

Formative Assessment: Policy, Perspectives and Practice

Ian Clark
University of Washington

*Florida Journal of Educational
Administration & Policy*

Spring 2011
Volume 4, Issue 2

Proponents of formative assessment (FA) assert that students develop a deeper understanding of their learning when the essential components of formative feedback and cultural responsiveness are effectively incorporated as central features of the formative assessment process. Even with growing international agreement among the research community about the benefits of FA in improving student learning, examination performance and promoting life-long learning, the standards-based reform agenda found in the US has created a politically inhospitable climate for the assessment-driven transformation of classroom practice. This article draws from some 75 sources on instruction and assessment in order to make sense of a) the debate surrounding FA in the US and b) the cultural and individual developmental issues encapsulated within the extensive conceptual territory that is FA. The article has five main sections: the emergence of the term 'formative assessment'; FA in the context of the US; the strategies and principles on which formative practices are founded; the debate in the US on formative assessment; formative assessment and culturally responsive pedagogy.

Key words: formative assessment, culturally responsive pedagogy, life-long learning

A Global Evolution

Despite being hailed as a ‘quiet revolution’ (Hutchinson & Hayward, 2005) it is perhaps more accurate to see the growth of the global awareness regarding formative assessment (FA) as evolutionary in nature. Before the term ‘formative’ existed, the earliest allusion to formative approaches may be traced back to 1963 and Cronbach’s seminal article on the improvement of course content. Four years later, Scriven (1967) originated the term ‘formative’ applying it in a manner consistent with Cronbach’s approach to the evaluation of whole programs. A crucial development in the traditions of FA is the progressive replacement of Scriven’s original term of ‘formative evaluation’ of educational programs by the term ‘assessment’ when the object is student learning in the classroom (Allal & Lopez, 2005). Since that time, educational researchers have emphasized balanced classroom assessment practices that support academic achievement *and* the cultural development of the ‘whole child’ (Bloom, Hastings, & Madaus, 1971; Sadler, 1989; Black & Wiliam, 1998a, 1998b, 2006a; Assessment Reform Group [ARG], 1999; Organization for Economic Co-operation and Development [OECD], 2005; National Association of State Boards of Education [NASBE], 2009). The consequence has been the implementation of long-term policy initiatives on an international scale. One example of the remarkable expansion in awareness regarding the benefits of FA is a 2005 Organization for Economic Co-operation and Development (OECD) study that features exemplary cases from secondary schools in Canada, Denmark, England, Finland, Italy, New Zealand, Queensland in Australia, and Scotland. Despite the growing global adoption of FA practice there is a relatively weak policy agenda for such transformation in the US.

Formative Assessment in the US

Black & Wiliam (2005) observe that the effective integration of formative and summative assessment will require a different change-management strategy depending on national circumstances and in some cases may be very challenging indeed. The Council of Chief State School Officers (CCSSO) present the “core” challenge to the effective implementation of FA in the US:

...despite the pioneering efforts of CCSSO and other organizations in the U.S., we already risk losing the promise that formative assessment holds for teaching and learning. The core problem lies in the false, but nonetheless widespread, assumption that formative assessment is a particular kind of measurement instrument, rather than a process that is fundamental and indigenous to the practice of teaching and learning. (Heritage, 2010, p. 1)

Current policies founded upon ‘scientism’ prompted The National Association of State Boards of Education (NASBE, 2009), which undertook a study on assessment systems for the 21st century learner to remark that, “a growing majority of testing experts and analysts now believe that education cannot be transformed under the constraints of the current state assessment and accountability systems” (p. 3). Consequently, circumstances do not favor the

inception of FA in the US. As L. A. Shepard noted at the 2005 Educational Testing Service (ETS) invitational conference:

The arrival of formative assessment in America was ill timed. This potentially powerful classroom-based learning and teaching innovation was overshadowed almost immediately by the No Child Left Behind Act (NCLB) (January 2002) with its intense pressure to raise scores on external accountability tests. (p. 2)

The advent of NCLB (which reauthorized the Elementary and Secondary Education Act [ESEA, 1965]) is particularly important in setting the current political tone for the continuing discussion on the applicability of FA in American classrooms. At present, the political terrain in the US creates an inhospitable environment for the transformation of education. Prevailing scientific rationalism ('scientism') requires the clear definition and measurement of an instructional process in order to establish its validity. With the passage of NCLB and subsequent infusion into state education systems throughout the US, standardized assessments went from being a method of determining the academic achievement of the student, school or district to being the only metric that mattered. The current emphasis on standardized testing has created a source for doubt regarding the future of FA in the US on the grounds of: a) the lack of an agreed upon (or standardized) definition of the term 'formative assessment' (Dunn & Mulvenon, 2009; Dorn 2010, Bennett, 2011); b) the inconsistent adoption of FAs in practical settings (Dorn 2010; Young & Kim, 2010) and; c) the poor quality of quantitative data used to support claims that FA practices improve academic achievement (Dunn & Mulvenon, 2009; Bennett, 2011).

It is ironic then that the underlying post-structuralist values (cf. Foucault; Bourdieu) of the formative process are: the explicit rebuff of 'scientism'; the transformation from passive to active learning; the democratic values of equality, representation and consensus; mutual discourse and the circulation of discursive power among students; an inquiry into the existing order and monolithic characterizations of history, society and culture. These values constitute an essential philosophical foundation, with which instructional practices and processes should be consistent. The consequence is to neglect the values of FA risks instruction which neither supports development of autonomous learning strategies among students (Cornford, 2002; White & Frederiksen, 2005) or culturally responsive teaching. This trend was confirmed by a US Department of Education commissioned review by the Mid-Continent Research for Education and Learning [McREL] (Kendall et al., 2008). The review of seven Central Region states (Colorado, Kansas, Missouri, Nebraska, North Dakota, South Dakota and Wyoming), titled "21st Century Skills: What do we expect of students?" analyzed the extent to which school standards documents referred to meta-cognitive skills and motivation essential to FA. The McREL report found that there was 'some evidence' that standards' documents supported the use of meta-cognitive strategies. Including, "setting their own learning goals, monitoring their progress toward learning goals, monitoring their thinking processes for accuracy and for clarity" (Kendall et al., 2008, p.3). However, the standards were found to be inconsistently applied across states and students were rarely expected to set their own goals except in the English language arts. In terms of motivation the report noted, "there were no clear statements in the documents reviewed that students should examine their motivation to clear" (p.4). In addition, the analysis did not reveal the inclusion of statements about the importance of culturally responsive teaching or student self-

efficacy. Self-efficacy is fundamental to FA because students are more likely to actively engage with their work, even lengthy, difficult or tedious tasks, if they believe they are competent at monitoring and regulating their thinking skills (Bandura, 1997; Brophy, 2004). Kendall and his colleagues at McREL found a small number of references to persistence (particularly in mathematics), however, the analysis did not find any reference to what teachers should do to create conditions in which persistence may exist.

