Stocks and Prospects: Research on Formative Assessment

In Secondary Classrooms

Robin Tierney & Julie Charland

Faculty of Education,
University of Ottawa,
Ottawa, Ontario, Canada

Please cite this work as: Tierney, R. D. & Charland, J. (2007, April). *Stocks and prospects: Research on formative assessment in secondary classrooms.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

This work is part of an ongoing review. If you are aware of any research that meets the inclusion criteria, or that would extend the review beyond the current date range, please do not hesitate to contact us at rtier017@uottawa.ca or jchar119@uottawa.ca

Stocks and Prospects: Research on Formative Assessment in Secondary Classrooms

Classroom assessment that supports student learning, or formative assessment, is strongly favored in current educational literature. Formative assessment has been championed by assessment specialists (e.g., McTighe & O'Connor, 2005; Stiggins & Chappuis, 2005), and it is increasingly endorsed by professional organizations (e.g., Joint Committee on Standards for Educational Evaluation, 2003; Miller, 2005). Formative assessment is thought to have "intrinsic acceptability" (Black and Wiliam, 2003, p.634) to teachers, but system wide implementation has met with some resistance at the secondary level (Hayward & Hedge, 2005; Smith & Gorard, 2005). The Centre for Educational Research and Innovation (CERI, 2005) notes that "powerful bureaucratic constraints" (p.28) limit the implementation of formative assessment in secondary schools, even though its use with adolescent students is particularly justified. In their international case studies, CERI (2005) concludes that the benefits of formative assessment outweigh the barriers to its implementation. Although they are negative, these barriers are also a natural part of the transformation currently underway in education. Black & Harrison (2001a) point out that "the development of formative assessment has led to more radical changes in the ways of working of many of the teachers involved" and that "it takes time and patience to achieve changes of this type" (p.7). As part of that process, empirical research on formative assessment captures not only ways of working and learning in classrooms, but also the ongoing dialogue between researchers about the changes.

The American abolitionist, Wendell Phillips, said that "a revolution is as natural a growth as an oak; It comes out of the past. Its foundations are laid far back" (1852, in Bartlett, 1968, p.658-659). Current assessment reforms can be traced back to past inquiry, and by taking stock

periodically, we can build up the foundation for further change. A significant increase in the quantity and accessibility of educational research, resulting from rapid technological advances in recent history, provides additional imperative for the thoughtful review and analysis of existing work. This paper identifies and discusses empirical research on formative assessment for two main purposes. The first is to understand, more comprehensively than is permitted with a single study, what has been learned recently about formative assessment in the secondary classroom context. The second purpose is to consider direction for further research in this area. With these purposes in mind, a methodical analysis of recent work was undertaken. Three main questions guided the research process:

1. What is the nature of empirical research done recently on formative assessment in the secondary school context?
2. What can be understood about formative assessment from this body of work?
3. What priorities are indicated in and by this body of work for further inquiry?

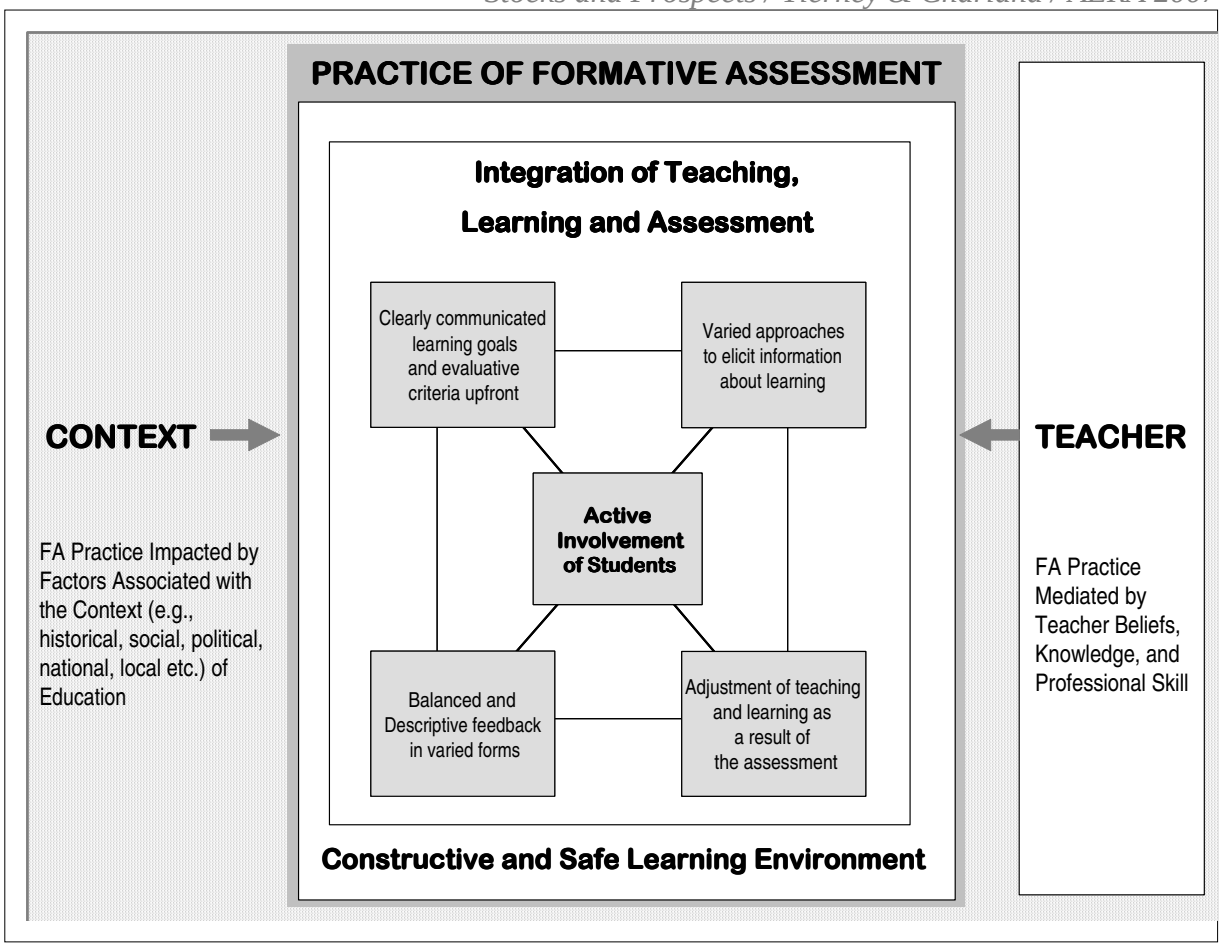
In the balance of this paper, the conceptual framework for the review is explained, the methodology is described, and results are discussed in relation to the research questions.

