

**CATCH-22: ASSESSING ONTARIO'S PROPOSED FORMATIVE ASSESSMENT
POLICY WITHIN ONTARIO'S STUDENT ACHIEVEMENT FRAMEWORK**

by

Mike Foley

A paper submitted to
Professor Skip Hills
in conformity with the requirements for
the degree of Masters of Education

Faculty of Education

Queen's University

Kingston, Ontario

January, 2010

Abstract

This paper examines the efficacy of Ontario's proposed formative assessment policy in the context of Ontario's student achievement framework which is founded upon large scale assessments for accountability. This paper comprises an extensive review of the literature on formative assessment policy, drawing heavily on the British experience, as well as an examination of the empirical evidence conducted on formative assessment in the 'real' classroom. This paper argues that the evidence clearly indicates that policies which are focused on measuring student achievement through large scale testing for accountability inhibit the practice of formative assessment in the classroom. It is also suggested that policies focused on increasing student scores on large scale assessments have not been as successful as anticipated thereby increasing the visibility and promotion of formative assessment in recent policy texts. Highlighting the similarities between student achievement frameworks found in Britain and that of Ontario this paper suggests that Ontario's student achievement framework does not fully address the complexity of formative assessment thereby preventing discussion on current education policies that may inhibit the successful implementation of the proposed formative assessment policy. Formative assessment and accountability is addressed throughout the paper.

ACKNOWLEDGMENTS

I wish to thank Professor Skip Hills for being open to my perspectives and providing the necessary support to see this project through. I also want to thank Professor Scott Johnston for his advice and analysis. I need to also thank my wife Jennifer for being my sounding board through good times and bad during this time. It is also necessary to thank my children, Hee Soo and Lena, for their mostly successful attempts at trying not to bother me too much during my long hours in front of the computer.

TABLE OF CONTENTS

CHAPTER 1 - OVERVIEW.....	5
What is Formative Assessment? A Theoretical Framework.....	7
CHAPTER 2 - ORIGINS: FORMATIVE ASSESSMENT AS EDUCATION POLICY: THE BRITISH EXPERIENCE	12
A False Start in England and Wales	12
Political Ideology and ‘Other Factors’	14
Scotland	18
Northern Ireland	20
CHAPTER 3 - A NEW PLAN OF ATTACK	23
Another Kick at the Can	25
Scotland: A Renewed Commitment.....	26
Success?	27
Wales: All in	31
Success?	33
Northern Ireland	34
England: Standing Fast?	35
The Expanding Definition of Formative Assessment	39
How the Strategies Undermine Formative Assessment	48
The Lessons From Great Britain	55
CHAPTER 4 - ONTARIO’S FORMATIVE ASSESSMENT POLICY.....	57
Ontario’s Large Scale Assessments	59
Ontario’s Student Achievement Framework	63
Building Teacher Capacity	63
What Data do the EQAO Assessments Provide?	65
Teaching to the Test	66
Classroom Responses to the EQAO Results	68
Teachers’ Acceptance of the EQAO Assessments	74
Formative Assessment: Why Now?	76
CHAPTER 5 - BEYOND LARGE SCALE ASSESSMENTS.....	79
Formative Assessment and Teachers’ Assessment Skills	79
Principals’ Assessment Skills.....	80
Full Disclosure: Formative Assessment in the Real Classroom.....	81
Concerns in the Subsystem.....	86
Unintended Consequences	88
CHAPTER 6 - IMPLICATIONS	92

	4
Conclusion	96
REFERENCES	101

Chapter 1

Overview

The province of Ontario is moving to make formative assessment, also known as assessment for learning, a key feature of the Ontario curriculum as of September, 2009 (Ontario Ministry of Education, 2008). This paper is an examination of Ontario's proposed formative assessment policy in the context of Ontario's student achievement framework. The purpose of this paper is to determine if Ontario's proposed formative assessment policy can be successfully implemented within Ontario's existing student achievement framework which includes, in no small part, responding to the Education Quality and Accountability Office (EQAO) assessments (Earl & Levin, 2000; Levin, 2004; Levin, Glaze, & Fullan, 2008; Campbell & Levin, 2009). In this paper I shall argue that Ontario's current, and arguably successful, student achievement framework which emphasizes improving student scores on the EQAO assessments will undermine the proposed formative assessment policy thus inhibiting its successful implementation.

The British experience with formative assessment policy provides a good starting point for discussion. Here, the origins of formative assessment as educational policy can be traced aiding in the understanding of the contexts under which current formative assessment policy initiatives have been developed. This paper will illustrate how and why formative assessment has come to the fore of assessment policy in Scotland and Wales as well as reemerging in policy texts in England and Northern Ireland. Of course education policy does not appear out of a

vacuum and it will be pointed out that Britain's current renaissance with formative assessment has also coincided with reductions and modifications of large scale assessments for accountability (Leung & Scott, 2009; Elwood, 2009; Whetton, 2009). Therefore, a large portion of this story involves exploration into the relationship between formative assessment and standardized tests for accountability. The literature in this area suggests a tension exists between the two that inhibits the practice of formative assessment (Troman, 1989; James, 2000; Hall, 2000; Hayward, Priestley & Young, 2004; Hutchison & Hayward, 2005; Daugherty & Ecclestone, 2006). The notion that large scale assessments for accountability can be considered as facilitating formative assessment practices, in what is defined as system wide formative assessment (Earl & Torrance, 200; Wideman, 2002; Johnson 2005; Campbell & Levin, 2009), will also be considered.

The British experience with formative assessment policy and large scale tests for accountability is of direct relevance to recent developments in Ontario and Ontario's proposed formative assessment policy. A comparison between Ontario and Britain can be justified since the history of education accountability in Ontario appears to have mirrored developments in Britain. The introduction of large scale assessments in Ontario and Ontario's student achievement framework as a response to such assessments represented in Ontario's literacy and numeracy strategy, in particular, appears to be based on the British model (Earl, Watson, & Torrance, 2002; Levin, 2004; Levin, Glaze, & Fullan, 2008) and therefore provides valuable insights relevant to the purpose of this paper.

Throughout this paper, however, empirical evidence on formative assessment will be addressed in context to the claims made on the efficacy of formative assessment as educational

policy. As will be explained in this paper the evidence in this area exposes a gap between formative assessment policy, purported to be evidence-based, and the ‘real’ classroom. The empirical evidence suggests, therefore, that formative assessment is a particularly difficult and complex evaluation technique not easily acquired by teachers even after extensive training. I want to suggest then that current policy instruments within Ontario’s current student achievement framework are insufficient for the successful implementation of the proposed formative assessment policy. Issues regarding formative assessment as educational policy in relation to accountability will also be addressed.

What is Formative Assessment? A Theoretical Framework

As Gareis (2007) points out formative assessment is not a new concept and has been around since the days of Socrates. He notes that the Socratic method “essentially amounts to using questions to assess understanding, to guide learning, and, ultimately, to foster critical thinking” (p. 18). The Socratic method then can be seen as the bare bones upon which the framework of formative assessment is structured leading to the students’ self-realization of where they are and what they need to do to move forward. Although Davies and Ecclestone (2007) note that “there is currently no watertight definition of formative assessment” (p. 72), Gareis (2007) suggests that “formative assessment is any means by which a teacher figures out what students are getting and what they are not getting-in the classroom, for the purpose of teaching and learning, but not for purposes of grading” (p. 18).