In addition to the analysis of state standards, recent studies have addressed the achievement gap in the US. “The McKinsey Report” (2009), compiled in collaboration with the US Chamber of Commerce and the Bill and Melinda Gates Foundation, provides very useful analysis on the relevance of the achievement gap in economic terms and makes a most compelling statement:

The persistence of these educational achievement gaps imposes on the United States the economic equivalent of a permanent national recession. The recurring annual economic cost of the international achievement gap is substantially larger than the deep recession the United States is currently experiencing. The annual output cost of the racial, income, and regional or systems achievement gap is larger than the US recession of 1981–82. (p. 6)

The report indicates that, despite the implementation of the NCLB Act (2002), a significant multi-dimensional (international, racial, income, and systems-based) achievement gap remains. A finding of particular significance was that 15 year old American students exhibit the lowest mathematics skills of any industrialized nation; the age at which they may be considering entering the competitive job market.

The Program for International Student Assessment (PISA) provides a measure of the challenges facing American public schools in the 21st century. The PISA is a respected triennial international comparison of 15-year-olds by the OECD that measures real-world (applied) learning and problem-solving ability. In December 2010, Angel Gurría (OECD Secretary General) summarized the performance of the US in 2009 via an ‘interactive town hall’ hosted by The Innovation Economy partners (The Aspen Institute, PBS NewsHour, Intel Corporation and The Information Technology & Innovation Foundation [ITIF]) and attended by Education Secretary Duncan. Secretary General Gurría revealed the US comes out as an average performer in reading (rank 14 in OECD) and science (rank 17) but the US drops below the OECD average in mathematics (rank 25) (The Innovation Economy, 2010). Secretary General Gurría also noted that, “unlike most other federal nations, the U.S. does not yet collect PISA data for individual states, but we understand that there are important regional differences in performance” (The Innovation Economy, 2010). This relatively poor ranking is attributed to the improvements in academic achievement experienced by other nations over the past 40 years, during which time the US exhibited no gains in performance in math, science or reading (McKinsey, 2009). The general inefficacy of standards documents to support instructional practices that actively engage students in the planning, monitoring and reflection of their work (Cornford, 2002; Kendall *et al*, 2008) has created the circumstances for critical theorists to advance their argument for a transformation of classroom practice.

One such transformative practice is best known in the US as formative assessment, the potential of which is recognized with increasing regularity by a growing number of

education researchers (ARG, 1999; Stiggins, 2002, 2007; Harlen & Deakin-Crick, 2003; OECD, 2005; Cauley & McMillan, 2010; Bennett, 2011). It is usual for the literature on assessment to compare the post-structuralist values of FA with the values of 'scientism' which underpin summative evaluation in three main areas: a) high-stakes testing creates anxiety and disaffection among students (ARG, 1999; Harlen & Deakin Crick, 2003; Harlen 2006); b) the political emphasis on school-level data narrows the instructional content taught to students distorting how and what students are taught; and c) the intense focus on testing de-skills, de-motivates and de-professionalizes teachers (Shepard, 2000; Abrams, 2007). Noddings (2004) contends that standardized tests "are loaded with trivia" and students are "being fed intellectual junk food" (p. 494) to succeed on these tests. Noddings's critical tone is perhaps some indication of the current level of political tension over student assessment in the US. The negative effects of high-stakes testing have been empirically demonstrated by various US studies. For example, Pedulla et al. (2003) conducted a national survey and found that 90% of teachers reported feeling pressure from district administrators and 79% felt pressured by their principal to improve test performance. The study also revealed that one-third of teachers wanted to transfer out of tested grades.

Formative Assessment: Strategy & Principles

Formative Feedback

The starting point for the work on FA was the idea of providing feedback (Clark, 2010b). Hattie and Temperley (2007) contribute to an understanding of feedback as a powerful instructional approach in their meta-study that derived effect sizes for different kinds of feedback. They obtain high effect sizes when students are given 'formative feedback'; feedback on *how* to perform a task more effectively and far lower effect sizes when students are given praise, rewards or punishment. It is the notion of students being given strategic guidance on how to improve their work that resides at the center of FA practice. Torrance & Pryor (1998) argue, "many teachers focus on praise as a form of 'feedback' because of the efficacy of behaviorist reinforcement systems" (p. 40). However, when feedback is used effectively it is, "the most powerful single moderator that enhances achievement" (Hattie, 1999 p. 9). Feedback becomes formative when students are provided with scaffolded instruction or thoughtful questioning that serve as prompts for sustained and deeper discussion. This instructional approach closes the gap between their current level of understanding and the desired learning goal. Simply telling a student to 'try again' or 'reconsider your work' does not possess the qualities of formative feedback because it does not strategically guide (or scaffold) learning by telling the student how or why they need to do this. Feedback, therefore, becomes formative when learners: a) are engaged in a process which focuses on meta-cognitive strategies that can be generalized to similar problems of varying degrees of uniqueness; b) are supported in their efforts to think about their own thinking; c) understand the relationship between their prior performance, their current understanding, and clearly defined success criteria; and d) are activated as owners of their own learning.

Principles

Formative classrooms are founded upon key principles which, when taken together, provide the starting point for what FA looks like in practice. For effective FA to take place a classroom must exhibit climactic conditions that permit high quality interactions between students and teachers. Mutuality and equality are key features of the FA classroom, and essential predicates for culturally responsive teaching (Damon & Phelps, 1989; Bishop & Glynn, 1999; Black & Wiliam, 2006a). The principles that appear here are derived from a variety of sources (Assessment Reform Group [ARG], 1999; Black & Wiliam, 1998a, 1998b; Sadler, 1989; Stiggins 2007; AAG/APMG, 2002-2008; OECD, 2005). The development of FA strategies is connected by two closely related assessment themes: assessment *for* learning (A f L) and assessment *as* learning (A a L). A f L is primarily a teacher initiated student-centered aspect that focuses on the progress of the learner toward a desired goal, seeking to close the gap between a learner's current status and the desired outcome. This can be achieved through processes such as sharing criteria with learners, effective questioning and feedback (AAG/APMG, 2002-2008) A a L is an aspect which structures peer engagement and supports autonomous learning and assessment. An emphasis is placed upon the co-reflection on evidence of learning. It is a process where pupils set learning goals, share learning intentions and success criteria, and evaluate their learning through dialogue and self and peer assessment (AAG/APMG, 2002-2008). Taken together A f L and A a L are carefully designed to encourage a real-time feedback loop between teacher and student and between peers.

There are several key principles underpinning the A f L aspects of FA, the precise expression of which is contingent on national political and cultural circumstances. National circumstances aside, all standards and policy documents should be consistent with post-structuralist values in the pursuit of a classroom environment in which students are: a) able to understand clearly what they are trying to learn, and what is expected of them; b) given immediate feedback about the quality of their work and what they can do to make it better; c) given advice about how to sustain improvement; d) fully involved in deciding what needs to be done next; and e) aware of who can give them help if they need it and have full access to such help. In addition, there are 2 key principles underpinning A a L, which hold that students should: a) be able to build knowledge of themselves as learners and become meta-cognitive, and b) take more responsibility for their learning and participate more in the process of learning with their teacher as their advisor and with their peers in a climate of equality and mutuality (Clark, 2010a, 2010b).

Moments of Contingency

Black and Wiliam's (2009) theory of formative assessment departs from other conceptions of instruction. They explain that, "it is clear 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" (p. 10). They go on to emphasize that these moments of contingency can be either synchronous or asynchronous. Asynchronous feedback includes the use of evidence gathered from homework or from the students' summation of their understandings at the end of a lesson that is used to plan the next lesson. As such, the feedback loop can be various lengths ranging from Shepard's (2005) synchronous use of the word "immediately" or Harlen's (2006) transitional reference to the

use of evidence in “the near future” to Black and Wiliam (2009) who concur with both Shepard and Harlen and include the long-range asynchronous use of insights gained from previous years.

