Building on Learner Thinking: A Framework for Assessment in Instruction
(Commissioned paper for the Committee on Highly Successful STEM Schools or Programs for K-12 STEM Education: Workshop, May 10/11, 2011)

Jim Minstrell and Ruth Anderson, FACET Innovations
Min Li, University of Washington

Classroom assessment includes all actions taken by a teacher for the purpose of gathering information about student learning. Classroom assessment becomes *formative* in nature when the information is used to adjust instruction and provide students with information to advance their learning. The literature has established abundant empirical evidence on the positive impacts of formative assessment with regard to student achievement and motivation in various subject domains by shaping the design of instruction (e.g., Black, 2000; Black & Wiliam, 1998; Brookhart, 2004; Crooks, 1988; Herman & Heritage, 2007; Kanjee, 2000; Pryor & Torrance, 1997; Shepard, 2005; Shepard et al., 2005; Stiggins, 1988, 2002; Torrance & Pryor, 1998, 2001; Wiliam, 2005). And while findings from most empirical studies are moderate at best, effective implementation of formative assessment has also been found to have a more significant impact on student performance than other powerful instructional interventions, including one-on-one tutoring (Leahy, Lyon, Thompson, & Wiliam, 2005; Shepard, 2005).

As educators across the U.S. and abroad continue to pursue the publicized positive outcomes of formative assessment, disagreements persist in the research community as to its conceptualization and specific tools and strategies considered critical to its effective implementation. In this paper, we view formative assessment as a process rather than an instrument, which is consistent with most definitions (Black & Wiliam, 2009; CCSSO, 2008; Nichols et al., 2009; Popham, 2006; Shepard, 2009; Wiliam & Thompson, 2008). For example, CCSSO (2008) defines formative assessment as “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” (p. 3). We are aware that the time frame of such practice can vary from a teacher interacting with her students and making minute-to-minute decisions, to teachers using benchmark assessments to adjust the subsequent lessons, to teachers collaboratively analyzing their students’ science fair projects and revising the curricular materials for next year.

There is also some dispute regarding the strength of the research findings of the impact of formative assessment on student learning. Much of the research remains descriptive in areas other than literacy (i.e., the original focus of formative assessment research). Researchers disagree regarding the strength of empirical findings as they question some studies’ sample size and selection (Bennett, 2010; Dunn & Mulvenon, 2009), measures of learning and even the definition of formative assessment in some of the earlier studies completed before the turn of the last century.

While researchers puzzle at the mixed empirical findings, they also acknowledge that effective formative assessment rarely occurs in classrooms, regardless of how it is specifically characterized within the study. Myhill and Brackley (2004) found that teachers rarely probed students’ prior knowledge. Teacher interviews and video analysis confirmed that teachers only considered prior knowledge as “simply about children’s prior knowledge of facts, or children’s prior social and cultural experiences” (Myhill & Brackley, 2004, p. 271) instead of seeing the cognitive and conceptual connections that they could build upon. In a survey study of high school teachers, Noonan and Duncan (2005) report that high school science and mathematics teachers are less likely to implement self or peer assessments than their social studies and language art colleagues.

Despite the growing availability of formative assessment tools and an increasing number of teachers who are engaging in formative assessment activities (e.g., “making student thinking visible,” looking for learner misconceptions, monitoring progress through probes, science notebooks and discussions, etc.), we are not typically seeing the kind of student effects that research has promised in STEM education (Herman et al., 2005; Shavelson et al., 2008; Yin et al., 2008). Instructional decisions continue to be based largely on what “should” come next according to the curriculum, anticipated learning progressions or the teacher’s lesson plan, rather than on what actually emerges from student responses to assessment (Hall & Burke, 2003; Ofsted, 1998; Torrance & Pryor, 2001).

Part of the problem, as Black and Wiliam (1998) have pointed out, is underdeveloped teacher knowledge and skills related to formative assessment and little understanding of how to make implementation feasible in real classrooms. Others emphasize persistent “misconceptions” on the part of educators, such as considering formative assessment as an “add on” to instruction rather than an ongoing process of making sense and using of data or as a “low stakes” test or a particular kind of assessment (e.g., pre-test or probe, etc.) (e.g., Darling-Hammond, 2010; Hattie & Timperley, 2007; Heritage, 2010; Stiggins & Conklin, 1992).

Our own research in K-12 science classrooms confirms these problematic issues, and has identified two dominant instructional stances assumed by teachers who engage in a formative assessment cycle that seem to be correlated with the effectiveness of classroom practice of formative assessment (Minstrell, Li, & Anderson, 2009). The less effective but typical stance is primarily teacher or curriculum-driven, focused on how much students have learned (i.e., “got it” or “didn’t get it”). The other, more effective, stance appears to be learner- and learning-driven, focused on not just what, but also *how* students are learning. While teachers of both stances may appear to enact similar classroom actions, the thinking behind those actions is quite distinct (see Figure 1 below), which can ultimately yield vastly different impacts on student learning.

Figure 1: Two Enactments of a Formative Assessment Cycle

Teacher and teaching focus (less effective and most typical)	**Learning and learner focus (more effective)**
Gather data - How much have my students learned of what I have taught?	**Collect data intentionally -** What and how are my students learning in relation to the learning goal?
Evaluate - How many “got it”? Did enough of them get it so I can move on or do I need to slow down?	**Interpret -** What are the strengths and problematic aspects of their thinking? What experience or particular cognition do they need next to deepen their learning?
React - Do I re-teach to the entire class or assign a review to a few? How can I teach more effectively next time?	**Act intentionally -** What specific learning experience or feedback will address the learning needs I just identified?

