Formative Assessment in K-8 Science Education: A Conceptual Review[1]

Erin Marie Furtak
School of Education
Stanford University

Introduction

The assessment that occurs every day in science classrooms is often overlooked. Recently, the *National Science Education Standards* (NRC, 1996), *Inquiry and the National Science Education Standards* (National Research Council [NRC], 2001a), and others (Atkin & Coffey, 2003) proposed to move daily assessment to the foreground in debates involving accountability. These proposals are strengthened by evidence that shows ongoing formative assessment has a positive effect on student learning (Atkin & Coffey, 2003; Black & Harrison, 2001; Black & William, 1998a).

The original intent of this paper was to examine research that has been performed to date on formative assessment in K-8 science classrooms for the purpose of identifying common practices and emergent models of classroom-based formative assessment. However, a careful review of available literature revealed that strikingly little research has been performed in this important area. Thus, this paper is more of a review of new and important conceptual issues in formative assessment practices in K-8 science education since Black & Wiliam's (1998a) review.

To begin, this paper describes the climate in science education in the United States, and describes and defines formative assessment. Next, Black & Wiliam's (1998a) review and two other important empirical studies will be summarized. Then, a framework characterizing different forms of formative assessment is presented. Non-empirical studies are organized

[1] Commissioned paper by the National Research Council for Science Learning K-8 consensus study

according to this continuum. Finally, the paper describes limitations in the implementation of formative assessment in K-8 science, and summarizes assessment practices that show promise for improving student learning. The important contribution of the paper is in organizing the review into a continuum of formal and informal formative assessment.

Science Education

The present climate of science education in the United States can be characterized by three major influences. First, international assessments of student learning have indicated that students in the United States are, at best, performing only in the mid-range internationally in science (Doran, Lawrenz, & Helgeson, 1994; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999). Furthermore, detailed analyses of science education in the US have revealed a curriculum that is "a mile wide and an inch deep," giving students only superficial understanding of critical scientific concepts (Schmidt, McKnight, & Raizen, 1997). Second, the *National Science Education Standards* (NRC, 1996) set forth the content that students are expected to learn at each level in science classrooms in the US. Third, the No Child Left Behind Act of 2002 mandated that all states measure students' science achievement against these standards at least once in each of three grade spans each year starting in 2007. The combination of these three events has brought prominence to the degree to which students in the United States are learning science. The NRC (2001a) has stated that effective assessment by teachers and students is among the topics of highest priority in fulfilling the mission of the *National Science Education Standards*.

Formative Assessment

In contrast to summative assessment, which makes overall judgments about the learning achieved during a period of time for the purposes of accountability, formative assessment has learning as its purpose and places agency for the improvement of learning on both the teacher and student (Shavelson & SEAL, 2003). The formative assessment literature frames the importance of better understanding classroom assessment in the move to raise standards and improve learning for all students, so that high standards may be achieved (Black & Wiliam, 1998a).

Teachers commonly view assessment as something apart from their regular teaching, serving the purpose of assigning grades (Shavelson & SEAL, 2003). However, in order for instruction to be effective, teachers must also assess their students *while learning is in progress* to gain information about their developing understanding so that instruction can be adapted accordingly (Black, 1998). Teachers have the most direct access to information about student learning, and are thus in a position to interpret and use information about student learning to provide students with timely feedback (Shepard, 2003; Wilson, 2005). Teachers can also use the information to monitor the effectiveness of their own teaching (NRC, 2001a); however, formative assessment also involves students since they need to recognize, evaluate, and react to their own learning and/or others' assessments of their learning (Bell & Cowie, 2001; Sadler, 1989).

Black & Wiliam (1998a) defined formative assessment as "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p. 8). This kind of

assessment, called *formative assessment*[2], can be conceived as assessment *for* learning and not *of* learning (Black & Wiliam, 1998b; Pellegrino, Chudowsky, & Glaser, 2001).

Assessment becomes formative in nature – informing teaching and learning – only when the teacher uses that information to adapt instruction, and/or the student uses the information to influence his or her learning (Black, 1998). For example, a teacher asking a planned sequence of questions might find out that students had not understood the concept to be learned in a particular lesson, and as a result the teacher might use that information to modify the subsequent lesson to reinforce the prior learning goal. In another situation, a student comparing his or her own work to an exemplar shown by the teacher might make modifications on the basis of reaching the goal made explicit in the form of the exemplar. Therefore, whether assessment is formative hinges on a criterion of *use*; that is, assessment can be considered formative when information is used to take action to advance students toward learning goals (Bell & Cowie, 2001; Black & Wiliam, 1998a; Shavelson, Black, Wiliam, & Coffey, 2003).

Formative Assessment and Feedback

Formative assessment can be summarized in three central questions to be answered by the student or teacher (NRC, 2001a):

> Where are you going?
> Where are you now?
> How are you going to get there?

This three-step process summarizes what has been called the "feedback loop" in formative assessment; that is, setting a learning goal, determining the gap between the learning goal and the

[2] While the term *classroom assessment* is often used interchangeably with the term *formative assessment*, this paper will use the latter term to isolate the type of assessment that is used to inform instruction while learning is in progress from other forms of classroom-based assessment (e.g. teacher-authored unit tests, quizzes provided for the purpose of assigning grades).

student's present state of understanding, and formulating feedback to close the gap. Each step will be described in more detail below.

Articulating clear criteria and goals. Although they may often be tacit, teachers have goals for their students as they conduct learning activities. Sadler (1989) described the process of goal-setting in detail. These goals may come in the form of what type of product they are looking for, the quality of their argument, or the clarity of an explanation provided by a student. Teachers can make goals explicit to students through descriptive statements, which detail the different aspects of the goal; or exemplars, which show gradations of quality up to the desired standard. Despite a teacher's best efforts, a goal only becomes important to the students when they adopt the goal internally for themselves. In many educational settings, goals that are specific rather than vague have been shown to be most effective at capturing students' attention and increasing mobilization on a task. These learning goals, while often viewed as conceptual in nature, can also be spread across the other domains of scientific inquiry, comprising not only conceptual understanding, but also knowing how knowledge is generated in science (epistemic) and how knowledge is communicated and negotiated (social) (Duschl, 2003).

The Gap. The literature offers a metaphor of a gap to help conceptualize the role assessment can play in helping students to achieve learning goals (Sadler, 1989; Black & Wiliam, 1998b). If one side of the gap represents student learning goals (point B) and the other represents the current place where students sit with respect to those goals (point A), the distance between points A and B comprises a gap that needs to be bridged.