For practitioners to implement FA practices in their classrooms it is necessary for them to develop instructional skills that create and capitalize upon moments of synchronous contingency. This is a very challenging first step, or as Black & Wiliam (2009) put it, “a formidable problem for teachers” (p. 13). It is formidable because a) it exposes teachers to the many ways in which students argue, evaluate and synthesize information for problem solving purposes and b) it often requires a significant change in instructional approach (Black & Wiliam, 2006a; Black & Wiliam, 2009). The skill is then to be found in how teachers shape and regulate the interaction as one may work on any creative task or project. That is by more effectively applying their skills by planning for appropriate environmental conditions and constantly monitoring and evaluating the process. In later sections, the observable data arising from sustained interactions between teachers and their students will be analyzed. It will be seen that gathering evidence of student learning by engaging in synchronous dialogue goes to the very heart of teaching as a social accomplishment.

The Formative Assessment Debate

Dr. Mary Kay Sommers (former President of the National Association of Elementary School Principals [NAESP]) expressed her concerns about the emphasis attached to high-stakes testing at the Miller-McKeon draft of the reauthorization of the Elementary and Secondary Education Act (ESEA): “As in current law, standardized test scores are raised to a level of importance that education research and practice indicate is unwarranted,” (2007, p.3). It will be seen in this article that many researchers concur with Dr Sommers in her criticism of the disproportionate political emphasis placed upon school-level accountability protocols designed to obtain high performance outcomes (grades) as opposed to the emphasis on classroom-level instructional processes designed to improve student learning which in turn realize better examination performances. Dr. Sommers continued: “This is not about avoiding accountability; it’s about addressing the needs of the whole child and making sure that all is done that should be done to help every student succeed without penalty or fear of failure” (2007, p.4).

While Sommers does not refer to formative assessment specifically, there is a clear parallel with the vision of a school where everyone is learning together and assessment is integrated into the process of instruction. In addition to the claim that high-stakes testing causes psychological harm, NASBE (2009) raise a further problematic aspect of NCLB: “Overall, standards-based reform has been disappointing: standards that emphasize complex reasoning and problem-solving skills are often ignored in the classroom because they are not measured by most state accountability tests” (pp. 3 – 4). Abrams (2007) confirms the criticism of NASBE by reporting the concerns of teachers who responded to NCLB by sacrificing instructional approaches that emphasized the depth of knowledge in order to cover the breadth of content. Abrams remarks, “the pressure to ‘cover’ materials resulted in the elimination of what teachers characterize as ‘good’ lessons and important topics” (2007, p. 83).

Those who express skepticism about FA do so from the realization that the provisions of NCLB are rigorous and potentially punitive for those schools that do not meet Annual

Yearly Progress (AYP) targets. They launch their critiques from two closely related platforms. Firstly, the term formative assessment is so flexible so as to render it meaningless in practice (Young & Kim, 2010; Bennett, 2011), consequently the definition of FA is “fuzzy” (Dorn, 2010) and “ethereal” (Dunn & Mulvenon, 2009). Secondly, Dorn (2010) points to the inconsistent adoption of FA practices. He remarks that there is a lack of regularity, structure and uniformity that is required in a high-stakes testing environment as found in the US. In sum, FA lacks the statistical reliability expected of assessment practices in the US at the current time. A third criticism arises from the methodological problems associated with quantitative studies (Dunn & Mulvenon, 2009; Bennett, 2011) and the qualitative data collection methods used to support the claim that FA improves achievement (Bennett, 2011). The remainder of this section will discuss these criticisms.

Defining Formative Assessment

The definition of formative assessment developed by the *Formative Assessment for Teachers and Students* (FAST) group was meant to be accessible to US educators. After considering various existing definitions and revisions to their own version it was determined that: “Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes” (cited in Popham, 2008, p. 5).

Young and Kim (2010) are, therefore, accurately applying the adjective ‘flexible’ and correct in seeing it as a characteristic worthy of closer analysis. For FA to be successfully implemented, the term ‘formative assessment’ should not be understood as static or rigid in any sense. It is important to recall that FA takes place in the locality of each individual classroom. As such the process concerns itself exclusively with the activities undertaken by teachers and their students during the process of classroom instruction. If that statement has the ring of familiarity about it then it is because it is to be found in perhaps the most quoted explanation of FA as provided by Black and Wiliam, who include a ‘what’ element and a ‘when’ element’ within their conception:

‘What?’: all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to *modify* the teaching and learning activities in which they are engaged. *‘When?’* ‘Such assessment becomes ‘formative assessment’ when the evidence is used to *adapt* the teaching work to meet the needs. (1998b, p. 2)

In a relatively short time, therefore, we have answered key questions regarding FA: *Where* it takes place (in the classroom); *when* it takes place (during the process of instruction); *what* it is (it is a strategic instructional approach, strategic in that it is adapted to meet the needs of individual students) and finally, *when* assessment becomes formative (when it is so adapted to close the gap between a student’s current level of understanding and the target). However, no matter how clearly FA is delineated it has yet to bring unity to the wide variety of perspective on FA to be found within the Educational Testing Service (ETS) (Wylie, 2011, personal communication). Bennett (2011), a distinguished scientist in the Research and Development Division at the ETS, consistently points to the difficulty in validating claims that FA improves examination performance. The problem begins with the lack of a unitary

definition, without which one cannot establish what FA is designed to achieve. This then makes measuring its effectiveness as an intervention, which improves attainment problematic because it is unclear exactly what kinds of observational data should be gathered, and by which qualitative data collection methods (Bennett, 2011).

Young & Kim's (2010) reference to the nebulous nature of formative assessment is expressed by the verbs of action that reside at the heart of FA - 'modify' and 'adapt'. Adaptation is a dynamic and flexible process and a process that is most central to FA practice. Advocates reason that the principles of FA are not prescriptive structures because each classroom has unique conditions to which the principles are applied as each practitioner deems appropriate. FA is anchored in classroom practice and clearly characterized by the modification and adaptation of teaching practice to meet the individual needs of students in the social setting of the classroom. Wiliam and Leahy (2007) remind us that the fundamental need for adaptation means that any prescription for teacher action, which seeks to define the 'best' course of action given certain conditions, is impossible even in principle. Young & Kim's concerns regarding the excessive flexibility of the term closely relate to the lack of an agreed upon 'gold standard' definition. Dorn (2010) writes of FA, "its definition is fuzzy, its adoption is inconsistent, and the prognosis for future use is questionable" (p. 325). Similarly, Dunn and Mulvenon (2009) remark that it is an "ethereal construct...perpetuated in the literature due to a lack of agreed upon definition" (p. 2).