Conceptual Framework

The concept of formative assessment evolved from the early definitions by Bloom, Hastings and Madaus (1971), Ramaprasad (1983), and Sadler (1989). The term formative assessment has been used in conjunction with the popular term, assessment-for-learning (Assessment Reform Group [ARG], 1999; Black, 2003; Earl, 2003), and the two are now considered "conceptually identical" (Threlfall, 2005). However, considerable confusion remains regarding the nature of the concept (Black, 2003). As Yorke (2003) observes, "formative

assessment is a concept that is more complex that it might appear at first sight" (p.478). It is sometimes described as a linear sequence, involving teacher-directed instruction, feedback and "correctives" (Guskey, 2005). This definition fits well with behaviourist or early cognitive theories of learning (Allal & Ducrey, 2000; Yorke, 2003), but its limitations are seen in classroom-based research. In their analysis of formative assessment events, Pryor and Torrance (1998) found that a "purely cognitive" approach minimized the "complexity of the situation" (p.170) by overlooking the social aspects of classroom learning. Extending beyond prescribed instruction, formative assessment is "fundamentally a collaborative act" (Yorke, 2003, p.496) that necessitates interaction between teachers and students. By nature, it is inherently social, and thus more consistent with social constructivist theory (Allal & Ducrey, 2000; Gipps, 2002; Shepard, 2000, 2005; Torrance & Pryor, 2001). The use of assessment to support learning is part of the shift from a testing culture toward constructivist ideals in education (Gipps, 1994; Shepard, 2000). This notion is supported by the numerous studies that refer explicitly to constructivism in relation to changing assessment practices (e.g., Borko, Davinroy, Bliem & Cumbo, 2000; Briscoe & Wells, 2002; Hand & Prain, 2002; Mabry, Poole, Redmond & Schultz, 2003; Torrance & Pryor, 2001). The strength of this association is made explicit by Roos and Hamilton (2005) who dub formative assessment as the "constructivist assessment" (p.10). In short, formative assessment can be considered a practice that is socially situated as a form of classroom interaction, and historically situated as part of an ongoing theoretical shift in the field of education.

Several recent definitions detail the characteristics and elements of formative assessment (ARG, 2002; Cowie & Bell, 1999, CERI, 2005; Crooks, 2001; Leahy, Lyon, Thompson &



*1*Figure 1: *Illustration of formative assessment practice.*

Wiliam, 2005; Shepard, 2005; Stiggins, 2002; Torrance & Pryor, 2001). In synthesizing these, formative assessment is described as a composite practice, involving: a) clearly communicated learning goals and evaluative criteria, b) varied approaches to elicit information about learning, including questioning and observation, c) balanced and descriptive feedback in varied forms, d) the adjustment of teaching and learning as a result of the assessment, and e) the active involvement of students. This last element is central to the concept, and is accordingly placed in the visual model (see Figure 1). Two key ideas are also reflected in the model. First, assessment, teaching, and learning are ideally integrated within a safe learning environment. Second, a

multitude of internal and external factors, including teachers' knowledge and beliefs, impact the elements of formative assessment in practice. This model is presented, not as a static or permanent definition of formative assessment, but as a pragmatic framework for further analysis and discussion around the concept, as it is presently understood.

Methodology

Overview

We drew methodological guidance for this analytical study from several sources on qualitative content analysis (Hsieh & Shannon, 2005; Mayring, 2000), systematic reviews (Davies, 2000; Evans & Benefield, 2001), and qualitative meta-synthesis (Sandelowski, Docherty & Emden, 1997). This study differs from a traditional literature review in that the process used to select and analyse texts is more focussed, systematic and transparent. We collected data in two phases. The purpose of the first was to identify empirical research in an organized and comprehensive manner, through abstracts that we located using several different databases and journal collections. In the second phase, we ensured the relevancy of the selected studies following a reading of the full text. Data analysis involved both deductive and inductive approaches undertaken directly in relation to the research questions. We used a number of methods to ensure research quality, including detailed documentation of inclusion and exclusion criteria, multiple readings to permit immersion in the texts, and repeated categorizing. We discussed differences in the information we gathered from the texts, but we found that our inter-researcher interpretation and analysis of the texts was fairly consistent**.**

Data Sources, Collection and Analysis

To locate empirical research, seven databases and journal collections were initially

searched: CBCA Education, ERIC, PsycInfo, Sage Education, Scholars' Portal E-Journals, Scopus, and Wilson Education, along with one online journal, Practical Assessment Research and Evaluation (PARE). To mitigate against the inconsistencies in database indexing (e.g., key journals not indexed for full date range in certain databases), three journals were also searched individually (Assessment in Education, Educational Assessment, and Curriculum Journal). Although this process mainly produced duplicates, several articles were added as a result. In addition, to ensure that North American journals were represented in the selected texts, two Canadian journals (Canadian Journal of Education and Alberta journal of Educational Research) and several American journals (American Educational Research Journal, Educational Evaluation and Policy Analysis and Educational Researcher) were searched for any further articles relating to classroom assessment. No relevant additions were found in this last check. Several keyword combinations were used because the term formative assessment has yet to be added to the "existing controlled vocabularies" (Evans & Benefield, 2001). These were formative assessment, assessment for learning, assessment as learning, embedded assessment, interactive assessment, and constructive assessment. In addition, related terms, such as portfolio assessment, feedback, assessment criteria, self-assessment, and peer-assessment, were used to find relevant studies.

In the first data collection phase, a checklist was created to standardize the selection of texts (see Appendix A, 1 of 2), and it was completed for each study identified through the database searches. The checklist contains a list of inclusion criteria, as well as details for exclusion. Selection criteria included document type (peer-reviewed journal), publication date range (2000 to 2005), research type (empirical), relevancy (relating to at least one aspect of formative assessment), and context (secondary classroom assessment). The latter was somewhat

difficult in that there are variations in which grades or years are considered secondary in different educational contexts. Although our intention was to concentrate on the high school grades (9 to 12 in Ontario), some of the texts also include students from lower secondary (e.g., years 7 and 8 in New Zealand and England) or elementary grades. Studies that did not include *any* students in grades 9 to 12 (e.g., North American middle schools) were, however, excluded from the final set.

During the first data collection phase, we identified 362 texts using the search features in the different databases and journal collections. From this, 45 articles met the inclusion criteria based on the information provide in the abstract. The final set was reduced to 30 articles following full text readings in the second phase. Where the same study was reported in multiple articles, a single, most relevant article was retained for review, unless the texts contained different information. For example, Black and Harrison (2001a, 2001b) and Wiliam, Lee, Harrison and Black (2004) all refer to the same study, but the first two articles include teachers' comments on different topics, while the latter focuses more specifically on the effects of the intervention. All texts in our final selection were published in peer-reviewed journals between 2000 and 2005, and they all relate to at least one aspect of the synthesize model of formative assessment. In the second data collection phase, another form was used to gather and organize information from, and about the studies, such as their context details (e.g., grade levels, academic subjects, etc.) and their methodology (see Appendix A, 2 of 2).

Two approaches to analysis were taken. First, the central aspects of formative assessment illustrated in the synthesized model were set as a priori categories, for which each text was examined. Each researcher then worked on a specific aspect, which was subsequently checked by the other, and followed by discussion regarding similarities and differences in the results. A

series of working tables was created to aid the analysis and focus the discussion. Second, each text was considered more conventionally, using an inductive process, with patterns across the texts emerging through the readings. Again, the interpretations made by the researchers, as readers of research texts, were discussed, and the results thus became a joint construction.

Results and Discussion

Three research questions were used to guide this review: What is the nature of empirical research done recently on formative assessment in the secondary school context? What can be understood about formative assessment from this body of work? What priorities are indicated in and by this body of work for further inquiry? The results of our analysis are reported and discussed directly relation to these three questions.