More effective implementers of formative assessment tend to collect assessment information that is well aligned with a specific learning goal. Student responses are interpreted for strengths to be built upon and problematic aspects to be addressed rather than a simple identification of right and wrong. In the most sophisticated examples, we find teachers are able to identify the cognitive or experiential need suggested by the problematic responses (e.g., the need for students to test their hypothesis or differentiate between two or more related ideas). These teachers are consequently better able to identify subsequent instructional actions tightly linked to those identified learning needs, rather than simply topically related ones.

In less effective formative assessment practice, teachers tend to gather information on how much their students have learned (typically declarative knowledge stated as fact) or simply on the extent to which students have completed the activity. These teachers then evaluate the extent to which the students got all the information correct. Subsequent instructional actions by these teachers often tend to be associated with pacing (e.g., spend more time on topic or move on) or long-term adjustments rather than being tightly related to specific learning needs uncovered through assessment.

The two enactments of formative assessment in Figure 1 not only highlight distinct conceptions of formative assessment (i.e., periodically "checking in" vs. systematically honing in on more specific learning needs), but also suggest underlying fundamental differences in teacher held beliefs on the nature of teaching and learning. The "typical" enactment, for example, suggests an accretion model (i.e., more teaching or re-teaching should result in more facts learned). The "effective" enactment, on the other hand, implies learning involves a process of constructing and reconstructing knowledge in new contexts or to address new learning situations.

At the heart of this effective formative assessment lies the need to access and build upon student thinking as it develops from naïve to more sophisticated. This is particularly vital in the realm of conceptual learning as in STEM where feedback to students and next steps in instruction cannot rely upon exemplars and repeated practice as can be done more readily in teaching skills and procedures. Research has shown that students need learning experiences as interactions with phenomena and ideas to test and revise their own initial or developing ideas so that they can eventually arrive at those goal science ideas themselves: "…Only by keeping a very close eye on emerging learning through formative assessment can teachers be prospective, determining what is within the students' reach, and providing them experiences to support and extend learning through which students can then incorporate new learning into their developing schema" (Heritage, 2010, p. 8).

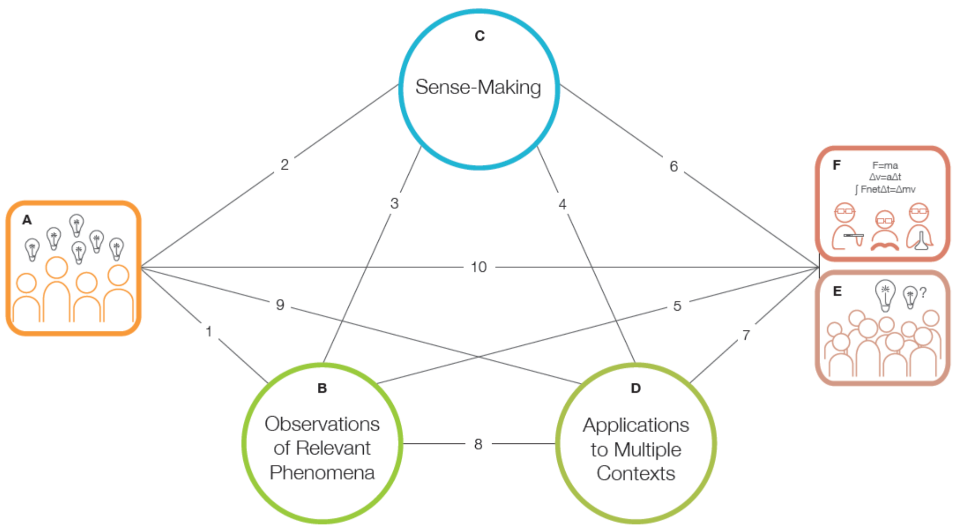
How can teachers be helped to integrate their (recently acquired) knowledge and skills of formative assessment into a classroom practice that effectively follows, monitors, and supports students' developing ideas? We conceptualize the assessment and instruction as one entity—not within a measurement paradigm, but within a larger paradigm of learning as some researchers have suggested (Heritage, 2010; Shepard, 2005). We believe this may be an important first step in addressing the gaps in perspectives and practices of formative assessments among teachers and researchers alike.

From Formative Assessment to Building on Learner Thinking

In this section we present a framework for Building on Learner Thinking (BOLT) in science, which aims at broadening our perspective on formative assessment. BOLT, represented in Figure 2 below, attempts to re-frame the familiar principles of *How People Learn* in a way that models learning and prioritizes the processes of *coming to know* and embed assessment within the learning and teaching cycle. It is a convergent framework, comprised of several research-based components that have individually or collectively been shown to positively impact student learning in STEM classrooms. BOLT is not tied to any particular curriculum or teaching strategy. Rather, it is intended to guide teachers to select, modify or order existing curricular materials to create and facilitate learning experiences that coherently foster, support and monitor the development of student ideas toward goal science or scientists' ideas.

The lettered boxes in the diagram represent ideas: Students' Ideas, Class Consensus, and Scientists' Ideas. The lettered circles represent three typical learning experiences for learners: Observations of Relevant Phenomena, Sense-Making, and Application to Multiple Contexts. The numbered segments joining pairs of components represent ongoing connections to be drawn, by students and their teachers, between the component sources of information throughout the learning process. The numbers and letters associated with the shapes and segments in this figure are for reference and are ***not* meant to imply a particular order**. As in the case of scientific investigation, the sequence of "steps" may vary according to the learning situation.