Yet others question the value of defining educational terminology. Simply sharing definitions with practitioners is unhelpful because "all research findings are generalizations and as such are either too general to be useful or too specific to be universally applicable" (Wiliam & Leahy, 2007 p. 39; Wiliam, 2003). The essence of FA does not reside in an agreed upon definition, but in a principled application of formative practice to the specific learning interactions taking place. To seek a standardized definition of any strand of dynamic assessment is counter-intuitive and unrealistic in the diverse practical settings found in thousands of unique classrooms. No matter how elegantly we formulate definitions, which seek to standardize teaching practice across classrooms, they would remain inoperable in practical settings and are somewhat reminiscent of the 'scientific management' movement of the late-19th century, which sought to maximize production efficiency. Brian Gong (2007), Executive Director of the National Center for the Improvement of Educational Assessment alludes to retrospective "19th century" policy in his response to the Miller-McKeon Discussion Draft of the Reauthorization of the ESEA:

I can imagine that unless an investment is made, educational testing in 20 years will be as hobbled by a lack of imagination and by 19th century measurement theories as it is today. Reauthorization should look to the future as well as try to make mid-course corrections to the present. (p. 4)

The professional development of classroom assessment practice requires teachers to understand the potential for the social construction of knowledge to improve student learning, particularly teaching strategies that emphasize high-quality interactions.

Inconsistent Practice

Dorn (2010) remarks, “although formative assessment is appealing in theory, its practice as well as its definition is inconsistent” (p. 328). His concerns over the inconsistent adoption of FA closely relate to the discussion found in the previous sub-section because consistency is best facilitated by a well-defined system, such as Response To Intervention (RTI), which, at least in theory, minimizes the error in determining the most proximal intervention. Dorn, therefore, suggests the development and incorporation of what he calls ‘*structured* formative assessment’ (p. 326) into existing standardized frameworks such as RTI with a view to improving the system. Dorn along with Young and Kim (2010) are prepared to accept FA when it means the regulated response to data on student performance when compared against a set of documented micro-standards as found on benchmark assessments such as RTI or the US Department of Education’s ‘Reading First’ initiative. Dorn argues that this ensures “fidelity, regularity and use of *structured* formative assessment” (p. 328). He also implies that teachers are more likely to “uniformly respond to data by changing instruction” (p. 328) when formative assessment data is used within a structured framework.

In contrast, Shepard (2005) and Wiliam (2004) believe that benchmark assessments such as RTI are not directly formative: “What makes formative assessment formative is that it is *immediately* used to make *adjustments* so as to form new learning” (Shepard 2005, p.4). The conception of FA as discussed in Black & Wiliam’s (1998b) seminal contribution is anchored in the world of classroom interaction, where FA is not just linked to instructional goals but is synchronously (Black & Wiliam, 2009) integrated into the process of instruction itself. Teachers should therefore understand that opportunities for supporting learning wink in and out of existence in spans of time (‘moments of contingency’, Black & Wiliam, 2009) that make standardized measurement practices (e.g. RTI) appear pedestrian and uneventful.

The word ‘inconsistent’ is indeed a prominent feature of FA practice, although less pejorative terminology (e.g. responsive) could be applied without partisan bias. The key notions of modification and adaptation to meet the needs of specific learners or groups of learners, is a key feature and one that undermines consistent application and standardization. As already noted, each classroom is unique and teachers will interact with students in ways that are consistent with their collective conceptions of FA but which will be inconsistent with the practice of their colleagues engaged as they are in entirely different learning interactions, with their own students, in different contexts. Consistent practice is very difficult to ensure in any circumstance. Consider the National Association of State Boards of Education (NASBE) statement regarding the standards imposed on schools by NCLB: “teaching of the standards varies widely from classroom to classroom even within the same school” (2009, p. 4). Perhaps no further proof is required to assert *de facto* that public school classrooms are so unlike each other so as to defy *any* attempt to standardize teaching practice. Perhaps more realistically classrooms are understood as complicated and dynamic settings in which learning is improved when teaching is contingent on the unique interactions taking place. This makes the task of becoming an effective practitioner of FA very challenging indeed. Black and Wiliam (2006a) note the size of that challenge when they observe:

The task of developing an interactive style of classroom dialogue required a radical

change in teaching style for many teachers, one they found challenging not least because it felt as if they were losing control. Some were well over a year into the project before such changes were achieved. (p. 14)

Benchmark and interim assessments have been adopted by many school districts to help monitor progress during the school year toward meeting state standards and NCLB performance goals. Typically these assessments are formal and provide teachers with information about the strengths and weaknesses of individual students against content standards. Wiliam (2004) calls such information “early warning summative”, and Shepard (2005) remarks that the individual profile data from these assessments are not directly formative because a) the data available are at too gross a level of generality and b) feedback for improvement is not part of the process.

Quantitative Research Methods

It is against the backdrop of NCLB implementation that Dunn and Mulvenon (2009) take a critical view of the empirical studies, which claim that FA improves achievement. They are particularly critical of meta-analyses, which combine studies of mixed methodological quality. An article by Wiliam, Lee, Harrison, and Black (2004), “Teachers Developing Assessment for Learning: Impact on Student Achievement” is specifically mentioned in their critique. Although Dunn and Mulvenon describe the results as “promising”, the 2004 report did little to convince quantitative scholars or ETS scientists that FA creates measureable increases in performance, due to a number of concerns regarding internal validity. Wiliam et al. (2004) do not attempt to defend their methodology and in fact, concede that, “the quantitative evidence provided here is difficult to interpret. The comparisons are not equally robust.” (p. 62). As such, each of their results reflects a separate ‘mini-experiment’ and “care needs to be taken in drawing general conclusions about the net effect of the adoption of formative assessment” (p. 60).

The difficulty of researching questions related to the impact upon motivation and achievement from a quantitative perspective arises from the “flexible and unplanned” nature of “high-level formative assessment” (Cauley & McMillan, 2010, p. 2) and the necessary conception of each classroom as a unique practical setting unlike any other. The endeavor to devise categories from which FA processes can be accurately coded to create precise numerical patterns is confounded by the quantity and complexity of the dynamic, multiplicity of interactions taking place between student and teacher, the grain of which may only be comprehensively understood by discourse analyses. Scotland provides a good example of large-scale implementation of FA practices that took place in some 1500 schools between 2002 and 2009. Charlie Penman (Head of Assessment & Recognition of Achievement) at the Scottish Government, remarks that:

The Scottish Government does not hold data on the impact of formative assessment on performance of participating students. We have, however, conducted surveys of teachers involved in the Assessment is for Learning Programme five years after the inception to gather views. These views are, of course, subjective and not statistically reliable for performance data. (personal communication, 2009)

The preference was to engage in a detailed qualitative study designed to capture the experiences and perspectives of participating teachers; in effect each participating teacher presented their own case study on the impact in their own classroom. Compare this approach with the views of Randy Bennett at the ETS (stated earlier) and it is clear that there is a profound degree of difference between the political culture in Scotland and that found in the US; a situation confirmed by the words of Dr. Sommers, the then President of the NAESP:

We are disappointed to see that it retains core provisions, most notably a reliance on high-stakes testing...NAESP does not support these core provisions because we believe them to contradict our aims as chief architects of learning within our school *communities*, and in fact, believe them to be harmful to children. (2007, p .3)

Teachers in Scotland are listened to carefully and consulted as key architects in long-term national policy development. This is a sharp contrast with the situation in the US, where the professional opinions of practitioners are seen as little more than a marginal note.