Black and Wiliam (1998a) define formative assessment as “all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify

the teaching and learning activities in which they are engaged” (p. 8). Furtak (2003) describes this process as a “feedback loop” (p. 4). The feedback loop is the process where feedback is provided by the teacher based on the teacher’s interpretation of where the student currently is in comparison to the goal to be achieved. The size of the gap is determined by the teacher’s observations of the students classroom performance in oral responses, in-class assignments, homework, or conversations in small groups. Such activities “make the students’ thinking visible so that their level of understanding can be compared to the goal” (p. 5). Furtak explains this difference as the ‘gap’ in where the student is (point A) to where the teacher wants the student to go (point B). Heritage (2007) maintains that once the student’s level is identified, measuring the distance to where the student wants to go is critical. If the distance is too short, the student may view an increased effort as inconsequential. On the other hand if the gap is too large, the student may feel that it is unattainable and become discouraged. Heritage explains that formative assessment is a process that identifies the “just the right gap” and is what educational psychologists term “the zone of proximal development (ZED)” (p. 141).

Heritage (2007) notes that a vital component of the ‘feedback loop’ is the students’ involvement in the process. The teacher-student relationship facilitates the progression to the predetermined goal by increments in terms of scaffolding. Scaffolding is the process that moves the learner forward “from what they already know to what they can do next” (p. 142). The predetermined goal is approached incrementally by the student and teacher together. During this step by step process students’ continually reflect and monitor their progress during the learning process gauging what they presently know and when to acquire more information to move forward in collaboration with the teacher. Theoretically, students begin to acquire “self -regulation

strategies and are able to adapt their learning tactics to meet their own learning needs” (p. 142). With every movement forward the students’ motivation and self-efficacy are increased providing the fuel that continues to propel the student forward (Heritage, 2007).

Heritage (2007) identifies three strategies that teachers can use to promote formative assessment learning in the classroom. On-the-fly assessment is a spontaneous occurrence where the teacher quickly addresses student misconceptions that were perhaps noticed during class discussions, or when listening to student conversations, before preceding to the planned lesson. The second strategy is planned-for-interaction. Here teachers preplan how they will elicit student responses during instruction about the ideas presented in the lesson. The third strategy is curriculum-embedded assessments. Heritage divides embedded assessments into two types. The first type is developed by teachers and curriculum developers who plant questions or activities in the curriculum at “key points in a learning sequence” to elicit student responses and allow for teacher feedback if misconceptions are detected (p. 141). The second type of embedded assessment is described by Heritage as activities that are ongoing within the classroom such as in-class assignments or in the checking of student notebooks from time to time.

Central to the successful use of formative assessment is the quality of the feedback that teachers provide to their students. Heritage (2007) explains “effective feedback provides clear, descriptive, criterion-based information that indicates to the students where they are in a learning progression” and “how their understanding differs from the desired goal, and how they can move forward” (p. 142). Here too, however, just the right amount of feedback is required. Torrance (2007) observes that too much teacher feedback can lead to instrumentalism. This is where students come to rely on the teacher as a crutch, in moving from point A to point B in the

feedback loop rather than acquiring autonomy over their learning. When this occurs, the learning that takes place poses little challenge to the student so the results of the outcomes become less valid in terms of the quality of the learning that took place (Torrance, 2007).

Since teachers are guiding students to learning goals dictated by curriculum expectations it is essential that the feedback offered by teachers is criterion referenced. However, Torrance and Pryor's (2001) conception of *convergent* and *divergent* formative assessment suggests that effective formative assessment is much more than applying a simple formula of sharing learning aims with students when feedback is provided by the teacher. Convergent formative assessment follows a behaviourist approach to learning following strict planning using formative assessment as a series of summative assessments closely following the established curriculum using closed or pseudo open-ended questions that focus on fixing errors with correct responses. For instance a teacher may introduce material to students followed by questions that encourages them to recall the material rather than providing an opportunity for a process of student engagement and discovery. Absent here is a collaboratively constructed understanding resulting from the interaction between the teacher and the student or the class in general. The student is described as the recipient of assessment. On the other hand, divergent formative assessment follows a constructivist approach to teaching centering on student and teacher interaction and collaboration where teaching takes place in the zone of proximal development. The focus here is on how the student interacts with the curriculum and explores how the students' work can bring insights into their current understanding and on aiding metacognition. Through student involvement in their own assessment they are both the recipient and the initiator of the assessment. The authors note that as a starting point for adopting divergent assessment processes teachers must "be able to

investigate and reflect upon their own classroom practices—particularly the way they question and give feedback to students” (p. 629). As Black and Wiliam (2003) note, good feedback “causes thinking” (p. 631).

The theoretical framework guiding this paper then views formative assessment as a partnership between the teacher and the student, albeit one where the teacher takes the initial lead role in bringing the student into the learning process. It is a complex hybrid of both a teacher-centred and student-centred approach. It is a process that involves the teacher interacting with the student for the purpose of enabling increases in learning occurring within the ‘feedback loop’ described above. Successful formative assessment requires a partnership between the teacher and the student working collaboratively, producing a shared understanding of where the student is and how they will move forward. Formative assessment has to be understood by the teacher and the student. The teacher must execute formative assessment with skill and precision to be effective in the classroom. Formative assessment is not a formula to be implemented but an acquired skill through professional development in a process of “knowledge creation” (Black & Wiliam, 2003, p. 631).

Chapter 2

The Origins of Formative Assessment as Education Policy: The British Experience

A False Start in England and Wales

Throughout the 1970's and 1980's debates emerged in Britain on how well the education system was serving its students and the nation. There were growing concerns that the education system was becoming too child centered or that it was not meeting the demands of a changing global economy witnessed in the economic downturns of the time (Whetton, 2009; Wyse & Torrence, 2009). British Prime Minister James Callaghan, in a speech at Ruskin College in 1976, asserted that although the education system maintained high standards improvements were necessary linking the education system to future economic success. There were also growing perceptions by some government officials that “schools were too free to do as they please, with little apparent accountability” (Wyse & Torrance, 2009, p. 15). Noting these growing concerns Callaghan warned educators that “if the public is not convinced then the profession will be laying up trouble for itself in the future” (Whetton, 2009, p. 139). A warning, Whetton (2009) notes, that went unheeded.

Legislation was drafted in 1988 establishing the very first common curriculum for England and Wales. Although Scotland and Northern Ireland had been devolved from England for some time, this was largely administrative in nature as education policies were essentially “London-based” with a “certain amount of local adaptation” (Leung & Scott, 2009, p. 64).

Secretary of State for Education Kenneth Baker requested the Task Group on Assessment and

Testing (TGAT), chaired by Professor Paul Black, to create an assessment system for the new National Curriculum that would serve both informative and diagnostic purposes (James, 2000). The new assessment system proposed by the TGAT Report (1988) envisaged teachers assessing their students based on a criterion referenced assessment tasks that would provide teachers with evidence of what students have learned. It was expected that teachers would use the information from the assessment tasks to guide future teaching. Formative assessment then was primarily a process of relying on the teachers' use of the assessments tasks that would provide foci for teachers to use the results to "both feed back to the pupil and feed forward to the next teacher or institution" (para 32), and would "develop from existing good practice" (para 88). A ten-point scale would monitor student progress through the new curriculum. Summative assessments, included in TAGT's plan, of students at the age of 7, 11, 14 and 16 were intended to be used sparingly and designed as extended tasks rather than tests to provide parents information on how their children were progressing (Daugherty & Ecclestone, 2006). The system designed satisfied both formative and summative purposes of assessment but with formative assessment placed in a position of dominance (James, 2000).