To establish the size of the gap between points A and B, the teacher must in some way make the students' thinking visible so that their level of understanding can be compared to the goal. This can include eliciting students' thinking through verbal or written prompts, reviewing

students' notebooks or homework, or listening to small-group conversations. In many conversations about assessment, the focus stops with an inference concerning student understanding (point A), and at times includes how much it falls short of point B or the goals.

Pellegrino, Baxter, & Glaser (2000) summarized the work of Minstrell, who developed a computer-based method for identifying students' ability to separate fluid/medium effects from gravitational effects (Minstrell, 1992). Minstrell identified a group of students' common misconceptions (i.e., facets of student understanding) on the way to understanding gravitational effeccts. By asking students specific questions, Minstrell could map their understanding onto his framework of facets, and thus diagnose the gap that needed to be bridged between the students' current state of understanding and the ultimate learning goal.

While the gap metaphor lacks the complexity inherent in any classroom activity, it does capture the possibility of how assessment can provide teachers and students with information that can inform actions that bridge the gap. The optimum gap size is hypothesized to be not too large or too small, so as to create sufficient determination for the student to adopt and reach the goal.

Feedback: Closing the Gap**.** The process that connects the teacher's goals or criteria with the students' current state of understanding (that is, the process that closes the gap), is the "feedback loop" or the process of the teacher providing feedback to students. The extent that any such information serves to inform teaching and influence learning depends in a large part on how it is used. Teachers must not only interpret and make meaning of the information; they must also use the information to adapt their teaching to meet the needs of their students (Black & Wiliam, 1998a).

Assessment that facilitates learning not only helps the teacher know where the student is starting from (point A), but also highlights for the student where he or she is headed (point B) and provides actions to help them reach this point. A feedback loop from assessment to teaching and learning is a primary mechanism by which the gap between point A and point B is bridged. Teachers can use feedback to make decisions about diagnosing levels of student understanding and preparing for remediation when it is necessary, whereas students gain information about the strengths and weaknesses of their performances so that they can maintain those aspects that are of high quality and focus their efforts on those in need of improvement (Sadler, 1989). In order to deliver feedback effectively, the teacher must have set clear goals and have some kind of interpretive framework for student understanding (Black & Wiliam, 1998a; Minstrell, 1992); however, the teacher must also capitalize on opportunities to elicit student thinking and provide feedback based on the goal and framework. The feedback provided by the teacher is dependent upon the particular learning goal; for example, action can help to refine students' understanding of important concepts, to identify aspects related to the process or nature of science (White & Frederiksen, 1998), or to support the development of students' scientific communication skills (Duschl, 2003).

The medium for delivery of feedback, like formative assessment, can take many forms, from written comments on a student's paper to informal conversations during class. Instructional feedback is not intended to be evaluative, but is a qualitative evaluation of a student's progress at a point in time. This aspect is a characteristic of all formative assessment, where student work is not evaluated on a right or wrong basis, but as part of a continuum of growth toward increasing quality or degree of expertise (Sadler, 1989).

There are many actions a teacher can take to close the gap, from describing new procedures, to explaining how a sentence could be edited for more clarity, to planning another activity to re-teach a certain concept. While teachers are commonly engaged in the activity of critiquing the work of others, students are often not involved in this activity. Allowing students to review the work of peers provides them the opportunity to see how the work of others might be improved, and is an important step to helping them learn to self-assess (Black et al., 2002; Sadler, 1989).

The manner in which feedback is communicated to students is essential, since the application of an evaluative statement, such as “you’re right” which implies the existence of correct or incorrect criteria can defeat the purpose of the continuum described above. Other comments may be lacking in specificity, like saying “yes!” Students may not be expected to make progress if their teachers are providing them with evaluative or nonspecific feedback on the basis of looking at their work. When more specific comments are provided to the student, they should be based upon a clear description of what the underlying criteria are; for example, a student needs to know what “clarity” means in terms of their own work (Sadler, 1989). The effectiveness of feedback depends on the quality of the feedback rather than existence or absence (Black & Wiliam, 1998b; Black, 1998; Crooks, 1988). This includes the quality and saliency of the information gathered in the first place and the appropriateness and relevance of subsequent actions.

Research on Formative Assessment: Black & Wiliam’s Review

Research into the effectiveness of formative assessment suggests compelling results.

In an extensive review of the literature that included more than 250 articles, Black and Wiliam (1998a) placed the effect size for learning gains in interventions involving aspects of formative assessment between 0.4 and 0.7[3]. While gains were seen across student achievement levels, gains were highest for lower achieving students. Black (1998) summarized the findings of the 1998 review into four features:

- Formative assessment will require new teaching practices and thus calls for significant changes in classroom practice;
- Students must be actively involved in their learning;
- For assessment to function in a formative manner, results have to be used to modify teaching and learning;
- Assessment has the potential to affect not only student learning, but also motivation, self-esteem, and participation in self-assessment.

Despite these encouraging findings, Black & Wiliam also found that few quantitative studies on formative assessment existed:

> "Individual quantitative studies which look at formative assessment as a whole do exist…although the number with adequate and comparable quantitative rigour would be of the order of 20 at most (p. 53)."

In addition, of the few studies that did exist in 1998, most were performed in disciplines other than science or in content-free problem situations. Furthermore, of the studies in science, most were qualitative or descriptive in nature, and did not provide clear links to student learning. Thus, while the impetus for formative assessment in all aspects of learning is clear, many well-developed models of formative assessment exist, and several descriptive accounts have been performed, the field of science education has yet to determine the ways in which formative assessment may be effectively integrated into instruction. Black & Wiliam's (1998a) review has

[3] Effect size derived only from studies with pre and post measures of student learning.

since inspired educational researchers to explore in more detail how best to realize effective formative assessment in science classrooms.

Influence of Classroom-Based Assessments on Student Learning: Empirical Studies

Shavelson & Towne (2002) emphasized the importance of quantitative, experimental studies to determine causality in educational research. Unfortunately, as Black & Wiliam (1998a) found, very few empirical studies have been performed on the effects of formative assessment on student learning. Of those studies, only one took place in the context of science education (White & Frederiksen, 1998).

In a controlled study, White & Frederiksen (1998) explored how peer and self-assessment could help to build students' understanding of scientific inquiry. Students from four middle school science classes were randomly assigned to conditions: half to complete the reflective assessment process, and the other half to serve as a control. Students in both groups were provided with criteria for scientific inquiry processes; for example, "being systematic" and "reasoning carefully." Two of the classes used regular time during class to reflect on what they were learning and how they were learning it (e.g. using evidence from their work to support their evaluations) while the other two classes spent the same amount of time talking about how the activities could be changed. In this way, students in the reflective assessment (i.e. formative assessment) group monitored their own progress and the progress of their peers through verbal and written feedback, and then were provided with opportunities to improve their performance later in the unit. The two classes of students that engaged in the reflective assessment process performed better on both project work and the unit test. Perhaps most notable, however, is the

fact that lower performing students in the experimental class (as designated by CTBS score) showed the greatest improvement in performance when compared to the control class.