Despite the challenges encountered by teachers in the Scottish region of the UK as they implemented FA practice, they responded very positively on an evaluative rating scale: 100% of respondents reported that the project was 'very successful' or 'successful' (VS/S) in increasing teacher understanding of the role of assessment and in improving teacher motivation. Further, 91% of respondents indicated that the project was VS/S in changing classroom practices; 92% indicated the project was VS/S at changing assessment practices; 94% the extent of questioning; 95% in increasing the level of discussion, and 97% of respondents assigned a VS/S rating to an improved teacher focus on the learning needs of students (Hallam, Kiston, Peffers, Robertson & Stobart, 2004). Undeniably, there is a lack of statistical inference regarding the impact on academic motivation and achievement. However, the essential point remains that quantitative ambiguities are not sufficient to obscure or undermine the deep cognitive and meta-cognitive processes germinated by high-quality interaction and effective feedback; a point widely understood and accepted (Piaget, 1965; Vygotsky, 1978; Sadler, 1989; Butler & Winne, 1995; OECD, 2005, AAG/APMG, 2002-2008).

Culturally Responsive Teaching

FA techniques arise from a creative process dependent on the values of active participation, equality, representation and consensus (Black & Wiliam, 2006a; Clark, 2010b). The relationship between discourse and social power relations is of paramount importance to FA and, therefore, to culturally responsive teaching (Bishop & Glynn, 1999). When discussing how FA 'encapsulates' culturally responsive education in practice, it is necessary to examine the ideas of progressive critical theorists. For example, Barabara Rogoff's (2003) inquiry into cultural communities and their development provides a starting point for teachers as they seek to understand individual development in the cultural context of their own classrooms: a) children do not develop in a uniform way through stages regardless of the culture in which they are situated; b) culture is not monolithic and uniform; and c) neither is it separate from the individual. Vygotsky (1987) expressed the inseparability of individuals from the cultural context in which they learn, observing that in order to understand how water extinguishes a flame one does not attempt to reduce it to its constituent elements of oxygen and hydrogen. Vygotsky remarks that the scientist will

“discover, to his chagrin that hydrogen burns and water sustains combustion. He will never succeed in explaining the characteristics of the whole by analyzing the characteristics of the elements” (1987, p. 45). In addition, to the rebuff of dichotomous conceptualization of individual or group constructs, the generalizing of theory on the influence of culture across communities of students is not possible because of the cultural diversity found in America’s classrooms. Dewey (1916) also rejects generalizations in the form of “universal rules and other invariant instructional elements” (p.13) and argues for the careful consideration of how individual students learn in situational or local conditions.

Exploring the ‘Open Classroom’

Dewey, like Rogoff, was intrigued by the relationship between tacit aspects of the individual (experiences, ideas, beliefs, etc.) and their contribution to the various communities that constitute society. Dewey was aware that ‘formal schooling’ is not based on the concept of ‘community learning’ (cf. Lave & Wenger, 1991) and proposed alternative models of adult control and child freedom by contrasting teachers’ roles in schools with roles in a community of learners (Black & Wiliam, 1998b; 2006a; Rogoff, Goodman Turkanis, & Bartlett, 2001). Dewey (1938) challenged that the primary relationship between adult and child should be asymmetrical and controlling. He felt that the democratic process should be accessible to all regardless of race or socio-economic status. This entailed the collaboration between adults and children as mutual partners who share responsibility and play different cultural roles. Evidently, Dewey thought it was the cultural responsibility of educators of all kinds to listen to the opinions, ideas and beliefs of students and collaborate with them instead of inferring their needs. While qualified teachers are *de jure* professional educators, it should be considered that this right does not extend to *de facto* expertise which permits educators to infer the individual needs of the ‘novice child’. Indeed, progressive critical theorists (e.g. Dewey) explicitly refute that effective teaching is in any way connected to control over others. Consider the Open Classroom environment found in the Salt Lake City Charter School (OCCS) in Utah:

Students at the Open Classroom learn to assume responsibility for their own education by planning their activities for the week, balancing “must do” assignments with optional activities including independent exploration. During the course of the week, the student, parent and teacher will track work together so that, if necessary, the student can be guided toward making choices to support the completion of work. (OCCS, n.d. p. 2)

Productive discourse arising from the ideas of students is fundamental to the ‘formative classroom’ and to the creation of a culturally responsive classroom. Maori education scholars Bishop and Glynn (1999) assert that teachers need to develop insights into their own “preconceptions, goals, aspirations and cultural preferences” and to “be prepared to listen to others in such a way that their previous experiences and assumptions do not close them off from the full meaning of the student description of their experience” (p. 158).

There is a growing interest in the notion of a situated community of learners arising from ideas on the co-construction of cultural texts expounded by Dewey and Vygotsky and their contemporaries: Black & Wiliam (1998b; 2006a), Lave & Wenger (1991), Polanyi

(1967), Schön (1987) and Rogoff (2001; 2003) to name but a few. In the 2001 book *Learning Together: Children and Adults in a School Community*, Rogoff and her colleagues, Goodman Turkianis and Bartlett focus on the OCCS and how it provides a unique learning culture which implicitly draws from the post-structuralist values that underpin the theory of FA. The approach focuses on developing students who will be lifelong learners able to work in a collaborative environment. The 'open classroom' environment of OCCS occasions an opportunity to analyze the democratic community ideas of Dewey and the importance of 'tacit knowledge' (cf. Polanyi, 1967). Such elements are inherent to the theoretical frame of FA as the 'theory of formative assessment' draws considerable energy from research that emphasizes mutual learning relationships (Storch, 2002), positive interdependence (Johnson & Johnson, 1975) and a more open classroom which values the fluidity of tacit knowledge (Polanyi, 1967; Schön, 1987).

Rogoff and Guitierrez (2003) observe that many school systems advocate a single way of teaching and learning without accounting for individuals' past experiences with certain local or situational practices. This stands in stark contrast to the strategic forms of assistance and robust learning communities, which routinely carry out formative assessments. For Polanyi (1967), creative thinking actualizes scientific discovery and cognitive growth. Such thinking is charged with personal opinions and beliefs. It is these personal inner thoughts that the culturally responsive classroom seeks to reveal. Participants share the learning activity, which mediates the collaboration, and each interactant possesses unique tacit knowledge (Polanyi, 1967) as a result of their different cultural experiences. Polanyi (1967) and Schön (1987) both note the personally empowering nature of tacit knowledge:

...it is personal, in the sense of involving the personality of him who holds it...but there is no trace in it of self-indulgence...His act of knowing exercises a personal judgment in relating evidence to an external reality, an aspect of which he is seeking to apprehend. (Polanyi, 1967, pp. 24-25)

In a FA classroom, value is placed on the students' personal beliefs, opinions, guesses and misconceptions. When students are permitted to express who they are, they develop a strong sense of self-efficacy, a central requirement for autonomous learning strategies (cf. Bandura, 1997; Zimmerman, 2002). The social context edifies them and personalizes the learning experience for students. Schön (1987) expresses the same idea slightly differently by noting that learners engage with the process of learning when they are guided to realize that they too hold opinions and theories on similar themes, "they tend to think differently about the theories offered by researchers when they realize that they hold comparable tacit theories of their own" (p. 324). Polanyi (1967) and Schön (1987) emphasize the formative and reflective purpose of discourse and the social context characterized by an open classroom community where ideas and opinions are exchanged and differences are seen as opportunities to co-construct shared meanings. It is this variability in their experience that makes it possible for students to scaffold each other's understandings in at least some aspects of the activity. In such culturally dynamic learning communities students receive multiple forms of individual assistance and participate in rigorous learning activities that extend their initial approaches to learning (Rogoff et al., 2001). As a result, students have ongoing opportunities to assume new roles and learn new approaches. Taken together these propositions clarify the way in

which cultural and individual development are both interdependent and rooted in particular, situated occasions of joint activity.