James (2000) observed that although TGAT's plan had yet to work out the "technical details" of their proposal she attributes the main reason for its eventual demise was due to the perception by the Conservative government that the plan represented an "intolerance of liberal progressivist educational ideologies" (p. 356). Similarly Daugherty and Ecclestone (2006) note that TGAT's plan was undermined by "several factors" emphasizing "the influence of the New Right" (p. 154) as it was "perceived as the work of an insidious left-leaning education establishment intent upon subverting the government's best intentions to raise student

achievement” (p. 154). As then Prime Minister Margaret Thatcher stated in her memoirs, the fact that it (the TGAT Report) was welcomed by the Labour Party, The National Union of Teachers and the Times educational Supplement was enough to confirm for me that its approach was suspect. It proposed an elaborate and complex system of assessment - teacher dominated and uncoded. (Thatcher, 1993, p. 594 as cited in Daugherty & Ecclestone, 2006, p. 154)

Thatcher’s comments reflected a neo-conservative ideology that was primarily focused on raising student scores by standardized tests as different cohorts of students passed through the system (James, 2000; Daugherty & Ecclestone, 2006). This thinking dominated government policy makers and reflected a culture of “performativity” (Daugherty & Ecclestone, 2006, p. 156) as this was believed to provide an incentive for teachers and students to try harder (James, 2000). Similarly, in Northern Ireland, Hall (2000) suggests that preliminary support for formative assessment in Ireland assessment policy was discarded due to the “more conservative, Fianna Fail element” within the Fine Gael/Labour Coalition and closely mirrored the Conservative agenda in England by focusing on standardized tests (p. 86).

Political Ideology and ‘Other Factors’

The importance of examining the origins of formative assessment policy in the United Kingdom is that it reveals formative assessment as primarily an issue of political ideology for government support for, and reliance upon, standardized tests as the overriding factor determining its adoption as educational policy. Daugherty (2007) argues that government ministers rarely relied upon, at least during that time, on academic research as their main source

of information regarding what policy would be most effective. The assumption here of course is that if educational research did inform policy formative assessment would have come out on top and the assessment proposal of TGAT would have been realized. However the “technical details” and “other factors” alluded to by James (2000), Hall (2000), and Daugherty and Ecclestone (2006) that also contributed to the disappearance of TGAT’s proposals and the prioritizing of standardized tests in Britain deserve further scrutiny.

One of these issues regarded time. As James (2000) noted, at the time, TGAT advised that to prepare and implement the new assessment system would take a minimum of five years. This time line was politically inexpedient. James (2000) observed there were very few, if any, politicians willing to wait the five years it would take to put the policy into practice. Shavelson, Black, Wiliam, and Coffey (2004) reflect that the authors of TGAT generally understood that even a five-year time frame was overly optimistic admitting that a more realistic time line for the new proposal to be properly implemented was more in the area of ten years. Most importantly however TGAT did not offer any suggestions or recommendations on how their proposals would be implemented (Whetton, 2009, p. 146).

TGAT was also rather vague on exactly what formative assessment was. Black later reflected, regarding formative assessment, “little was known at the time about either its precise meaning or about how to develop its potential” (Shavelson, Black, Wiliam, & Coffey, 2004, p. 11) adding that,

the TGAT report said very little about the link between the formative and summative aspects of teachers’ work and, whilst it listed the various ways that teachers might collect data, it had little to say about how teachers should make judgments on the evidence they

might collect. (p. 11).

The reason for the ambiguity regarding TGAT's proposals, suggests Torrance (1993), is that TGAT did not address the theoretical aspects of formative assessment or provide much in the way of empirical evidence to support their proposal. Torrance notes that although the TGAT report "crystallised a good deal of emergent professional opinion" on the potential of formative assessment to increase student achievement the published research up to 1987 did not support their arguments for formative assessment over standardized testing (p. 5). He suggests that the formative assessment research that was published was "at the level of teacher perception - what the teacher reports is the case - and has been generated in the context of innovative projects, usually staffed by interested volunteers" (p. 5) and largely "derived from very small scale studies, often of non-formal educational settings" (pg. 4). Torrance concluded that formative assessment, at that time, needed to be "articulated more fully and explicitly, and built on more than taken-for-granted assumptions about what constitutes good practice" (p. 6) questioning the efficacy of such studies as grounds for a national policy.

Given then the somewhat obscure nature then of TGAT's proposal Secretary of Education Kenneth Baker, although sympathetic, described the proposed system as "complicated and costly" (as cited in Shavelson et al. 2004, p. 11). Although government political ideology was certainly an underlying factor in the phasing out of TGAT's proposal it was not, evidently, the only reason. The weak presentation of formative assessment and the lack of supporting empirical evidence and the limited recommendations for its implementation were critical in undermining formative assessment as a workable national policy.

Troman (1989) argued that the government was faced with two options in the move to

provide accountability and raise standards deciding to opt for the one which was “neat, quick, cheap and quantifiable and away from the complex, slow, expensive and qualitative” (p. 289). This observation was somewhat premature. Standard assessment tasks were implemented in 1991 following TGAT’s blueprint but were deemed impractical in terms of “manageability and reliability arguments” (Whetton, 2009, p. 143). As James (2000) notes this was the era when teachers “cupboards were stuffed with bulging portfolios of children’s work” (p. 344) and was not what TGAT had in mind. Yet, as noted above, TGAT did not have much to offer as an alternative. Soon after it was initiated, the new Secretary of Education Kenneth Clarke described the assessment tasks as “elaborate nonsense” and replaced them with written external examinations. Black reminisced that “no convincing alternative” to standardized tests was offered (Black, 1997, p. 41 as cited in James, 2000, p. 352).

Troman (1989) argued calls for standardized tests would result in school effectiveness being determined by test results detracting teachers attention from assessment as a diagnostic and developmental process. He speculated that “the tension,” between formative and summative assessments, “may eventually be resolved by the gradual atrophy of formative assessments and the ascendancy of summative forms (at the termination of stages of schooling)” (p. 292). Of course this implied that formative assessment was already taking place which was not the case. Teachers may have increased their recordings and documentation of what their students were doing in attempts to be compliant with the new system but there is little evidence that they were altering their instruction as a result. As Black and Wiliam (2003) inform “the debate over the relative weights to be applied to the results of external tests and teacher’s judgements obscured the fact that both of these were summative assessments” (p. 626). However, the tension between

large scale summative assessments and formative assessment observed by Troman (1989) can be seen, in hindsight, as quite prophetic when applied to formative assessment in Scottish educational policy.