Although not performed in science specifically, Black (1998) identified the work of Butler (1998) as one of the most important quantitative studies in formative assessment, and since it was conducted in a curriculum-free environment, its results can be generalized, to a certain degree, to the area of science education; thus, it is included here. Butler (1988) studied 11-year old students from four schools in Israel, 24 from the top quartile of their own class in tests of mathematics and language, and 24 from the bottom quartile. Students completed written tasks that were not related to the regular curriculum, and were then provided with one of three types of feedback on their work: 1) tailored written remarks addressing criteria that they had been made aware of before taking the assessment, 2) grades derived from scoring of previous work, or 3) both grades and comments. Post-test performance indicated that scores on the tasks increased most significantly for students who received comments only across all three sessions, while scores declined across the three sessions for those who received both comments and grades. Students receiving grades only declined and then increased between the second and third sessions. The only significant difference between the high and low-performing students was found in terms of interest; students with lower scores also showed lower interest when they received grades on their work. Although revealing important information about the effects of the type of feedback on performance, this study lacks ecological validity, as it was done as a multi-week intervention with material that was not related to the school's curricula. Another aspect of Butler's research involved student attitudes about themselves as students and about subject matter (Butler & Neuman, 1995). Attitudes improved among students who received comments only. Among students who received grades and comments, students who performed well

maintained positive attitudes, while students who performed poorly demonstrated negative attitudes.

A Framework for Formative Assessment

Although few empirical studies have been performed in the area of formative assessment in science education, additional studies of a more qualitative, descriptive, or conceptual nature still provide a broad view of research since Black & Wiliam's (1998a) review. This section presents important studies in that category. First, a framework for discussing these studies will be presented, spanning a continuum from *formal* to *informal formative assessment* (Shavelson & SEAL, 2003).

Formative assessment can be *formal* -- a planned act designed to provide evidence about students' learning, or *informal* – where evidence of learning is generated in the course of a teacher's day to day classroom activities (Bell & Cowie, 2001; Duschl, 2003; Shavelson et al., 2003). Each can be characterized in a different manner. Formal formative assessment usually starts with students doing/carrying out an activity designed or selected in advance by the teacher so that information may be more precisely collected (gathering). Typically, formal formative assessments take the form of curriculum-embedded assessments that focus on some specific aspect of learning (e.g., students' knowledge about why objects sink or float), but they can also be direct questioning, quizzes, brainstorming, generation of questions, and the like (Bell & Cowie, 2001). The activity enables teachers to step back at key points during instruction, check student understanding, and plan on the next steps that they must take to move forward their students' learning.

Informal formative assessment, in contrast, is improvisational and can take place in any

student-teacher interaction. Teachers engaging in informal assessment cannot anticipate exactly when, where, or how these opportunities to obtain assessment information will arise; thus the process of informal formative assessment is flexible and interactive. It can arise out of any instructional/learning activity at hand, and is "strongly linked to learning and teaching activities" (Bell & Cowie, 2001, p 86). The information gathered during informal formative assessment consists of students' and teachers' verbal questions and comments (Bell & Cowie, 2001), but can also be non-verbal (based on teacher's observations of students during the course of an activity). The time frame for interpreting and acting is more immediate when compared with formal formative assessments. A student's incorrect response or unexpected question can trigger an assessment event by making a teacher aware of a student's misunderstanding. Acting in response to the evidence found is usually quick, spontaneous, and flexible since it can take different forms (e.g., responding with a question, eliciting other points of view from other students, conducting a demonstration when appropriate, repeating an activity).

Although any particular instance of formative assessment can fall in any location along the continuum between formal and informal formative assessment, the continuum can be divided into three basic categories: on-the-fly assessment, which is basically informal formative assessment; planned-for assessment, or assessments that are planned or anticipated in advance by the teacher; and curriculum-embedded assessments, or those which are a formal, written element in a curriculum or unit of study. Figure 1 illustrates a continuum of types of formative assessment.

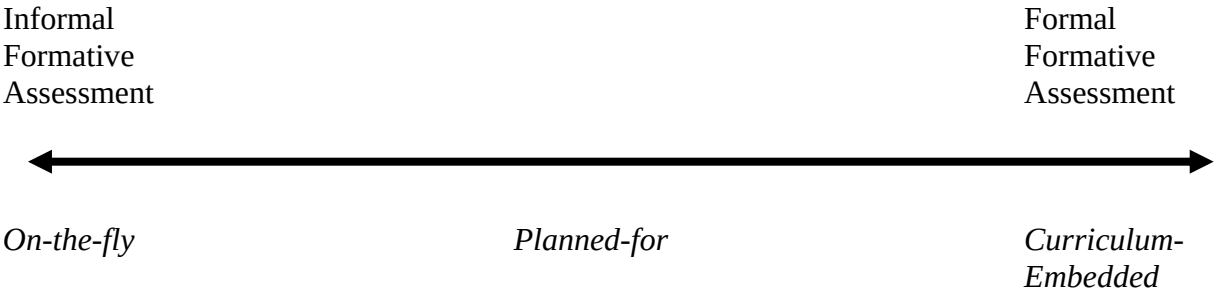

Figure 1. Continuum of formative assessment (from Shavelson & SEAL, 2003).

On-the-Fly Formative Assessment

On-the-fly formative assessment occurs when "teachable moments" unexpectedly arise in the classroom. For example, a teacher may overhear a conversation in a small group in which a student claims, "Density is a property of a material. No matter the mass and/or volume of that material, the property of density stays the same for that material." The teacher seizes the opportunity to challenge the student's thinking, and asks the student's group mates to try to test this idea by measuring the density of a new material of various sizes or masses. This kind of teaching action is seamless with everyday teaching practice; in fact, some may consider the example above as an instance of "good teaching" rather than of formative assessment. However, opportunities to gather information about students' thinking and to take action to move students toward learning goals arise continuously during instruction and should be taken as opportunities for on-the-fly formative assessment (Shavelson & SEAL, 2003).

