What the Student Evaluation Standards say about Formative Assessment

E. Caroline Wylie

Cynthia Tocci

ETS, Princeton, NJ

Paper presented at the annual meeting of the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME) held between April 13 to 17, 2009, in San Diego, CA.
Abstract

The Joint Committee on Standards for Educational Evaluation published the Student Evaluation Standards in 2003 (JCSEE, 2003) presenting “the first unified, national discussion of how to plan, conduct, use, and judge classroom-based student evaluations (p.1).” In this paper we review the Student Evaluation Standards with the particular view of formative assessment. Formative assessment is described in some detail to provide a framework for the review and remainder of the paper discusses each of the 28 standards. The standards are grouped in terms of those that are specifically related to the core activities of formative assessment, those that relate to formative assessment although the connections may not be as clear as for the first group as the standards are currently written, and finally those that are seldom applicable to formative assessment.
Introduction

The Joint Committee on Standards for Educational Evaluation published the Student Evaluation Standards in 2003 (JCSEE, 2003) presenting “the first unified, national discussion of how to plan, conduct, use, and judge classroom-based student evaluations (p.1).” Many people contributed to writing of the standards, and participated in both a national and international review, along with a field test of the semi-final draft. Calls for improved assessment literacy within the educational system have been ongoing (Stiggins, 1991, McMunn, McColskey, Butler, 2003; Brookhart, 2003), and the Student Evaluation Standards make a useful contribution to that end.

Although the topic of student evaluation is a broad one, the JCSEE chose to limit it to classroom evaluation, noting that their focus “stems from the belief that strong student learning requires consistent, persistent – indeed, daily – attention to effectively gathering, analyzing and using evaluation information to guide student learning (p.2).” Given the focus on classroom evaluation, it is not then surprising that the primary audience for the Student Evaluation Standards is teachers, those with the every-day responsibility for assessment, although others within a school or district will find it useful. The introduction further explains the focus on student evaluation and teachers’ roles since it is critical that teachers know “what each student knows and can do competently and what must be done to further develop and encourage the student’s educational development (p.3).”

The standards themselves are divided into four categories: propriety, utility, feasibility and accuracy. These four major categories provide clear articulations of the various aspects of evaluation in order to meet the JCSEE desire “to help teachers and others who evaluate students plan and conduct sound, trustworthy evaluations and report results of these evaluations in an appropriate, accurate, and credible manner (p.3).” Altogether there are 28 standards with seven propriety standards, seven utility standards, three feasibility standards and eleven accuracy standards. It is important to note that the JCSEE was clear from the outset that not all standards are applicable in every evaluation situation.

As the title of this paper suggests, we have reviewed the Student Evaluation Standards with the particular view of formative assessment. The standards use the term evaluation to refer to the “systematic investigation of the worth or merit of a student’s performance in relation to a set of learner expectations or standards of performance (p.228).” In our work we focus on formative assessment – or assessment for learning – as a subset of assessment where the focus is on evaluating students’ performance at a point during instruction when adjustments can still be made by either teachers or students to improve the performance (Wylie, 2008). In this paper we will use evaluation and assessment synonymously.
The task of reviewing the Student Evaluation Standards suggests two questions. First, to what extent do the standards support someone who wishes to incorporate more formative assessment opportunities into his or her practice? Second, are there important aspects of formative assessment practice either not reflected in or perhaps contradicted by the standards? Although the term formative evaluation only appears in the glossary of the Student Evaluation Standards, the concept of formative assessment is certainly part of the classroom evaluation process as described by the standards. As already noted above, these standards are inclusive of the kinds of assessment that may take place on a daily basis. Formative assessment will be described in more detail in the next section in order to provide a clear foundation for the review of the standards; the review will form the bulk of this paper.

**Formative Assessment**

In education dichotomies are often used when it comes to discussing assessment: objective versus subjective assessments; summative versus formative; high stakes versus low stakes. While there is some value defining a concept in stark terms of what it is and is not, important nuances can be lost. In this section we will present several definitions of formative assessment without resorting to dichotomous contrasts. Although the term “formative assessment” is not found in the Student Evaluation Standards, the phrase “informal evaluation” is used. However, this lacks some clarity in terms of what counts as an informal assessment and what counts as formal. In the discussion of several of the standards we will draw attention to how this language may lead a reader to treat formative assessment as something that requires less care and attention than a summative assessment.

Formative assessment has been defined in various ways. One definition comes from the work of Black, Harrison, Lee, Marshall, and Wiliam, (2003) who described it as occurring when “information about learning is evoked and then used to modify the teaching and learning activities in which teachers and students are engaged [emphasis in the original]” (page 122). The Council of Chief School Officers (CCSSO, 2008) have defined 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.” For the purpose of this paper we also present a definition that has been used with our colleagues in the Learning and Teaching Research Center at ETS as we have engaged teachers in thinking about formative assessment. We define the overarching principle of formative assessment as “Students and teachers using evidence of learning to adapt teaching and learning to meet immediate learning needs minute-to-minute and day-by-day”(ETS, 2007).
There is significant agreement across the three definitions of formative assessment proffered, in large part due to the fact that each one was developed with the same research basis for formative assessment in mind. The extensive literature review conducted by Paul Black and Dylan Wiliam (Black & Wiliam, 1998) along with other work such as that of Brookhart (2005) and Nyquist (2003) illustrate that there is a strong and positive connection between teachers’ use of formative assessment in everyday teaching and improved student learning. From a review of the research supporting formative assessment, and more cursorily from the definitions presented above, formative assessment is tightly connected to what happens “in the moment” of teaching. The ETS definition has the most explicit reference to time, but all the definitions suggest that in order to really influence the teaching and learning in a classroom, the formative assessment process must happen far more frequently than just quarterly or monthly.

The CCSSO definition of formative assessment is further explicated through what they define as five attributes that clarify and expand on the concepts inherent in the definition. These attributes include:

- Learning progressions
- Learning goals
- Specific feedback
- Collaboration
- Self- and peer assessment.

Based on the Black and Wiliam research, the ETS definition of formative assessment is broken down into five key strategies (ETS, 2007), which are as follows:

- Clarifying and sharing learning intentions and criteria for success
- Engineering effective classroom discussions, questions, and learning tasks that elicit evidence of learning
- Providing feedback that moves learning forward
- Activating students as owners of their own learning
- Activating students as instructional resources for one another

The purpose of breaking apart or expanding on a definition of formative assessment is to help practitioners think about the scope of what it is, and also to identify what it is not. For example, learning intentions, or learning goals form a critical map for teachers: to illustrate what kinds of learning are important prerequisites for new knowledge and skills, to measure against to determine where learning is currently located, and to indicate where learning should be headed. Discussions, questions, learning tasks all provide opportunities to collect evidence of student learning, and it is critical that teachers have enough time and supporting resources as they work to engineer these events to ensure that they get the
precise information they need. In order to impact instruction, teachers often have to make instructional adjustments. This can take the form of feedback to students to help them see both where they are in their learning and to nudge them along to the next stage. Students have critical roles: both learning to be self-reflective of their own learning and to be supportive of the learning of their peers. Engaging students in this way can increase the number of learning and feedback opportunities. In our work at ETS we also refer to what we call the “formative assessment cycle” of collecting, analyzing and using evidence to inform instruction. We have found this three phase division of formative useful for teachers since the challenges of each phase are different.

William (2004) noted that “in order for assessment to function formatively, it needs to identify where learners are in their learning, where they are going, and how to get there. (p.5). Although roles and responsibilities differ, teachers and students alike need to know where learning is intended to go, have ways of identifying current position with respect to those goals, and most importantly, when there is a gap between current position and desired goals, make use of the assessment information to help close the gap. As we discuss the Student Evaluation Standards we will use these questions as a framework – Where is the intended learning going? Where are students currently? How can the gap be closed? – to consider how the standards relate to formative assessment.

None of the discussion above on formative assessment contradicts what is presented in the Student Evaluation Standards and much of it resonates closely with how the JCSEE present classroom assessment. Before beginning a review of the standards, we will conclude this section with one description of a teacher’s use of formative assessment that illustrates how several of the strategies listed above can be combined. This example originally appeared in a document that one author collaborated on for CCSSO (CCSSO, 2008) although the analysis has been tailored for this paper.