The Nature of the Research

The recent, empirical work reviewed in this paper varies considerably in scope, and the articles can be clustered into five different groups according to their focus. Articles in the first group are concerned explicitly with formative assessment as a practice (Bell & Cowie, 2001; Black & Harrison, 2001a; Davies, Durbin, Clarke & Dale, 2004; Hodgen & Marshall, 2005; William et al., 2004), whereas articles in the second group focus on one aspect of formative assessment (Black & Harrison, 2001b; McDonald, 2002; McDonald & Boud, 2003; Noonan & Duncan, 2005). This second group is concentrated, in this review, in the area of student-involvement in assessment, and the studies relate specifically to peer and self-assessment. Articles in the third group investigate an associated assessment tool or strategy, including portfolios (Barootchi & Keshavarz, 2002; Clark et al., 2001; Nunes, 2004; Simon & Forgette-Giroux, 2000; Torres Pereira de Eca, 2005), writing-to-learn (Hand & Prain, 2002), a time-scale

assignment (Hermann & Lewis, 2004), mysteries (Leat & Nichols, 2000), invention activities (Schwartz & Martin, 2004), and several computer-based tools (Thissen-Roe et al., 2004; Vendlinski & Stevens, 2002). Articles in the fourth group describe classroom assessment practices in which the formative function plays a significant role (Brookhart, 2001; Cowie, 2005; Doppelt, 2003; Hickey & Zuicker, 2005; Kirkwood, 2000; Stokking, van der Schaaf, Jaspers, Erkens, 2004). Articles in the fifth group are concerned with changing assessment cultures such that formative practices increase (Dori, 2003; Verhoeven & Devos, 2005; Hayward, Priestly & Young, 2004; Yung, 2001), and they emanate from systems (Israel, Belgium, Scotland and Hong Kong, respectively) where large-scale reform in assessment is either underway or being studied. Taken as a whole, this body of work is clearly international in nature. Countries represented are the United States (7 of 30 articles), England (6 of 30) and Scotland (2 of 30), Barbados (2 of 30), Canada (2 of 30), Israel (2 of 30), New Zealand (2 of 30), Portugal (2 of 30), Australia (1 of 30), Belgium (1 of 30), Hong Kong (1 of 30), Iran (1 of 30), and the Netherlands (1 of 30).

A considerable variety of school subjects are involved in these studies, and many focus on more than one subject. Only two articles do not mention specific school subjects, and their studies appear to be cross-curricular. Of the 28 articles that do specify school subjects, more than half involve the Sciences (17 of 28). Language Arts (10 of 28), Mathematics (7 of 28), and Social Studies, such as Geography, Economics and History (7 of 28), are also fairly well represented. A smaller number are associated with Technology (4 of 28), Art (3 of 28), Business (1 of 28) and Physical Education (1 of 28). Studies are not concentrated in any single grade, and they span across the secondary years (see Appendix B). In many of the articles reviewed, teachers are given voice (Bell & Cowie, 2001; Black & Harrison, 2001a, 2001b; Hand & Prain, 2002;

Noonan & Duncan, 2005; Torres Pereira de Eca, 2005; Yung, 2001), or their perspectives play a central role in the study (Dori, 2003; Hayward et al., 2004; Verhoeven & Devos, 2005; Wiliam et al., 2004). Although the majority of these studies include students as participants, and many draw on students' responses, fewer actually highlight students' voices (Brookhart, 2001; Clark et al., 2001; Cowie, 2005; Torres Pereira de Eca, 2005).

The articles reviewed describe a wide variety of research methodologies. Categorizing broadly, 16 of 30 (53.3%) involve qualitative methods, 7 of 30 (23.3%) are quantitative, and 7 of 30 (23.3%) employ mixed methods. Of the different methods used, many are traditional, such as the use of survey questionnaires (e.g. McDonald, 2002; Noonan & Duncan, 2005), classroom observations and interviews (e.g. Brookhart, 2001; Clark et al., 2001), video-taped lessons (e.g., Hodgen & Marshall, 2005; Leat & Nichols, 2000), with formal examinations or external tests often acting as a measure of the impact of an experiment (e.g., Torres, 2005; William et al., 2004; Doppelt, 2003, Davies et al., 2004; Verhoeven & Devos, 2005). Some innovative or unusual methods were also noted, such as the use of an expert panel (Stokking & al, 2004), college level students for control purposes (Schwartz & Martin, 2004), computer simulation (Vendlinski & Stevens, 2002), and a comparative design with three teams taking different epistemological approaches (Hickey & Zuicker, 2005). Professional development and collaboration between teachers and researchers also feature in many of these studies (e.g. Hand & Prain, 2002; McDonald & Boud, 2003; Wiliam, et al., 2004). The multiplicity of approaches seen in these studies reflects a current ideal in educational research (Larabee, 2003), and may be especially beneficial for understanding the diversity of issues associated with formative practices in secondary classrooms. However, qualitative methods are most frequently used, and this

dominance is consistent with a broader trend in educational research (Gorard, 2002). The studies here are also consistent with those in a previous review of research on changing classroom assessment practices in that they are concerned with "processes that involve a high degree of human activity, and in many cases, their dynamic nature is acknowledged through the choice of methodology" (Tierney, 2006, p.256). Overall, this body of work can best be described as international and diversely focused, but a stronger interest in the subject of Science, and a reliance on qualitative methods are seen.

Formative Assessment in the Secondary School Context

To answer the second question, we examined the texts in relation to the five central elements of formative assessment in the synthesized model: student involvement, feedback to students, adjustment of teaching, explicit learning goals or assessment criteria, and varied approaches to elicit learning. Some elements of formative assessment do appear in the results and discussion sections of the articles more than others. Student involvement (80%), feedback to students (70%), and explicit goals or criteria (53%) are more frequently mentioned than the use of assessment to inform instruction (47%), or varied assessment approaches (40%).

At least three elements are mentioned in 20 of the articles (67%), supporting the notion that formative assessment is a complex activity (Cowie & Bell, 1999). What can be understood about formative assessment from our analysis of these works is discussed below in relation to each of the five central elements.

Feedback to Students. There is considerable focus in this set of articles on giving feedback to secondary students. Different types of feedback are discussed, such as comment-only marking by teachers (Black & Harrison, 2001; Wiliam et al., 2004), oral feedback offered

informally and responsively during classroom activities (Bell & Cowie, 2001), or computer-generated feedback that is tailored to specific errors (Thissen-Roe et al., 2004). Rubrics are used as a feedback tool, to direct student attention to specific dimensions of an assignment (Hermann & Lewis, 2004), or to guide "feedback conversations" that involve peers in discussion about learning (Hickey & Zuicker, 2005, p.297). While these studies do not give indication of the relative merits of these different methods of feedback, positive consequences are generally seen. Feedback is described as an effective means of scaffolding learning (Leat & Nichols, 2000; Hodgen & Marshall, 2006), and encouraging greater student autonomy (Kirkwood, 2000). However, feedback is rarely considered in isolation from other elements of formative assessment, and only one study attributes increased student achievement specifically to feedback. Hickey & Zuicker (2005) encouraged greater use of feedback by students through the design of their study in its second year, finding that "the improved learning outcomes over time appear to be mostly due to continued enhancement of participation in the feedback conversations" (p.298). To improve the length and quality of the feedback conversations, the teacher modeled the use of the feedback rubric for students, and the successful outcome of the study can, therefore, be associated with the teachers' pedagogical skill. Several studies conclude that teachers need guidance in this area. For example, Yung (2001) argues that teachers in Hong Kong should be provided with professional development on the use of feedback to motivate students and support learning. In their survey of assessment practices in the Netherlands, Stokking and colleagues (2004) note a wide range in the type, form, and quality of feedback, and they observe that some of the reported practices are less than ideal, especially for learning purposes.