Figure 2: BOLT Framework



Descriptions of Framework Components

Box A on the left represents the learners' ideas. This core component of BOLT is the very place where instruction that builds on learner thinking typically starts. Anticipating or knowing the common useful and problematic ideas that students exhibit related to the content allows the teacher to actively listen and watch for those ideas to come up in class activities. The valid and useful ideas can be built upon and the problematic ideas can be addressed by exploring contexts that lead to cognitive disequilibrium (Posner, Strike, Hewson, & Gertzog, 1982). If these problematic preconceptions are not addressed in instruction, they are likely to persist (Bransford et al., 2000).

Box F on the right represents scientists' ideas. Together with the operationalized version in Box E as student consensus, these represent the learning goal(s). While these are important for the planning stage of instruction, instruction would rarely (if ever) begin here. Ideas here refer not only to conceptual ideas, such as theories, explanatory models, or principles but also the process skills through which scientists come to know. Recognizing that there are enormous, worthwhile ideas, we stress the need for limiting the main learning goals to a core set of science content so that teachers can prioritize, organize, and order them into productive learning activities for learners to construct a deep understanding of these goal ideas (Michaels et al., 2008).

Box E next to Box F represents ideas that are a consensus product of the activities of the class; in other words, a shared understanding by the group of learners regarding what they know and the supporting experiences and rationale to conclude these ideas. The consensus requires engaging students in the same intellectual processes that scientific understanding has been accumulated through the collaboration among scientists. Ideas from consensus are then "owned" by the class since they were built from the experiences and initial ideas of individuals and arrived at by collaboratively generating, testing and evaluating ideas and making collective sense of a common set of experiences. While these class derived ideas are similar to scientists' ideas, the language and form are often more operational and accessible to the learners so that they can make use of and draw connections to their own ideas (Lemke, 1990).

Circle B represents the students' experiences with phenomena in the form of observations and measurements. It can appear as the scenario or set of tasks, experiments, problems from which learners can easily identify entry points to bring their prior thinking and will know or obtain the observable "facts." It is the real world data that students need to interpret and explain rather than de-contextualized examples (Osborne, Duschl, & Fairbrother, 2002). For example, it may consist of a probe of students' initial thinking along with an experiment in which students can test and evaluate different ideas as alternative hypotheses.

Circle C represents the sense-making experience for the learners. Simply doing hands-on activities, recording the observation, or noticing what happens is not sufficient. Learners need to mentally process the observations to create their inferences, make meaning, and determine implications (e.g., NRC, 1996, 2001; Duschl, Schweingruber, & Shouse, 2007). This process of sense-making involves constructing explanations based on the collected data (Duschl, Schweingruber, & Shouse, 2007; Sandoval, 2003; Sandoval, & Reiser, 2004), organizing information to form an argument and test new understanding (Bell & Linn, 2000; Linn & Hsi, 2000; Michaels, Shouse, and Schweingruber, 2008), as well as model-based reasoning (Lehrer & Schauble, 2006).

Circle D represents the multiple other contexts and representations that promote learners to generalize and transfer the ideas they produced through the learning experiences. Transfer can be enhanced by modifying features of problem situations that involve potential use of the same concept and helping learners look for and see similarities that cue relevant principles (Wagner, 2010).

Curricular activities and classroom discourse guide the connections between components in BOLT. The numbered connecting segments in the framework represent instructional moves to deliberately initiate or develop these connections. In high functioning BOLT, these are typically questions by the teacher or suggested by the curriculum guide, but could also be by learners. For example, **1** and **2** ask the learners for their observations (B) and inferences (C) in relation to their initial ideas (A). **3** and **4** ask learners to make a connection between their observations of B and D and their inferences (C) about the meaning, interpretations, generalizations, or explanations for the observed phenomena. They also ask for "how do you know" or "why do you believe" those conclusions. **8** might represent a question about the similarities and differences between the observations of B and D in order to promote generalization and application of ideas. **5, 6, 7** and **10** represent questions probing learners to reflect on where they started and how they came through the experiences and thinking in B, C, and D to the arrive at the consensus E (diSessa and Minstrell, 19xx; Rivet and Krajcik, 2008).

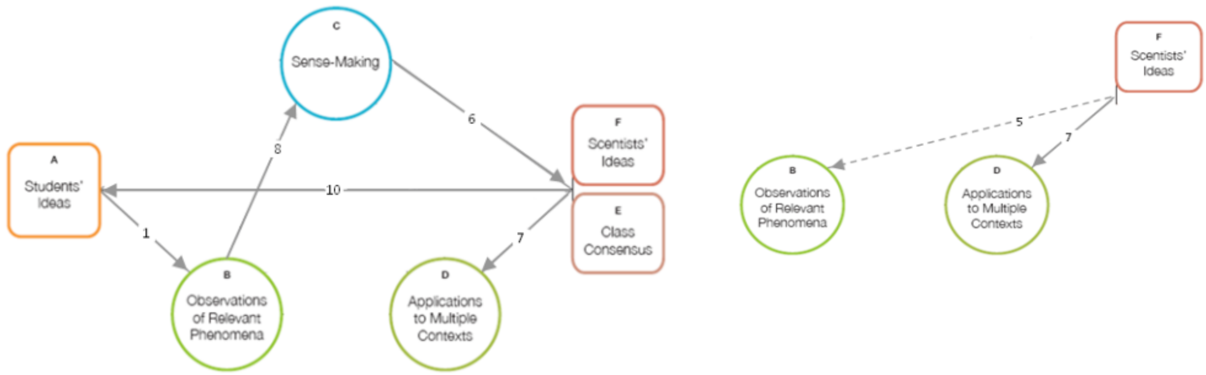
What Might Tracking Instruction through BOLT Look Like?

