Education, 19*(2), 211–230.

Phelan, J., Kang, T., Niemi, D., Vendlinski, T., & Choi, K. (2009). *Some aspects of the technical quality of formative assessment in middle school mathematics* (CRESST Report No. 750). Los Angeles, CA: CRESST.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*, 459–470.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Ringstaff, C., Timms, M., & Tiu, M. (2013, April). Investigating the impact of embedded assessment in elementary science. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119–144.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., & Ruiz-Primo, M. A. (2008). On the impact of curriculum-embedded formative assessment on learning. *Applied Measurement in Education, 21*, 295–314.

Shavelson, R. J., Yin, Y., Furtak, E. M., Ruiz-Primo, M. A., Ayala, C. C., Young, D. B., … Pottenger, F. III. (2009). On the role and impact of formative assessment on science inquiry teaching and learning. In R. Coffey, R. Douglas, & C. Sterns (Eds.), *Assessing science learning: Perspectives from research and practice* (pp. 21–36). Washington, DC: NSTA Press.

Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership, 63*(3), 66–70.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progressions for matter and atomic-molecular theory. *Measurement: Interdisciplinary Research and Perspectives*, *14*(1-2), 1–98.

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251–296.

Thorndike, E. L. (1918). The nature, purposes and general methods of measurements of educational products. In G. M. Whipple (Ed.), *The measurement of educational products : Seventeenth yearbook of NSSE FOSS* (pp. 16–24). Bloomington, IN: Public School Publishing Co.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.

VanLehn, K. (1996). Cognitive skill acquisition. In J. Spence, J. Darly, & D. J. Foss (Eds.), *Annual review of psychology* (Vol. 47, pp. 513–539). Palo Alto, CA: Annual Reviews.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. Dwyer (Ed.), *The future of assessment* (pp. 53–82). Mahwah, N.J: Lawrence Erlbaum.

Wilson, M. R., & Bertenthal, M. W. (2005). *Systems for state science assessment*. Washington, DC: National Academies Press.

Wilson, M., & Sloane, K. (2001). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, *13*(2), 181–208.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, *17*(2), 89–100.

Zimmerman, B. (2002). Becoming a self regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.

The Center for K–12 Assessment & Performance Management at ETS creates timely events where conversations regarding new assessment challenges can take place, and publishes and disseminates the best thinking and research on the range of measurement issues facing national, state and local decision makers.

K-12 Center at ETS

Invitational Research Symposium on
Science Assessment