Assessment Informing Teaching. The teachers in many of these studies benefit from

sustained support in learning how to use assessment to inform teaching (e.g. Doppelt, 2003; Wiliam et al., 2004). Even experienced teachers can be surprised by student's misunderstandings (Bell & Cowie, 2001; Thissen-Roe et al., 2004), and in learning formative strategies, teachers are better able to use assessment information (Bell & Cowie, 2001; Hand & Prain, 2002). Teachers in these studies draw on a variety of assessment sources to inform their teaching, from students' responses to oral questions (Bell & Cowie, 2001; Black & Harrison 2001a), student interaction with computer simulations (Vendlinski & Stevens, 2002), discussion in group problem solving (Leat & Nichols, 2002), individual products (Hermann & Lewis, 2004), and portfolios (e.g. Barootchi & Keshavarz, 2002). The teachers in Bell and Cowie's (2001) study describe different types of action under the umbrella of formative assessment, from proactive and planned to reactive and spontaneous, and they list a host of ways in which assessment can support a range of teaching activities, from planning to reporting.

The teachers' experiences in Bell and Cowie's (2001) study are reflected in many of the classrooms that are described in these articles. Pedagogical change is emphasized, highlighting the possibility of responding to the needs of an individual learner (Nunes, 2004) or a group (Thissen-Roe, et al. 2004), adjust unit plans (Hand & Prain, 2002) or shift curricular goals (Barootchi & Keshavarz, 2002; Dori, 2003). Again, the consequences are generally portrayed as positive. Assessment information provided by students can be "invaluable" (Nunes, 2004, p.333) for teachers, and it can be used intentionally to improve the relevance and effectiveness of instruction (e.g., Vendlinski & Stevens, 2002). Improvements in student learning are linked to greater use of assessment information by teachers (Barootchi & Keshavarz, 2002; Dori, 2003), and improvements in student engagement are also suggested as teachers are able to "design future

instructional strategies, materials and activities that are more meaningful and valuable to the learners" (Nunes, 2004, p.333).

Student Involvement in Assessment. The most frequently mentioned aspect of formative assessment in these texts is student involvement. The articles in this large group (24 of 30) involve students through portfolio assessment (Barootchi & Keshavarz, 2002; Clark et al., 2001; Nunes, 2004; Simon & Forgette-Giroux, 2000; Torres Pereira de Eca, 2005), or they focus more specifically on peer and student self-assessment (Black & Harrison, 2001b; Davies et al., 2004; McDonald, 2002; McDonald & Boud, 2003; Noonan& Duncan, 2005). In some, peers support or mediate the learning and assessment process (Bell & Cowie, 2001; Cowie, 2005; Hickey & Zuicker, 2005; Kirkwood, 2000; Wiliam et al., 2004; Yung, 2001), and in others, student self-assessment plays a strong role in the learning and assessment activities (Brookhart, 2001; Hand & Prain, 2002; Hermann & Lewis, 2004). The importance of student involvement in assessment is also suggested in some of the texts that take a broader look at the context in which classroom assessment occurs (Dori, 2003; Hayward et al., 2004; Stokking et al., 2004).

Student is involved in their own assessment in two different ways in these studies. The first is a reflective process with a retrospective orientation. By looking back and reflecting on past efforts, change becomes visible for learners (Bell & Cowie, 2001; Clark et al., 2001; Hand & Prain, 2002), and they become aware of themselves as learners (Barootchi & Keshavarz, 2002; Brookhart, 2001; Nunes, 2004). The second way students are involved in their own assessment can be "triggered" (Clark et al., 2001, p.221) by the first, but it is more forward-looking, and action oriented. Learners "take charge of their own learning" (Barootchi & Keshavarz, 2002, p.286), and take responsibility for learning (Davies et al., 2004; Hayward et al., 2004; Kirkwood,

2000; McDonald and Boud, 2003), especially when they are able to make decisions about, or have choices in the process (Bell & Cowie, 2001; Clark et al., 2001; Dori, 2003; Nunes, 2004; Torres Pereira de Eca, 2005).

Students may also be involved in the assessment of others, and many of these studies they are seen to play supportive roles in classroom assessment processes, either as individuals or in group activities. In some of the classrooms, a supportive peer culture seems to emerge as a consequence of the opportunity for interaction between students (Cowie, 2005; Kirkwood, 2000). In other studies, student activities are organized to encouraged students to give each other feedback in small groups. Although peer assessment does not appear to be used to a great extent (Noonan & Duncan, 2005; Stokking et al., 2004), it seems to be valued as a means of assessment for learning. For example, the teachers who are given voice by Black and Harrison (2001b) described several strategies that involve students in small group discussion about their work, and collaborative groups are purposefully established for "feedback conversations" (p.295) in the study by Hickey and Zuicker (2005). The importance of a supportive peer group for some students is clearly illustrated by the experience of one student in the study by Clark and colleagues (2001). They note that "without this group, she floundered and lost her motivation to write and even her skillfulness as a writer" (p.230). Concern about the consequences of withdrawing peer support is also expressed by one of the teachers in a pilot project in Scotland. Although the results are ultimately positive in both these cases, the relationship between fostering autonomy and encouraging constructive collaboration is not well explored in this body of work. Although the benefits of allowing students to "confer in groups" (Hand & Prain, 2002; p.748; also Yung, 2001) is clearly appreciated by some teachers, and the merits of peers as an authentic

audience are noted (Dori, 2003; Hand & Prain, 2002), the risks involved are not often addressed. The role of student disclosure, and the need for trust and respect in classroom dynamics, are repeatedly stressed by Cowie (2005), but in most of these articles there is scant mention of the potentially negative consequences of peer assessment. Our own experience as teachers tells us that classroom assessment is not a private endeavor, and that a supportive learning environment must be actively nurtured. Several of these studies capitalize on the public nature of classroom learning in their assessment strategies (Hand & Prain, 2002; Hickey & Zuicker, 2005), but this is also an area that is not well illuminated.

There are a host of contextual factors that can impede or facilitate the involvement of students in assessment for learning, such as time limitations due to curricular requirements (Hayward et al., 2004), the familiarity of students with formative practices (Hermann & Lewis, 2004; Nunes, 2005; Torres Pereira de Eca, 2005), and teachers' ideas about its feasibility or value (Hayward et al., 2004; Noonan & Duncan, 2005; Stokking et al., 2004). While the importance of interaction between students, and between the teacher and students is highlighted through many of these articles (Barootchi & Keshavarz, 2002; Bell & Cowie, 2001; Cowie, 2005; Hermann & Lewis, 2004; Nunes, 2004; Wiliam et al., 2004), there is also strong evidence that training is needed for assessment to be constructive. It is suggested that students need to understand the rationale for self-reflection (Kirkwood, 2000), and that they need to be "prompted to elaborate" on their thoughts (Hermann & Lewis, 2004). In one study, the teachers use a variety of tools and strategies, from checklists to writing prompts, to support students' self-reflection (Simon & Forgette-Giroux, 2000), but it is not clear which of these is more effective, and in what circumstances. Making the formative process explicit for students does, however, seem to be a

key factor, even in the varied contexts of these studies. Torres Pereira de Eca (2005), for example, found that art students who had more experience with critical self-reflection were more successful with a new portfolio assessment system, and Hickey & Zuicker (2005) had more success when the use of feedback was modeled by the teacher in an introductory genetics class. A large-scale program to train students in self-assessment across the curriculum in Barbados had a strong positive impact, not only on academic achievement, but also on students' attitudes about self-assessment (McDonald & Boud, 2003). The students in this study reported feeling more "independent" and "empowered" (p.215), and they were more confident in preparing for examinations and setting goals for the future.