Stronger Example of BOLT- Day 1: The teacher may begin by asking students to draw and write their responses to two situations: First, the teacher pushes a book across the table with a constant speed and secondly, he pushes the book in such a way that it accelerates more or less at a constant rate as he pushes it along the length of the table. On a handout there is a first set of five drawings of the book equally spaced (representing constant speed), and a second set of images of the book where the spacing between adjacent images gets larger uniformly (constant acceleration). “On each image draw and label the forces acting on each book at that time.” For the first set of images typically 90+% of the class draws images that suggest there is a constant, net unbalanced force in the direction of the uniform speed. For the accelerating object the images typically show a net force that is constantly increasing as the object goes across the table. After the students have finished their drawings, the teacher conducts a class discussion around the student responses without evaluating the responses. Two popular hypotheses emerge very quickly: To explain constant speed, you need to have a constant extra push (net force), and to explain constant acceleration you need a constantly increasing net force (A). These become the hypotheses that motivate the students to find out, if they are correct. The teacher directs them to conduct an experiment keeping the net force constant, each group using an identical cart (same mass) but with each group getting a different system to pull the cart and so different amounts of net force acting on the different carts (B).

Day 2: Continues with analyzing the results of the experiment. Part way through, a couple of students go to the teacher saying that their equipment is not working (i.e., not coming out the way they expected), can they use a different set of equipment. With new equipment they are getting the same result, inconsistent with their initial idea. In a subsequent large group discussion, various groups share their results—the cart sped up but the force scale reading stayed more or less constant. The class ends up concluding that their original idea expressed in the hypothesis did not work and that a *constant extra force will produce acceleration*. The discussion goes on with sharing results of different amounts of net force acting on the equal mass carts. The class concludes that the smaller the constant net force, the smaller the rate of acceleration. Someone asks what would happen if we had zero net force after we got it started. After some discussion of the resulting measurements that show *the smaller the net force, the smaller the acceleration* the class generally comes to the logical conclusion that *no net force would yield no acceleration*. These two new conclusions become the new tentative class consensus (E). Homework encourages reflection on how our understanding has changed (1, 2, 10), how we know our consensus ideas make sense (3, 5, 6), practice and extension (4, 7, 8) to move toward transfer of ideas (D). Problems are also assigned to practice and extend the contexts and representation for the consensus ideas (D).

Figure 3. Stronger and Weaker Examples of Instruction with Respect to BOLT

Stronger Example | Weaker Example



Weaker Example of BOLT- Day 1: The teacher engages student interest by sharing some of the accomplishments of Newton's work and some about his quirky personality. He hands out a copy of Newton's Second Law from the *Principia* by Newton and justifies the importance of knowing this law in understanding and applying physics to explaining all kinds of situations from movement of planets to movement of sub-atomic particles. He then expresses the Law with the equation F_{net}=ma and explicates the parts of the equation with examples of what happens to the third quantity in the equation if one quantity is doubled or tripled while a second quantity is kept constant. He finishes his presentation by giving example problems of simple situations involving force and various accelerated real objects (D). For homework, he assigns more similar problems to practice the ideas introduced during class (D). The teacher also hands out a sheet with the experimental procedures for the next day's lab activity and asks the students to be prepared to do the experiment the next day.

Day 2- Students begin the class by asking questions regarding the procedures in preparation for conducting the laboratory activity, measuring the mass of objects with three different masses and the force (three different pulling forces), and measuring the resulting acceleration (B). During the lab, the teacher moves from table to table monitoring how far along the students have progressed. The students' homework for that evening asks them to verify that the results of their experiment are consistent with F_{net}=ma (B and F). Additionally they are asked to complete some end-of-chapter problems (D).

Note that in the weaker example there is virtually no relation to students' ideas (A) or to sense-making (C) but only an attempt to connect results of the experiment (B) to the equation representation of scientists' ideas (F). In contrast the stronger example elicits learners' ideas (A) which become the hypotheses to test in the activity (B) and having students make sense (C) of the observations in B as they relate to students' initial ideas (A). Finally consensus is reached (E) and related back to A, B, and C.

The comparison of these two examples exemplifies the need for BOLT to initiate instruction from the left side of the diagram (student ideas, interacting with phenomena, and justifying ideas or explaining situations) and proceed to the right (reaching consensus and scientists' ideas), rather than a "top down" movement that begins with science goal ideas and moves to confirmation of those ideas. Moreover the stronger example highlights the potential within

BOLT to forge assessment and instruction into one seamless learning system. As instruction follows the development of learners' thinking, each learning experience naturally becomes an assessment opportunity for teachers and/or students, an experience that next learning can be built upon.

Bringing How People Learn to Instruction

BOLT is essentially the instantiation of the three key findings of *How People Learn* and related principles for the design, implementation and evaluation of learning environments. The four interrelated attributes of learning environments summarized in *How People Learn* are re-framed in BOLT to emphasize the learning-driven nature of the framework.

- Assessment-centered: Emphasis on diagnostic formative assessment that continually identifies strengths of student ideas to build on and problematic aspects to address in the next steps of learning and instruction for learners and teachers.
- Community-centered: Emphasis on learners collectively coming to consensus regarding what is known as an operationalized version of scientists' ideas or goal science ideas. Students build their understanding and come to own their ideas with and within a community of learners. Students have an explicit and active role in establishing a classroom culture conducive to building understanding both individually and collectively.
- Knowledge-centered: Emphasis on how—the process by which—we come to know, not just what we know. Students come to own the science ideas as they constantly test initial and developing ideas against observable phenomena and against ideas and arguments suggested by other learners.
- Learner-centered: Emphasis on constant attention to building on learner thinking. Instruction actively values and uses students' initial and developing ideas made visible by assessments. The student's role in the learning and assessment processes is made explicit and they are supported in developing their capacity.