An elementary language arts teacher began the lesson by first reminding the students about their reading learning goals for the week which focused on identifying the main idea and supporting details within a story. The teacher then asked a series of planned questions about a story that students had just finished reading. Her questions required careful analysis from the students, so the teacher structured the task by having the students think about their answers as individuals first, and then discuss them in small groups to select the best answer. This group answer was then shared with the rest of the class using Whiteboards which designated students held up. With this questioning and group work approach, the teacher was able to identify several groups of students who seemed to be struggling. Student summaries of the main idea of the story varied widely in accuracy and clarity, and the teacher was surprised that students were unclear about it.
As the lesson was nearing the end, she asked the students to look at the various groups’ answers about the main idea, to select the one that they thought was the best answer, and to write down why they made the choice they did. She had students answer using an Exit Ticket – index cards on which students wrote their individual answers and then handed to her as they left the classroom. This approach provided her with a quick way to review student thinking at the individual level, thus providing information that she could use to shape the next day's lesson, depending on how many students within the class were struggling with a particular topic under study.

This example of formative assessment incorporates various aspects of formative assessment practice during a single lesson. The analysis of the evidence deepened the teacher’s understanding of where her students were in their learning, and informed her instructional decision making. This teacher’s practice embodied several of the formative assessment strategies listed previously (ETS, 2007). The teacher had clear learning intentions that she communicated to her students. Ahead of time she developed a series of questions to ask her students to ensure that the classroom discussion went in the direction she wanted to support student learning. One aspect of engineering a classroom discussion is to consider how students should respond to questions. The teacher planned a systematic way to allow students to think deeply about the questions, and to share their thinking with her. In order for this to be accomplished, the teacher had, over time, established a learning environment that emphasized collaboration: students were used to working in small groups and using whiteboards was part of the classroom routine so that the teacher could use these evidence-gathering approaches with little explanation required. Finally, the teacher was able to elicit evidence of student learning using the exit ticket, to support the impression that she had of group difficulties during the lesson, an approach that required little time to review. With this student-level evidence of understanding she was able to tailor her lesson the following day to capitalize on those students who had a deeper understanding of the learning goal.

Returning to the three questions that Wiliam (2004) raised as central to formative assessment, the teacher in the example above understood where she wanted her students to go and communicated those learning goals to her students, she identified where they were with respect to those learning goals through a variety of approaches, and at the end of the lesson she had enough information to support her decisions for how to structure the next day’s lesson to help close the gap for some students.
Review of the Student Evaluation Standards

The purpose of this paper is to examine the Student Evaluation Standards (JCSEE, 2003) and how the previously mentioned definitions of formative assessment relate to these standards. As we reviewed the standards we identified three groups. The first group contains those standards that are specifically related to the core activities of formative assessment. The second group of standards contains those that relate to formative assessment although the connections may not be as clear as for the first group as the standards are currently written. The final group of standards contains the ones that are seldom applicable to formative assessment. The appendix provides a table summarizing our judgments. The fact that the standards could be classified in this way was not surprising, given that the document itself early on stated that the writers did not expect all standards to be applicable to all situations.

In the next three sections we will review the standards as they fell into the three groups, beginning with those standards that specifically relate to the core activities of formative assessment. Although a number of the standards apply to the entire formative assessment process, we will draw attention to specific standards as they address the three guiding questions of formative assessment that were identified earlier – Where is the intended learning going? Where are students currently? How can the gap be closed?

Student Evaluation Standards Specific to Core Formative Assessment Activities

Across the 28 Student Evaluation Standards, we considered the majority of the standards (17 in all) to relate to the core of formative assessment and critical for someone focusing on their formative assessment practice. Each one is discussed below.

P1: Service to Students

The first standard in the Propriety group focuses on the overarching goal of student evaluation, that it be in service to students and should thus “promote sound education principles, fulfillment of institutional missions and effective school work (p.29).” The standard breaks this goal into two distinct aspects, first to help students and others understand where they are located with respect to instructional goals and second to “help students, parents/guardians and teachers plan future instruction, and where needed, appropriate follow-up remedial action (p.29).” The first aspect seems to have a summative assessment orientation, although not exclusively. The second aspect has a strong formative assessment bent, in particular the teachers’ and students’ use of the assessment information for planning future instruction, and all the more so, if the “future” refers to the very near future,.
The caveat provided for this standard is a useful one. All too often the process of assessment and grading becomes one of threats and pressures attempting to goad students into better performance (Stiggins, 2002). The caution that evaluation ought to “serve instruction and learning needs of students” is an important one, although in the service of formative assessment, we would reword the second sentence of the caveat to read, “teachers must ensure that formative evaluations primarily serve as guidelines for teachers’ next instructional steps and as signals to students illustrating where they currently are on the learning continuum and provide feedback to support thinking on how to improve.” All too often assessment, given in the form of grades, only confirms a student’s opinion of themselves for either good or ill, and the grade alone does nothing in terms of directing next steps.

Although the assessment in the first illustrative case is described as learner centered, it does not seem to have a formative assessment component. In the analysis, the assessment is described as an opportunity for students to “demonstrate their knowledge and skills.” This approach provides students with a positive attitude to the assessment process and multiple opportunities, but does not suggest that the teacher views the assessment data as valuable with respect to modifying instruction. Below is a brief description of a teacher, observed by one of the authors, engaging in a formative assessment in which her actions were very transparent to her students.

**Description:** Ms Dunn was part way through a physics unit on properties of light. At the start of the lesson students were asked to respond to one question posted on the white board addressing a major idea that had been covered in several previous lessons. Students wrote their response on a post-it note and stuck it to the white board. Without indicating who provided a particular response, the teacher reviewed each response and engaged the class in a discussion. Responses were then classified by the students as “correct”, “almost there,” or “not quite.” The majority of the responses fell in the “not quite” category. After the review was complete, the teacher asked the class, “So what does this tell Ms Dunn?” Several students mumbled things like “we don’t get it yet” or “we’re stupid” and the teacher quickly interrupted those responses and said “No! It gives me information to help me decide what to do next and I think I need to get you to complete one more activity to help you understand this concept a little better before we move on.”

**Analysis:** The teacher had a very specific goal in mind with this evaluation: to determine whether to move on to the next part of the unit or to spend additional time reinforcing the concept that students had been working on. She also made this very clear to the students.
They were not being graded on their responses. The purpose of the assessment was entirely to inform the teacher. In addition, it is worth noting that the teacher had come to the lesson prepared to either continue with the unit as planned, or to complete an additional activity to reinforce the concept if it turned out that students were still struggling. Furthermore, the teacher conveyed her attitude that it was her responsibility to ensure that students had sufficient learning opportunities, rather than suggesting that it was the students’ fault that they had not yet learned this particular concept.

The teacher in the example above used the opening question in class to help her decide the best possible use of the class time to meet students’ learning needs. She conveyed to students where they were currently in their learning with respect to the goals that she had articulated for them, and provided an additional learning opportunity to deepen their understanding. Her actions exemplified formative assessment and service to students. The teacher was able to answer all three formative assessment guiding questions: she understood where her students’ learning was headed, used the question at the start of the lesson to determine where they were with respect to that goal, and adjusted her instruction to close the gap.

**P4: Treatment of Students**

The Treatment of Students standard focuses on the ethical and fair treatment of all students in the assessment process. This standard certainly is important for all assessment processes, and no less so for formative assessment. One of the guidelines under this standard focuses on encouraging students to take responsibility for their own learning.

In our work, we have seen these ideas of fairness and responsibility play out in a variety of ways as teachers incorporate formative assessment into their practice. Making learning intentions clear to students, in language that they can understand can certainly be considered an aspect of fairness: ensuring that all students are meaningfully accessing the instruction. Many of the teachers with whom we work, in an effort to increase the level of participation during class discussions are using Popsicle sticks – students names are each written on a Popsicle sticks and all are stored in a container. During a discussion, the teacher will ask a question, allow all students to consider their response and then draw a Popsicle stick to select a student at random to respond. In addition to ensuring that the teacher consistently is hearing from a wider range of students in the class, and often increasing attention and participation, students tend to respond favorably to what they see as the inherent fairness of this approach. (As an aside, some teachers initially have concerns about this approach for fear of selecting a student unable to respond. However,
with thoughtful attention to possible student response options (e.g., phone a friend for help), and a classroom climate where students feel safe to admit uncertainty in an answer or to provide an answer that proves to be incorrect, those teachers have found ways to use this technique in a way supports student learning.)

Encouraging students to engage in self-reflection, an important aspect of formative assessment, provides evidence of student perceptions of their learning for teachers, and engages students in the assessment process. While a teacher should not rely on student self-assessment as the only evidence source a great deal can be learned just from asking students at the end of a lesson whether they think that they have mastered whatever was the learning goal for that lesson. This self-assessment could take the form of “thumbs up, thumbs down” to indicate an overall comfort level with the topic, or by asking students in turn to state one thing that they learned in the lesson. Engaging students in this process, letting them know that the teacher really cares to know how they are doing with the learning, and demonstrating a willingness to do something with that information creates an assessment climate that is both fair and teaches responsibility.

