

Formative assessment and contingency in the regulation of learning processes

Paper presented in a Symposium entitled *Toward a Theory of Classroom Assessment as the Regulation of Learning* at the annual meeting of the American Educational Research Association, Philadelphia, PA, April 2014

Dylan Wiliam, Institute of Education, University of London

Introduction

The relationship between instruction and what is learned as a result is complex. Even when instruction is well-designed and students are motivated, increases in student capabilities are, in general, impossible to predict with any certainty. Moreover, this observation does not depend on any particular view of what happens when learning takes place. Obviously, within constructivist view of learning, the mismatch between what is taught and what is learned is foregrounded but it is also a key feature of associationist views of learning. If learning is viewed as a process of making associations between stimuli and responses, then it is impossible to predict in advance how much practice will be required before the associations are established, so establishing what has been learned and then taking appropriate remedial action is essential. Within situated perspectives on learning, failures to demonstrate learning in a given context might be attributed to the extent to which the environment affords certain cognitive processes, but the important point here is that any approach to the study of human learning has to account for the “brute fact” (Searle, 1995) that students do not necessarily—or even generally—learn what they are taught.

Of course, the idea that effective instruction requires frequent “checks for understanding” has been around for a very long time, but about 50 years ago a number of researchers and writers involved in education began to think of this process of “checking for understanding” explicitly as a form of assessment. Indeed, assessment can be thought of as the bridge between teaching and learning—only through some kind of assessment process can we decide whether instruction has had its intended effect. Such assessment can be conducted at the end of the instructional sequence, but in recent years there has been increasing interest in the idea that assessment might be used to improve the process of education, rather than simply evaluating its results.

There does not appear to be any agreed definition of the term formative assessment. Some (e.g., Shepard, 2008) have argued that the term should be applied only when assessment is closely tied to the instruction it is intended to inform, while many test publishers have used the term “formative” to describe tests taken at intervals as long as six months (e.g., Marshall, 2005). Others (e.g., Broadfoot, Daugherty, Gardner, Gipps, Harlen, James, & Stobart 1999) have argued for the term “assessment for learning” rather than “formative assessment” although Bennett (2011) has argued that there are important conceptual and practical differences between the two terms.

Other issues on which there appears to be little consensus are:

- a) Whether it is essential that the students from whom evidence was elicited are beneficiaries of the process;
- b) Whether the assessment has to change the intended instructional activities;
- c) Whether students have to be actively engaged in the process.

This paper reviews the prevailing debates about the definition of formative assessment and proposes a definition based on the function that evidence of student achievement elicited by the assessment serves, thus placing instructional decision-making at the heart of the issue. This definition is then further developed by exploring the way these decisions occur at moments of contingency in the instructional process, so that whether an assessment functions formatively or not depends on the extent to which the decisions taken serve to better direct the learning towards the intended goal. In other words, assessments function formatively to the extent that they regulate learning processes. The paper concludes with a discussion of different kinds of regulatory mechanisms suggested by Allal (1988), specifically whether the regulation is proactive (measures taken before the instructional episode), interactive (measures taken during the instructional episode (interactive regulation) or retroactive (measures taken after the instructional episode to improve future instruction), and explores briefly the relationship of formative assessment and self-regulated learning.

The origins of formative assessment

It appears to be widely accepted that Michael Scriven was the first to use the term “formative,” to describe evaluation processes that “have a role in the on-going improvement of the curriculum” (Scriven, 1967, p. 41). He also pointed out that evaluation “may serve to enable administrators to decide whether the entire finished curriculum, refined by use of the evaluation process in its first role, represents a sufficiently significant advance on the available alternatives to justify the expense of adoption by a school system” (pp. 41-42) suggesting “the terms ‘formative’ and ‘summative’ evaluation to qualify evaluation in these roles” (p. 43).

Two years later, Benjamin Bloom (1969, p. 48) applied the same distinction to classroom tests:

Quite in contrast is the use of “formative evaluation” to provide feedback and correctives at each stage in the teaching-learning process. By formative evaluation we mean evaluation by brief tests used by teachers and students as aids in the learning process. While such tests may be graded and used as part of the judging and classificatory function of evaluation, we see much more effective use of formative evaluation if it is separated from the grading process and used primarily as an aid to teaching.