Blending Assessment and Learning

Formative assessment is subsumed within the BOLT Framework and therein takes on a very diagnostic nature—going for depth of understanding by integrating cycles of eliciting student thinking, interpreting learner responses, and contingent actions based on identified needs (Bell & Cowie, 1998). This cycle is crucial in the realm of conceptual learning. Formative assessment research emerged from the field of literacy, focused on developing skills of reading and writing. Within that context, mechanical solutions and targeted feedback or use of exemplars are reasonable ways to address students' learning difficulties. That approach does not hold true for conceptual learning. Viewing a scientific model may not fundamentally affect the problematic models that learners may hold in their heads; in the process of constructing conceptual understanding, students may not be able to identify their own learning needs because frequently "you don't know what you don't know."

The success of BOLT oriented instruction necessitates the need for constant following and building on learner thinking. Each instructional move in the BOLT system needs to take account of the learners' thinking, such as initial ideas, different interpretations of the data, so that the instruction can be effectively designed and/or modified to address students' learning needs and move toward the desirable learning goals. For example, students' initial thinking is made visible through elicitation activities and discourse during which teachers and students can test ideas and seek alternatives; on the fly assessment happens in the dialogic connections between the components of BOLT when students clarify their observations, their inferences and their sense-making arguments. These assessments offer to teachers and students the needed information to monitor and make corrections to instruction. This sort of assessment practice is perfectly aligned with the essential nature of formative assessment: the assessment data has to be interpreted and taken up by the teacher and students to shape the process of learning (e.g., Bell & Cowie, 2008; Brookhart, 2004; Ramaprasad, 1983; Salder, 1998).

BOLT assessment also assumes high quality tools and strategies that make student thinking visible and track-able. One such approach is diagnostic assessment which includes carefully crafted tasks that can efficiently activate students' problematic thinking. For example, Diagnoser.com provides elicitation questions for opening up students' initial ideas relevant to the unit (Minstrell, Anderson, Kraus, and Minstrell, 2008). Often these are divergent questions to "get the lay of the land" for the class with respect to student thinking. This gives the teacher information from which to make instructional decisions about the class. The tools also include sets of questions which are more convergent around the consensus and scientists' ideas. Since the sets are delivered online, and since each option in multiple choice and numerical items is coded with a particular facet of thinking, the system diagnoses the response and gives the learner feedback to promote further learning. The teacher also gets a report from which to make instructional decisions for the class or for individual students. Suggestions for prescriptive lessons allow the teacher to address identified needs.

BOLT and Classroom Culture

Actualizing the BOLT framework requires teachers and students to establish and maintain together a "culture of learning" in the classroom. This means cultivating classroom norms, relationships, expectations and actions conducive to students openly expressing their conceptual thinking as it unfolds, challenging emerging ideas and engaging in reflection and revision of those ideas. Cultivating the right classroom culture needs to be an explicit and ongoing agenda (something we are mostly accustomed to at the elementary level) for the classroom. And, as the term "culture" implies, it necessarily requires the collaboration of both students and teacher. Students, like their teachers, develop firm ideas and expectations of what constitutes learning and teaching. Both will need to take on new roles and interact differently. Two key areas of shared responsibility that emerge for us are classroom discourse and knowledge of learners and learning.

<u>Classroom Discourse</u>. One particularly challenging area is discourse--a key element in the BOLT components of sense-making and consensus. Discourse is a vital component of science and scientific communities, but rarely a centerpiece of science classrooms (Michaels, 2007). Classroom "discussions" often take the form of an I-R-E progression (teacher **I**nitiates a question, student gives a **R**esponse and teacher **E**valuates the response) that leads toward a conclusion that the teacher draws for the class based on a series of "correct" responses from individual students. On the other hand, a format that prompts students to make claims, provide evidence or offer new interpretations or rebuttals can help students to build on scientific thinking, to become aware of discrepancies between their own ideas and those of others (including the scientific community) and to improve logical reasoning skills. This kind of discourse requires active and deep engagement on the part of teacher and students as they drill down to deeper conceptual understanding. Most teachers and students, however, have little

experience with such group discussions and find it challenging to maintain a group conversation that values a broad range of thinking, while respectfully challenging the science ideas.

Knowledge of learners and learning. Teachers need to be aware of the prior knowledge, experience, lives, and cultures of their learners, such as what cultural intellectual resources students draw upon to attack learning activities, what motivates students, or how students interpret assessment tasks. This knowledge of students is directly associated with how students perceive assessments and how they interpret and use assessment information or teacher feedback (Haertel et al., 2005; Sadler, 1998; Solano-Flores & Li, 2005). Teachers also need to have a strong understanding of both anticipated learning progressions (i.e., the "logical" sequence of learning that builds toward a specific knowledge goal) and *actual* learning progressions (i.e., the many varied paths or sequences of learning that students forge during the learning process). The former may be easily learned through studying the standards or good curricula. The latter usually requires significant and extended interaction with students' thinking, through written work and dialogue, in order to become familiar with what is more and less productive or problematic among learners' ideas.