**P5: Rights of Students**

The Rights of Students standard, similar to the previous one, cuts across all assessment events. The standard provides guidelines for what students have a right to know for various evaluations. As for other standards, the burden is higher for evaluations that carry greater consequences. From a formative assessment perspective, the guideline that seems most relevant is the one that advises informing students about “the benefits and consequences of evaluations.” This concept is closely related to the content of the P1 Standard focused on how an assessment can be of service to students. For some students, assessment has become a negative experience because they see it as confirming what they think they know already: they aren’t smart, they can’t learn etc. Formative assessment is an opportunity to change that perspective, to use assessment for informational purposes at a point during instruction when the outcomes can be changed. Students have a right to see that their teachers are engaged in helping them learn and that they are continually seeking out information to help them do it better. They deserve nothing less.

**P6: Balanced Evaluation**

The Balanced Assessment standard focuses on the importance of student evaluations to provide evidence of both strengths and weaknesses “so that strengths can be built on and problem areas addressed (p.55).” This standard speaks to a critical aspect of formative assessment: it is hard for a teacher to adjust instruction or students to adjust learning strategies if they do not know where they are going. The text
provides some frames of references that can be used for assessments, but it seems that for formative assessment the more useful frames of reference will be those that allow performance to be understood in relation to specified standards or the amount of improvement. Comparisons to peers seem much less useful and potentially problematic (Dweck, 2006). Clear, and meaningfully connected, learning intentions provide a road map for teachers and students to understand where instruction is going. Helping students understand their strengths with respect to the road map provides the encouragement to continue then to work on the weaknesses with continued support.

Below is a brief description of teacher engaging in a formative assessment in a manner in which balanced evidence of learning is provided.

*Description:* Ms Clements’ 4th grade students were engaged in a writing project to develop short articles for the school magazine. Her students had never written something that was going to be published and were both excited and nervous at the thought of writing for an audience other than their teacher. She was helping them write initial drafts, providing them feedback to help them think about what to do in their second draft of their writing. She introduced them to the concept of “two stars and a wish” where she provided two comments on positive aspects of the writing, such as how the writer had grabbed the reader’s attention, and one comment that was something she wanted them to especially work on for their next draft.

*Analysis:* The teacher provided feedback for students that both encouraged them by drawing their attention to the strengths of their writing, but also providing some concrete guidance for how they might improve their writing in the next draft.

Returning again to the three questions that guide formative assessment, both the teacher and students in this example understood that the purpose of the writing was to engage the readers of the school magazine, the teacher read the writing drafts to gauge current status of her students, and provided feedback in terms of two stars and wish in order to help the students close the gap.

*U1: Constructive Orientation*

The Constructive Orientation standard identifies the importance of student evaluations being constructive in order to “result in educational decisions that are in the best interests of the students (p.67).” This standard points to the need for timely information that “enables students to direct their
energies effectively, provides parents/guardians with information they can use to assist their children, and helps educators plan and implement appropriate opportunities to learn (p.67).” In addition under the guidelines is specific advice about providing “timely evaluation information and findings to all stakeholders (p.68)” Three stake-holder groups are clearly identified, but no distinction is made between the groups and what “timely” might mean for each group. From a formative assessment perspective, timely information for a teacher could very well be at the end of, or even during, every lesson, for students it could be just as frequent with assessment information taking the form of feedback from the teacher, self reflection on their own level of understanding, or comments from a peer reviewing a piece of work. For parents/guardians timely information might be less frequent, unless a significant issue emerges that requires immediate attention.

One caveat provided in the text states “Evaluations that identify weaknesses but do not assist in correcting deficiencies contribute little to the students, their parents/guardians, or other legitimate users (p.67).” Here again it seems that the teacher’s use of the information is relegated, obliquely at that, to third on the list. It does however identify a critical issue that teachers often struggle with. Collecting evidence of student learning and analyzing it is one thing, but taking that information and making instructional adjustments can often prove challenging. Below is a brief description of teacher engaging in a formative assessment of his students where information was collected and analyzed in a timely way but where the teacher struggled to connect what was learned back to future instruction.

**Description:** Mr. Caldwell was teaching a pre-Algebra course and ended a lesson with an exit ticket question with the problem $\frac{1}{x} + \frac{1}{y} = ?$. Only a couple of students were able to answer it, even though every student in the class at the start of the period had been able to answer the question $1/3 + 1/5 = ?$. When a colleague asked the teacher what he did in the next lesson, Mr. Caldwell responded with “Used another starter.” He colleague probed to see if Mr. Caldwell had done anything with the results and found out that he had not.

**Analysis:** The teacher was interested in finding out whether the students were able to apply what they knew about adding fractions to a problem that incorporated variables. He analyzed their responses and came to a correct understanding that in fact his students were probably only able to use an algorithm to solve the fraction problem and did not have sufficient conceptual knowledge to apply what they knew to the algebraic version of the problem. While the teacher had timely information he did not consider incorporating that knowledge into what he would do next in class. In fact, as a result of the continued
conversation with his colleague he was able to think about various ways the information
could have been used (e.g. having students who understood the concepts explain their
thinking to the whole class.) and gained some new understanding about the formative
assessment cycle – collect-analyze-use – and how using evidence of learning would help
him adapt his teaching to meet immediate learning needs.

In the example above, the teacher had addressed the first two guiding questions of formative
assessment: he knew where his students ought to be and he collected evidence of their location with
respect to the learning position. However, it had not occurred to the teacher or he was unable to take any
action that would help close the gap between desired and current location of his students.

Thinking about a constructive orientation to assessment also requires a clear and in-depth
understanding of the learning intentions, and not just what activities that a teacher wants students to do.
Without a clear vision of the intended learning, it is difficult to collect appropriate evidence of that
learning.

The importance of professional development is addressed in Standard F3, Evaluation Support, but
has relevance here also. Supporting teachers as they change their attitudes to assessment, develop new
assessment strategies, and consider how they present it to their students can sometimes require significant
support.

U2: Defined Users and Uses

The Defined Users and Uses standard focuses on both the users and uses of evaluation
information and provides a range of decisions that students, teachers and parents/guardians might need to
make based on appropriate evaluation information. While not describing it as formative assessment, the
text mentions how evaluation information can “be used to make informal, low-stakes decisions in day-to-
day classroom teaching (p.72).” This is not the only place where what we would consider to be formative
assessment is described as “informal assessment.” The one caution we would urge is that informal not be
taken to mean “casual” or “haphazard.” In Popham’s definition of formative assessment (2008) he
describes it as “a planned process (p.6)” which is of critical importance. Deciding on a single question,
one that gets at the heart of the main idea or concept just taught, to ask students as they are finishing a
lesson can rarely be achieved on-the-fly, but requires some forethought and planning.
U6: Effective Reporting

The Effective Reporting standard focuses on reporting of evaluation results so that the information is of maximum use to students, parents and teachers. The need for reports to be “clear, timely, accurate and relevant (p.98)” is true regardless of what form the assessment takes. There can be a temptation to consider this standard to apply primarily to summative reports although the text does not explicitly say so, but “reports” tends to bring to mind end-of-quarter or end-of-year report cards. However, the second illustrative case deals with a teacher providing feedback to students after a particular lab, although in the example, the teacher is focused on giving grades, albeit with clear explanations to justify the grades, rather than giving feedback to improve future performance.

When thinking about feedback to students, the four qualities of reports still apply. Feedback needs to be clear – in language that students can understand; timely – with an opportunity for students to internalize the information and apply it to their work again before a final grade; accurate – a consistent set of standards or expectations should be applied to all work and where possible the rubric shared with students so that those expectations are clear and transparent; and finally, relevant – connected clearly to current learning, perhaps reflecting language from learning progressions.

Effective feedback is an important part of formative assessment. It is the bridge between where students ought to be and their current location. However, if neither the teacher nor the students have a clear understanding of the learning goals, it may in fact be a “bridge to nowhere.” This standard on effective reporting is tightly connected to the next one which focuses on follow-up. Feedback, or reporting, can truly only be effective if in fact there is follow-up. If evaluation information only results in students’ reaction to the information, rather than action as a result of the information, it is not effective.