Learning Goals and Assessment Criteria. More than half of the articles reviewed mention learning goals or assessment criteria. Goal setting by students is discussed as a necessary, or prerequisite part of assessment for learning (Barootchi & Keshavarz, 2002; Black & Harrison, 2001b), and teachers' awareness of students' goals can enhance student engagement (Clark et al., 2001). However, Brookhart (2001) found that students' goals are not necessarily tied to specific learning targets, and they relate more generally to the improvement that students believe they need for future achievement. The use of clearly specified goals or criteria may also be more appropriate in certain subjects than others. For example, in the English lesson described by Hodgen and Marshall (2005), the teacher has students compare two versions of a text (Henry V) as a means of "sharing the criteria" (p.161) in an assignment that ultimately asks students not only to critique existing literary productions, but also to generate their own ideas for production. Hodgen and Marshall (2005) write that unlike the mathematics lesson they also analyze, the learning process in the English lesson involves "heading more towards a horizon than a tightly

defined goal (p.166). Although this teacher exhibits considerable skill in "apprenticing" students into the "guild knowledge" (p.172) of the discipline, the criteria for success with the assignment were implied, but sometimes "unspoken" (p.163). Implicit criteria is also seen in Yung's (2001) study of Biology teachers in Hong Kong, where one teacher repeatedly warns students that the amount of help they ask for during practical work may affect their evaluation. As Yung (2001) notes, the teacher persists in using autonomy as evaluative criteria, even though it conflicts with the learning purpose of the assessment, because of her desire to be 'fair' in the process. Similar conflicts or inconsistencies are also seen in a survey of teachers' practices done by Stokking and colleagues (2004) in the Netherlands. Not only was there a mismatch between learning goals and criteria, in some cases the "assessment criteria were lacking or were not sufficiently explicit" (p.109) for use by students. This is unfortunate as positive consequences are associated with the use of explicit criteria in several of the studies reviewed here (Barootchi & Keshavarz, 2002; Davies et al., 2004; Hermann & Lewis, 2004), and a variety of methods are used to focus student learning, including criteria in rubric format (Hand & Prain, 2002; Hermann & Lewis, 2004) and exemplars of student work (Wiliam et al., 2004).

Varied Approaches to Elicit Learning. This final aspect of formative assessment is not one that is explicitly discussed with great frequency in the articles reviewed. Those that do mention the use of varied approaches to elicit learning are often studies where the teachers were not strictly bound by the terms of the research, but were encouraged to develop a range of strategies following some form of professional development with the researchers (Bell & Cowie, 2001; Hand & Prain, 2002; Simon & Forgette-Giroux, 2000; Wiliam et al., 2004). There are also descriptions of classroom projects that include multiple methods of eliciting learning (Hermann

& Lewis, 2004; Kirkwood, 2000), and some studies where different approaches are compared in more experimental-type research (e.g., Dori, 2003; Schwartz & Martin, 2004). In general, the use of varied methods has a positive effect on the students involved in these studies. For example, the alternative embedded assessments used in the experimental schools in the Matriculation 2000 project described by Dori (2003) included a wider variety of tasks and activities than the traditional assessments used in the control schools. Students in the experimental groups participated in laboratory activities, group projects, and individual self-assessments, and their exposure to higher-order activities had a positive impact on their attitudes, as well as on their choice of more challenging summative tasks. Teachers in some of the studies analyzed also benefit from the use of varied assessment methods. Hand and Prain (2002) note that the "participant teachers believed that their repertoire of assessment strategies had been greatly enhanced by their learning from the program, and that their students had reacted very positively to more diverse assignments and classroom procedures" (p.752). While it is evident that some of the teachers in these studies are using a variety of strategies (e.g. Wiliam et al., 2004), we don't know from this body of work to what extent the assessment approaches in most classroom are varied. We also don't know a great deal about which of these varied approaches might be more effective in which contexts. The different questioning strategies proposed by the teachers in Black & Harrison (2001a), or the variety of self-reflection tools used by the teachers in Simon & Forgette-Giroux (2000) provide an array of choices. Additionally, innovative methods that go beyond traditional measures of achievement, such as the problem-solving assessment with an embedded learning resource (Schwartz and Martin, 2004), open possibilities for classroom assessment practice that could bear further empirical and comparative study.

Priorities for Further Inquiry

To answer our third question, we looked first at the manifest content, or the recommendations given explicitly in the texts. These include recommendations for educational policy, professional development, and classroom practice, but our interest here is in those that relate to research. Of the 30 articles analyzed, over half (17, or 57%) clearly contain research recommendations. To facilitate our discussion, we grouped the recommendations into three large areas that are well represented in the texts: educational context, assessment methods, and student engagement. We also discuss two areas that are less clearly addressed in this body of work: classroom interaction and the feasibility of assessment.

Educational Context. Predictably, calls for additional research in the study area are made, either in general terms (Hickey & Zuicker, 2005; Nunes, 2004; Thissen-Roe et al., 2004), over time (Cowie, 2005), or on a broader scale (Dori, 2003; Noonan & Duncan, 2005; Schwartz & Martin, 2004; Wiliam et al., 2004). There are numerous recommendations relating to the setting or subject area, with several made for extension into different contexts. Dori (2003) suggests extension to classrooms that are more representative than the exemplary schools included in the Matriculation 2000 project; McDonald and Boud (2003) recommend similar work on the impact of self-assessment training in countries where "conditions may be less favourable" (p.219), and Simon and Forgette-Giroux (2000) suggest looking at their selection framework (portfolios) for different assessment purposes. Recommendations regarding academic subject focus on extension to other curricular areas (Kirkwood, 2000; Schwartz & Martin, 2004), or they are concerned with the influence of academic subjects on assessment (Hodgen & Marshall, 2005; Kirkwood, 2000; Noonan & Duncan). The number of recommendations relating to the educational context

highlights the situated nature of classroom assessment, and reminds us of Gipps' (1999) statement that classroom assessment can be understood "only by taking account of the social, cultural, economic, and political contexts in which it operates" (p.355). While the transferability of results from one educational context to another can not be assumed, these studies do enlarge our understanding such that the options, or possibilities for formative assessment at the secondary level in any educational system or academic subject may be broadened.

Assessment Methods. Another area in which multiple calls for further inquiry are seen can be categorized as assessment methods. There are recommendations for further work to compare the "relative merits" of specific methods (Noonan & Duncan, 2005, p.6; also Hand & Prain, 2002), as well as suggestions that are concerned with the impact or effectiveness of specific methods, such as student self-assessment (McDonald & Boud, 2003; Noonan & Duncan, 2005), or computer-based teacher intervention (Vendlinski & Stevens, 2002). Some recommendations indicate a need for tools and tasks that are better suited to the methods of assessment for learning, such as self-reflection sheets or rubrics (Simon & Forgette-Giroux, 2000) and tests with embedded learning resources (Schwartz & Martin, 2004). Stokking and colleagues (2004) note the challenge in developing new tools and tasks that "meet such diverse requirements as authenticity and controllability" (p.113), and they are particularly interested in establishing standards of quality for classroom assessment. While we agree with their concern regarding the "relevance of psychometric qualities" (p.112) for the classroom, we don't believe it necessary to develop entirely new standards of quality. At this point considerable work has been done by the educational community in this area, which is reflected in several collaborative documents, such as the *Principles for Fair Student Assessment Practices for Education in Canada* (Joint Advisory

Committee, 1993) and *The Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2003). We suggest that calls for further inquiry into methods of assessment for learning be taken up with reference to the standards of quality that are currently considered suitable for classroom assessment.