Benjamin Bloom and his colleagues continued to use the term “formative evaluation” in subsequent work and the term “formative assessment” was routinely used in higher education in the United Kingdom to describe “any assessment before the big one,” but the term did not feature much as a focus for research or practice in the 1970s and early 1980s, and where it did, the terms “formative assessment” or “formative evaluation” generally referred to the use of formal assessment procedures, such as tests, for informing future instruction (see, e.g., Fuchs & Fuchs, 1986).

In a seminal paper entitled “Formative assessment and the design of instructional systems,” Sadler (1989) argued that the term formative assessment should be intrinsic to, and integrated with, effective instruction:

Formative assessment is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student's

competence by short-circuiting the randomness and inefficiency of trial-and-error learning (Sadler 1989 p. 120)

He also pointed out that effective use of formative assessment could not be the sole responsibility of the teacher, but also required changes in the role of learners:

The indispensable conditions for improvement are that the student comes to hold a concept of quality roughly similar to that held by the teacher, is able to monitor continuously the quality of what is being produced during the act of production itself, and has a repertoire of alternative moves or strategies from which to draw at any given point. In other words, students have to be able to judge the quality of what they are producing and be able to regulate what they are doing during the doing of it. (p. 121)

The need to broaden the conceptualization of formative assessment beyond formal assessment procedures was also emphasized by Torrance (1993):

Research on assessment is in need of fundamental review. I am suggesting that one aspect of such a review should focus on formative assessment, that it should draw on a much wider tradition of classroom interaction studies than has hitherto been acknowledged as relevant, and that it should attempt to provide a much firmer basis of evidence about the relationship of assessment to learning which can inform policy and practice over the long term. (Torrance, 1993 p. 341)

It seems clear, therefore, that while the origins of the term formative assessment may have been in behaviorism and mastery learning, for at least two decades there has been increasing acceptance that an understanding of formative assessment as a process has to involve consideration of the respective roles of teachers and learners.

Defining formative assessment

Black and Wiliam (1998a) reviewed research on the effects of classroom formative assessment intended to update the earlier reviews of Natriello (1987) and Crooks (1988). In order to make the ideas in their review more accessible, they produced a paper for teachers and policy makers that drew out the implications of their findings for policy and practice (Black & Wiliam, 1998b). In this paper, they defined formative assessment as follows:

We use the general term assessment to refer to all those activities undertaken by teachers—and by their students in assessing themselves—that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes formative assessment when the evidence is actually used to adapt the teaching to meet student needs. (p. 140)

Some authors have sought to restrict the meaning of the term to situations where the changes to the instruction are relatively immediate:

“the process used by teachers and students to recognise and respond to student learning in order to enhance that learning, during the learning” (Cowie & Bell, 1999, p. 32)□

“assessment carried out during the instructional process for the purpose of improving teaching or learning” (Shepard, Hammerness, Darling-Hammond, Rust, Snowden, Gordon, Gutierrez, & Pacheco, 2005, p. 275)

“Formative assessment refers to frequent, interactive assessments of students’ progress and understanding to identify learning needs and adjust teaching appropriately” (Looney, 2005, p. 21)

“A formative assessment is a tool that teachers use to measure student grasp of specific topics and skills they are teaching. It’s a ‘midstream’ tool to identify specific student misconceptions and mistakes while the material is being taught” (Kahl, 2005, p. 11)

The Assessment Reform Group—a group of scholars based in the United Kingdom and dedicated to ensuring that assessment policy and practice are informed by research evidence—acknowledged the power that assessment had to influence learning, both for good and for ill, and proposed seven precepts that summarized the characteristics of assessment that promotes learning:

- it is embedded in a view of teaching and learning of which it is an essential part;

- it involves sharing learning goals with pupils;

- it aims to help pupils to know and to recognise the standards they are aiming for;

- it involves pupils in self-assessment; □

- it provides feedback which leads to pupils recognising their next steps and how to take them;

- it is underpinned by confidence that every student can improve;

- it involves both teacher and pupils reviewing and reflecting on assessment data (Broadfoot *et al.*, 1999, p. 7).