Scotland

The difficulties on how to make formative assessment work was not lost on Scottish policy makers and seemed to support Troman's prophecy. Since 1991 formative assessment termed 'assessment for learning' in order to, as Hayward (2007) points out, make the ideas surrounding the summative and formative assessment process sound less complicated, was placed at the fore of Scottish assessment policy. The formative assessment policy in Scotland stated that student "feed back should help pupils to identify what they have learned, what they still have to learn and their next steps in learning" (SOED, 1991a, p. 16, as cited in Hutchinson & Hayward, 2005, p. 227). Teachers were directed to make judgements about students' achievement levels each year based on their own assessments and then, when the teacher felt the student was ready, test the student using a national test held in a national assessment bank to moderate their judgement. National accountability was derived from an annual survey using a 5 per cent sample at year four and seven at the primary level and the second year of secondary school in English, Math and Science (Hutchison & Hayward, 2005). Teachers in Scotland supported this set up (Hayward, 2007). This system can be seen to mirror, in spirit, TGAT's proposal that teacher assessment would play a primary role which would be mediated, to a degree, by large scale assessments.

Hutchison and Hayward (2005) claim that the formative assessment policy was "robust"

since it was based on empirical research. The robustness of research evidence is questionable however since it is the same research outlined earlier in this paper and refers to studies conducted in conditions removed from a 'real' classroom. In terms of implementation of the policy, they note it was assumed that teachers needed only to be informed of the ideas behind it. Policy guidelines provided for the availability of materials to support teacher professional development in planning and conducting formative assessment. Although teachers supported the policy, they did not bother with the materials that supported it. The materials available were to be distributed upon request, requests that were not forthcoming by many schools and therefore were largely ignored by teachers (Hayward, 2007). A review of the system at the end of the decade found that teachers were using the results from the national tests as the prime indicator of student achievement. As Hayward notes, "rather than being used to confirm or to challenge teachers' professional judgement, National Tests were replacing it" (2007, p. 256). Despite an assessment policy that advocated formative assessment and was "informed by research evidence in assessment" this had little impact on what occurred in the classroom (p. 258).

Hayward and Hutchinson (2005) attribute teacher reliance on standardized tests primarily to the "macro political climate with a concern for performativity" (p. 258). They observe that since part of the government's assessment policy included national test results the HMI used the results to confirm teachers' judgements in reading, writing and math. They attribute this focus on test scores to the increasing forces of the accountability movement that was defined by "a culture of mistrust of teachers' professionalism" (p. 229). They note that the influence on classroom practice was "for the time being, overlooked" (p. 230). Hayward (2007) argues that teachers felt that the "system demands that they give significant attention to summative assessments" (p. 257)

that resulted in, as Hayward, Priestly and Young (2007) argue, teachers feeling “obliged to focus on what they believed to be the real priority, national testing” (p. 399). Therefore, despite having a national policy advocating formative assessment, Scotland experienced the same situation in England where the large scale assessments ruled the day in terms of expectations for student achievement.

By 1998 a new system was established to provide readily available scores that the test when ready system did not fulfill (Hutchinson & Hayward, 2005). In 1998 the Scottish Office Education and Industry Department (SOEID) began to gather aggregate attainment information from schools for reading, writing and mathematics confirmed by the use of national tests and was reported at the local authority aggregate level in order to prevent school by school comparison, or league tables, as was the case in England. Hutchinson and Hayward (2005) note though that the increased emphasis on test scores lead to the perception by teachers that national test results were what mattered. Therefore, through their experiences teachers had little incentive to consider other forms of evidence or challenge students’ test results. “The ‘driving up standards dream’ (Black, 2001) appeared to have driven out developments in formative assessment” (p. 231).

Northern Ireland

The focus on standardized assessments over professional development in formative assessment can also be seen in Ireland. In the early 1990's the National Council for Curriculum and Assessment (NCCA) made a number of influential publications and recommendations that were incorporated into the Irish assessment policies (Hall, 2000). Notably the NCAA outlined that assessments for accountability “should not be used for teacher or school accountability (para

5.1.4, as cited in Hall, 2000, p. 88) but emphasized the identification of learning needs of the students especially in the core subjects such as reading, writing and math. Ireland's assessment policy paralleled England's in that both had tests for 14 year olds at key stage three. However, unlike England, teacher assessments were to be the primary means of assessment at the end of key stage one for seven year olds and key stage two for eleven year olds (Elwood, 2007). Yet as was seen in Scotland, the use of standardized tests overshadowed teacher assessments by the teachers' voluntary reliance on the standardized tests (Hall, 2000). Although the assessment policy in place claimed the use of norm-referenced and criterion-referenced assessments were to be used formatively she argues this was neither "sufficiently elaborated or explained" (p. 92). Hall speculates there was a belief among policy formulators that teachers were already practicing formative assessment. Therefore, she suggests, large scale tests were emphasized in policy texts as the means to identify learning needs at the expense of encouraging alternative methods. Hall conjectures that teachers already knew how their students were doing and did not need further tests evidently implying that standardized tests cannot be used for formative purposes. Hall concludes that teachers need to be encouraged to use other forms of assessment especially in "interpreting and responding to assessment in a formative way" (p. 94).

It is evident then that formative assessment had been introduced to governments in Britain and acknowledged within policy texts in varying degrees. Clearly however the emphasis was on large scale assessments for accountability since there was an assumption that the feedback from large scale assessments could be used for formative purposes in the classroom. I would like to also point out that it also appears evident that the emphasis on standardized tests within policy texts has had a negative influence on the growth and development of formative

assessment practices in the classroom. Black and Wiliam and other formative assessment advocates were not content though to sit idly by and watch large scale assessments dominate the educational system. They may have lost the first battle but the war was far from over and preparations were being made for a more strategically planned assault on assessment policies in Britain.

Chapter 3

A New Plan of Attack

The Assessment Policy Task Group of the British Educational Research Association, later called the Assessment Reform Group (ARG) led by Paul Black and Dylan Wiliam undertook a comprehensive literature review that culminated with their now famous *Assessment and Classroom Learning* (Black & Wiliam, 1998a). This time the authors would not repeat the mistakes made a decade earlier by TGAT in their advocacy of formative assessment, at least as far as their presentation of the evidence was concerned. Black and Wiliam's *Assessment and Classroom Learning* (1998a) and their summarized version *Inside the Black Box* (1998b) can be seen as the bait for governments to adopt formative assessment by claiming that not only did formative assessment increase learning in the classroom but it also showed how this would also improve international test scores.

Perhaps the most appealing feature for governments in the age of large scale assessments and the links of such assessments to national competitiveness was in the authors reporting of a meta-analysis the authors conducted on 40 studies selected from their review and other selected reviews. Black and Wiliam (1998b) report that they found an effect size on student achievement between 0.4 and 0.7. They note that a 0.4 effect size would place the average student from a class where formative assessment is used into the top 35 per cent in a class where formative assessment was not being used. They conclude that, "formative assessment helps low achievers more than other students and so reduces the range of achievement while raising achievement

overall” (p. 141). They also point out that a gain of 0.7 would move countries like the United States and Britain from an average ranking on international tests such as the Trends in International Math and Science Survey (TIMSS), to near the top of the list. The significance of this meta-analysis is especially tantalizing for governments since, if adopted as policy, formative assessment will narrow the achievement gap between students of different socioeconomic (SES) backgrounds and at the same time increase test scores on international tests. Packaged in this way there is a direct appeal to governments concerned with improving not only equity in education but also as a means to ensure national economic security.