U7: Follow-up

The Follow-Up standard focuses on procedures that provide opportunities for all those who receive evaluation information have an opportunity to review it, understand it and take follow-up actions. This standard is critical when it comes to formative assessment: there is no value in a teacher providing formative feedback to students if students are not given an opportunity to directly act on the feedback and to apply the comments to the work that the comments were based on. On occasion, a teacher might spend considerable time providing feedback to students but fail to put structures in place to ensure that students benefit from that feedback.

In the first illustrative case that accompanies this standard, a teacher realized as she graded a homework assignment that students were struggling with a particular concept, so she corrected the error on the homework, provided additional instruction and problems in class, and assigned a second
homework to assess student progress. From responses to this homework she concluded that the majority of students then understood the concept and she could follow-up separately with those few individuals who were still struggling. Interestingly, in this case it appears that the feedback came to the teacher from her marking the homework, and although she corrected errors on the homework there was no opportunity for students to review her corrections. In other words there seems to have been limited value in her actually making corrections since there was no structure in place for students to review their work and take follow-up action on their own. The example as it currently stands does illustrate how a teacher can respond to information, and take appropriate action to remediate before trying to progress to the next topic, but it does not illustrate how students can be supported to apply feedback to their own work to deepen their learning.

Below we suggest an alternative approach that would make a clearer use of her comments to students and help them take additional responsibility for deepening their own understanding.

**Description:** Mrs. Johnson realized that many students in her class were struggling with an algebra concept. She marked students’ homework either by circling errors, and for some of her weaker students scaffolded her feedback more by providing the next step for some problems. The next day in class she grouped students together in such a way that in each group students had a variety of difficulties and at least one student (not always the same student) got each problem correct. The teacher directed the students to look at what she had circled on their homework, to refer back to their class notes and textbook and to talk to each other in the group to work through the homework problems and to make corrections. As students worked together the teacher circled the room to help students explain the concepts to each other and to provide additional help where needed. As she listened to discussions, she realized there was one problem that all students were struggling with, even those who had answered correctly were not sure of what they had done, so she pulled the class together for some whole-class teaching on that one issue.

**Analysis:** The teacher ensured that students had an opportunity to review the corrections she made to their work. She structured the groups so that students could learn from each other, which meant that she did not have to review concepts that some students had mastered. It also provided her an opportunity to see whether those students who had answered correctly had a true understanding of the concepts and when she discovered an area of confusion she was able to step in with some targeted direct instruction.
Returning to the three questions that guide formative assessment (Wiliam, 2004), the third one asks “How can the gap between intended learning and current learning be closed?” To answer this question often requires action on the part of the teachers, such as making an instructional adjustment, but can also require student action. If a teacher provides feedback to students, in order for it to be beneficial students must be provided with an opportunity to do something with it.

**F3: Evaluation Support**

The Evaluation Support standard speaks to the need for “adequate time and resources (p.119) for student evaluations and although not directly mentioned in the standard summary, in the body of the text for this standard, the importance of professional development is noted. In relation to a previous standard we alluded to the fact that much of our own work with respect to formative assessment has in fact been in the context of professional development.

In the introduction to this paper, we introduced the five strategies that we use to define the domain of formative assessment. From them, it is clear that there is a close relationship between formative assessment and instructional practice. For a teacher to incorporate those strategies into his or her practice, to find ways of translating general principles to specific practice, requires ongoing support and assistance. As teachers gain experience much of what they do becomes routine and automatic: classes begin following certain routines, classroom discussions have a particular ebb and flow; teacher expectations find a particular level. The five strategies might require a teacher to interrupt those habits and introduce something new. Just like learning to use a new computer program or increasing the amount of daily exercise, time is required for new habits to become routine, and often people slip back to old habits. At the beginning, certain practices may not even feel comfortable – such as deliberately giving students additional wait time in order to improve the quality of their thinking and responses – and teacher need encouragement to persevere with new approaches until they start to see some returns for their efforts. Finally, increasing the amount of formative assessment may alert teachers to the fact that some of their students did not understand a particular concept first, or even second time around. Teachers may need help identifying new ways to approach their teaching as they realize that knowing that they need to adjust instruction is quite different from knowing how to adjust instruction. Furthermore, teachers may need support reevaluating a pacing guide: some content may need to be omitted in order to have enough time to ensure students understand the critical content. The non-critical content ought not to just be the last two chapters that no-one gets to, but should be determined through careful review and discussion.
In our work with teachers around the issue of formative assessment we have emphasized the role of teacher learning communities as a critical component. For teachers to make changes to their practice as described above they need to have opportunities to consider their practice, reflect on changes they are attempting to make, and receive feedback from colleagues. Thompson and Wiliam (2008) identify this type of teacher learning as requiring “repeated cycles of learning, practice, reflection, and adjustment (p.12)””. When professional development is in the form of day or two during the school year these cycles of learning are not supported. Teacher learning communities that are school-based and meet throughout the year are much more suitable. The kinds of thinking and reflection that the Metaevaluation standard (A11) talked about can be supported in the teacher learning communities.

A1: Validity Orientation

The validity orientation standard speaks to the need to develop student evaluations such that the interpretations or conclusions made are valid. This is an important standard since the value in the formative assessment lies not in the assessment but in the interpretations made in the light of the evidence of student learning and the actions that a teacher takes based on the interpretations: to slow down, re-teach, move more quickly, divide students into groups, to name but a few possible actions. Ensuring that interpretations are valid requires careful attention to a chain of events: a clear understanding of the assessment target; an appropriate means for collecting information; opportunities for students to demonstrate their knowledge or skills; scoring carried out with due care; the information analyzed appropriately; and reasonable inferences made about student learning. This chain of events can be broken at any point resulting in flawed assessment results.

For formative assessment, the stakes are lower than might be suggested by the blanket statement “invalid inferences or judgments can do great harm (p.128).” If a teacher makes an incorrect assumption about the level of student learning and moves on to a the new topic in a sequence of learning even though students are struggling, unbeknownst to her, the consequences of this action might just be that students struggle with the concept, requiring her to back-track, find the earlier source of confusion and re-teach. On the other hand, the cumulative consequences of a teacher consistently making inappropriate judgments over the course of a year, or never looking for evidence of student learning in the first place, could be much higher.

Although it is hinted at in this standard, it seems that a foundational piece of information that teachers must have, before beginning any instruction or assessment, is a clear map of the domain being taught. Not all state standards are of a quality to provide what is needed: it may need to be in the form of a learning progression (Heritage, 2008) or learning targets (Stiggins, Arter, Chappuis & Chappuis, 2004).
Formative assessment can help a teacher bridge the gap between where a student is now and where the teacher intends them to be. But without a clear learning map of the territory, it is difficult for a teacher to move students from “here to there.” Thus, while all the requirements for valid interpretations may still hold true in one form or another, the need for a clear understanding of the domain is of utmost criticality.

In illustrative case one, while all aspects of the validity chain are in place in that instruction is connected to the learning objectives which are in turn connected to the assessments, we would contend that the process is somewhat flawed because the learning objectives are very weak. Many of them are not learning intentions but activities, for example, observe the conditions…, develop a testable hypothesis…, identify the manipulated variable… (p.131) [emphasis added].” Some of what are listed as potentially success criteria that a teacher might use to determine if the learning had taken place. However, the specifics of the intended learning have not been spelled out. It is not clear why students should learn about bats other than the fact that the teacher is clearly interested in them. Without knowing what the major learning outcomes are from this unit, it is impossible to guide students through and to know what is a critical gap in learning that must be addressed in this unit, and what could be reviewed on later as the concept appears in future learning even if the context of bats is different.

For a low stakes, formative assessment the need for a full-blown validity argument is not warranted, but grounding in clear learning intentions is critical. Thus, being able to answer the first question that ought to drive formative assessment thinking – what is the intended learning? – is a vital starting point for ensuring that appropriate formative assessment data are solicited from students. The second guiding question – where are students currently? – requires attention to how evidence from students is interpreted.

**A2: Defined Expectations for Students**

The Defined Expectations for Students standard calls for students to know “what they are to learn and how they are to demonstrate what they have accomplished. In our work we refer to this as “learning expectations” and “success criteria” respectively, and they form the basis of one of the five key strategies of formative assessment (ETS, 2007). Under the list of common errors, while we agree that an error is failing to use some form of pre-assessment with entering students, we would go further than that to encourage teachers to assess students against the learning expectations all through the instructional period and not just at the beginning. The illustrative cases show both commendable and concerning examples of how teachers related their course descriptions and learning goals to final assessments. However, from a formative assessment perspective there are further benefits to be had from clearly defined expectations.
In our work with teachers, we encourage them to use clear learning expectations for several reasons. First, the process of writing a learning intention can serve as a test of the importance of a particular lesson. If a teacher struggles to identify the learning of a particular lesson it may be that it is a “neat activity” but that the time could be spent on a more worthwhile instructional activity. Second, research shows enabling students to understand learning expectations or success criteria with students has a demonstrable benefit (White & Frederiksen, 1998; Fuchs, Fuchs, Karns, Hamlett, Dutka, Katzaroff, 2000). Third, it provides a touchstone for a teacher to return to at the end of the lesson, to either engage the class in self-reflection or discussion about what was learned in relation to the stated learning expectations, or to pose an exit ticket question or some other formative assessment in order to gauge student learning. Evidence of learning at the end of one lesson may impact planning for the next.