By the same token, students need to be better aware of their own learning process (metacognition) in order to act on feedback in ways that move their learning forward (Sadler, 1998; Black & Harrison, 2001). There is also growing evidence that raising student awareness of how people learn and of basic neuroscience (e.g., plasticity of the brain and the non-fixed nature of intelligence) can increase student engagement in the learning process and have a significant impact on learning outcomes (Blackwell, Trzesniewski, & Dweck, 2007).

In the previous sections we have alluded to some of the challenges associated with building on learner thinking. It is extremely challenging, but the rewards are worthwhile. As one high school physics teacher describes it:

> "You don't have total control over the agenda, but you are ultimately in charge. You've got to monitor where you are going, but constantly taking feedback from students who are directing you and...trying to identify where they are coming from. What is their conception at this moment? Why would they ask this question? Why did they make that statement? And then judge your next question to solve their problem… It's a balancing act of where are they? Where are they going? And all of this is coming without a lesson plan that says this is what I do next and next and next. And simultaneously you're dealing with a 50 minute period and [saying to yourself] 'I've got to be done with something and bring closure to this at the end of the 50 minutes. How do I get there?' You are on your feet thinking, constantly. It is draining. You become so intensely involved with them [the students] and 'what is your question? What do I do? How do I pull you in?'…you really do get physically tired afterwards…you have to be in charge of not [just] what's going into but what is *going on inside* [their heads]. That is the struggle."

BOLT: Looking Back and Looking Forward

We referred to BOLT early on as a convergent framework that can frame the work of many other researchers and classroom teachers. It is also a framework that has a history of over thirty years with one of the authors of this paper, Jim Minstrell. BOLT as an approach to science instruction first emerged from Minstrell's individual and collaborative efforts with colleagues (teachers and researchers) to explore and contribute to the then new research on student misconceptions in science (Minstrell, 1982a, 1982b). This and subsequent research involved engineering instruction that began a unit by eliciting students' initial thinking and motivating students to test their initial ideas both by experiment and through rational argument. With the support of successive funding from the National Science Foundation and the former National Institute of Education and later the McDonnell Foundation, Minstrell and various collaborators were able to conduct a series of studies over the years that essentially tested and helped to revise what we are now labeling BOLT.

The earliest study was a pre-post comparison conducted with two comparable classes of physics students, one of which was taught in a traditional format and the other with a more reformed approach. Baseline differences between the two essentially disappeared after the teachers implemented a common unit that was designed to engineer BOLT like instruction. Student performance in the traditional teacher's class improved approximately 30% on explaining accelerating cases and 50% better on explaining constant velocity cases than students the previous year so that the students in both classrooms were performing at the same higher level. (Minstrell, 1984)

When the experiment was extended to two more colleagues over the course of an entire year of physics (rather than a single unit) the results were even more interesting. Minstrell recruited and trained two math teacher colleagues to teach physics in his rapidly growing department. The teachers were provided with training and curricular activities designed consistently with BOLT and were closely coached by Minstrell (who was teaching his own course at the time). Within the first year of that three year study, all three teachers were getting like results (Hunt & Minstrell, 19xx). During this time frame, the teachers were also assisted by other researchers in documenting their investigation of the classroom dialogue and collaborative culture while building on learner thinking (vanZee & Minstrell, 1997). Their research also described features of the curricular activities to be integrated with the dialogic interaction (diSessa & Minstrell, 1998).

Challenged by the funding agency to test his reformed approach beyond close coaching within the same school, Minstrell and co-researcher Hunt (a cognitive scientist from the University of Washington) recruited twelve physics teachers from across the state of Washington for a two-year study using a pre- and post-test design. During the summer these teachers participated in a 5-day workshop focusing on the curricular adjustments and dialogue to elicit student ideas and motivate subsequent investigation and sense-making dialogue and the construction of scientific argument. The teachers were also given print and electronic versions of diagnostic assessments and sample curricular activities (referred to by the study group as the Physics Pedagogy Program) as supplements that could be used to elicit and attend to problematic conceptions as well as learning goals. During the academic year teachers met face-to-face about every two months and individually (and collectively) via email as questions or interesting experiences arose. At the end of the first year, the student performance on the end-of-course physics test was on average 15% higher (range=5% to 35% improvement) than student performance the previous year (prior to teacher participation). At the end of the second year, students performed 19% better than the teacher's students had prior to the study (Minstrell & Matteson, 1994).

Following the study, the twelve involved teachers (who have mostly remained in contact with one another) continued in their BOLT-like instruction and most went on to influence the practice of their colleagues. Out of the group's early research efforts and the student "facets" of understanding that were coded came the practical online tools available today on Diagnoser.com. Recognizing the daily challenge and practical difficulties of building on learner thinking in real classroom environments, these tools were designed to support teachers in each stage of the

BOLT framework. Over the last eight years, more than four thousand teachers and one hundred forty thousand students have used the tools. Research conducted on teacher use of Diagnoser (Minstrell & Kraus, 2007) has suggested that the tools work best for classes wherein the teacher already has a perspective of, and genuine interest in listening to students and addressing and building on what they were thinking.

Results from Minstrell's research experiences along with the summary design principles from *How People Learn* have supported our contention that a system that builds on learner thinking can achieve positive and significant improvement in STEM learning. However, it takes carefully crafted curricular activities to engage important observations and ideas in the content, dialogue that keeps students responsible for creating evidenced-based arguments for making sense of experiences, and formative assessment to monitor learning progress and make adjustments in learning and instruction. Our contention is also supported by results from other efforts with which we have had close contact. We have seen aspects of BOLT reflected in the curricular design of Physics by Inquiry, the Constructing Physics Understanding (CPU), and Physics for Elementary Teachers (PET) curricula; in the questioning and classroom discourse reflected in the Disciplinary Literacy Project, the UW Discourse Tools and Contingent Pedagogy Project; in the sense-making activities of Thinker Tools and Modeling Physics Approach; and in the development of consensus and classroom culture depicted in professional development of the SPU-Energy Project and the North Cascades and Olympic Science Partnership activities.