It is significant here to add here however, that there is little offered in terms of implementation. It is also important to realize that there were not any new ‘breakthrough’ studies regarding formative assessment between 1987 and 1998 but rather the interpretations of the findings were more forcefully presented by being repackaged with considerably more care and presented with considerably more finesse in order to, make it as attractive as possible for governments.

Although Black and Wiliam (1998a) state that, “formative assessment shows conclusively that formative assessment does improve learning” (p. 61) they also added that all the studies examined in their review exhibited, “some degree of movement away from ‘normal’ classroom” (p. 16). It is beyond the scope of this paper to examine all 250 research studies presented in Black and Wiliam’s review but this same observation was made by Sebatane (1998) who, in a general observation of the studies surveyed in Black and Wiliam’s (1998a) literature review, questions if the results could be, “replicated in a different contextual setting” (p. 4). He implies that the success of the studies may have been, “determined by contextual considerations” of the particular

classroom in which the studies took place (p. 4). The above observations by Black and Wiliam and Sebatane suggest that the evidence, although consistent in the findings and impressive regarding their effect sizes on the effectiveness of formative assessment to increase student learning, had not moved much beyond the evidence presented in 1987 and echoes the same concerns previously noted by Torrance (1993) a decade earlier.

The above observations on the shortcomings of the evidence for formative assessment were later confirmed as an ongoing issue by Black and Wiliam (2003) who reflected that the studies selected for the meta analysis was based on, “a best evidence synthesis” (p. 627). Furthermore they admit they, “inevitably, at some points, went beyond the evidence” when noting the implications of their literature review on formative assessment in terms of policy and practice (p. 628). If they limited the implications to only the evidence of formative assessment research, they found that they, “would have very little to say” (p. 628). They also admit that *Inside the Black Box*, “represents our opinions and prejudices as much as anything else, although we would like to think these are supported by the evidence” and the, “success of *Inside the Black Box* has been as much due to its rhetorical force as to the evidence that underpins it” (p. 633). From these statements it can be concluded that the formative assessment was indeed packaged and presented to sell. In many respects it has become a best seller in terms of its influence not only in Great Britain but internationally as well, although it must be stressed there is little offered in terms of how to implement formative assessment as policy.

Another Kick at The Can: The Re-emergence of Formative Assessment as Education Policy

Around the same time Black and Wiliam’s (1998a) literature review was published, the

“new post-devolution administrations in Scotland and Wales, were making overt commitments to ‘evidence-informed policy’ (Daugherty & Ecclestone 2006, p. 166) that the authors suggest allowed for an increase in the receptiveness to new ideas. Black and Wiliam targeted Scottish policymakers and Daugherty targeted Wales policymakers, and these efforts had an, “explicit influence on the assessment policy decisions announced in 2004” (p. 166) where formative assessment gained prominence in Wales and became re-emphasized in Scotland.

Scotland: A Renewed Commitment

Highlighting Black and Wiliam’s (1998) *Assessment for Learning* the Scottish Executive Education Department (SEED) refocused their attention on formative assessment with new vigor establishing the Assessment is for Learning (AifL) project in 2002 (Hayward, 2007). This program was developed in recognition of the absence of formative assessment practices despite, as previously noted in this paper, having guidelines in place advocating formative assessment. The Scottish Minister of Education set up an investigation committee to gain insights on how to get formative assessment to work in the classroom and to, “reconcile the relationship between assessment for learning and assessment for accountability” (p. 256). This program aimed to explore methods that would facilitate changes in teacher pedagogy in terms of adopting formative assessment practices. The project attempted to overcome barriers between policy and practice such as a lack of clarity between research findings and real classroom conditions, support, costs, or the threat of negative inspector reports, that were argued as undermining formative assessment in the 1990's.

The project therefore entailed a supportive and collaborative approach between policy

makers, researchers and classroom teachers that attempted to create teacher capacity and will to adopt formative assessment practices and to influence pedagogical approaches to teaching suggested by Black and Wiliam's (1998a) literature review (Hayward , Priestly & Young, 2007). The study involved 35 schools including elementary, middle, and secondary, and included all local education authorities in Scotland. Teachers were educated on a number of strategies in terms of enhancing questioning techniques and teacher feedback to students, promoting self and peer assessments and sharing the criteria of learning objectives with the students. Teachers involved in the study were directed to develop planned interventions in the classroom and to evaluate its effectiveness.

Success?

Kirtron, Hallam, Peffers, Robertson, and Stobart's (2007) analyses of the study suggest the program was extremely successful. Increases in student meta-cognitive skills were reported with students having awareness, "of their learning needs...what was good about their work and what they needed to do next time to improve" as well as increases in student motivation (p. 620). Despite some "small difficulties" dramatic changes in teacher pedagogy were reported (p. 620). Some teachers noted that formative assessment allowed them to focus on the process of learning instead of the summative outcomes. Teachers noted that the process promoted reflection on their teaching and prompted them to try new methods. Teachers also noted that they enjoyed the sharing of ideas with other teachers during the project and being more informed.

Kirtron et al. (2007) temper these findings noting that those involved in the study were volunteers which may have influenced the reported results. As well, Hayward , Priestly and

Young (2007) acknowledge that the teachers' self reports may or may not have been an accurate reflection of their use of formative assessment practices. Hayward et al.'s (2007) analysis informs that 14 schools "embraced" formative assessment "wholeheartedly," 14 schools made "good progress" and five schools did not do as well (p. 402). They suggest that conventional teacher practices are difficult to overcome but that a similarity between the successful and less successful schools were in differences in terms of support structures and levels of communication within the schools. Kirton et al. (2007) note,

the project itself, especially the funding, the conferences, meetings, networks and opportunities for sharing ideas, were an important element in supporting the teachers in changing and sustaining their practice. Without such support and funding it may be difficult for teachers to continue or develop the use of formative assessment. (p. 625)

Hayward et al. (2004) found that the most significant inhibitor identified by teachers was in their perception that formative assessment practices would not be appreciated by the school inspectors (p. 410). Kirtrou et al. (2007) echo this observation noting that teachers, "expressed concern that there was a tension between HMI requirements for concrete evidence of monitoring and assessment and the principles of the project" (p. 624). For example some teachers found that formative assessment practices lead to a slower delivery of the curriculum and expressed concern that they could not cover the curriculum in time. There was also concern that formative assessment conflicted with the summative demands of "teaching to the test" and preparing students for examinations (p. 624). One teacher summed up this feeling noting that the "the two don't mesh easily" (p. 620). Hayward et al. (2004) concluded that formative assessment will not, "thrive in a climate dominated by evidence-driven assessment for measurement" (p. 413). The

pedagogical change for the teachers involved in the project can be seen as fragile. Teachers expressed concern that they, “could slip back into older habits” (Kirtrou et al. 2007, p. 625). The transitory or superficial internalization of teachers’ use of formative assessment from the study appeared to be confirmed when later evaluations discovered “a lower level of commitment” by the teachers involved in the study (p. 625).