Within the groupings of the standards, imposed for the purposes of this paper, they have been discussed in the order in which they appear in the Student Evaluation Standards. This standard feels out of order since we have already discussed Standards U6 and U7 which focused on effective reporting and follow-up. In that earlier discussion we noted the importance of having clear learning expectations for students as bedrock on which to build instruction. However, the concept is worth repeating. Given the three driving questions for formative assessment – What is the intended learning? Where are students currently? How can the gap be bridged? – this Standard addressed the heart of the first question.

**A3: Context Analysis**

The Context Analysis standard focused on the need for teachers to understand student contexts and the impact that those contexts might have on assessment. The guidance is clear that understanding the context is not the same as altering assessment results, but rather the understanding of context should be used to help the student, parents or others to understand the results.

While this standard is an important part of formative assessment, what seems missing is any guidance for teachers regarding how they should act during instruction. It seems that understanding student context is not just something applied to the interpretation of summative assessment results, but goes to the heart of good instruction also. Formative assessment results should be interpreted in the light of student context. A poor student response to a learning log at the end of a lesson perhaps indicates that the student was struggling with the content, but it could also be due to the student’s reading difficulty, language capability or some other student factor. The teacher’s knowledge of context will influence how the teacher interprets the results and the resulting action that might be taken. It might even impact how the teacher sets up the task in the first place. Thus context analysis is at the heart of the instructional and
formative assessment processes, not just something applied at the end after assessment results have been generated.

**A4: Documented Procedures**

The Documented Procedures standard speaks to the importance of clearly documented, communicated and understood assessment procedures for all parties involved, and that having these procedures will help ensure the equity and fairness of the assessment process. In the explanation for this standard there is mention of the fact that the level of documentation required will differ according to the consequences for a particular assessment (p.149). However, a peculiar dichotomy is established with the comment that “much of the teacher’s evaluation of students, particularly in the early grades, focuses on student learning and feedback during the instructional process (emphasis added) [p.149]” which seems to suggest that for later grades there is, or perhaps should be, less focus on assessment to support student learning and feedback.

Research studies that support the use of formative assessment as integral to instruction include students of all ages including high school and beyond (King, 1992; Vogt & Mazur, 2005). In our work at ETS, we have worked with teachers at all grades levels and are currently engaged in project in which math and science high school teachers are implementing formative assessment. From our work, we would suggest that formative assessment has an important role to play as students get older. As students can potentially become more disengaged with school, it becomes increasingly important to provide clear learning intentions so that students can see how lessons connect and where they are headed. Older students are also better able to engage in self-reflection and support the learning of their peers as they are perhaps more able to work independently.

Some of the documentation that a teacher might create as part of the procedures can also be useful to share with students to help them understand the learning expectations. For example, sharing scoring keys or rubrics along with samples of student work can provide a tangible way for students to understand what quality work can look like and increase the likelihood that they can produce it (White and Frederiksen, 1998).

**A5: Defensible Information**

The Defensible Information standard speaks to the importance of attending to issues of reliability and validity in order to have information that is defensible. However, it goes beyond that and touches on issues of students cheating, plagiarizing or engaging in other misconduct so that results are not defensible
Furthermore results can be indefensible if the teacher is biased or somehow has a conflict of interests (p.156).

In the first illustrative case for this standard, while it is not explicitly stated the teacher seems to value formative assessment: there is continuous assessment and she tells the students that she will use it “to help them (p.158)”. As a result students are very clear about what they are learning and can articulate it to others. As we have worked with teachers who are learning to implement more formative assessment in their classes one concept that we address is the notion of the changing “classroom contract.” Thompson and Wiliam (2008) define the classroom contract as “the complex network of shared understandings and agreed ways of working that teachers and students arrive at in classrooms (p.5)” (Thompson & Wiliam, 2008). They describe several changes that can occur between students and teachers as a result of increased attention to formative assessment, but one seems particularly relevant to this example given the emphasis that the teacher in the illustrative example placed on articulating to students that one of her roles was to use assessment information to help them. Thompson and Wiliam (2008, p.5) describe the change as follows:

The final change is a change in the student-teacher relationship from adversaries to collaborators. Many of the teachers commented that their relationship with the students changed. Whereas previously the teacher had been seen as an adversary who might or might not award a good grade, increasingly classrooms focused on mutual endeavor centered on helping the student achieve the highest possible standard. (Thompson & Wiliam, 2008, p.5)

When students and teachers act together in a classroom where assessment is viewed primarily as something that is there to help learning progress, everyone is likely to take it more seriously. As a result, assessment information should be more defensible.

**A6: Reliable Information**

This standard addresses the issue of reliability, that is, “the degree of consistency of the score or information obtained from an information-gathering process (p.161). The standard (pg. 162) suggests a variety of ways (internal consistency, stability, observer agreement, decision consistency) that reliability can be estimated given different forms of assessment, and provides an important reminder that all forms of assessment will have some level of unreliability, and importantly for higher stakes decisions assessments with low levels of reliability should not be used.

A concern about this standard is the use of language that parallels that which was used in Standard A4. In this case the statement reads, “*In early grades* much teacher evaluation focuses on
student learning and feedback during the instructional process [emphasis added] (p.162).” As before, we would challenge the notion that using assessment to inform feedback to students or next instructional steps is somehow only appropriate at lower grade levels.

Both of the illustrative cases provided for this standard deal with scoring responses and the lack of or quality of a scoring guide. In the second example, the teacher worked with students to develop scoring criteria collaboratively. In addition to providing guidance to ensure that the teacher and her colleague scored the student work using common understandings, the development of the scoring criteria with the students was important from a formative assessment perspective. It helped students develop a common understanding of what was important to the project, what the various components were and what quality work would look like. This process then would help them judge their own progress and final piece of work, or assist them in giving meaningful feedback to a peer.

While the ability for a teacher to make reliable judgments about student learning is critical, it may well be that we do have not have ways to calculate reliability estimates that reflect what happens in a classroom. The text suggests that these teacher judgments may be based on limited student information band therefore “are likely to have low reliability (p.162).” Considering an example of formative assessment, a teacher may need to make a decision like “do my students understand how to calculate the median in a data set sufficiently well enough for me to proceed to the next form of central tendency that I want to introduce?” A knowledgeable teacher will be aware of two common student misconceptions in this domain: first, students may believe that the median is the middle number and disregard the importance of ordering the numbers in the data set; second, because it is easier to grasp the concept of the median with an odd-numbered data set students may then conclude that a data set with an even number of numbers does not have a median.

A teacher could then construct two multiple choice items to address common misconceptions specific to the particular topic at hand. This specific content pedagogy knowledge could have been developed by reading about student misconceptions or from her experience as a teacher seeing common student mistakes made year after year. By asking students to add an explanation of their answer she has additional insight into whether students were guessing or not. Furthermore the pattern of incorrect answers gives her direction for her next instructional steps. Students may have been able to sort both datasets into order before attempting to select the median, but a group of students could be identified as being confused by the second question with the even number of numbers. The teacher is now able to confidently decide that her next steps will be to work with this small group of students to provide additional instruction, while having the rest of the class engage in an extension activity. It is unlikely however, that calculating Cronbach’s alpha for a two-item test will result in a high reliability coefficient,
and yet the alpha value alone is a disservice to the value of the information collected. While no-one would suggest that a teacher should calculate an alpha value for every assessment that is used, the example above is provided to illustrate how difficult it is to estimate reliability. How can a teachers’ knowledge of misconceptions and the decision to incorporate them into a targeted question be estimated? How can a reliability estimate be calculated for an interpretation that a teacher might make that most students in a class understand a concept given 23 out of 25 students have the correct answer? The assumption at the student level will be less accurate since each student represents only a single data point, but at the class level the evidence is stronger.

It is important to ensure that assessment literacy is not confused with statistical literacy. It is critical that teachers understand the content domain, know how to craft questions and tasks that get at the breadth of the desired knowledge and understandings, and that they have the instructional repertoire to re-teach a concept in another way if students struggle to understand.