The processes of independent goal-setting, prioritizing, the collaborative monitoring of learning progression and 'scaffolding' for further development are culturally responsive strategies that create the conditions for autonomy and self-regulated learning (SRL) associated with lifelong learning (Butler & Winne, 1995, Zimmerman & Pons, 1986; Organization for Economic Cooperation and Development [OECD], 2005). Accordingly, the OCCS state: "Our goal is that all students, with practice, will learn to make those choices independently" (n.d. p. 2).

Formative Assessment Techniques

Practitioners may be trained to teach students in a way which capitalizes on individual experience and supports autonomous learning (McCombs & Whistler, 1989). Clark (2010a) noted 16 FA teaching strategies: higher order questioning techniques; use of problem solving techniques; jot time (students are afforded an extended period of time to commit their ideas in writing before the beginning of an interaction); use of misconceptions; wait time; traffic lighting; group work and pair work; discussions; feedback as comments and not grades; oral feedback; sharing assessment criteria; peer assessment; redrafting of work; developing peer-peer communication skills (relational skills training); collaborative goal setting; and reflective learning.

Case Study: New Zealand

In the OECD report on formative assessment in secondary schools, Looney and Poskitt (2005) note that Aotearoa/New Zealand is a bi-cultural nation where the Maori community has become increasingly influential. The Ministry of Education found that "there are significant disparities in achievement evident throughout New Zealand's schools" (2002, as cited in Looney & Poskitt, 2005, p. 178). In order to redress the imbalance the Ministry developed and implemented formative assessments, which in New Zealand are embedded in multiple nation policies. One such initiative is the Maori Mainstream Programme (MMP). The MMP arose from the unequal power relations within society and "within this framework, the importance of culture and relationships is paramount" (Looney & Poskitt, 2005, p. 179). To this end, the MMP supports teachers as they reflect on their own cultural preconceptions and establish a learning environment where children can safely express their personalities (tacit knowledge). Bishop and Glynn (1999) advise that a simple individual – group dichotomy is not enough. Indeed, such dichotomous conceptualizations are subject to post-structuralist critique (Dewey, 1938; Vygotsky, 1987; Rogoff, 2003).

Consequently, teachers need to create a social context in which students can bring 'who they are' to the learning process and where a range of discourses and learning strategies occur. MMP has been referred to as being founded on cooperative learning relationships between teachers and students and the use of proverbs and *karakia* (prayer). The sharing and recirculation of discursive 'power' is fundamental to this process and it is a process that requires a significant change in approach that practitioners often find very challenging (Bishop & Glynn, 1999; Black & Wiliam, 2006a), typically requiring a year or more of support before they become professionally adapted. Closely related to changes in power

relationships is the sense of emotional and psychological safety. Maori (and other) students feel safe in an open classroom where they can bring their tacit knowledge (what they know and who they are) into the learning relationship (Polanyi, 1967; Bishop & Glynn, 1999).

MMP and Formative Assessment Techniques

MMP teachers use a variety of FA techniques: a) Feedback: How are we doing? A question, which monitors and assesses learning progression, either for a specific task or more generally; b) Feed-forward: Where to next? This question relates to the next steps required for improvement on a specific task/project or more generally across time (Looney & Poskitt, 2005; Hattie & Temperley, 2007); c) Scaffolding: a collaborative process which differs from other forms of classroom talk because it should occur only when there is clear evidence that the learner is unable to progress without assistance. If the intervention is a timely one and one that provides the correct amount of assistance, then the learner may cross the conceptual spectrum of achievement known as the ‘zone of proximal development’ (ZPD) (Vygotsky, 1978; Shepard, 2005). These techniques engage students in reflective thinking and problem solving. Some teachers referred to their attempts at resolving the endemic ‘unfairness’ associated with differing styles of learning by providing several tasks from which students may choose. The nature of discourse is also different. Teachers focus on open-ended discussion style questions, provide positive feedback and scaffold questions (e.g. “can you think about what might happen if you do [...insert...]”). The dialogue is very animated, to the extent that non-MMP teachers often remark on the level of noise emanating from MMP classrooms (Looney & Poskitt, 2005).

The MMP innovation was provided with strong leadership by the principal and deputy principals. Participants in the MMP are given very substantial professional development by the Ministry of Education: joint training and observation; monthly meetings; four-day intensive cultural immersion programs; innovation grants; meetings with Maori parents; and conferences on a range of topics for MMP teachers, principals and deputy principles (Looney & Poskitt, 2005).

Concluding Summary

The final irony is that it is precisely the demand for accountability, which has produced unprecedented pressure to improve education systems that is likely to be the biggest impediment to achieving that improvement (Black & Wiliam, 2005, p. 260).

Impact on Students

The irony referred to above, is best expressed by the boldly named intention of the ESEA itself: ‘No Child Left Behind’. The performance-oriented provisions of NCLB while intended to support low achievers may all too easily exaggerate academic disaffection and the achievement gap revealed in “The McKinsey Report” (2009). The connection between a high-stakes culturally unresponsive learning environment and academic disaffection has been clearly established by numerous studies (e.g. Bishop & Glynn, 1999; EPPI-Centre, 2002; Harlen & Deakin Crick, 2003). The consequence of which are significant disparities in achievement (McKinsey, 2009). Harlen (2006) underscores the “serious impact of summative

assessment and tests on lower achieving students” (p. 78) caused by repeated failure and public comparison. In a US national survey, 73% of teachers noted the intense pressure exerted upon their students and 76% indicated that high-stakes testing created extreme anxiety (Pedulla et al, 2003). Turner, Thorpe and Meyer (1998) remark that students of low self-efficacy and high performance goal orientation typically exhibit high levels of test anxiety and that such students do not rise to meet the demands placed upon them by high-stakes tests. On the contrary, they divide their attention between worrying about their performance and thoughts about their negative characteristics, “as a result, anxiety influences memory and test performance in these students” (Turner et al., 1998, p. 759). In the UK, Condie, Livingstone, and Seagraves (2005) investigated the outcomes of FA and also used questionnaires as a part of their method. They issued the same basic template in two phases spaced 16 months apart. Forty-four teachers and 21 head-teachers participated in phase one and 56 teachers and 26 head-teachers participated in phase two of the survey. In phase one, 75% of participating practitioners reported that formative assessments enhanced student learning; 71% that students’ confidence and self-esteem had increased; 70% reported an increase in student motivation to learn and 81% indicated that students had become more actively involved in the learning process. In phase two, the impacts of FA on engagement and participation were perceived to have plateaued at similar levels.

It is well documented in various research that student experience high levels of test anxiety when facing summative tests (Turner et al., 1998; EPPI-Centre, 2002) and much prefer other forms of assessment (EPPI-Center, 2002; AAG/APMG, 2002-2008). The ARG (1999) encourage the use of FA techniques in the classroom because they were found to mitigate the problems identified with summative testing and create large positive increases in the level of student engagement. High-stakes testing creates disaffection among students who do not achieve their full potential. As with all systemic phenomena the outcomes are cyclical and reinforcing. Those students who exhibit disaffection often experience an ever-deepening spiral of self-defeating learning behavior and worsening motivational problems (Covington, 1992), thus widening the achievement gap.