Given the need to more seamlessly integrate assessment into the instruction system, new approaches are needed. We need models of assessment that reflect the complexity of learning and instruction/learning environments, including the multiplicity of learning routes within the context of the opportunities available through the learning system. Basically, how might we characterize what learners know and what learners can do in light of their instruction? How can an assessment of the instructional system also serve as an assessment of the student learning? This calls for the research on instructional validity with respect to assessment giving good information from which to make next instructional decisions as well as innovations of psychometrics to model student learning that is complex, dynamic, collective, and context-dependent. Moreover, moving from a primary focus on accountability to a focus on learning requires professional development with supporting tools and experiences for teachers and for students and possibly for administrators and other stakeholders.

References

Bell, P. and Linn, M. (2000). Scientific Arguments as Learning Artifacts: Designing for Learning from the Web with KIE. International Journal of Science Education. 22(8), 797-817.

Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht: Kluwer Academic Publishers.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education*, *18*(1), 5-25.

Black, P. (2000). Research and the development of educational assessment. *Oxford Review of Education*, *26*(3&4), 407-419.

Black, P., & Harrison, C. (2001). Self- and peer-assessment and taking responsibility: The science student's role in formative assessment. *School Science Review*, *83*(302), 43-49.

Black, P., & McCormick, R. (2010). Reflections and new directions. *Assessment & Evaluation in Higher Education*, *35*(5), 493-499.

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

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5-31.

Blackwell, L., Trzesniewski, K., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*, 246-263.

Bransford, J., Brown, A., and Cocking, R., Eds. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington DC, National Academy Press.

Brookhart, S. M. (2009). Editorial: Special issue on the validity of formative and interim assessment. *Educational Measurement: Issues and Practice, 28*(3), 1-4.

Council of Chief State School Officers (2008). *Attributes of effective formative assessment*. Washington, DC.

Crooks, T. J. (1988). The impact of classroom evaluation practices on student. *Review of Educational Research, 58*(4), 438-481.

Darling-Hammond, L. (2010). *Performance counts: Assessment systems that support high-quality learning*. Washington, DC: Council of Chief State School Officers.

diSessa, A., & Minstrell, J. (1998). Cultivating conceptual change with benchmark lessons. In J. G. Greeno & S. Goldman (Eds.), Thinking *practices in learning and teaching science and mathematics.* Mahwah, NJ: Lawrence Erlbaum Associates.

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

Haertel, E. H., Beach, K. D., Gee, J. P., Greeno, J. G., Lee, C. D., Mehan, H., et al. (2005, April). *Assessment, equity, and opportunity to learn*. Paper presented at the AERA annual meeting, Montreal, Canada.

Hall K., & Burke, W. (2004). *Making formative assessment work: Effective practice in the primary classroom*. Maidenhead: Open University Press

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81-112.

Heritage, M. (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?* A project of Margaret Heritage and the Council of Chief State School Officers (Paper prepared for the Council of Chief State School Officers). Washington, DC: Council of Chief State School Officers.

Heritage, M., & Niemi, D. (2006). Toward a framework for using student mathematical representations as formative assessments. *Educational Assessment*, *11*(3-4), 265-282.

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.

Herman, J. L., & Heritage, M. (2007, June). *Moving from piecemeal to effective formative assessment practice: Moving pictures on the road to student learning.* Paper presented at the annual Council of Chief State School Officers (CCSSO) National Conference on Large Scale Assessment, Session 143, Nashville, TN.

Herman, J. L., Osmundson, E., Ayala, C. C., Schneider, S., & Timms, M. (2005, April). *The nature and impact of teachers' formative assessment practices*. CSE Technical Report 703. National Center for Research on Evaluation, Standards, and Student Testing (CRESST); Paper prepared for the Annual Meeting of the American Educational Research Association (Montreal, Canada).

Hunt, E., & Minstrell, J. (1996). Effective instruction in science and mathematics: Psychological principles and social constraints. *Issues in Education: Contributions from Educational Psychology*, *2*(2), 123-162.

Kanjee, A. (2000). Investigating formative assessment: Teaching, learning, and assessment in the classroom. *Assessment in Education*, *7*(1), 160-163.

Leahy, S., Lyon, C., Tompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, *63*(3), 19-25.

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 371-388). Cambridge, MA: Cambridge University Press.

Lemke, J. (1990). Talking Science: Language, Learning, and Values. Ablex Publishing Corporation.

Linn, M. and Hsi, S. (2000). Computers, Teachers, Peers: Science Learning Partners. Mahwah, NJ: LEA.

Macintyre L. M., Buck, G., & Beckenhauer, A. (2007). Formative assessment requires artistic vision. *International Journal of Education & the Arts*, *8*(4), 1-23.

Michaels, S. (2007). Taking science to school. Washington DC: The National Academies.

Michaels, S., Shouse, A. W., & Schweingruber, H. A. (2008). Ready, set science! Putting research to work in K-8 science classrooms. Washington, DC: The National Academies Press.

Minstrell, J. (1982a). Conceptual development research in the natural setting of a secondary school science classroom. In M. Rowe (Ed.), Education for the 80's: Science: NEA.

Minstrell, J. (1982b). Explaining the "at rest" condition of an object. The Physics Teacher, 20(10).