In looking for a term to describe such assessments, they suggested that because of the variety of ways in which it was used, the term “formative assessment” was no longer helpful:

The term ‘formative’ itself is open to a variety of interpretations and often means no more than that assessment is carried out frequently and is planned at the same time as teaching. Such assessment does not necessarily have all the characteristics just identified as helping learning. It may be formative in helping the teacher to identify areas where more explanation or practice is needed. But for the pupils, the marks or remarks on their work may tell them about their success or failure but not about how to make progress towards further learning. (Broadfoot *et al.*, 1999, p. 7)

Instead, they preferred the term “assessment for learning,” which they defined as “the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there” (Broadfoot, Daugherty, Gardner, Harlen, James, & Stobart, 2002, pp. 2–3).

The earliest use of the term “assessment for learning” appears to be as the title of a chapter by Harry Black (1986). It was also the title of a paper given at AERA in 1992 (James, 1992)—and three years later, was the title of a book by Ruth Sutton (1995). In the United States, the origin of the term is often mistakenly attributed to Rick Stiggins as a result of his popularization of the term (see, for example, Stiggins, 2005), although Stiggins himself has always attributed the term to other authors.

Most recently, an international conference on assessment for learning in Dunedin in 2009, building on work done at two earlier conferences in the UK (2001) and the USA (2005), adopted the following definition:

Assessment for Learning is part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning. (Klenowski, 2009, p. 264)

The phrase assessment for learning has an undoubted appeal, especially when contrasted with assessment *of* learning, but as Bennett (2009) points out, replacing one term with another serves merely to move the definitional burden. More importantly, as Black and Wiliam and their colleagues have pointed out, the distinctions between assessment for learning and assessment of learning on the one hand, and between formative and summative assessment on the other, are different in kind. The former distinction relates to the purpose for which the assessment is carried out, while the second relates to the function it actually serves. Black, Wiliam and their colleagues clarified the relationship between assessment for learning and formative assessment as follows:

Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students’ learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence. An assessment activity can help learning if it provides information that teachers and their students can use as feedback in assessing themselves and one another and in modifying the teaching and learning activities in which they are engaged. Such assessment becomes “formative assessment” when the evidence is actually used to adapt the teaching work to meet learning needs. (Black, Harrison, Lee, Marshall, and Wiliam, 2004, p. 10)

Five year later, Black and Wiliam restated their original definition in a slightly different way, which they suggested was consistent with their original definition, and those others given above, including that of the Assessment Reform Group. They proposed that an assessment functions formatively:

to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black & Wiliam, 2009 p. 9)

One important feature of this definition is that the distinction between summative and formative is grounded in the function that the evidence elicited by the assessment actually serves, and not on the kind of assessment that generates the evidence. From such a perspective, to describe an assessment as formative is to make what Ryle (1949) described as a category mistake—ascribing to something a property it cannot have. As Cronbach (1971) observed, an assessment is a

procedure for making inferences. Where the inferences are related to the student's current level of achievement, or to their future performance, then the assessment is serving a summative function. Where the inferences are related to the kinds of instructional activities that are likely to maximize future learning, then the assessment is functioning formatively. The summative-formative distinction is therefore a distinction in terms of the kinds of inferences that are supported by the evidence elicited by the assessment rather than the kinds of assessments themselves. Of course, the same assessment evidence may support both kinds of inferences, but in general, assessments that are designed to support summative inferences (that is, inferences about current or future levels of achievement) are not particularly well-suited to supporting formative inferences (that is, inferences about instructional next steps). It is, in general, easier to say where a student is in their learning than what should be done next. It might be assumed that assessment designed primarily to serve a formative function would require, as a pre-requisite, a detailed specification of the current level of achievement, but this does not necessarily hold. It is entirely possible that the assessment might identify a range of possible current states of achievement that nevertheless indicate a single course of future action—we might not know where the student is, but we know what they need to do next.