Planned-for-Interaction Formative Assessment

At the center of the continuum is an area in which there is some level of deliberate planning on the part of the teacher to conduct formative assessment. In contrast to on-the-fly

opportunities, planned-for formative assessment is deliberate, but is not as formal as curriculum-embedded assessments. Rather than waiting for opportunities to arise in the course of normal classroom interactions, teachers conducting planned-for assessment plan in advance the kinds of questions that will maximize their acquisition of information. That is, teachers realize the value of good questions (and other pedagogical actions for eliciting information) and spend time planning these pedagogical moves prior to class (Black, Harrison, Lee, Marshall, & Wiliam, 2002; Shavelson & SEAL Group, 2003). For example, a teacher might consciously plan to integrate longer wait-time into her questioning practices to give students more opportunities to intellectually engage in discussions (Black et al., 2002 ; Rowe, 1974). In another case, a teacher might move away from writing simple feedback in the form of a brief comment ("good"), a happy face, a check, or, a grade toward providing more thoughtful and constructive statements (Black, Harrison, et al., 2002; Ruiz-Primo, Li, Ayala & Shavelson, in press).

Formal, Curriculum-Embedded Formative Assessment

At the other end of the continuum, teachers or curriculum developers may embed assessments in the ongoing curriculum to intentionally create "teachable moments." In simplest form, assessments might be embedded after every 3 or so lessons to make clear the progression of subgoals needed to meet the goals of the unit and thereby provide opportunities to teach to the students' problem areas. In its more sophisticated design, these assessments are based on a "theory of knowledge in a domain" or an assessment framework (Black & Wiliam, 1998a). The assessments are then embedded at critical junctures, and crafted so feedback on performance to students is immediate and pedagogical actions are immediately taken to close the learning gap (Shavelson & SEAL, 2003). For example, the Stanford Education Assessment Laboratory

(SEAL) created a set of assessments designed to tap declarative knowledge ("knowing that"), procedural knowledge ("knowing how") and schematic knowledge ("knowing why") and embedded them at four natural transitions or "joints" in a 10-week unit on buoyancy. The assessments served to focus teaching on different aspects of learning about mass, volume, density and buoyancy. Feedback on performance focused on problem areas revealed by the assessments (Shavelson & SEAL, 2003; Yin, 2005).

Studies on Formative Assessment in K-8 Science Education

The section below presents studies grouped according to the framework above. For the major studies reviewed, the grade level of student participants is provided, the types of formative assessment strategies employed in each study are described, and the influence of formative assessment on instruction and students is identified.

Informal, On-the-Fly Formative Assessment

Questions are an common element to teacher-student interactions, often following the traditional IRE/F sequence where the teacher *Initiates* a question, the student *Responds*, and the teacher provides an *Evaluation* of the student's response or some kind of generic *Feedback* (Lemke, 1990). Informal, on-the-fly formative assessment is a step beyond these traditional classroom interactions; it becomes a method of genuine probing for understanding, rather than simply checking and evaluating the state of students' understanding (White & Gunstone, 1992). This point is especially relevant in the context of science education, where teachers of scientific inquiry need to continuously elicit student thinking and help students consider their developing conceptions on the basis of scientific evidence.

Bell and Cowie (2001) derived a model of on-the-fly formative assessment from a study of classroom-based assessment in eight New Zealand science classrooms. Students ranged in age from 11-14 (years 7 to 10 in school). Data collection focused upon students' and teachers' ideas about assessment, classroom-based case studies, and investigations of individual teachers' development as practitioners of formative assessment. In Bell & Cowie's study, on-the-fly formative assessment is viewed as taking place during everyday student-teacher interactions. Their model of on-the-fly formative assessment consists of three steps oriented around a central purpose for the lesson: noticing, recognizing, and responding. First, the teacher pays attention (notices) information about student learning in the form of asking questions or simply listening or a particular student; second, the teacher compares the information that has been noticed to the purpose of the lesson or learning goal (recognizes); third, the teacher responds to the student in an immediate manner. Bell & Cowie concluded that interactive, informal formative assessment allowed teachers to focus upon student development, draw upon their own pedagogical content knowledge, increase the amount of interaction involved with everyday lessons, and was an integral part of teaching and learning, not a separate element. In this way, on-the-fly assessment can be perceived as synonymous with existing descriptions of scientific inquiry teaching (NRC, 2001a).

In a comparison of students' scientific reasoning processes in peer- and teacher-guided discussions, Hogan, Nastasi, and Pressley (2000) identified several aspects of teacher questioning practices endemic to situations in which the teacher does not provide information directly to students, but rather supports them while they are constructing their own understanding – an element in certain types of scientific-inquiry oriented teaching (NRC, 2001b). These targeted questions, used for the purpose of extending students' present level of understanding,

make Hogan, Nastasi, and Pressley's study an example of on-the-fly formative assessment. The focus of the study was an 8th grade teacher in a suburban, upstate New York school. The curriculum focused students on theory building and the construction of mental models through design. The researchers found that the teacher typically began interactions with small groups by asking a question that revealed the status of students' thinking, followed by repeating and elaborating what the student said – similar to the first two steps of *noticing* and *recognizing* in Bell & Cowie's (2001) model. Hogan, Nastasi, & Pressley's study is an example of how on-the-fly formative assessment can be seamless with science instruction: the teacher used questions to determine the current level of the students' understanding, and then asked follow-up questions intended to help students make their explanations more complete or to phrase them in more acceptable scientific terms. The authors also found that the independent small group discussions were more "generative and elaborative than discussions with teachers," suggesting that listening to students as they work in small groups may be more fruitful for teachers trying to determine the state of students' understanding; however, the extent to which students' conceptual understanding increased did vary between small groups in the classroom.

Ruiz-Primo and Furtak (2004) explored the on-the-fly formative assessment practices of three middle school science teachers and compared them to student performance. These practices were described as ESRU cycles, based on Bell & Cowie's (2001) model - the teacher Elicits a question, the Student responds, the teacher Recognizes the student's response, and then Uses the information collected to support student learning. *Eliciting* information focuses on the teacher's strategies, such as asking questions, that allow students to share and make explicit their thinking (e.g., ask the students to relate evidence to explanations). *Recognizing* students' thinking requires the teacher to listen and acknowledge students' responses, explanations, or mental models (e.g.,

teacher repeats the student's comment to make sure it has been understood appropriately). *Using* information involves taking action on the basis of student responses to help students move toward learning goals (e.g. by responding with another question, eliciting alternate points of view, conducting a demonstration, or repeating an activity). For example, a teacher might ask a student to provide an example (*Eliciting),* the student provides an example (student *Responds),* the teacher repeats the statement to confirm that she has understood it correctly (*Recognizing),* and then the teacher encourages the student to share his idea with another student who has a different example for the same idea (*Acting)* (Furtak & Ruiz-Primo, 2005). Most of the cycles observed in the study were classified as focusing on making predictions, interpreting graphs, and other epistemic factors, with only a few cycles observed across the three teachers that focused on conceptual development. The study found that while students' performance varied across questions and teachers, the highest level of student performance was observed in the class of the teacher with the most complete questioning cycles. However, the study also raises the question of whether the differences observed between teachers was attributed to their on-the-fly formative assessment practices, or was simply a part of the teachers' overall differences in everyday science teaching skills.