What is evident then, at least from the point of view of the teachers in the AifL programme, is that standardized tests offered little in the way of feedback that they could constructively use to help with their teaching and was the source of incredible interference when attempting formative assessment practices in the classroom. Hayward et al. (2004) report that for many teachers the concept of formative assessment and its promises of enhanced learning “resonate with many teachers’ ‘memories’ of better times; particularly of less pressured classrooms where there was more space for learning” (p. 411). Kirtrou et al. (2007) note that for some teachers in the project the ‘ideas’ of formative assessment were not new but had become, “lost in the increased bureaucracy teachers faced in their roles and the dominance of summative assessment on teachers practice” in a culture focused on test scores (p. 622). It must be noted however that such reminiscent for ‘the good ole days’ belies the fact that previous research on teachers formative assessment indicated a “poverty of practice” (Black & Wiliam, 1998b, p. 5).

It was argued then that the tension between formative and summative assessments has to be resolved so that accountability does not extinguish formative assessment. Hayward and Hutchinson (2005) point out that teachers are still being encouraged to use national tests for the data such tests provide, which they suggest could be useful,

if, as intended in Scotland, benchmarking is taken to mean the intelligent use of a range

of dependable data, whose strengths and limitations are clearly understood, as a framework for comparing one's own provision and performance with that of others in similar or different circumstances, then it becomes a powerful tool for informing improvements. (p. 242)

Yet how to achieve the balance between the two is unclear. For the government, the reduction in the tension between formative and summative assessments for accountability appears to be lessening the primacy of large scale assessments and modifying their usage. The Scottish Executive Education Department circulated the following advice in 2005:

The children's results on National Assessments should be only part of a range of evidence teachers consider to arrive at judgments about levels of attainment. No decision about a child's attainment or future learning should be made or reported on the basis of a single assessment or test score, as it will not, on its own, be sufficiently reliable for that purpose. All assessments and tests used to monitor children's progress and attainment should be demonstrably fit for their purpose. It is unlikely that widespread reliance upon standardized tests will be a common feature within the new arrangements. (as cited in Leung & Scott, 2009, p. 71)

Clearly the above guidelines suggest that formative assessment should be emphasized with standardized/national tests serving a subservient role as was the case in 1991. Whether or not this can be interpreted as the seeds to the demise of public accountability, in terms of a reliance on national assessments in Scotland remains to be seen. For now, Leung and Scott (2009) note, the majority of schools in Scotland still rely on standardized tests. The concern here is that the progress witnessed in the AifL project "remains at risk of being undermined" by the

difficulties of meeting the, “twin demands of assessment for learning and assessment for accountability” (p. 73).

Wales: All in

Daugherty (2008) provides an analysis of a government directed review of assessment in Wales that came to be known as the ‘The Daugherty Review’. Here it is interesting to note that it was understood by the Education Minister and the review committee that the recommendations would be transmitted directly to the policymakers. In this sense, Daugherty provides a relatively straightforward, but important example of the incredible influence formative assessment advocates had on government policy in Wales. This was due to the Education Minister’s continued public declaration that policy would be determined by research. Not surprisingly formative assessment was recommended and accepted by the government. Daugherty argues that the review findings were influenced by the evidence rather than favoring the views of any individual or party must be taken cautiously. Daugherty, an admitted advocate for formative assessment, reveals that the findings of the review were, “grounded in a research review by Black and Wiliam (1998)” and the ARG Group (2008, p. 81) although, as mentioned earlier in this paper, the evidence for formative assessment has still not progressed beyond that of 1987. Daugherty acknowledges that the evidence gathered by the Review Group may appear to have been “inclined to favour one person or party,” namely the formative assessment advocates, the “quality of the evidence,” he argues, was enough to compensate for this lack on inclusiveness (p. 84). The evidence then was largely derived by Black and Wiliam’s (1998a) *Assessment and Classroom Learning*. He suggests that the recommendations of the review were not so much

based on “the findings of that research” but rather on, “a sharing of ideas, variable backed by the evidence” (p. 84). This he notes, led to the construction of, “a third type of knowledge-a ‘where do we go from here’ type of knowledge” (p. 84). This, ‘where do we go to from here’ seems to suggest a sober reflection on how to implement formative assessment. The Daugherty Report stated,

it is clear that test preparation and practice, a narrowing of curriculum coverage and styles of learning that contribute to good test performance have become prominent features of the Year 6 experience of pupils in many schools. . . . The Group has considered whether end-of-key-stage testing, in terms of the “hard” data it gives us on pupil attainments and the targets it gives some pupils to aspire to, is of sufficient value to compensate for the evident impoverishment of pupils’ learning that is occurring at a critical stage in their educational development. (Daugherty Assessment Review Group, 2004, section 3.4 as cited in Leung & Rhea-Dickens 2007, p. 14)

Within months of the publication of the Daugherty Report in 2004 the government announced that teachers for students aged five to seven were directed to continue to use the national tests in reading, writing and math but the results would be used by the teacher in tandem with their classroom assessments. The teacher’s assessment, not the result of the standardized test would be reported nationally (Leung & Rea-Dickens, 2007). It was also announced in 2004 that in 2007 standardized tests would also be abandoned for 11 and 14 year olds. Standardized tests have all but disappeared in Wales for Key Stages 1, 2 and 3 being replaced with teacher assessments except for the final year of secondary school (Leung & Rea-Dickens, 2007).

Success?

The Department for Children, Education, Lifelong Learning and Skills (DCELLS), Welsh Assembly Government established the Developing Thinking and Assessment for Learning programme in 2005 (Welsh Assembly Government Website, 2009). The three-year program was put in place to, as the title of the program suggests, develop thinking skills in students and promote formative assessment in the classroom. BMG Research (2008) provided an ‘independent’ evaluation of the program. Forty-two schools across Wales, that were admittedly enthusiastic about the project, were “hand picked” to implement the program (p. 4). The materials provided in the program consisted of example analyses questions teachers could use, the use of traffic lights (where students hold up a hand to signal understanding of a concept before the teacher moves on), and an assortment of learning games facilitating group work. Teachers were offered assistance only when it was requested. Overall the program appeared successful as qualitative increases in learning were reported by teachers and students. Some increases in student engagement and increases in participation and discussions were noted but quantitative reports of increases in student achievement were not evident. BMG Research made the following summation;

although significant impact on attainment is yet to be observed this is not surprising at this early stage; there is a view from both teachers and LEA advisers that associated changes in attainment will take several more years, and it is difficult to disentangle the impacts of other changes that will also have a bearing on attainment. However, evidence from both learners and teachers indicates that learners are more engaged in learning activities and have a greater understanding of learning approaches, which in turn will

undoubtedly lead to improvements in learner outcomes. (p. 5)

It was recommended that teachers continue to receive support to increase their understanding of formative assessment concepts and to attend professional development programs. School managers, “reported that the majority of the funds provided had been used to allow teachers to have time outside of the classroom which was considered critical to the effective implementation of the programme” (BMG Research, 2008, p. 53). From this study it remains to be seen how well formative assessment as national policy will fulfill policy expectations that formative assessment will lead to increases in student achievement in Wales. Currently then the Wales experience can be seen to be a work in progress.

Daugherty (2009) recently concluded that teacher assessments are now, “reliable enough for the purposes for which summative data is required (p. 249) although, as noted, quantitative increases in student achievement have yet to materialize. A Welsh inspectorate recently, “expressed concern that the proportion of five to seven-year-olds with good levels of reading and writing has stopped rising over the past five years” (Eason, 2009, p.1).