Minstrell, J. (1984). Teaching for the development of understanding of ideas: Forces on moving objects. In C. W. Anderson (Ed.), AETS Yearbook, Observing science classrooms: Observing science perspectives from research and practice. AETS.

Minstrell, J. (1989). Teaching science for understanding. In L. B. Resnick & L. E. Klopfer (Eds.), Toward the thinking curriculum: Current cognitive research (Vol. Yearbook of the Association for Supervision and Curriculum Development) (pp. 129-149). Alexandria, VA: Association for Supervision and Curriculum Development.

Minstrell, J. (1992). Constructing understanding from knowledge in pieces: A practical view from the classroom. Paper presented at the Annual Meeting of the American Educational Research Association.

Minstrell, J. (2000). Student thinking and related assessment: Creating a facet assessment-based learning environment. In J. Pellegrino, L. Jones & K. Mitchell (Eds.), Grading the nation's report card: Research from the evaluation of NAEP. Washington, DC: National Academy Press.

Minstrell, J. (2001). Facets of Students' Thinking: Designing to cross the gap from research to standards-based practice. In K. Crowley, C. Schunn, & T. Okada (Eds.), Designing for science Mahwah, NJ: Lawrence Erlbaum Associates.

Minstrell, J. (2001). The role of the teacher in making sense of classroom experiences and effecting better learning. In D. Klahr & S. Carver (Eds.), Cognition and instruction: 25 Years of progress Lawrence Erlbaum Associates.

Minstrell, J. and Kraus, P. (2005). *Guided Inquiry in the Science Classroom.* In Donovan & Bransford, J. D. (Eds.). How students learn: History, mathematics, and science in the classroom. Washington, DC: The National Academies Press.

Minstrell, J. and Kraus, P. (2007). *Applied Research on Implementing Diagnostic Instructional Tools.* Final Report to National Science Foundation. Facet Innovations.

Minstrell, J., & Matteson, R. (1994). *Adopting a different view of learners*. In Annual Report video to the James S. McDonnell Foundation, Cognitive Studies for Educational Practice. A.C.T. Systems for Education.

Minstrell, J., & van Zee, E. (2003). Using questioning to assess and foster student thinking. *Everyday assessment in the science classroom.* National Science Teachers Assoc.

Minstrell, J., Anderson, R., Kraus, P., & Minstrell, J. E. (2008). Bridging from practice to research and back: Perspectives and tools in assessing for learning. In J. Coffey, R. Douglas & C. Stearns (Eds.), *Assessing science learning.* Arlington, VA: National Science Teachers Association.

Minstrell, J., Li, M., & Anderson, R. (2009). *Annual Report for Formative Assessment Project*. Seattle, WA: Facet innovations.

Minstrell, J., Li, M., & Anderson, R. (2010). *Annual Report for Formative Assessment Project*. Seattle, WA: Facet Innovations.

Myhill, D., & Brackley, M. (2004). Making connections: Teachers' use of children's prior knowledge in whole class discussion. *British Journal of Educational Studies*, *52*, 263-275.

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research & Evaluation*, *19*(17). http://pareonline.net/pdf/v10n17.pdf.

Osborne, J. F., *Duschl*, R., & Fairbrother, R. (2002). *Breaking the mould: Teaching science for public understanding.* Nuffield Foundation.

Popham, W. J. (2006). Phony formative assessments: Buyer beware! *Educational Leadership*, *64*(3), 86-87.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66,* 211-227.

Pryor, J., & Torrance, H. (1997). Making sense of formative assessment. *International Studies in Educational Administration*, *25*(2), 115-125.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28,* 4-13.

Rivet, A. and Krajcik, J. (2008). Conceptualizing Instruction: Leveraging Students' Prior Knowledge and Experiences to Foster Understandingg of Middle School Science. *Journal of Research in Science Teaching.* 45(1), 79-100.

Sadler, R. (1998). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119-144.

Sandoval, W. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences*, *12*(1), 5-51.

Sandoval, W., & Reiser, B. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*, 345-372.

Shavelson, R. J., Yin, Y., Furtak, E. M., Ruiz-Primo, M. A., & Ayala, C. (2008). On the role and impact of formative assessment on science inquiry teaching and learning. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing science learning. Perspectives from research and practice* (pp. 21-36). Arlington, VA: National Science Teachers Association Press.

Shepard, L. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, *63*(3), 66-71.

Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice*, *23*(3), 32-37.

Solano-Flores, G., & Li, M. (2005). *The use of Generalizability (G) theory in the analysis of wide open-ended interviews in assessment research with multiple cultural groups*. Paper presented at the 11th EARLI Biennial Conference.

Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest priority. *Phi Delta Kappan*, *January*, 363-368.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *June*, 758-765.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practice of classroom assessment*. Albany: State University of New York Press.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching and learning in the classroom*. Buckingham: 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*, *26*(5), 615-631.

van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *Journal of the Learning Sciences*, *6*(2), 227-269.

Wiliam, D. (2005, April). *The integration of formative classroom assessment with instruction*. Paper presented at the AERA annual meeting, Montreal.

Wiliam, D. (2007). Keeping Learning on Track: Formative assessment and the regulation of learning. In F. K. Jr. Lester (Ed.), *Second Handbook of Mathematics Teaching and Learning*, 1053-1098. Greenwich, Conn.: Information Age Publishing.

Wiliam, D., & Thompson, M. (2008). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-82). Mahwah, NJ: Erlbaum.

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., Tomita, M. K., & Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, *21*(4), 335-359.