In their discussion of this definition, Black and Wiliam (*ibid.*) make a number of further points:

1. Anyone—teacher, learner or peer—can be the agent of formative assessment.
2. The focus of the definition is on decisions. Rather than a focus on data-driven decision-making, the emphasis is on decision-driven data-collection. More precisely, we could contrast data-driven decision-making with decision-driven *evidence* collection, on the grounds that evidence is simply data associated with a claim (Wainer, 2011). This is important, because a focus on data-driven decision-making emphasizes the collection of data first without any particular view about the claims they might support, so the claims are therefore accorded secondary importance. By starting with the decisions that need to be made, only data that support the particular inferences that are sought need be collected.
3. The definition does not require that the inferences about next steps in instruction are correct. Given the complexity of human learning, it is impossible to guarantee that any specified sequence of instructional activities will have the intended effect. All that is required is that the evidence collected improves the likelihood that the intended learning takes place.
4. The definition does not require that instruction is in fact modified as a result of the interpretation of the evidence. The evidence elicited by the assessment may indicate that what the teacher had originally planned to do was, in fact, the best course of action. This would not be a better decision (since it was the same decision that the teacher was planning to make without the evidence) but it would be a better *founded* decision.

Black and Wiliam then suggested that one consequence of their definition is that formative assessment is concerned with “the creation of, and capitalization upon, ‘moments of contingency’ in instruction for the purpose of the regulation of learning processes” (2009, p. 6).

Of course, these moments of contingency do not occur in a vacuum. The way in which teachers, peers, and the learners themselves, create and capitalize on these moments of contingency involves consideration of instructional design, curriculum, pedagogy, psychology, and epistemology. However, the focus on these moments of contingency in learning does restrict the

focus to aspects of instruction that could reasonably be regarded as “assessment” and thus prevents the concept of formative assessment from expanding to subsume all of learning, and thus losing any useful focus.

Wiliam (2009) points out that moments of contingency can be synchronous or asynchronous. Synchronous moments include teachers’ real-time adjustments during teaching, or the way a teacher, following a class poll of students’ responses, suggests that students discuss their responses with a neighbor (Crouch & Mazur, 2001). Asynchronous examples include those situations where teachers get students to provide feedback for each other using a protocol such as “two stars and a wish” (Wiliam, 2011b), or the use of evidence derived from student work (e.g., homework, students’ summaries made at the end of a lesson) to plan a subsequent lesson. Most commonly, the evidence would be used to modify the instruction of those from whom the evidence was collected. However, evidence about difficulties experienced by one group, and used to modify instruction for another group of students at some point in the future, would also qualify, although there would, of course, be an inferential leap as to whether the difficulties experienced by one group would be relevant to a different group of students.

As Allal (1988) has pointed out, the regulation can be pro-active, interactive, or retro-active. Pro-active regulation of learning can be achieved, for example, through the establishment of “didactical situations” (Brousseau, 1997) where the teacher “does not intervene in person, but puts in place a ‘metacognitive culture’, mutual forms of teaching and the organization of regulation of learning processes run by technologies or incorporated into classroom organization and management” (Perrenoud, 1998 p. 100). Such didactical situations can also be planned by the teacher as specific points in time when she will evaluate the extent to which students have reached the intended understanding of the subject matter, for example through the use of “hinge-point questions” (Wiliam, 2011b) as specific parts of the lesson plan. While the planning of such questions takes place before the lesson, the teacher does not know how she will proceed in the lesson until she sees the responses made by students, so this would be an example of interactive regulation, in which teachers use formative assessment in “real time” to make adjustments to their instruction during the act of instruction. Those examples above in which teachers reflect on instructional sequences after they have been completed, whether for the benefit of the particular students concerned, or others, would be examples of retro-active regulation of learning.

Formative assessment and self-regulated learning

From the foregoing discussion, while many authors have focused on formative assessment largely as a process in which teachers administer assessments to students for the purpose of ensuring that the intended learning has taken place (see, e.g., Ainsworth & Viegut, 2006), it is clear that for at least a quarter of a century, some authors have regarded the role of the learner as central. Wiliam (1999a; 1999b; 2000) suggested that formative assessment consisted of teacher questioning, feedback, and the learner’s role (essentially, understanding criteria for success, peer-assessment and self-assessment), and a number of other authors have proposed similar ways of understanding formative assessment or assessment for learning. For example, Stiggins, Arter, Chappuis and Chappuis (2005) proposed that assessment for learning consists of seven strategies:

1. Provide students with a clear and understandable vision of the learning target.
2. Use examples and models of strong and weak work.

3. Offer regular descriptive feedback.
4. Teach students to self-assess and set goals.
5. Design lessons to focus on one learning target or aspect of quality at a time.
6. Teach students focused revision.
7. Engage students in self-reflection and let them keep track of and share their learning.