Northern Ireland

Black and Wiliam’s (1998a) review and the wide spread dissemination of their research also influenced assessment policy in Ireland. To address the “perceived shortcomings” of the previous policy’s heavy reliance on standardized tests and the minimal support for the professional judgement of teachers Ireland has made recent modifications to their assessment policy (Elwood, 2009, p. 252). Elwood notes that the key stage 3 test for 14 year olds has been eliminated and formative assessment is now being strongly encouraged for integration into the

assessment practices of teachers. Northern Ireland's curriculum and assessment body the Council for Curriculum, Examinations and Assessment (CCEA), are persuading teachers in official documentation to use formative assessment by providing, "in-service training and the inspection frameworks to use generic AfL strategies" such as sharing learning aims with students, providing effective feedback, promoting student self assessment and peer assessments (p. 252). For accountability teacher assessments will comprise the main component in reporting student achievement which will be supported by computer-based ability tests in the core subjects of English, math and science with an emphasis on thinking skills, especially in the primary grades. Elwood concludes that Northern Ireland is in the process of improving student learning by, "stepping away from the negative consequences of one assessment system and embracing the benefits of another" (2009, p. 253).

England: Standing Fast?

Around the same time Scotland, Wales, and Northern Ireland were exploring alternative methods in the use of standardized tests England was reasserting the primacy of such tests. Although Tony Blair's Labour government had similar goals as the Thatcher and Major governments which was to improve education outcomes to address economic interests the means were different. As James (2000) explains, Thatcher was concerned with creating competition within the education system to increase test scores whereas Blair's government was concerned with raising achievement overall. Leung and Rhea-Dickens (2007) note that since 1997 special emphases have been placed on increasing literacy and numeracy as these skills were and continue to be believed to be in demand for the new millennium. They point out that although

standardized tests and teacher assessments are given ‘equal status’ in policy documents, standardized tests have been officially privileged since they are used for measurement and accountability purposes and as a way to “lever up” achievement (p. 10). They note that the government, which observed that only 63% of students were receiving a level 4 on the national curriculum tests in 1997, set an achievement target of having 80% of 11 years old obtaining level four by the year 2002. Therefore, an army of coordinators and consultants was sent forth to assist and guide schools to achieving those targets.

Leithwood, Jantzi, Earl, Watson, Levin and Fullan (2004) note that all the Local Education Authorities (LEA’s) and schools were made aware of the Strategies. School participation in the strategies was not mandatory but a carrot and stick approach to encourage schools to comply seems to have been utilized. As Leithwood et al. observe, if schools did not join,

the pre-existing national inspection process and accountability system meant that LEAs and schools would need to be confident that they could meet the performance levels without implementing the Strategies. Along with the pressure inherent in establishing targets, the DfES also provided significant extra funding to support the Strategies with materials and personnel. (2004, p. 67)

Earl, Watson and Torrence (2002) argue, after reviewing England’s literacy and numeracy strategies, that standardized tests have “jolted” educators to take steps to increase student achievement adding those, “mandates for accountability have served the utilitarian purpose of encouraging engagement where it would otherwise be absent” (p. 39). They further suggest the majority of teachers have come to welcome the strategy and “feel grateful” for the tools they

have been provided to reach the targets of achievement set by the government (p. 46). Leithwood et al. (2004) argue that England's literacy and numeracy strategy is, "one of the most ambitious and successful examples of large-scale school reform in the world to date" (p. 59). At the time they noted gains in literacy and mathematics has risen "from 53% in 1996 to 74% in 2002" as well as decreasing the gap between the most and least successful schools districts (p. 59).

Others are less enthusiastic. There is growing evidence that increases in student achievement as a result of standardized tests have run their course. Whetton (2009) suggests that up to the year 2000 a, "significant proportion of the apparent improvement in national results may have arisen from variation in test standards" (p. 150). Although he notes that there were increases in student scores they were not "as strong as the official data suggested" (p. 150). Wyse and Torrance (2009) observe that, despite the literacy and numeracy strategy, increases in scores in literacy and numeracy have slowed down considerably since 2000 and have not yet been able to reach the 80 per cent target of students achieving level 4 set by the government.

Earl, Watson and Torrance (2002) noted that although teachers were using materials provided by the Strategies they also observed that much of the gains resulted from, "teachers using their existing capacities more fully, rather than having developed substantial new skills and knowledge" (p. 38). They advised that since England was "data rich" in terms of large scale assessments the use of such data would enable "wiser" evidenced-based planning at the classroom level (p. 48). They suggested that formative assessment was being practiced in many classrooms in England, although they provided no evidence for this claim, and went on to argue that the "benefits from strengthening formative assessment as an integral part of the Strategies" that "are consistent with the aims of the Strategies" (p. 49) was needed in order to see continued

improvement.

The evident topping out of scores on the national assessments was a source of frustration for the government and this, argues Leung and Rea-Dickens (2007), resulted in the government taking more notice of formative assessment. They note that formative assessment has appeared on the government radar over the past few years “gaining credibility and momentum” (pg. 13).

The government’s chief examination officer stated in 2001 that,

most teachers were incapable of making reliable judgements about pupils’ performance a decade ago” and that the curriculum and testing reforms in the 1990s had “transformed teachers’ understanding of what is to be taught and learned and their competence to assess what pupils achieve,” and for that reason teachers should be given a bigger role in assessment. (Henry, 2001, p. 2 as cited in Leung & Rea-Dickens, 2007, p. 13)

As well Leung and Rea-Dickens (2007) note that since 2001 the Assessment Reform Group’s guidelines for formative assessment at the classroom level have been published on England’s QVC website. In 2004 England’s curriculum authority announced that teachers of students between five and seven years of age would continue to use the national tests for reading, writing and math, but that teachers’ assessments would be reported for accountability (Leung & Rea-Dickens, 2007). A pamphlet, published by the Assessment Reform Group (ARG, 2006), informs us that in 2005 the Chief Executive of England’s Qualifications and Curriculum Authority, “predicted that external summative tests for 11-year-olds and 14 year olds will eventually be replaced by moderated teacher assessment” (p. 2). Whetton (2009) notes as well that in 2005, for the first time since standardized tests were implemented, the Key Stage 1 assessments emphasized teacher assessments over standardized test scores. Although teachers still had to use

the national assessments, they could decide when to use them, although noteworthy here is that most teachers used the tests as before. This parallels the findings from Ireland and Scotland, noted earlier in this paper, that when teachers' use standardized tests to confirm their judgements the large scale assessments take prominence.

A continuing and notable point of contention remains. The issue centres on over how formative assessment was being defined in the assessment discourse. Although formative assessment was being acknowledged by the government Daugherty and Ecclestone (2006) point out teacher unions, who support formative assessment, are not pleased with how formative assessment in England has been defined by the government in policy related texts. That is to say in terms of teachers use of data provided by large scale assessment which is "in direct contradiction...to assessment for learning adopted by King's College, London and the National Assessment Reform group" (ATL, NUT and PAT, 2004, p. 4, as cited in Daugherty & Ecclestone, 2006, p. 164). They note that there was little discussion on the development of students taking responsibility for their learning which they note is central to formative assessment. Instead, focus was placed on a "teacher-led and target-dominated usage of the term" (p. 165) providing materials for schools and teachers that consisted of regular use of summative assessments, not formative assessment. Daugherty concludes that although official assessment policy in England acknowledges the research by Black and Wiliam (1998a) and personalized learning for the student there remained a, "mind set that sees target-setting by teachers and schools as the only route to higher achievement" (p. 165).