Impact on School Staff

As Abrams (2007) observes, NCLB is intended to motivate school administrators and teachers to achieve optimal performance levels. In challenging schools to improve standards on a continuous basis, legislators expect the use of innovative strategies as they seek to meet performance targets. Yet, in his response to the reauthorization of the ESEA, Gong (2007) expresses his reservations about the rationale behind NCLB:

The reauthorization and any school improvement plan must have a better theory of action than saying “Clear goals and strong sanctions will motivate schools and districts to solve this problem.” I simply do not believe that is true; it is not a helpful characterization of the problem or the solution to improving American education. (p. 2)

When the words of Dr. Sommers, who as President of the NAESP spoke on behalf of its 30,000 members, are considered, it is no surprise that studies have demonstrated that a high-stakes testing environment can create dissatisfaction among teachers. When conducting a

study in North Carolina, Jones et al. (1999) found that 77% of respondents reported a decrease in morale and 76% an increase in stress. In Kentucky, 75% of the participants in the survey reported a decrease in morale (Koretz et al., 1996a, 1996b). In contrast, practitioners involved in the implementation of FA in the UK valued the experience: 100% of respondents (n = 72) assigned a rating of 'very successful' or 'successful' to the project improving their motivation (Hallam et al., 2004, ¶ 1.5.18). Condie et al. reported numerous positive reactions to FA by teachers, for example: "re-energised, satisfied, confident, renewed enjoyment" (2005, ¶ 4.4). When teachers are engaged in culturally responsive teaching innovation they consistently report the sense of reward and a commitment to the idea that teachers can make a difference in learning outcomes (Looney & Poskitt, 2005). Teachers on the MMP program, for example noted that they had abdicated too much responsibility based on the belief that socio-economic factors determined student success, a view also disputed by studies in achievement and student autonomy (e.g. Zimmerman & Pons, 1986).

The discourse gradient among practitioners appears to favor a reduced emphasis on teaching strategies, which prepare students for frequent high-stakes summative tests. Traditional teacher-fronted methods were found to be ineffective when compared with collaborative learning systems which emphasize culturally responsive, high quality interactions between teachers, students and parents/caregivers (ARG, 1999; OECD, 2005; AAG/APMG 2002-2008). Perhaps more significantly, it remains difficult to see how students benefit from a high-stakes testing environment; how current policy develops the characteristics and strategies for lifelong learning; closes the international, racial, income, and systems-based achievement gap, or, how it will achieve the stated objective of proficiency in reading and mathematics for students in grades 3 - 8 by the year 2014.

References

- AAG/APMG (Assessment Action Group / AiFL Programme Management Group). (2002 - 2008). *AifL--Assessment is for learning*. Retrieved from <http://www.ltscotland.org.uk/assess>
- Abrams, L.M. (2007). Implications of high-stakes testing for the use of formative classroom assessment. In H. McMillan (Ed.), *Formative assessment classroom: Theory into practice* (pp. 70-98). NY: Teachers College Press.
- Allal, L., & Lopez, L. M. (2005). Formative assessment of learning: A review of publications in French. In Office of Economic Co-operation and Development, *Formative assessment: Improving learning in secondary classrooms*. Paris: OECD Publishing.
- Assessment Reform Group (ARG). (1999). *Assessment for learning: Beyond the black box*. Cambridge University: Cambridge School of Education.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. NY: Freeman.
- Bennett, R. (2011). Formative assessment: a critical review. *Assessment in Education*, 18(1), 5-25.
- Bishop, R., & Glynn, T. (1999) *Culture counts: Changing power relations in education*. Palmerston North, New Zealand: Dunmore Press Ltd.
- Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-73.
- Black, P. & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London: School of Education King's College.
- Black, P. & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal* 16(2), 249-261.
- Black, P. & Wiliam, D. (2006a). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning* (pp.9-25). London: Sage.
- Black, P. & Wiliam, D. (2006b). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp.81-100). London: Sage.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31.
- Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). *Handbook on formative and summative evaluation of pupil learning*. New York: McGraw-Hill.
- Brophy, J. (2004). *Motivating students to learn* (2nd Edition). Mahwah, NJ: Lawrence Erlbaum Associates.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245 -281
- Cauley K.M. & McMillan, J.H. (2010). Formative assessment techniques to support student motivation and achievement. [*Clearing House: A Journal of Educational Strategies, Issues and Ideas*](#), 83(1), 1-6.
- Clark, I. (2010a). The development of 'Project 1': Formative assessment strategies in UK schools, *Current Issues in Education*, 13(3), 1-34. Retrieved from <http://cie.asu.edu/ojs/index.php/cieatasu/article/viewFile/382/27>
- Clark, I (2010b). Formative assessment: There is nothing so practical as a good theory. *Australian Journal of Education*, 54(3), 341-352.
- Condie, R., Livingstone, K., & Seagraves, L. (2005). *Evaluation of the assessment is for learning*

- programme: Final report and appendices*. The Quality of Education Centre, University of Strathclyde. Retrieved from www.scotland.gov.uk/Publications/2005/12/0210792641/26428
- Cornford, I.R. (2002). Learning-to-learn strategies as a basis for effective lifelong learning. *International Journal of Lifelong Education*, 21(4), 357-368.
- Covington, M. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. New York: CUP.
- Cronbach, L. (1963). Course improvement through evaluation, *Teacher College Record*, 64, 672-683.
- Damon, W. & Phelps E. (1989). Critical distinctions among three approaches to peer education. *International Journal of Educational Research* 58, 9-19.
- Dewey, J. (1985/1916). Democracy and education. In J. A. Boydston (Ed.), *John Dewey; The Middle Works, 1899–1924, Vol. 9*. Carbondale, IL: SIU Press. [Originally published as Dewey, J. (1916). Democracy and education. New York: Macmillan].
- Dewey, J. (1988/1938). Experience and education. In J. A. Boydston (Ed.), *John Dewey: The Later Works, 1925–1953, Vol. 13*. Carbondale, IL: SIU Press. [Originally published as Dewey, J. (1938). Experience and education. New York: Macmillan].
- Dorn, S. (2010). The political dilemmas of formative assessment. *Exceptional Children* 76(3), 325-337.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Assessment, Research & Evaluation*, 14 (7). Retrieved from <http://pareonline.net/getvn.asp?v=14&n=7>
- EPPI-Centre. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning*. Social Sciences Research Unit. Institute of Education (IoE), University of London. Retrieved from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=108>
- Gong, B. (2007). Testimony of Brian Gong, Executive Director, The National Center for the Improvement of Educational Assessment before the Committee on Education and Labor hearing on the “Miller-McKeon Discussion Draft of the Reauthorization of the Elementary and Secondary Education Act” Washington, D.C. September 10, 2007. Retrieved from <http://edlabor.house.gov/testimony/091007BrianGongTestimony.pdf>
- Hallam, S., Kiston, A., Peffers, J., Robertson, P., & Stobart, G. (2004). *Evaluation of Project 1 of the AiFL Development Programme: Support for professional practice in formative assessment*. Institute of Education (IoE), University of London. Retrieved from www.scotland.gov.uk/Publications/2004/10/19947/42988
- Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education*. 10(2), 169-207.
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp.103 - 117). London: Sage.
- Hattie, J. (1999). *Influences on student learning*. University of Auckland, New Zealand: Inaugural professorial lecture. Retrieved from www.citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.114.8465.pdf
- Hattie, J., & Temperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.