Planned-for Formative Assessment

Ongoing formative assessment occurs in a learning environment that helps teachers acquire information on a continuing and informal basis, such as within the course of daily classroom talk. This type of classroom talk has been called an assessment conversation (Duschl & Gitomer, 1997; Duschl, 2003), or an instructional dialogue that embeds assessment into an activity already occurring in the classroom. When planned deliberately, assessment

conversations become an example of planned-for assessment. Assessment conversations permit teachers to recognize students' conceptions, mental models, strategies, language use, or communication skills and allow them to use this information to guide instruction. In classroom learning environments in which assessment conversations take place, the boundaries of curriculum, instruction, and assessment should blur (Duschl & Gitomer, 1997). For example, an instructional activity suggested by a curriculum, such as discussion of the results of an investigation, can be used as an opportunity for the teacher to conduct an assessment conversation.

In the Science Education through Portfolio Instruction and Assessment (SEPIA) project, these assessment conversations are used to help teachers provide scaffolding and support for students' construction of meaning by carefully selecting learning experiences, activities, questions, and other elements of instruction (Duschl and Gitomer, 1997). Project SEPIA uses modeling and explicit teaching to help students "learn how to learn in science " (p. 41). Duschl & Gitomer explored how two middle school teachers worked with Project SEPIA's model of instruction. Developing a portfolio as they complete the unit, students are presented with authentic problems and proceed through an established sequence of investigations to develop their conceptual understanding, reasoning strategies related to ways of knowing in science, and communication skills. A central element of the assessment conversation is a three-part process that involves the teacher receiving student ideas through writing, drawing, and sharing orally, so that students can show the teacher and other students what they know. The second step involves the teacher recognizing students' ideas through public discussion, and the third has the teacher using ideas to reach a consensus in the classroom by asking student to reason on the basis of

evidence[4]. Project SEPIA also provides teachers with criteria for guiding students during these conversations, including a focus on relationships, clarity, consistency with evidence, use of examples, making sense, acknowledging alternative explanations, and accuracy. Engaging students in assessment-related conversations about their work provides a context where standards and criteria of quality are negotiated and discussed publicly (Duschl & Gitomer, 1997). The authors concluded that teachers should focus less on tasks and activities and more upon the reasoning processes and underlying conceptual structures of science.

Minstrell & vanZee (2003) describe questioning as a form of planned-for formative assessment by using questions both to diagnose the state of students' thinking and to prescribe an appropriate next step for students to take in their learning. VanZee and Minstrell's (1997) study explored how the "reflective toss" strategy Minstrell used in his high-school physics classroom gave students responsibility for monitoring their own thinking and making their meanings clear. A reflective toss is defined as a question that "catches" the meaning of a student's statement and then "throws" responsibility for thinking back to the student. For example, if a student made a particular assertion, the teacher would respond with another question such as "Now what do you mean by…" or "If you were to do [that]…, what would you do?" (p. 245). In this way, the teacher (in this case, Minstrell) used questions to find out what students were thinking, to consider with his students how their thinking fits with what physicists think, and to place responsibility for thinking back on the students. While the study took place in the high school classroom of only one teacher, it raises the important point for all levels of science instruction

[4] Duschl & Gitomer's (1997) description of a three-step questioning process is very similar to that previously described in Bell & Cowie (2001) and Ruiz-Primo & Furtak (2004) as examples of on-the-fly, informal formative assessment. However, Duschl & Gitomer's study is considered an example of planned-for formative assessment because the questioning process is intended to take place in the context of planned assessment conversations. In contrast, Bell & Cowie and Ruiz-Primo & Furtak observed the questioning process in the course of everyday, on-the-fly classroom interactions.

that a simple planned-for questioning strategy can be an effective tool for formative assessment. The "reflective toss" forced students to take ownership of their ideas and to think about them further, and also allowed the teacher to react and take action on students' ideas as they were offered to the class.

The Classroom Assessment Project to Improve Teaching and Learning (CAPITAL) was a 4-year, NSF-funded study that explored how teachers shape and modify their teaching practices to "create the conditions for the kind of assessment that fosters learning" (Atkin, Coffey, Moorthy, Sato, & Thibeault, 2005, p. 3). The study investigated how 25 middle school teachers of varying levels of experience evolved in their formative assessment practices through collaboration with each other and university researchers. The change process was deeply personalized, as each teacher had different objectives for improving his or her own practice, but with the goal of helping to cultivate a reflective orientation toward teaching in all participants. For example, one teacher learned that even if students could show they understood a concept verbally, they would often still have difficulty in expressing their understanding in writing. Through discussions of student work samples with other teachers and CAPITAL researchers, the teacher developed strategies for engaging her students in discussions about writing, including what makes a particular piece of writing "good." The teacher's new perspective on the importance of writing in science also allowed the teacher to provide more meaningful feedback on students' written work. While the details of the individual case reports varied widely, several features emerged as common themes:

- Making room for each teacher to identify their own starting point for change that they cared about;
- Cultivating a culture of collaboration by allowing teachers time to get to know each other, and facilitating relationships among group members;
- Supporting a culture of professionalism by valuing teachers' priorities, and allowing teachers to exchange ideas with colleagues, administrators, and parents;

- Focusing on students;
- Drawing upon personal beliefs and expectations;
- Supporting the development of a reflective stance toward teaching; and
- Allowing for time for exploration, reflection, and change.

Daws & Singh (1996; 1998) argued that formative assessment strategies can deepen student learning by encouraging reflection upon learning in a structured manner, discussion of progress with teachers to focus on steps toward improvement, and development of greater confidence in their scientific knowledge. Daws & Singh found that formative assessment was generally not being practiced in the secondary schools in Essex that they studied; however, they did find evidence that teachers found formative assessment to be a desirable element to integrate into their teaching. In a series of pilot studies, the authors explored instances of 10th grade students marking their own work, 8th grade students using self-assessment sheets, and 7-9th grade students recording homework and end-of-unit tests. In a series of case studies, the authors explained how a teacher led students through scoring their own work, asking for the reasoning behind the "official" answers to the questions, and shared their own reasoning behind the responses they provided on the test. The goal was to prepare students to do well on future examinations. In the 8th grade class, students were provided with sheets of learning goals for the unit, and as the unit progressed, students were asked to report on evidence they were achieving the learning goals on the sheet. In the third case study, students reported that completing record sheets at the end of units helped them to monitor their levels of achievement. The authors cite these three examples as evidence of how planned formative assessment strategies can be integrated into everyday teaching and learning.