A7: Bias Identification and Management

This standard addresses the issue of bias and the importance of ensuring that bias does not creep into the assessment process (p.167). It goes beyond what might be initially considered potential sources of bias such as gender or racial stereotyping, to also categorizes assessments with construct under-representation or construct irrelevant aspects as biased. We place considerable importance on the role of learning intentions for teachers and students to clearly understand what is being taught and learned. When learning intentions are meaningfully connected across lessons to identify what is truly important in a domain, they also provide a critical tool for helping to ensure that the assessments capture the breadth of the domain (avoiding construct under-representation) and that they do not bring in extraneous material (avoiding construct irrelevance).

The second illustrative case is interesting from a formative assessment perspective. The case is presented as an example of a teacher recognizing the potential for bias when reviewing results of students from a special admissions program on the first writing assignment. He made adjustments to the assessments and also provided additional assistance. This example could also be viewed from a formative assessment perspective. The teacher collected evidence of students’ writing ability from the first assessment and identified weaknesses. Adjustments where made to instruction as a result both in the form of modifying future assignments and providing additional instruction that focused on areas of particular weakness. This is not to suggest that formative assessment alone can take care of all sources of bias, but when a teacher adopts a mindset of collecting evidence of student learning and using it to make immediate instructional adjustments it can reduce some issues.
One form of bias, not identified in the standard, can arise when a teacher makes a judgment of about the level of understanding across a class of students on the basis of incomplete information. Often classroom questioning can come in the form of a teacher asking a question and students who know the answer raising their hands. The teacher receives a correct answer, and then decides that the class as a whole is ready to move in. As mentioned earlier, one of the techniques that we encourage teachers to use is all-class response systems: individual student white boards for short answers or ABCD cards for multiple choice questions (a low-tech approach to the increasingly popular electronic clickers). The teacher now can ask a question, provide enough time for students to write or select a response, and provide a cue so that all students show their answer simultaneously. Now the teacher has information from every student in the class without requiring more time to collect it. On the basis of what the teacher can see, an informed decision can be made whether to ask a follow-up question, to pursue a student’s thinking behind a question, or to provide additional instruction.

A9: Analysis of Information

In the introduction we mentioned the formative assessment cycle and its three distinct phases: collect-analyze-use information. The final standard to be discussed in terms of its applicability to formative assessment focuses on the middle phase of that process and touches in the third phase as the text notes that “analyses serve teachers in planning, monitoring, and modifying their teaching (p.181).” The text identifies various forms of reference that may be used in an analysis process, and it would seem that from a formative assessment perspective, analysis of performances in relation to prescribed content standards or perhaps aptitude would be most useful. The list of additional steps (p.183) provides some helpful pointers for how to look at data. From a formative assessment perspective is not enough just to arrive at a total score on an assessment, even with assurances of reliability, validity and freedom from bias. Rather data should be looked at in a variety of ways. For example, are there patterns of incorrect answers which suggest that students who do not yet understand the content are struggling in similar ways? Are there groups of questions that address the same content that the majority of students answered correctly or incorrectly? Do students who perform well on constructed response items also do well on multiple-choice items and vice versa? And for each of these questions, the teacher should further ask “what impact does this have on my immediate instruction?” and “what impact might this have on my instruction the next time I teach this?” One of the common errors listed for this standard is “Using item analyses only to assess the technical quality of assessment methods and not for understanding student performance or for instructional planning.” However, teachers often struggle with both analyzing their
data and translating their conclusions to instructional next steps and require ongoing support for these tasks.

In the caveats for this standard it is suggested that the day-to-day evaluations that a teacher will make do not require the “detailed operations” described in this standard, but that “the type of information collected and the data themselves should be defensible (p.184)” As noted for other standards previously, we encourage teachers to continually ask themselves three questions: Where are my students headed? Where are they right now? And how do I close the gap? Careful attention to these questions can help ensure that data collected and inferences made are defensible.

### Student Evaluation Standards Related to Formative Assessment

Across the 28 Student Evaluation Standards, we identified five standards that related to formative assessment, although not all the connections are as clearly drawn out as they could be to support someone focusing on their formative assessment practice. Each is discussed below.

**P2: Appropriate Policies & Procedures**

The Appropriate Policies and Procedures standard addresses the need for schools to have well-developed, publicly available policies that define the purposes and procedures for student evaluations. The bar is set quite high in this standard as it notes “student evaluations conducted without a written policy and accompanying procedures risk being poorly performed (p. 34)” and the examples supplied by the illustrative cases refer to school policies such as using multiple assessment methods or ensuring fair treatment to all students.

However, an individual teacher or school may need to also think about what level of information parents and guardians might require depending on how formative assessment is playing out in individual classrooms or across the school. For example, a teacher might decide to change how some work is graded, based on research by Butler (1988) and others, which indicates that when students are provided with both grades and comments less learning takes place than when they are given comments only. As a result, rather than providing a letter grade with some comments, she will only provide comments to students along with an opportunity for them to improve their work in the light of the comments. Such a change in practice might be received better by parents if the process and rationale are explained ahead of time.

Other formative assessment practices that a teacher might use in a classroom, such as asking an exit ticket question at the end of the lesson (a single question that gets at the heart of the intended learning for the lesson that students respond to on index cards that can be handed to the teacher as they exit the classroom
and quickly reviewed by the teacher prior to planning for the next day’s lesson) are less likely to cause concern and a teacher deciding to use the approach would likely not need to provide advance explanation to parents and guardians.

**U3: Information Scope**

The Information Scope standard identifies the need for the evaluation to provide relevant and sufficient information in order to support whatever decisions are to be made. This standard, which advocates the articulation of a clear connection between what a teacher wants to know about her students and what can be learned from the assessment, is an important one, although as currently written the standard is directed more closely to higher stakes assessments, and so is only somewhat connected to formative assessment.

As part of the explanation for this standard, the text notes that for “low-stakes decisions, such as those made from observations of students during a single class period or evaluations of daily homework, typically require a smaller and narrower amount of information gathered in a brief span of time (p. 77).” A low-stakes assessment might be to evaluate the extent to which students remember the main ideas of a concept introduced the previous year, in order to select the most appropriate starting activity for a period of instruction that is to build on what was previously learned. A teacher might accomplish this pre-assessment by asking students to respond to one or two questions at the start of the lesson. If the questions are crafted carefully so that they require only short answers, the teacher can review responses at the start of the lesson and decide how to proceed or draw students into a discussion of answers as she reviews them. While the information gathered will necessarily be “smaller and narrower” than the information gathered at the end of the unit, the teacher needs to have a clear idea of what students have learned in the previous year, and be able to design targeted questions to get at the main concepts rather than use a more trivial question that perhaps only requires students to recall vocabulary. While this standard is written to address the larger evaluation questions in which a teacher might engage on a monthly or quarterly basis, thoughtful attention to the nuances of questions asked can result in higher quality assessment information.

In addition, while not directly addressed in this standard, attention to how evidence is collected is an important aspect of ensuring an appropriate scope of information. For example, continuing with the situation introduced above, had the teacher developed two questions to help her understand what students remembered from their previous exposure to the topic in the previous school year, and asked them orally to the class, she may have got responses from a few students, and typically those who tended to answer questions most frequently in class. On the other hand she might have had students write three-to-five word responses on white boards and then hold up on cue for her to review. Or as described above she
could have asked students to write their answers on index cards for her to collect. These all-student-
responses systems (Leahy et al., 2005) and other similar approaches ensure that a teacher can collect
evidence from every student in the class, at a level of detail that is sufficient to direct next steps, and yet
small enough that the teacher can process the information on-the-fly. So while the breadth of the question
might be narrow, how the question is asked can ensure that the teacher has a systematic understanding of
the thinking of all students in the class, which is critical given that the purpose of the questions was to
inform immediate instructional steps.

_F2: Political Viability_

The Political Viability standard provides advice to those planning and conducting student
evaluations on how to anticipate questions from, and obtain support and cooperation from students,
parents/guardians and other stakeholders (p. 113). The standard is written, it seems to us, from the
perspective that the higher the stakes of the assessments (for program placement, for example) the more
likely it is that the decisions could be challenged and the greater need for preparation for such challenges.

We categorized this standard as somewhat connected to formative assessment since we can
identify cases when attention to the process of garnering support from parents and other stakeholders is
important. However, we also recognize that much work by individual teachers can be done without a
great deal of attention to the broader political landscape.