Student Motivation and Engagement. A third area that draws numerous recommendations for ongoing research relates to student motivation and engagement. The need to motivate students in the classroom naturally arises, and several texts suggest that the relationship between student motivation and assessment methods should be explored further (Clark et al., 2001; Hand & Prain, 2002; Hickey & Zuicker, 2005). However, the interest is not in external motivators, but more on students' internal processes as they engage in formative assessment. Kirkwood (2000), for example, suggests further investigation into students' "actual learning" (p.532) within an intervention, and Schwartz & Martin (2004) argue that student production in learning and assessment is a "powerful mechanism" of which there is still not "a satisfactory account" (p.171). Vendlinski and Stevens (2002) recommend additional inquiry into the relationship between problem difficulty, student ability, and their strategy selection, and they note that although "such considerations may be less important for summative assessment, they are critical when using these results to formulate curricular interventions" (p.13). In more general terms, the shift of interest from assessment *of* learning to assessment *for* learning requires that greater attention be paid to the nature of student learning in the process of classroom assessment.

Classroom Interaction. Although this body of work is strongly focused on student involvement in assessment, very few recommendations are made for further inquiry into the dynamics of that involvement. Though the importance of classroom interactions and the influence

of the classroom environment are repeatedly seen (Barootchi & Keshavarz, 2002; Black & Harrison, 2001a; Cowie, 2005; Doppelt, 2003; Hand & Prain, 2002; Hermann & Lewis, 2004; Kirkwood, 2000; Leat & Nichols; Nunes, 2004; Torres Pereira de Eca, 2005), only two articles (Cowie, 2005; Schwartz & Martin, 2004) emphasize the need for further study in this area. Student voices, which are not strongly heard overall in this body of work, could play a more significant role. For instance, Kirkwood (2000) remarks that because secondary students have "daily exposure to teaching across a range of subjects, they may actually be in a stronger position than their teachers to assess the generalizable aspects of courses" (p.529). A greater use of students as informants about the experience of assessment for learning in a variety of secondary classroom contexts could certainly enhance current understandings in both research and practice.

Feasibility of Assessment. The influence of the educational context on assessment for learning practice is seen in broader terms in several of the studies that take system-wide perspectives (Hayward et al., 2004; Verhoeven & Devos, 2005), and some recommendations that deal with the feasibility of formative practices are made in this area. For example, Torres Pereira de Eca (2005) suggests further investigation into methods that can be used for dual assessment purposes (i.e., formative and summative). This is in keeping with Brookhart's (2001) findings about the dual use of assessment information by 'successful' students, and it might also help address the 'shortage of time' issue that is frequently mentioned in research on changing assessment practices (e.g. Black & Harrison, 2001a; Hayward et al., 2004; Noonan & Duncan, 2005; see Tierney, 2006 for fuller discussion). However, many of the recommendations relating to the educational context focus on policy, practice, or professional development, and they do not explicitly suggest further research. However, the number and frequency of context-related issues

that arise in this body of work, from curricular demands, reporting requirements, and cultural expectations to the need for training and ongoing support (Hayward et al., 2004; Hermann & Lewis,2004; Leat & Nichols, 2000; Noonan & Duncan, 2005; Nunes, 2004; Torres de Peirera de Eca, 2005; Verhoeven & Devos, 2005; Wiliam et al., 2004; Yung, 2001) does suggest that further research should be done on ways of enhancing the feasibility of assessment for learning.

Conclusions

In this review we aimed first, to take stock of existing work on formative assessment at the secondary school level, and second, to provide a view of the research prospects ahead. Overall, the pedagogical potential of formative assessment suggested by Black and Wiliam (1998) is reflected in these texts. Although practical challenges are often raised, the benefits of formative practices are stressed (e.g., Barootchi & Keshavarz, 2002; Davies et al., 2004; Doppelt, 2003; Dori, 2003; Hand & Prain, 2002; Hayward et al., 2004; Hickey & Zuicker, 2005; Kirkwood, 2000; McDonald & Boud, 2003; Thissen-Roe et al., 2004; Wiliam et al., 2004). However, the diversity exhibited in this collection of classroom-based research reminds us that there can be "no one-size-fits-all-package" (Leahy et al., 2005, p.20). With evidence of cultural differences in how learning and assessment are perceived (e.g., Yung, 2001), the contextual nature of formative assessment must be considered. Methods that are effective in one academic subject or geographic location may not be so in another. However, the international nature of current research on formative assessment broadens the spectrum of possibilities for classroom assessment, and numerous prospects for further research can be drawn from these studies despite the different educational contexts in which they are situated.

Certain limitations are inherent to this type of work. The first limitation became apparent

in the first phase of data collection for this study. The selection of relevant texts is extremely time-consuming, and not at all as straightforward as it might appear. Inconsistencies in and between databases (e.g. use of keywords, date range of indexed journals) meant that our efforts to standardize our search procedures were thwarted to a large degree, and we ultimately engaged with each database on its own terms. To maintain research rigour in these circumstances, we carefully documented every action taken, keeping detailed records of the search histories and the texts excluded from our review. As we worked on this project, we became acutely aware of a second limitation, which is best described as a tension between depth and breadth. While we wished to provide an overview of the research on formative assessment at secondary level by surveying across all the relevant studies, we also wanted to provide sufficient detail to make the results and discussion meaningful and useful to others. As qualitative methods currently dominate this area of educational research, the results of many current studies do not lend themselves well to more conventional means of research synthesis, such as meta-analysis. In structuring this study, we found the guidelines for qualitative reviews to be vague, and sometimes conflicting. Literature on methods for reviewing qualitative research would benefit from further dialogue amongst researchers who have experience in this type of work. A third limitation relates to the timeliness of the research. As we looked for research prospects within this body of work, we noted that some of the recommendations in earlier studies were already being addressed. For example, Simon and Forgette-Giroux (2000) call for further inquiry into the complementary natures of portfolio and traditional assessment, and although their work may not be in direct response, Barootchi and Keshavarz (2002) do deal with this issue just a few years later. In several studies, the authors also indicate areas where they propose to, or have already furthered their

research (Black & Harrison, 2001b; Vendlinski & Stevens, 2002). However, this final limitation may also be viewed in a positive light in that it does give indication of ongoing movement within the field of classroom assessment research.

We do believe that the benefits of this review outweigh its limitations. This study deepens the accumulated knowledge about formative assessment as a complex practice through the review of a meaningfully-clustered group of 30 articles. We have been able to understand more about specific aspects of formative assessment by analyzing the texts in relation to a model of formative assessment based on a synthesis of current definitions. The quiet zones in our analysis suggest several promising avenues, such as stronger use of student voices, deeper inquiry into the human dynamics that play into formative assessment, and greater consideration of feasibility issues. It is hoped that this information will be valuable for those who wish to improve classroom practices, assessment policy, or professional development, and especially for those who are interested in further research on formative classroom assessment.

References

* References marked with an asterisk indicate studies included in the review.

Allal, L. & Ducrey, G. P. (2000). Assessment of - or in- the zone of proximal development. *Learning and Instruction,* 10, 137-152.