Formal, Curriculum-Embedded Formative Assessment

In *Classroom Assessment and the National Science Education Standards,* embedded assessments are described as those that "occur as part of regular teaching and curricular activities" (2001a, p. 31). However, the term *curriculum-embedded formative assessment* as used in this paper has a more refined meaning; it refers to formative assessments created within the context of a curriculum that are designed to elicit student thinking, and which are referenced specifically to an interpretive framework. Few studies of curriculum-embedded assessments have been completed, but additional studies are in progress. For example, the Berkeley Evaluation and Assessment Research Group [BEAR] (2005) is creating embedded assessments for the Full Option Science System [FOSS]. The assessments are being developed to help teachers of students in grades 3-6 to assess, guide, and confirm student learning in science. These assessments make use of construct maps, which model levels of student understanding of a particular construct (e.g. students' ability to reason with evidence) on the way to developing proficiency (Wilson, 2005). BEAR has helped to develop and refine the associated assessment frameworks, items, scoring guides, and other elements of the system, and will later provide support in the process of psychometric data analyses.

In a recently completed study, the Stanford Education Assessment Laboratory explored Black & Wiliam's (1998a) contention that formative assessment would increase student learning by developing curriculum-embedded assessments for the Foundational Approaches to Science Teaching (FAST) curriculum (Yin, 2005). The first unit of FAST guides students through a series of investigations to culminate in an explanation of floating and sinking on the basis of relative density. As described in a previous section, assessments were embedded at key conceptual "joints" in the curriculum, following a developmental trajectory of understanding

density that students were expected to experience. Twelve 6th and 7th grade teachers were selected from a pool of FAST- trained volunteers. Teachers were matched in pairs according to school characteristics and one member of each pair was then randomly assigned to a control group, which would teach FAST as they normally did, while the other was assigned to an experimental group, which would implement the curriculum-embedded assessments. Experimental-group teachers attended a five-day workshop, where they were trained to implement the curriculum-embedded assessments following the interpretive framework for formative assessment. Multiple measures of student learning were administered to all students of teachers in both the control and experimental groups. Pre-tests consisted of a multiple-choice achievement test and a science motivation questionnaire. Post-tests included the achievement test and motivation questionnaire, as well as a performance assessment, a predict-observe-explain assessment, and an open-ended question assessment. Results of the study indicated that the teachers and their contexts were extremely influential on students' motivation, achievement, and conceptual change; teacher effects overshadowed the treatment effect. Possible interpretations suggest that some experienced teachers implemented their own informal formative assessment strategies regardless of the treatment group they belonged to; while some experimental teachers, despite the five-day workshop, could not implement the curriculum-embedded assessments as intended.

Stern & Ahlgren (2002) analyzed assessments provided in middle school curriculum materials. The study included only comprehensive middle school science programs; that is, those that covered three years of instruction, and which were widely in use by school districts and states. Two two-member teams independently analyzed the curriculum materials and accompanying assessments. With respect to curriculum-embedded assessments, the analysis

revealed that all materials received poor scores in terms of providing guidance for teachers to use students' responses to modify instruction. Those curriculum-embedded assessments that were aligned with the curriculum materials usually focused upon terms and definitions that could be easily copied from the text. Few questions were included that were able to sufficiently elicit students understanding, and even when those questions were included, the materials failed to provide interpretive frameworks for the teachers to interpret students' responses.

Factors impeding use and implementation of formative assessment practices in science education

Despite substantial evidence of its positive impact on student achievement (Black & Wiliam, 1998a), research indicates that meaningful formative assessment is, in general, not a key priority for teachers (Crooks 1988; Black and Wiliam, 1998b). Most teachers limit their assessment practices to assigning grades or norm referenced marks that are unrelated to criteria and with few accompanying details or comments (Butler, 1988; Daws and Singh, 1996; Ruiz-Primo et al., in press).

White & Frederiksen (1998) cite two important caveats to their findings related to reflective assessment: first, both students and teacher need to know that performance is being rated, not individuals; and second, students must be given the means to understand what it is they need to do well in their performance; otherwise, ratings may be damaging. These caveats, according to White & Frederiksen, relate to the important point that if students are not given explicit feedback on how to improve their performance, they are likely to fall back upon ability-related attributions for their performance– similar to Butler's (1988) findings. In addition, less-advantaged students may be further discouraged if performance criteria and steps to

improvement are not made clear. The authors caution that reflective assessment is an integral part of a curriculum and should scaffold the development of the skills being developed, and should not simply be "added on."

A limitation of teachers' ability to provide useful feedback to students in science classes may also be related to their own misconceptions about scientific inquiry teaching and the nature of science; for example, many teachers maintain "folk conceptions" about the scientific method as being linear and atheoretical (Windschitl, 2004). Furthermore, teachers view science as being "dominated by tasks and activities rather than conceptual structures and scientific reasoning" (Duschl & Gitomer, 1997, p. 65). Formative assessment practices such as asking students to argue and defend their ideas, reason from evidence, and develop consensus are based upon a complex, nonlinear model of science that is quite different from that commonly taught in schools (Duschl, 2003; Windschitl, 2004). Therefore, the incongruence between science as taught in schools and more complex models further complicates successful enactment of formative assessment strategies (Duschl & Gitomer, 1997).

Teachers also must have a very clear understanding of the subject domain in which they are working so that they may anticipate potential ideas that students may generate (Duschl & Gitomer, 1997). Such challenges further underscore the need to research common student ideas and misconceptions about science, and to develop interpretive frameworks to provide teachers to supplement their own understanding and predict potential responses from students (Chi, 1992; Driver, Guesne & Tiberghien, 1985). A further problem, identified by Stern & Ahlgren (2002), is that many of the curriculum materials commonly in use, at least in middle schools, do not provide the kind of curriculum-embedded assessments that elicit students' thinking, and do not provide interpretive frameworks for teachers to use. Such frameworks are needed to help

teachers “co-ordinate the many separate bits of assessment information in the light of broad learning purposes” (Black & Wiliam, 1998a, p. 19). Thus, teachers need access to quality formative assessment tools, training, and frameworks.

Doran, Lawrenz, & Helgeson (1994) found that teachers do not receive much training in teacher education programs in terms of how to conduct classroom assessment, formative or otherwise, and little technical help is offered to them in their daily practice. However, as Yin (2005) found, even when provided with quality assessment tools and training to implement them, teachers’ experiences and prior beliefs seemed to override efforts to change teachers’ practices to integrate formative assessment. Another limitation identified in Bell & Cowie’ (2001) case studies was that teachers’ successful implementation of on-the-fly formative assessment depended upon teachers’ skills of interaction with students, and the previous relationships that teachers had established with their students. Preparing teachers to implement formative assessment thus needs to give credit to teachers’ prior experiences and beliefs (Atkin et al., 2005).