While it could be argued that strategies 5 and 6 are not solely focused on assessment, it seems clear that a number of authors (see also Brookhart, 2007; Bailey & Heritage, 2008; Popham, 2008) on formative assessment have been addressing the same conceptual territory, although dividing it up in different ways.

Of course, where formative assessment is presented as a number of strategies, it not clear whether the list is in any sense complete. To address this, Leahy, Lyon, Thompson and Wiliam (2005) proposed that formative assessment could be conceptualized as five “key strategies,” resulting from crossing three processes (where the learner is going, where the learner is right now, and how to get there) with three kinds of agents in the classroom (teacher, peer, learner), as shown in figure 1. This model could also be criticized on the grounds that the strategies are not solely concerned with assessment process. That said, provided that the two strategies that involve learners and peers (activating students as learning resources for one another and activating students as owners of their own learning) are interpreted as specifically focusing on moments of contingency in the regulation of learning processes, then the framework provided in Figure 1 provides a reasonable conceptual basis for formative assessment.

	Where the learner is going	Where the learner is now	How to get there
Teacher	Clarifying, sharing and understanding learning intentions and success criteria	Engineering effective discussions, tasks and activities that elicit evidence of learning	Providing feedback that moves learning forward
Peer		Activating students as learning resources for one another	
Learner		Activating students as owners of their own learning	

Figure 1: Five “key strategies” of formative assessment (Leahy et al., 2005)

Detailed explanations of the derivation of the model in Figure can be found in Wiliam (2007) and Wiliam (2011a). For the purpose of this symposium, the remainder of the paper focuses on the strategies of “Clarifying, sharing and understanding learning intentions and success criteria” and “Activating students as owners of their own learning” and their relationship with self-regulated learning (SRL).

Formative assessment and self-regulated learning

Self-regulated learning (SRL) as defined by Butler and Winne (1995), “is a style of engaging with tasks in which students exercise a suite of powerful skills: setting goals for upgrading

knowledge; deliberating about strategies to select those that balance progress toward goals against unwanted costs; and, as steps are taken and the task evolves, monitoring the accumulating effects of their engagement” (p. 245). SRL therefore overlaps considerably with formative assessment, as defined by Black and Wiliam (2009) and in particular the “unpacking” of formative assessment as five “key strategies” proposed by Leahy et al., (2005) described above.

First, self-regulated learning involves, amongst other things, the means by which students come to adopt particular goals for their learning. It is often assumed that learning is optimized when the motivation for the learning task is intrinsic (that is the student engages in the task because it is inherently rewarding) but research on expertise (see, e.g., Ericsson, Krampe & Tesch-Römer, 1993) suggests that elite performance is the result of at least a decade of optimal distribution of what they term “deliberate practice”—practice that is neither motivating nor enjoyable, but rather is undertaken for the impact it will have on subsequent performance (Ericsson, 2006). As Deci and Ryan (1994) point out, there is a continuum of kinds of self-regulation, varying according to the degree of integration with one’s sense of self, and the extent to which the goal is internalized. However the student comes to adopt a particular goal in learning, the strategy of “clarifying, sharing, and understanding learning intentions and criteria for success” deals with the process by which learners become clear about the goals they pursue.

Second, once the learners have embraced a particular goal, then both SRL and the formative assessment (FA) strategy of “activating students as owners of their own learning” emphasize the way in which learners use a range of strategies to pursue the goal. While there may be differences in the way that different authors have described the perspectives, it seems that the differences are simply differences in emphasis, and where proponents of one perspective find aspects that are more fully explored in their preferred perspective, proponents of the other perspective are able to argue that their perspective. If there is one difference, it is that the SRL perspective is more theoretically grounded, while the FA perspective is more pragmatically focused on classroom practice. The specific benefit of the SRL perspective for FA is that it allows practical classroom techniques to be theorized so that they can be shared more widely amongst practitioners.