By way of an example, however, a teacher might consider engaging students as intellectual
resources for each other by developing opportunities for students to provide feedback to each other, using
rubrics or other structures. Some parents may view the teacher as the sole provider of feedback to students
and may not want other students “judging” their child’s work. Such a change may need to be negotiated
with parents, not just to inform them of the change, but to explain why such a change is of benefit to all
students: students are not grading each other’s work, the provision of more frequent opportunities for
feedback helps students improve their work, and the process of commenting on another student’s work
can help the student creating the feedback to really understand what quality work looks like.

While this standard suggests that written policies are important for every assessment, when it
comes to formative assessment it is not so much an individual formative assessment that needs
documentation, but the overall approach to and process of formative assessment that should be clearly
articulated. In our formative assessment research work, we have been providing teachers with sustained
professional development to support them in the process of examining their practice, expanding their
repertoire of formative assessment approaches and sharing their successes and failures with peers in order
to get support, feedback and advice. We have seen schools vary in how they begin this work: sometimes
one department or grade level seems particularly interested in starting, or the principal asks for volunteers for the first year, or even just one or two individuals begin the work. Less frequently does an entire school start an intensive professional development effort with all teachers simultaneously. In contrast, starting small and establishing local existence proofs among a few teachers most receptive to innovation has often been an important way to develop the political support among teachers.

A10: Justified Conclusions

The Justified Conclusions standard focuses on the importance of explicitly justifying evaluative conclusions so that all stake-holders can have confidence in them (p.197). This standard is written largely from the perspective of summative or overall evaluations and the illustrative cases reinforce this. One case provides an example of parents who were unable to get information about how their child’s grades were determined and what each grade really meant, while the second case provided a higher education example of how clear documentation and procedures proved to be necessary and valuable when a grade for an online course was challenged.

While this standard is not directed at the formative assessment process, the openness recommended does have value for all forms of assessment. Some of the guidelines, such as ensuring students understand the purpose of any assessment and providing them with rubrics or scoring guides are particularly helpful in a formative assessment context. Bringing students into the assessment process and being clear about expectations can contribute to increased engagement in the process and better learning as a result or better evidence to make instructional adjustments. The three questions mentioned as the start as a guide to formative assessment – Where is the intended learning going? Where are students currently? How can the gap be closed? – start with the requirement that teachers and students understand where the learning is going so that any conclusions about its current status are reasonable.

A11: Metaevaluation

The final standard that we considered to have a relationship to the needs of a teacher engaging in formative assessment was the metaevaluation standard. This standard recommends that teachers periodically reflect on their assessment practices in order to improve them over time. While this standard did not have a particular focus on formative assessment, in general the recommendation that teachers reflect on their practice is a sound one. This standard is closely related to one aspect of the Evaluation Support standard (F3) which addresses the need for professional development in the area of assessment. While reflection is an important process, it should come as part of a larger process that Thompson and Wiliam (2008) describe as “repeated cycles of learning, practice, reflection, and adjustment (p.12).”
Opportunity for reflection in the context of ongoing professional development was discussed in the context of the Feasibility Standard in the previous section.

**Student Evaluation Standards Seldom Applicable to Formative Assessment**

Across the 28 Student Evaluation Standards, we considered six of them to be seldom applicable to someone focusing on their formative assessment practice.

**P3: Access to Evaluation Information**

The focus of the Access to Evaluation Information standard is on preserving the confidentiality of students’ evaluation information. Caution is advised when determining who should be able to see reports of student progress, and sharing information with students about other students is never recommended. In the first illustrative case provided, a teacher read out scores on a practice test to his class in an ill-advised attempt to regain control of the class. By way of contrast, second example describes a case where parents hired a tutor and there was a respectful conversation, structured by the school policy, between parents, teacher and tutor regarding what information can be provided to best support the student’s needs. The two illustrative cases suggest that the focus on this standard is on the final evaluations and scores or grades that may be ranked or compared since this is the area of most potential sensitivity.

Given that the focus of formative assessment is about the collection of evidence of student learning that can be used to inform subsequent teaching and learning, it is most likely that such assessment will be done prior to a final evaluation. Furthermore, a teacher who had formatively evaluated student work and written comments that she would like students to consider as they make revisions might pair students together so that they can review each other’s comments and work together to pursue next steps on that piece of work. This level of sharing of information is different from sharing final grades, although some sensitivity is still required in terms of how students are paired, and the structures and norms that have been put in place for when students work together. In some instances it still may be both appropriate and important to preserve student anonymity, if, for example, student work is being used to illustrate both high and low scores on a scoring rubric.

**P7: Conflict of Interest**

The focus of the Conflict of Interest standard is on the need for those involved with evaluations to be aware of potential conflicts of interest that might arise. A conflict of interest is most likely to occur in an evaluation situation that is high stakes, as the two illustrative cases demonstrated. In one case while a
teacher was waiting to hear about a job promotion that would be decided by the school superintendent, she was grading a major exam for a class where she taught the superintendent’s son. In the second example, student test scores were going to be closely monitored by the school board and the principal overstepped ethical boundaries by strongly suggesting that teachers “do whatever they could” to ensure improvements in student test scores.

While conflicts of interest could arise in any teaching situation, given the low stakes nature of formative assessment, and the tight relationship between what a teacher learns about student learning and the next instructional steps that he or she might take, formative assessment does not have as long a reach into areas of potential unethical behavior as other assessments might have.

**U4: Evaluator Qualifications**

The Evaluator Qualifications standard focuses on the necessary knowledge and skills that are needed for competent evaluations and analyses to be carried out and reported. Under the guidelines section there is a list of knowledge and skills that teachers ought to have from both their pre-service and in-service education experiences. While the content of what is outlined for teacher knowledge and skills is indisputable, the standard is so broad that it is only applicable to formative assessment in the most general sense.

It is worth noting that the list of knowledge and skills does seem somewhat slanted towards the more formal end of the assessment spectrum and is missing any statement that might more obviously or directly capture the knowledge and skills needed for formative assessment. Two suggested additions might be “approaches for readily collecting systematic (class-wide) information about current learning levels that is sufficient to directly inform next instructional steps” and “approaches for supporting meaningful student reflection of their own learning, and appropriate student feedback to peers.”

**U5: Explicit Values**

The Explicit Values standard focuses on the need for those involved in evaluation to “identify and justify the values used to judge student performance.” The rationale for categorizing this standard as having little applicability to formative assessment is very similar to the previous standard – the breadth of this standard as it is currently written renders it less useful to the specific evaluation application of formative assessment although the principles are laudable. The two illustrative examples both focus on school-level policy decisions (inclusion of a math proficiency test as part of an high school graduation requirement and school wide reform resulting from concern about students’ low performance on a state assessment) rather than on how values might impact assessment at a classroom or formative level.
Explicit values are important in formative assessment situations. For example, a teacher might choose to be explicit about the fact that during any form of formative assessment the teacher sees her role as eliciting evidence of students’ current understanding of the topic at hand with a specific goal of modifying instruction if students were struggling rather than rendering any form of set-in-stone judgment. Similarly, a teacher might try to instill in students the value of formative assessment as providing an opportunity for them to evaluate their own level of understanding, with a view of acting on that knowledge. Both of these values might seem quite foreign to students accustomed to teachers whose assessment roles focus primarily on the assignment of final grades, and who are seen as the only arbiters of quality in a classroom.

*F1: Practical Orientation*

The Practical Orientation standard addresses the importance of student evaluations being of an appropriate scale so as not to unduly disrupt instruction. This standard points to the importance of finding the balance between providing sufficient resources for the evaluation to be sound, while ensuring that the assessment is not inefficient or unduly burdensome. Again, as for some of the previously discussed standards, the illustrative cases discuss assessments that are of a scale larger than most formative assessment processes: one illustrative case described an assessment system was developed as a district-wide response to a low writing scores on a state assessment and the other described a summative evaluation at the end of a particular course. From a formative assessment perspective, this standard calls for a level of documentation which may not be appropriate for every assessment, since every lesson could contain an aspect of formative assessment.