Implementation of formative assessment can also be limited by other conditions in the classroom. Hogan, Nastasi, & Pressley (2001) found that even if teacher moves from group to group and monitors learning with on-the-fly questioning practices, learning can still vary between groups, underscoring the importance of helping students to learn to question each other so that their interactions can be more conceptually fruitful in the teacher’s absence. Duschl & Gitomer (1997) found that teachers can become satisfied too quickly with their assessment conversations, and also can become frustrated by the time and effort necessary to conduct effective formative assessment. They concluded that while successful implementation of formative assessment is possible, it is a challenging prospect for teachers.

Classroom-based assessment practices showing promise for improving student learning outcomes

In the form that supports learning, assessment is a ubiquitous aspect of classroom activity, and is rarely a discrete event. It involves observing students at work and listening to what they say (Hogan, Nastasi, & Pressley, 2000), being clear with criteria, and making sure the criteria capture and reflect what counts in the subject area (Resnick & Resnick, 1991). It also involves analyzing student work in light of that criteria, and paying attention to what they are thinking, attending as much to their reasoning as what they don't understand. It involves engaging students as active participants in an assessment activity or conversation so that it becomes something they do, not merely something done to them (Duschl & Gitomer, 1997; White and Frederiksen, 1998). Finally, and most importantly, all kinds of formative assessment demand using that information in a way to inform teaching, learning and thus closing the gap (Black & Wiliam, 1998a).

Despite challenges to the successful implementation of formative assessment, the studies reviewed in this paper suggest several common practices and emergent models of formative assessment that show promise for improving student learning outcomes. These studies are summarized in Table 1.

Table 1.
Summary of K-8 formative assessment studies.

	On-the-fly Informal Formative Assessment	*Planned-for* Formative Assessment	*Curriculum-Embedded* Formal Formative Assessment
Influence on students	- Teacher with more "complete questioning cycles" had students with higher performance (Ruiz-Primo & Furtak, 2004) - Questioning helped students to make explanations more complete and phrased in scientific terms; small-group interactions became more generative (Hogan, Nastasi, & Pressley, 2000)	- Reflective toss strategy forces students to take ownership of ideas and make meanings clear (vanZee & Minstrell, 1997) - By developing portfolios and participating in discussions, students develop conceptual understanding, reasoning strategies, and communication skills (Duschl & Gitomer, 1997) - Students better able to monitor their own levels of achievement (Daws & Singh, 1996; 1998)	- Unclear; teacher effect overshadowed effect of curriculum-embedded assessments (Yin, 2005) - Other studies pending (e.g. BEAR, 2005)
Influence on instruction	- Allows teachers to focus upon student development, draw upon PCK, increase amount of interaction in everyday lessons (Bell & Cowie, 2001)	- Teacher becomes more able to diagnose state of students' learning through questions (vanZee & Minstrell, 1997) - Teacher develops formative assessment competence through collaboration and reflection (Atkin et al., 2005)	- Despite training, some teachers did not enact curriculum-embedded assessments as intended (Yin, 2005) - Other studies pending (e.g. BEAR, 2005)
Practices showing promise for improving student learning outcomes	- Viewing formative assessment as an integral part of everyday science instruction (Bell & Cowie, 2001) - Questioning should go beyond IRE/F patterns to determine state of students' thinking and to move students toward learning goals (Hogan, Nastasi, & Pressley, 2000; Ruiz-Primo & Furtak, 2004)	- Reflective toss (vanZee & Minstrell, 1997) - Holding assessment conversations (Duschl & Gitomer, 1997) - Referring students to learning goals and record sheets to monitor their own levels of achievement (Daws & Singh, 1996; 1998) - Providing teachers with time to explore, reflect, and change their formative assessment practices (Atkin et al., 2005)	- Taking into account contextual factors when preparing teachers to enact curriculum-embedded assessment (Yin, 2005) - Other studies pending (e.g. BEAR, 2005)

The majority of studies cited in this review were performed in middle school classrooms. Thus it is difficult to make any kind of claim about the differences in abilities of students of varying ages to participate in formative assessment. All that can be said is the strategies summarized in Table 1 suggest middle school students are capable of participating in and benefiting, to various degrees, from formative assessment. More research needs to be performed in K-5 classrooms to determine if the result is similar for students of that age.

The limitations of implementing formative assessment assembled in this study suggest that teachers often do not have access to quality assessment tools and interpretive frameworks. It is also possible that those teachers who do implement formative assessment effectively do so because their own beliefs about teaching and student learning are consistent with the values associated with formative assessment practices. Several of the studies reviewed in this paper suggest that simply training teachers to use formative assessments does not lead in a linear manner to effective implementation or increases in student learning.

In fact, the studies reviewed in this paper seem to point toward everyday questioning strategies, whether planned or on-the-fly[5], that elicit student thinking and take action to help to increase student learning. Future efforts to understand what "good science teaching" looks like should also consider formative assessment as part of the equation; that is, viewing science instruction to determine how effective, responsive teaching involves setting learning goals, finding out what students know, and taking targeted action to increase student learning. Furthermore, more educational researchers need to conduct more studies exploring the role of curriculum-embedded formative assessment in helping students to learn science.

[5] Only one completed, classroom-based study of curriculum-embedded formative assessment could be located for this paper. Thus, the conclusions of the paper focus upon the on-the-fly and planned-for studies reviewed.

It is worth nothing that in the process of writing this study, the author contacted several prominent researchers in the fields of science education and assessment, searching for studies performed in the area beyond the commonly cited works of Black & Wiliam, Butler, and others included in this review. Without exception, each researcher expressed discouragement at the few studies that have been performed, and the even smaller number of studies that have tested formative assessment in science education in controlled, randomized studies. If one conclusion is to be reached from this experience, it is that researchers have only begun to explore the role of formative assessment in science education.

References

Atkin, J.M. & Coffey, J.E. (Eds.). (2003). *Everyday Assessment in the Science Classroom* (pp. 41-59). Arlington, VA: NSTA Press.

Atkin, J. M., Coffey, J. E., Moorthy, S., Sato, M., & Thibeault, M. (2005). *Designing Everyday Assessment in the Science Classroom*. New York: Teachers College Press.

Bell, B., & Cowie, B. (2001). *Formative Assessment and Science Education*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Berkeley Evaluation and Assessment Research Center. (2005). *Assessing Science Knowledge (ASK).* Retrieved January 3, 2005, from http://bearcenter.berkeley.edu/research.php

Black, P. (1998). Formative Assessment: Raising standards inside the classroom. *School Science Review*, 90(291). 39-46.