This is valuable, because many teachers invent classroom techniques in their development of FA practice. Some of these techniques are idiosyncratic, and include details that are important to the individual teacher, but may be less important for the success of the technique, and may even be counterproductive for the purposes of dissemination because they are off-putting to other teachers. For example, one teacher wanted to introduce to a fifth grade math class the idea of students leading end of lesson plenary reviews of what the class had learned (Wiliam, 2011b). Because the teacher is a fan of the Star Trek TV show, he presented this to the class as the “Captain’s log.” At the beginning of the lesson, volunteers for the role are sought, and the selected individual becomes the “captain” of the lesson, and at the end of the lesson, comes to the front of the class to give the “Captain’s log, star date 2014-03-07...” In this classroom, the captain is also given a special “captain’s hat” to wear while conducting the end-of-lesson plenary. High school teachers have also found the basic technique of having students deliver the end of lesson plenary valuable, but many would be put off by the apparently childish nature of the task. Theoretical perspectives on these kinds of classroom techniques allow the techniques to be “optimized” for dissemination, stripping away the theoretically irrelevant features and thus

making them more widely available to others. The danger in all this, of course, is that to some practitioners, it is the “irrelevant” features that make particular techniques engaging to learners. This suggests that a productive way forward is to be clear about the theoretical core of a technique and what are its surface features. To the researcher, it may be the theoretical core that is most important, but to the practitioner, the surface features may provide the “hook” that engages students. However, the theoretical grounding of the technique provides a warrant for the teacher that the technique, with its surface features, is likely to be faithful to the underlying research evidence about the technique’s effectiveness in improving instruction (see Thompson & Wiliam, 2008).

Conclusion

The broad argument of this paper is that although the origins of formative assessment are in mastery learning, with most early conceptualizations emphasizing the role of the teacher, for at least a quarter of a century, formative assessment has also been considered within a “broader conceptual field” (Perrenoud, 1998), that takes into account the settings within which learning takes place, and the histories of those involved (see, for example, Black & Wiliam, 2012, which explores formative assessment in the context of cultural-historical activity theory). In particular, much research in formative assessment has recognized that a consideration of the role of the learners, and their peers, is absolutely essential for productive understandings of the potential of classroom formative assessment to improve learning. This paper has proposed that self-regulated learning can be thought of a key aspect of productive formative assessment, in particular in relation to the formative assessment strategies of “clarifying, sharing and understanding learning intentions and criteria for success” and “activating students as owners of their own learning”. It is also suggested that the most productive way forward for the relationship between formative assessment and self-regulated learning is to build on the strengths of each—the practical grounded nature of formative assessment and the theoretical perspectives afforded by self-regulated learning—for generating productive conversations between practitioners and researchers.

References

- Ainsworth, L. B., & Viegut, D. J. (Eds.). (2006). *Improving formative assessment practice to empower student learning*. Thousand Oaks, CA: Corwin.
- Allal, L. (1988). Vers un élargissement de la pédagogie de maîtrise: processus de régulation interactive, rétroactive et proactive. In M. Huberman (Ed.), *Maîtriser les processus d'apprentissage: Fondements et perspectives de la pédagogie de maîtrise* (pp. 86-126). Paris, France: Delachaux & Niestlé.
- Bailey, A. L., & Heritage, M. (2008). *Formative assessment for literacy grades K-6*. Thousand Oaks, CA: Corwin.
- Bennett, R. E. (2009). *A critical look at the meaning and basis of formative assessment* (Vol. ETS RM-09-06). Princeton, NJ: Educational Testing Service.
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles Policy and Practice*, 18(1), 5-25.
- Black, H. (1986). Assessment for learning. In D. L. Nuttall (Ed.), *Assessing Educational Achievement*. (pp. 7-18). London: Falmer Press.
- Black, P. J., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in*