The first illustrative case, the district wide writing assessment at four grade levels in the district, is an interesting example. The goal of the additional assessments was to improve writing performances since the district scores were currently below the state averages. Although the assessment system is district wide and only occurs once a year, it has some formative potential. However, from the description provided, it is unclear how the creation of the additional assessments would improve students’ writing skills although there are some hints: students were given an opportunity to see the rubric, essays were scored by both the classroom teacher and someone from a marking team. While not made explicit, the use of a rubric can help students understand what a quality essay looks like and that may not have been something that students had previously understood. In addition, having two scores provided for every student provides an opportunity for teachers to calibrate standards and expectations across a grade level. However, it may require some set-aside time for teachers to work together through a comparison of the scores they awarded their students compared to the scores from the marking team and to understand what
differences in scores might mean for them moving forward. Teachers could reflect on whether they had
tendencies towards leniency or severity and what that might mean for their instruction. Furthermore,
teachers might need support to review students’ scores and determine whether they needed to make
instructional adjustments in the near future. Beyond that, the timing of the assessment could be important,
impacting whether teachers would have time to adjust instruction in the light of the results. And finally,
attention should be given to how results are reported to students. If students are to improve their writing
they will need feedback in an appropriate form that helps them understand their strengths and weaknesses
and also on how to improve in the future. As this assessment system becomes more established it may be
worth revisiting the process to see how to best use the assessment data, both to provide important
information for teachers across the district, and to support students improve their writing.

**A8: Handling Information and Quality Control**

In the group of standards that seems least applicable to formative assessment, the Handling
Information and Quality Control standard focuses on data and information storage. There is a significant
overlap with the previously discussed P3 standard that addresses access to evaluation information.

For the most part an individual teacher will be dealing with information that only comes from his
or her students, needed to plan future instruction. While formative assessment practices can spread
throughout a school and can impact assessment practices such as a school’s grading policy for example,
this particular standard has little to add to what a teacher might consider. The scope of both illustrative
cases was much larger: one focused on an evaluation system used to determine who would be admitted to
the school’s advanced placement program, while the other case focused on how a school dealt with issues
that could arise from a computerized grade books. As was suggested for the P3 standard, while not of
direct relevance it suggests some level of thoughtfulness be applied to how student records are
maintained.

Looking across the set of six standards that we identified as being seldom applicable to formative
assessment practice one thing to note is that the illustrative cases largely focus on assessment issues that
go beyond immediate classroom assessment used to inform next instructional steps. While some of the
suggestions contained within the standards may be useful in a formative assessment context they do not
go the heart of the process.
Discussion

The Student Evaluation Standards equip those who have been calling for improved teacher assessment literacy (Stiggins 2002, McMunn, McColskey, Butler, 2003, Brookhart, 2003) with a valuable resource. Whether in pre-service or professional development contexts the Student Evaluation Standards (JCSEE, 2003) articulates ways that teachers and other decision makers need to attend to the multifaceted aspects of student assessment.

Certainly in a book focused on classroom evaluation, the role of formative assessment was identified although the term not used, but there were references to the role of assessment to support teachers in “planning, monitoring, and modifying their teaching (p.181).” However, while the use of evidence to inform instruction was acknowledged it was on occasion overwhelmed by other uses of assessment such as to inform grades or other high stakes decisions. Because of the large impact formative assessment can have on classroom teaching and learning, it will be important in the future to bring formative assessment issues to the forefront of the Student Evaluation Standards.

As we reviewed the Student Evaluation Standards we consistently asked ourselves what, if anything, was missing from them if they were read from a formative assessment perspective. One correction that we would make in the standards is the perhaps unintended message that formative assessment is something that is primarily useful in the early grades. Similar statements appeared in both Standards A4 and A6, that “much of the teacher’s evaluation of students, particularly in the early grades, focuses on student learning and feedback during the instructional process (p.149).” Granted, certain assessments might have higher consequences for older students such as admissions to particular classes or programs, but using formative assessment to provide feedback to students and make immediate instructional adjustments is something that is applicable at all grade levels. Indeed Bain (2004) identified the use of “nonjudgmental feedback on students’ work (p.35)” as one of the critical markers used to identify what the best teachers at the college level did to promote student learning.

Throughout the book, what we would refer to as formative assessment tends to be described as informal assessment. While it seems minor to quibble over a word choice we think it is important to consider using the term formative assessment where appropriate for several reasons. First, the term formative assessment itself acts a reminder of the purpose of the assessment: to inform the teacher or students regarding next steps. Second, the term “informal assessments” suggests something that is easy, relaxed, or casual in contrast to formal assessments where the need for reliability, validity, documentation etc. is much greater. However, to conduct formative assessment in such a way to move student learning forward in meaningful ways requires a crucial set of knowledge and skills. First, a teacher requires strong pedagogical content knowledge so that he or she can anticipate how students might misunderstand...
important content and how those misunderstandings could elicited. Second, for formative assessment to be of value, a teacher has to be able to act on that information. This might require knowing how to teach a particular topic in a different way, creating new analogies or metaphors to explain a concept or developing an alternative activity to help students deepen their understanding. Furthermore, if a teacher decides that some students need additional time on a particular concept or topic, but other students do not, the teacher needs to be able to provide extension activities for those students that will benefit them in meaningful ways. Third, a teacher has to be able to teach students to be reflective of their own learning and to support the learning of their peers. To do this a teacher needs to be able to arrange appropriate supports and structures for students to use so that their comments to each other are meaningful. Finally, a teacher needs to not only know the content being taught, but also to possess a broader view of the content as it relates to the overall learning goals for an entire year, and how it relates to future learning. As more attention is paid to ensuring that students understand critical content before moving on, a teacher needs to know what content is critical and what can be ignored if students need more time in some areas to understand a major concept. None of these requirements suggest the relaxed, casual approaches that might be suggested by the term “informal assessment”.

Regarding what is not covered by the Student Evaluation Standards, one omission is a standard devoted to the issue of the role of students in evaluation with a particular focus on self and peer assessment. In the introduction to the standards was a statement “as students grow older, we expect them to play a larger role in their evaluations (p.1).” However, this statement is not followed through in the standards other than in terms of how the teacher informs students about the process. There is a rich literature around these topics and the impact that engaging students in self and peer assessment can have on their learning (e.g., Fontana & Fernandes, 1994; Mercer, Wegerif & Dawes, 1999) and also some guidance in terms of the importance of supporting students with appropriate structures (King, 1992).

The final addition that we suggest for the standards, is one devoted to the use of assessment information to guide instruction. In the Accuracy set of standards, each step of the process is laid out including the collection of evidence of student learning, the handling of information and quality control, the analysis of the assessment data, and the conclusions that can justifiably be made from the data. However, for formative assessment the critical step is the last one: the action that a teacher takes based on the evidence. For formative assessment to be feasible, there seem to be two necessary conditions regarding the evidence. First it needs to be in a quantity that is “digestible.” Without even considering the load on students, a teacher does not have enough hours in a day to collect lengthy written assignments from students at the end of each lesson. What is more reasonable is that teachers review evidence of student learning in the form of responses to a single, well constructed question that gets at the core of the
learning for just that particular lesson, or that students are asked to reflect on what they thought was a key idea of the lesson. For example, if students hand in responses on a Post-it® note or an index card, the teacher can sort responses in groups of like responses and based upon this sorting, plan for the next lesson. Building of the notion of a well-constructed question, the teacher needs evidence that is actionable. In other words, if a teacher is to use evidence from one or two questions to determine next instructional steps the question itself has to clearly address the central learning intention of the lesson at a minimum, and may also help a teacher diagnosis what students are still struggling with rather than just that they are still struggling.

The purpose of this paper was to examine the Student Evaluation Standards (JSCEE, 2003) from a formative assessment perspective. In order to do this, we divided the standards into three groups: those that provide important advice and support for teachers currently engaged in formative assessment, those that are connected to formative assessment but that need the connections to be drawn more clearly, and those that seldom relate to the core of formative assessment. Across the standards, we identified ways in which the language of the standards could be strengthened to clearly reflect that work that is needed with respect to formative assessment. On some occasions we offered some additional examples of practice in an effort to demonstrate the kind of rigor that is required for teachers to engage in “minute-to-minute and day-by-day” formative assessment. We hope that our discussions provided a useful perspective on the standards.
References


Appendix

Standards specifically related to the core activities of formative assessment
  P1: Service to students  
P4: Treatment of students  
P5: Rights of students  
P6: Balanced evaluation  
U1: Constructive orientation  
U2: Defined users and uses  
U6: Effective reporting  
U7: Follow-up  
F3: Evaluation support  
A1: Validity orientation  
A2: Defined expectations for students  
A3: Context analysis  
A4: Documented procedures  
A5: Defensible information  
A6: Reliable information  
A7: Bias identification and management  
A9: Analysis of information

Standards related to formative assessment
  P2: Appropriate policies & procedures  
U3: Information scope  
F2: Political viability  
A10: Justified conclusions  
A11: Metaevaluation

Standards seldom applicable to formative assessment
  P3: Access to evaluation information  
P7: Conflict of interest  
U4: Evaluator qualifications  
U5: Explicit values  
F1: Practical orientation  
A8: Handling information and quality control