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box.* Retrieved July 28, 2004 from the Assessment Reform Group Web site: http: //www.arg.educ. cam.ac.uk/ publications.html

Assessment Reform Group.(2002). *Assessment for learning: 10 principles.* Retrieved June 2, 2005 from the Assessment Reform Group Web site: http: //www.arg.educ.cam. ac.uk/publications.html

* Barootchi, N. & Keshavarz, M. H. (2002). Assessment of achievement through portfolios and teacher-made tests. *Educational Research,* 44 (3), 279-288.

* Bell, B. & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education,* 85 (5), 536-553.

Black, P. (2003, April). *The nature and value of formative assessment for learning.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

* Black, P. & Harrison, C. (2001a). Feedback in questioning and marking: The science teacher's role in formative assessment. *School Science Review,* 82 (301), 55-61.

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

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education,* 5 (1), 7-74.

Black, P. & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal,* 29 (5), 623-637.

Bloom, B.S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on Formative and Summative Evaluation of Student Learning.* New York; McGraw-Hill.

Borko, H. , Davinroy, K. H., Bliem, C. L., & Cumbo, K. B. (2000). Exploring and supporting teacher change: Two third-grade teachers'experiences in a mathematics and literacy staff

development project. *Elementary School Journal,* 100 (4).

* Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education,* 8 (2), 153-169.

Centre for Educational Research and Innovation. (2005). *Formative assessment: Improving learning in secondary classrooms.* Retrieved March 2005 from the Organisation for Economic Co-operation and Development Web site: http://www.oecdbookshop.org/oecd/display.asp?sf1=identifiers&st1=962005021P1

* Clark, C. Chow-Hoy, T. K., Herter, R. J., Moss, P. A. (2001). Portfolios as sites of learning: Reconceptualizing the connections to motivation and engagement. *Journal of Literacy Research,* 33 (2), 221-241.

Cowie, B. & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education,* 6 (1), 101-116.

* Cowie, B. (2005). Student commentary on classroom assessment in science: A sociocultural interpretation. *International Journal of Science Education,* 27 (2), 199-214.

Crooks, T. (2001, September). *The validity of formative assessments.* Paper presented at the annual meeting of the British Educational Research Association, Leeds, England.

Davies, P. (2000). The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education,* 26 (3&4), 365-378.

* Davies, P., Durbin, C., Clarke, J. & Dale, J. (2004). Developing students' conceptions of quality in geography. *The Curriculum Journal,* 15 (1), 19-34.

* Doppelt, Y. (2003). Implementation and assessment of project-based learning in a flexible environment. *International Journal of Technology and Design Education,* 13, 255-272.

* Dori, Y. J. (2003). From nationwide standardized testing to school-based alternative embedded assessment in Israel: Students' performance in the Matriculation 2000 project. *Journal of Research in Science Teaching,* 40 (1), 34-52.

Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning.* Thousand Oaks, CA: Corwin Press, Inc.

Evans, J. & Benefield, P. (2001). Systematic reviews of educational research: Does the medical model fit? *British Educational Research Journal,* 27 (5), 527-541.

Gipps, C. V. (1994). *Beyond testing: Toward a theory of educational assessment*. London: The

Falmer Press.

Gipps, C. V. (2002). Sociocultural perspectives on assessment. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century: Sociocultural perspectives on the future of education* (pp. 73-83). Oxford: Blackwell Publishers, Ltd.

Guskey, T. R. (2005, April). Formative classroom assessment and Benjamin S. Bloom: Theory, research, and implications. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada..

* Hand, B. & Prain, V. (2002). Teachers implementing writing-to-learn strategies in junior secondary science: A case study. *Science Education,* 86, 737-755.

Hayward, L. & Hedge, N. (2005). Travelling towards change in assessment: Policy, practice and research in education. *Assessment in Education,* 12 (1), 55-75.

* Hayward, L., Priestley, M. & Young, M. (2004). Ruffling the calm of the ocean floor: Merging practice, policy and research in assessment in Scotland. *Oxford Review of Education,* 30 (3), 397-415.

* Hermann, R. & Lewis, B. (2004). A formative assessment of geologic time for high school Earth Science students. *Journal of Geoscience Education,* 52(3), 231-235.

* Hickey, D. T. & Zuiker, S. J. (2005). Engaged participation: A sociocultural model of motivation with implications for educational assessment. *Educational Assessment,* 10 (3), 277-305.

* Hodgen, J.& Marshall, B. (2005). Assessment for learning in English and mathematics: A comparison. *The Curriculum Journal,* 16 (2), 153-176.

Hsieh, H.-F. & Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research,* 15 (9), 1277-1288.

Joint Committee on Standards for Educational Evaluation. (2003). *The Student Evaluation Standards.* Thousand Oaks, CA: Corwin Press.

* Kirkwood, M. (2000). Infusing higher-order thinking and learning to learn into content instruction: A case study of secondary computing studies in Scotland. *Journal of Curriculum Studies,* 32 (4), 509-535.

Leahy, S., Lyon, C., Thompson, M. & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership,* 63 (3), 18-24.

* Leat, D. & Nichols, A. (2000). Brains on the table: Diagnostic and formative assessment through observation. *Assessment in Education,* 7 (1), 103-121.

Mabry, L., Poole, J., Redmond, L., & Schultz, A. (2003). Local impact of state testing in Southwest Washington. Education Policy Analysis Archives, 11 (22). Retrieved 28 July, 2004 from http://epaa.asu.edu/epaa/v11n22

Mayring, P. (2000). Qualitative content analysis. Forum: Qualitative Social Research [Online], 1 (2), Available at http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.html

McDonald, B. (2002). Self assessment skills used by high school students without formal training. *School Psychology International,* 23 (4), 416-424.

* McDonald, B. & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education,* 10 (2), 209-220.

McTighe, J. & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership,* 63 (3), 10 -17.

Miller, L. (2005). Assessing assessment. *Professionally Speaking*, March, 37-40.

* Noonan, B. & Duncan, R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research & Evaluation,* 10 (17). Available online: http://pareonline.net/getvn.asp?v=10&n=17

* Nunes, A. 2004. Portfolios in the EFL classroom: Disclosing an informed practice. *ELT Journal*, 58 (4), 327-335.

Principles for Fair Student Assessment Practices for Education in Canada. (1993). Edmonton, Alberta: Joint Advisory Committee.

Pryor, J. & Torrance, H. (1998). Formative assessment in the classroom: Where psychological theory meets social practice. *Social Psychology of Education,* 2, 151-176.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science,* 28 (1), 4-13.

Roos, B. & Hamilton, D. (2005). Formative assessment: A cybernetic viewpoint. *Assessment in Education,* 12 (1), 7-20.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science,* 18 (2): 119-144.

Sandelowski, M., Docherty, S. & Emden, C. (1997). Qualitative metasynthesis: Issues and

techniques. *Research in Nursing & Health*, 20, 365-371.

* Schwartz, D. L. & Matin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction,* 22 (2), 129-184.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher,* 29 (7),1-14.

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

* Simon, M. & Forgette-Giroux, R. (2000). Impact of a content selection framework on portfolio assessment at the classroom level. *Assessment in Education,* 7 (1), 83-101.

Smith, E. & Gorard, S. (2005). "They don't give us our marks": The role of formative feedback in student progress. *Assessment in Education,* 12 (1), 21-38.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan,*3 (10), 75-78.