- Heritage, M. (2010). *Formative assessment and next-generation assessment systems: Are we losing the opportunity?* Council of Chief State School Officers (CCSSO), Washington D.C. Retrieved from http://www.edweek.org/media/formative_assessment_next_generation_heritage.pdf
- Hutchinson, C., & Hayward, L. (2005). The journey so far: Assessment for learning in Scotland. *The Curriculum Journal*, 16(2), 225-248.
- Innovation Economy, The. (2010, December 7). Education for innovation. [Video file]. Retrieved from <http://www.aspeninstitute.org/video/education-innovation-digital-townhall-itifs-robert-atkinson-pbs-newhours-hari-sreenivasan>
- Johnson, D.W., & Johnson, R. (1975). *Learning together and alone: Cooperation, competition and individualization*. Englewood Cliffs, NJ: Prentice Hall
- Jones, M., Jones, B., Hardin, B., Chapman, L., Yarbough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81(3), 199-203.
- Kendall, J. S., Ryan, S., Weeks, S., Alpert, A., Schwols, A., & Moore, L. (2008). *21st Century skills: What do we expect of students?* Denver, CO: Mid-continent Research for Education and Learning.
- Koretz, D., Barron, S., Mitchell, K., & Keith, S. (1996a). *Perceived effect of the Kentucky Instructional Results Information System (KIRIS) (MR-792-PCT/FF)*. Santa Monica, CA: RAND.
- Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996b). *Perceived effect of the Maryland school performance assessment program (Technical Report 409)*. Los Angeles, CA: CRESST/RAND Institute on Education and Training.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.
- Looney, J., & Poskitt, J. (2005). New Zealand: Embedding formative assessment in multiple policy initiatives. In *Formative assessment: Improving learning in secondary classrooms* (pp. 177 – 184). Centre for Educational Innovation and Research. Paris: OECD.
- McCombs, B. L., & Whistler, J. S. (1989). The role of affective variables in autonomous learning. *Educational Psychologist*, 24(3), 277-306.
- McKinsey & Company. (2009). *The economic impact of the achievement gap on America's schools: Summary of findings*. NY: Social Sector Office. Retrieved from http://www.mckinsey.com/app_media/images/page_images/offices/socialsector/pdf/achievement_gap_report.pdf
- National Association of State Boards of Education (NASBE). (2009). *A call for balanced systems of assessment and accountability. NASBE Study Group on Assessment Systems for the 21st Century Learner*. Retrieved from http://www.k12.wa.us/assessment/ClassroomAssessmentIntegration/pubdocs/NA_SBEAssessmentReport_ExecSumry.pdf
- Noddings, N. (2004). High stakes testing: Why? *Theory and Research in Education*, 2(3), 263-269.
- Organisation for Economic Co-operation and Development, (OECD). (2005). *Formative Assessment: Improving learning in secondary classrooms*. Centre for Educational Innovation and Research. Paris: OECD.
- Pedulla, J., Abrams, L., Maduas, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national*

- survey of teachers*. Chestnut Hill, MA: Boston College, National Board on Educational Testing and Public Policy.
- Piaget, J. (1965). *The moral judgment of the child*. NY: Free Press.
- Polanyi, M. (1967). *The tacit dimension*. NY: Anchor Books.
- Popham, W. J. (2008). *Transformative assessment*. VA: ASCD.
- Rogoff, B., Goodman Turkkanis, C., & Bartlett, L. (2001). *Learning together: Children and adults in a school community..* New York: Oxford University Press.
- Rogoff, B. (2003). *The cultural nature of human development*. New York: Oxford University Press.
- Rogoff, B., & Gutierrez, K. D. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19-25.
- Sadler, D.R. (1989). Formative assessment and design of instructional systems. *Instructional Science*, 18, 119-144.
- Salt Lake City Open Classroom Charter School (OCCS). (n.d.). Retrieved from <http://ocslc.org/>
- Schön, D. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass
- Scriven, M. (1967). The methodology of evaluation. In R.W. Tyler, R.M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Shepard, L.A. (2005). Formative assessment: Caveat emptor. ETS Invitational Conference 2005. *The Future of Assessment: Shaping Teaching and Learning*, New York. Retrieved from http://www.cpre.org/ccii/images/stories/ccii_pdfs/shepard%20formative%20assessment%20caveat%20emptor.pdf
- Sommers, M.K. (2007). Statement of Mary Kay Sommers, Ph.D., President, National Association of Elementary School Principals (NAESP), "Reaction to the Miller-McKeon Discussion Draft of the Reauthorization of the Elementary and Secondary Education Act (ESEA)" Washington, D.C. September 10, 2007. Retrieved from <http://edlabor.house.gov/testimony/091007MaryKaySommersTestimony.pdf>
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Stiggins, R. (2007). Conquering the formative assessment frontier. In H. McMillan (Ed.), *Formative assessment classroom: Theory into practice* (pp. 8 – 28). NY: Teachers College Press.
- Storch, N. (2002). Patterns of interaction in ESL pairwork. *Language Learning*, 51(1), 119-58.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching and learning in the classroom*. Buckingham: OU Press.
- Turner, J.C., Thorpe, P.K., & Meyer, D.K. (1998). Students' reports of motivation and negative affect: A theoretical and empirical analysis. *Journal of Educational Psychology*. 90(4), 758-771.
- Vygotsky, L. (1978). Chapter 6: Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in Society* (pp. 79-91). Cambridge, MA: Harvard.
- Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber and A. S. Carton (Eds.). *The*

- collected works of L.S. Vygotsky, Volume 1: Problems of general psychology* (pp.39-285). New York and London: Plenum.
- White, B., & Fredericksen, J. (2005). A theoretical framework and approach for fostering metacognition development. *Educational Psychologist*, 40(4), 211-223.
- Wiliam, D. (2003). The impact of educational research on mathematics education. In A. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.). *Second international handbook of mathematics education*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Wiliam, D. (2004) Assessment and the regulation of learning. Paper presented at Invited Symposium 'What does it mean for classroom assessment to be valid? Reliable?' at the annual meeting of the National Council on Measurement in Education, April 2004, San Diego, CA. Retrieved from http://www.michigan.gov/documents/mde/WiliamAssessmentarticle_201263_7.pdf
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student learning. *Assessment in Education*, 11(1), 49-65.
- Wiliam, D. (2006). Assessment for learning: Why no profile in US policy? In J. Gardner (Ed.), *Assessment and learning* (pp.169-183). London: Sage.
- Wiliam, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In H. McMillan (Ed.), *Formative assessment classroom: Theory into practice* (pp. 29 – 42). NY: Teachers College Press.
- Young, V., & Kim, D. (2010). Using assessments for instructional improvement: A literature review. *Education Policy Analysis Archives*, 18(19), 1 – 40.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64-70.
- Zimmerman, B.J., & Pons, M.M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, 23(4), 614-628.

Ian Clark is at the University of Washington.