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., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working Inside the Black Box: Assessment for learning in the classroom*. London: King's College.

Black, P., & William, D. (1998a). Assessment and Classroom Learning. *Assessment in Education, 5*(1), 7-74.

Black, P., & William, D. (1998b). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan, 80*(2), 139-148.

Butler, R. (1988) Enhancing and undermining intrinsic motivation; the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*, 1-14.

Butler, R. & Neuman, O. (1995) Effects of task and ego-achievement goals on help-seeking behaviours and attitudes. *Journal of Educational Psychology, 87* (2), 261-271.

Chi, M. T. H. (1992). Conceptual Change within and across Ontological Categories: Examples from Learning and Discovery in Science. In R. N. Giere (Ed.), *Cognitive models of science: Minnesota studies in the philosophy of science* (Vol. 15). Minneapolis, MN: University of Minnesota Press.

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

Daws, N. & Singh, B. (1996) Formative assessment: to what extent is its potential to enhance pupils' science being realized? *School Science Review*, 77 (281), 93-100.

Daws, N. & Singh, B. (1998). Formative assessment strategies in secondary science. *School Science Review*, 80(293), 71-78.

Doran, Lawrenze & Helgeson (1994) Research on Assessment in Science. In D. Gabel, (Ed.), *The Handbook for Research on Science Teaching and Learning* (pp388-442). New York: Macmillan.

Driver, R., Guesne, E. & Tiberghien, A. (Eds.) (1985). *Children's ideas in science*. Buckinghamshire: Milton Keynes.

Duschl, R. A. (2003). Assessment of Scientific inquiry. In J. M. Atkin & J. Coffey (Eds.), *Everyday Assessment in the Science Classroom* (pp. 41-59). Arlington, VA: NSTA Press.

Duschl, R. A., & Gitomer, D. H. (1997). Strategies and Challenges to Changing the Focus of Assessment and Instruction in Science Classrooms. *Educational Assessment, 4*(1), 37-73.

Furtak, E.M & Ruiz-Primo, M.A. (2005, January). Questioning Cycle: Making Students' Thinking Explicit During Scientific Inquiry. *Science Scope,* p. 22-25.

Hogan, K., Nastasi, B. K., & Pressley, M. (2000). Discourse Patterns and Collaborative Scientific Reasoning in Peer and Teacher-Guided Discussions. *Cognition and Instruction, 17*(4), 379-432.

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

Minstrell, J. (1992, April). *Facets of students' knowledge: A practical view from the classroom.* Paper presented at the annual meeting of the American Educational Research Association. San Francisco.

Minstrell, J., & vanZee, E. (2003). Using Questioning to Assess and Foster Student Thinking. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday Assessment in the Science Classroom* (pp. 61-73). Arlington, Virginia: NSTA Press.

National Research Council. (1996). *National Science Education Standards*. Washington, D.C.: National Academy Press.

National Research Council. (2001a). *Classroom Assessment and the National Science Education Standards*. Washington, D.C.: National Academy Press.

National Research Council. (2001b). *Inquiry and the National Science Education Standards*. Washington, D.C.: National Academy Press.

Pellegrino, J. W., Baxter, G. P., & Glaser, R. (2000). Addressing the "two disciplines" problem: Linking theories of cognition and learning with assessment and instructional practice. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education, Volume 24* (pp. 307-353). Washington, DC: American Educational Research Association.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington D.C.: National Academies Press.

Resnick, L.B, & Resnick, D.P. (1991) Assessing the Thinking Curriculum: New Tools for Educational Reform. In B. Gifford (ed.) *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction*. MA: Kluwer

Rowe, M.B. (1974). Wait time and rewards as instructional variables, their influence on language, logic and fate control. *Journal of Research in Science Teaching, 11*, 81-94.

Ruiz-Primo, M. A., & Furtak, E. M. (2004). *Informal Assessment of Students' Understanding of Scientific Scientific inquiry*. Paper presented at the American Educational Research Association Annual Conference, San Diego, CA.

Ruiz-Primo, M.A., Li, M., Ayala, C., & Shavelson, R. J., (in press). Evaluating students' science notebooks as an assessment tool. *International Journal of Science Education.*

Sadler, D. R. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science, 18*, 119-144.

Schmidt, W. H., McKnight, C. C., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). *Facing the Consequences: Using TIMSS for a Closer Look at U.S. Mathematics and Science Education*. Dordrecht: Kluwer Academic Publishers.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Scott, P. (2004). Teacher talk and meaning making in science classrooms: A Vygotskyian analysis and review. In J. Gilbert (Ed.), *The Routledge Falmer Reader in Science Education* (pp. 74-96). London: Routledge Falmer.

Shavelson, R.J., Black, P.J., Wiliam, D., & Coffey, J.E. (2002) *On Aligning Formative and Summative Assessment*. Paper presented at the National Research Council's Assessment In Support of Instruction and Learning: Bridging the Gap Between Large-Scale and Classroom Assessment Workshop, Washington, DC, January, 2003.

Shavelson, R.J., & the Stanford Education Assessment Laboratory [SEAL] (2003). *On the integration of Formative Assessment in Teaching and Learning with Implications of Teacher Education.* Paper Presented at the Biannual Meeting of the European Association for Research on Learning and Instruction. Padova, Italy

Shavelson, R.J., & Towne, L. (Eds.) (2002). *Scientific research in education*. Committee on Scientific Principles for Education Research. Division on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Shepard, L.A. (2003). Reconsidering Large-Scale Assessment to Heighten its Relevance to Learning. . In J. M. Atkin & J. Coffey (Eds.), *Everyday Assessment in the Science Classroom* (pp. 41-59). Arlington, VA: NSTA Press.

Stern, L. & Ahlgren, A. (2002) An analysis of student assessments in middle school curriculum materials: Aiming precisely at benchmarks and standards. *Journal of Research in Science Teaching*, 39, 889-910.

vanZee, E., & Minstrell, J. (1997). Using Questioning to Guide Student Thinking. *The Journal of the Learning Sciences, 6*(2), 227-269.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction, 16*(1), 3-118.

White, R., & Gunstone, R. (1992). *Probing Understanding*. London: Falmer Press.

Wilson, M. (Ed.) (2004). *Towards Coherence Between Classroom Assessment and Accountability. Finish citation.*

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Windschitl, M. (2004). Folk theories of "inquiry:" How preservice teachers reproduce the discourse and practices of an atheoretical scientific method. *Journal of Research in Science Teaching, 41*(5), 481-512.

Yin, Y. (2005). *The influence of formative assessments on student motivation, achievement, and conceptual change.* Unpublished doctoral dissertation, Stanford University.