Stiggins, R. J. & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice,* 44 (1), 11-18.

* Stokking, K., van der Schaaf, M., Jaspers, J., Erkens, G. (2004). Teachers' assessment of students' research skills. *British Educational Research Journal,* 30 (1), 93-116.

* Thissen-Roe, A., Hunt, E. & Minstrell, J. (2004). The DIAGNOSER project: Combining assessment and learning. *Behavior Research Methods, Instruments, & Computers 2004,* 36 (2), 234-240.

Threlfall, J. (2005). The Formative Use of Assessment Information in Planning – the Notion Of Contingent Planning. British Journal of Educational Studies, 53 (1), 54–65.

Tierney, R. D. (2006). Changing practices: Influences on classroom assessment. *Assessment in education: Principles, Policy & Practice,* 13 (3), 239-264.

Torrance, H. & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal,* 27 (5), 615-631.

* Torres Pereira de Eca, M. T. (2005). Using portfolios for external assessment: An experiment in Portugal. *International Journal of Art & Design Education,* 24 (2), 209-218.

* Vendlinski, T. & Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *The Journal of Technology, Learning, and Assessment,* 1 (3). Retrieved November 25, 2005 from http://www.jtla.org

* Verhoeven, J. C. & Devos, G. (2005). School assessment policy and practice in Belgian secondary education with specific reference to vocational education and training. *Assessment in Education,* 12 (3), 255-274.

* Wiliam, D., Lee, C., Harrison, C., Black, P. (2004). Teachers developing assessment for learning:Impact on student achievement. *Assessment in Education,* 11 (1), 49-65.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education,* 45, 477-501.

* Yung, B. H. W. (2001). Three views of fairness in a school-based assessment scheme of practicalwork in biology. *International Journal of Science Education,* 23 (10), 985-1005.

APPENDICES

Appendix A: Data Collection Forms (1 of 2)

STUDY: _____

Inclusion Criteria

Date: Article published or thesis approved between 01 January 2000 and 31 December 2005

Language: English text

Quality: Refereed journal with editorial committee and peer review

Type: Empirical, defined here as classroom-based research in either primary form or reconceptualized from previous findings by author

Context: Secondary classroom assessment - Must includes grades 9, 10, 11 or 12, but may also include some lower secondary (7 to 9) or elementary (e.g.system survey)

Relevancy: Focus of inquiry clearly related to at least one aspect of the synthesized model in classroom practice:

$ Explicit learning goals and/or evaluative criteria
$ Varied approaches to elicit learning
$ Feedback to students
$ Adjustment of teaching and/or learning
$ Active involvement of students

$ Integration of teaching, learning and assessment
$ Constructive learning environment

Exclusion Criteria

Do not include in sample if inclusion criteria (date, language, quality, type, context, relevancy) is not met. Studies retained for analysis should also not:

$ Be Master's degree dissertations (inclusion in digital database fairly recent and representation still uneven; also more varied in quality, shorter, and not necessarily original contribution)
$ Involve special populations outside of regular classrooms, or relate to mental or physical health issues beyond those addressed in classroom curriculum
$ Relate to external assessment or program evaluation more than classroom assessment
$ Includes NO students in grades 9 to 12 (revised 14 02 07).

Sampling Decision

EXCLUDE INCLUDE

Rationale:_____

Appendix A: Data Collection Forms (2 of 2)

STUDY: _____

CONTEXT

Location: Years/Grades: Subjects:

_____ _____ _____

METHODOLOGY

Method: Number & Type of Participants:

_____ _____

ASPECTS OF ASSESSMENT FOR LEARNING		Intro/Lit	Results/Discussion
$	Communication of learning goals and/or assessment criteria	_____	_____
$	Varied assessment approaches to elicit learning	_____	_____
$	Integration of teaching, learning and assessment	_____	_____
$	Assessment informing teaching	_____	_____
$	Feedback to students for learning	_____	_____
$	Involvement of students in assessment	_____	_____
$	Other _____	_____	_____

RECOMMENDATIONS

$ None _____

$ For Classroom Practice _____

$ For Professional Development _____

$ For Further Research _____

$ Other _____

Appendix B: Context of Studies for Selected Articles

Author (Date)	**Country**	**Academic Subject**	**Level**	**Study Participants**
Barootchi & Keshavarz (2002)	Iran	English (EFL)	Sophomore (aged 16)	60 students
Bell & Cowie (2001)	New Zealand	Science	years 7 to10 (ages 11 to 14)	114 students 10 teachers
Black & Harrison (2001a)	England	Science	years 7 to 11	9 teachers
Black & Harrison (2001b)	England	Science	years 7 to 11	9 teachers
Brookhart (2001)	USA	Honors English Science Anatomy	grade 10, 11, 12	50 students 2 teachers
Clark et al. (2001)	USA	English	grades 9 to 12	3 cases exemplifying larger study n= ? teachers, but under 5
Cowie (2005)	New Zealand	Science	years 7 to 10	10 teachers 31 students (phase 1) and 75 students (phase 2)
Davies et al (2004)	England	Geography	year 8 and 9	128 students
Doppelt (2003)	Israel	Electrical and Electronics-Vocational	grade 10 to 12 (longitudinal)	54 students
Dori (2003)	Israel	Biology & Chemistry	grade 12	243 students
Hand & Prain (2002)	Australia	Science	years 7 to 10	2 teachers
Hayward et al (2004)	Scotland	Cross-curricular	Elementary and Secondary	Teachers in 33 school (n = ?)
Hermann & Lewis (2004)	USA	Earth Science	grade 9	140 students
Hickey & Zuicker (2005)	USA	Science	grade 9	36 students
Hodgen & Marshall (2005)	England	English and Math	year 10 and year 7	2 classes of student (n = ?)
Kirkwood (2000)	Scotland	Computing Studies	Secondary 3/4 (ages 14 to 16)	16 students

Leat & Nichols (2000)	England	Geography	Secondary (actual year not clear)	1 class in small groups n = ?
McDonald (2002)	Barbados	Cross-curricular	Form 5 (ages 14 to 16)	570 students
McDonald & Boud (2003)	Barbados	Cross-curricular	Form 5 (ages 14 to 16)	515 students
Noonan & Duncan (2005)	Canada	Cross-curricular	Secondary	118 teachers
Nunes (2004)	Portugal	English (EFL)	grade 10	14 students
Schwartz & Martin (2004)	USA	Mathematics	grade 9 and undergraduates	E1:100 students (g.9) and 25 undergrads E2: 102 students (g.9)
Simon & Forgette-Giroux (2000)	Canada	Varied: Writing, Geography, Math, Science	grade 1, 6, 7/8, 9	11 teachers
Stokking et al. (2004)	Netherlands	Varied: Physics, Biology, History, Economics and Geography	Upper secondary	214 teachers
Thissen-Roe et al (2004)	USA	Physics	Secondary	6000 (2002-03) students 1080 (2001-02) students
Torres Pereira deEca (2005)	Portugal	Art	Upper Secondary (ages 17+)	10 teachers 117 student
Vendlinski & Stevens (2002)	USA	Chemistry	Secondary	134 students
Verhoeven & Devos (2005)	Belgium	Dutch, Mathematics, History, Vocational & Technical	grades 7 to 12	1274 teachers
Wiliam et al (2004)	England	Mathematics and Science	years 7 to 11	24 teachers
Yung (2001)	Hong Kong	Advanced Biology	Secondary	3 teachers