- Education: Principles, Policy and Practice*, 5(1), 7-74.
- Black, P. J., & Wiliam, D. (1998b). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Educational evaluation: new roles, new means: the 68th yearbook of the National Society for the Study of Education (part II)* (Vol. 68(2), pp. 26-50). Chicago, IL: University of Chicago Press.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Gipps, C. V., Harlen, W., James, M., & Stobart, G. (1999). Assessment for learning: beyond the black box. Cambridge, UK: University of Cambridge School of Education.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge, UK: University of Cambridge School of Education.
- Brookhart, S. M. (2007). Expanding views about formative classroom assessment: a review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: theory into practice* (pp. 43-62). New York, NY: Teachers College Press.
- Brousseau, G. (1997). *Theory of didactical situations in mathematics* (N. Balacheff, Trans.). Dordrecht, Netherlands: Kluwer.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles Policy and Practice*, 6(1), 32-42.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2 ed., pp. 443-507). Washington DC: American Council on Education.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Crouch C.H. & Mazur E. (2001) Peer instruction: ten years of experience and results. *American Journal of Physics* 69, 970–977.
- Deci, E. L., & Ryan, R. M. (1994). Promoting self-determined education. *Scandinavian Journal of Educational Research*, 38(1), 3-14.
- Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683-703). Cambridge, UK: Cambridge University Press.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation—a meta-analysis. *Exceptional Children*, 53(3), 199-208.
- James, M. (1992). *Assessment for learning*. Paper presented at the Annual meeting of the American Educational Research Association, New Orleans, LA.
- Kahl, S. (2005, 26 October). Where in the world are formative tests? Right under your nose! *Education Week*, 25, 38.

- Klenowski, V. (2009). Assessment for learning revisited: an Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice*, 16(3), 263-268.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: minute-by-minute and day-by-day. *Educational Leadership*, 63(3), 18-24.
- Looney, J. (Ed.). (2005). *Formative assessment: improving learning in secondary classrooms*. Paris, France: Organisation for Economic Cooperation and Development.
- Marshall, J. M. (2005). *Formative assessment: Mapping the road to success. A white paper prepared for the Princeton Review*. New York, NY: Princeton Review.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22(2), 155-175.
- Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning. Towards a wider conceptual field. *Assessment in Education: Principles Policy and Practice*, 5(1), 85-102.
- Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: ASCD.
- Princeton Review. (2006). *Formative assessment*. Framingham, MA: Princeton Review.
- Ryle, G. (1949). *The concept of mind*. London, UK: Hutchinson.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 1, pp. 39-83). Chicago, IL: Rand McNally.
- Searle, J. R. (1995). *The construction of social reality*. London, UK: Allen Lane, The Penguin Press.
- Shepard, L. A. (2008). Formative assessment: caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: shaping teaching and learning* (pp. 279-303). Mahwah, NJ: Lawrence Erlbaum Associates.
- Shepard, L. A., Hammerness, K., Darling-Hammond, L., Rust, F., Snowden, J. B., Gordon, E., Gutierrez, C., & Pacheco, A. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: what teachers should learn and be able to do* (pp. 275-326). San Francisco, CA: Jossey-Bass.
- Stiggins, R. J. (2005). From formative assessment to assessment FOR learning: a path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328.
- Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: doing it right—using it well*. Portland, OR: Assessment Training Institute.
- Sutton, R. (1995). *Assessment for learning*. Salford, UK: RS Publications.
- Thompson, M., & Wiliam, D. (2008). Tight but loose: a conceptual framework for scaling up school reforms. In E. C. Wylie (Ed.), *Tight but loose: scaling up teacher professional development in diverse contexts* (Vol. RR-08-29, pp. 1-44). Princeton, NJ: Educational Testing Service.
- Torrance, H. (1993). Formative assessment: Some theoretical problems and empirical questions. *Cambridge Journal of Education*, 23(3), 333-343.
- Wainer, H. (2011). *Uneducated guesses: Using evidence to uncover misguided education policies*. Princeton, NJ: Princeton University Press.
- Wiliam, D. (1999a). Formative assessment in mathematics part 1: rich questioning. *Equals: Mathematics and Special Educational Needs*, 5(2), 15-18.
- Wiliam, D. (1999b). Formative assessment in mathematics part 2: feedback. *Equals:*

- Mathematics and Special Educational Needs*, 5(3), 8-11.
- Wiliam, D. (2000). Formative assessment in mathematics part 3: the learner's role. *Equals: Mathematics and Special Educational Needs*, 6(1), 19-22.
- Wiliam, D. (2007). Keeping learning on track: classroom assessment and the regulation of learning. In F. K. Lester Jr (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053-1098). Greenwich, CT: Information Age Publishing.
- Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). New York, NY: Taylor & Francis.
- Wiliam, D. (2011a). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 2-14.
- Wiliam, D. (2011b). *Embedded formative assessment*. Bloomington, IN: Solution Tree.