The Expanding Definition of Formative Assessment

Before continuing with the examination of formative assessment policy in England it is important to address the issue of how large scale assessments have been articulated to fit within the definition of formative assessment, not only in England, but in other nations as well. This problem fuels the debate over whether summative assessments can be utilized for the purpose of formative assessment. For instance, Taras (2007a) suggests formative assessment has been promoted in the academic literature while summative assessment has been downplayed, as evidenced in this paper. Also consistent with this paper Taras (2007a) observes that formative assessment has been presented in the literature as good assessment and summative assessment as bad assessment. Here she refers to the notion that formative assessment has been presented as supporting learning which “captures the desires of all educationalists” (p. 58). Summative assessment, on the other hand, particularly represented as large-scale assessments, has been relegated to the role of “validation and accreditation” and often accompanied with the negative educational connotations such as consequences of failure, judging, ranking, sorting, and labeling of students (p. 58).

Taras (2007b) argues however that excluding discussions on summative assessment prevents its use for formative purposes. She notes that the summative aspect of the learning process has often been, “considered absent and therefore not a necessary precondition for formative assessment” (p. 370). She argues then that formative assessment has not always been defined as having a clear beginning and a clear end and as such has been regarded as something different from summative assessment. She notes that the criterion in formative assessment is often presented as ambiguous and changeable since it is dependent upon the ‘interpretation’ of the teacher as they modify instruction based on the changing feedback from the student. Taras

(2007a) suggests that the separation between formative and summative assessment framework has, “created a dichotomy of socially and ethically charged assessment” that was unaware of the “contradictions within it” (p. 59). She points out that summative assessment can occur any time during a course of study since it is making a judgement on the work completed. The feedback to further learning based on the summative assessment is formative assessment.

In an evident contradiction to earlier statements by Wiliam noted in this paper, Wiliam and Thompson (2008) in their article *Integrating Assessment with Learning: What Will It Take to Make It Work?* suggest that time is not a critical issue in defining formative assessment. They argue that the formative assessment feedback loop can also occur on a system wide scale. They point out that large scale assessments, “can also be formative at the level of the school, district, and state, provided the assessments help to regulate learning” (p. 69). They note that the cycle of the feedback loop may occur in seconds or, as in the case of large scale assessments, may take up to a year as long as feedback from the assessment is “fed back into the system” (p. 61). Summative and formative assessment are in essence, as Taras (2007a) argues, part of the same cyclical process. She suggests that summative assessment contributes to the start of the process that represents the goal to be achieved. Formative assessment is the process that guides the student to achieving that goal.

It is necessary to note that the connection between large scale assessments as part of the formative assessment process has received wide dissemination in the literature. For example in the United States, Stiggins (2005) notes that formative assessment occurs when board, school, and teachers interpret and use assessment results that result in a culmination of resources and strategies that “guide all students toward attainment of the standards” (p. 32). Haycock (2008)

suggests that it is not that large-scale assessment prevents formative assessment but that the majority of States, “have too often essentially hidden the criteria on which they base their assessments” (p. 316) leaving teachers unsure of the standards or curriculum expectations they should be trying to have their students achieve. As Brookheart (2007) observes, the linkage between summative and formative assessment is defined by how close the relationship is between the criteria of the assessments and that both assessments must be focused on the same “attributes of quality of work” (p. 45).

According to Kennedy, Chan, Fok, Yu (2008) the Organization for Economic Cooperation and Development (OECD), which is largely an economic agency, have lent their support to formative assessment since they understand more than others the importance of learning to a nation’s economy. The OECD’s Centre for Educational Research and Innovation (CERI, 2005) in their series called “What Works” provide some literature reviews and “exemplary” case studies from a number of countries on what they see as the efficacy of formative assessment and note its great success. The OECD argues, based on the evidence provided by CERI, that the conclusions by Black and Wiliam (1998b) that the, “achievement gains associated with formative assessment have been described as ‘among the largest ever reported for educational interventions’ (p. 2). They continue that formative assessment “improves equity of student outcomes” and that schools implementing formative assessment, “show not only general gains in academic achievement, but also particularly high gains for previously underachieving students” (p. 2).

CERI (2005) suggests that the evidence they provide illustrates how formative assessment has been successfully adopted in context with standardized tests and/or standards-based reforms.

They do acknowledge that the “high visibility summative assessments are a significant barrier to formative practice” (p. 6) but go on to suggest that formative assessment is compatible with standardized testing and standards stating that,

in addressing tensions, and to ensure stronger validity and reliability of summative assessments, policy officials should consider multiple measures of student progress.

Ensure classroom, school and system level evaluations are linked and are used formatively to shape improvements at every level of the system. Policies that link a range of well-aligned and thoughtfully developed assessments at the classroom, school and system levels will provide stakeholders with a better idea as to whether and to what extent they are achieving objectives. Formative assessment, when applied at each level of the system, means that all education stakeholders are using assessment for learning. (p. 6)

Sebba (2006) notes that the OECD definition of formative assessment “arguably falls outside the ARG definition which places greater emphasis on the use to be made by learners of assessment information” rather than stressing “the adjusting of teaching in light of the assessment” (p. 188).

Yet an examination of the evidence provided by CERI appears wanting in terms of supporting their claims. The literature reviews or case studies they provide offer little in the way as adding support to their claims. The references for the French literature review, provided by CERI, are all pre 1998 and conclude, “the theoretical promise of French-language work on formative assessment is in need of considerably more empirical grounding. This is a major challenge for the researchers of this community in the coming decades” (Allal & Lopez, 2005, p. 256). Most of the references in the German literature also date pre 1998, although a few references are provided from 2002 and 2003. The German review is just as inconclusive as the

French review concluding that;

the previous sections of this literature report have shown that there is not very much German research on effects of formative assessment on educational outcomes. This is surprising to some extent, because there are many approaches of formative assessment described in the German literature. These approaches have not been sufficiently evaluated. (Koell, 2005, p. 275)

A review of some the case studies, at the elementary school level, is provided by Sebba (2006). She particularly focuses on Queensland but also gives notice to some Canadian provinces, Denmark and New Zealand. The limitations of the findings are acknowledged by Sebba given that they are based on the, “jetting in of experts to other countries for short periods” (p. 187) to a specific school and therefore can not be considered representative of that particular district, province, territory or country.

Interestingly Sebba (2006) notes that an important contextual variable in Queensland is a lack of large scale assessments for cohorts of students and a less prescriptive curriculum than England that may enable formative assessment as it reduces what teachers are to teach. Sebba (2006) reports that in two schools in Queensland some formative assessment practices have been successfully established. The evidence for the successful application of formative assessment was largely derived from student and teacher self reports. Formative assessment practices utilized were in the nature of student reflection journals, self and peer assessment, or teachers enquiring as to whether everyone understood enough of the lesson to move forward. It was noted however that the peer assessment was of an evaluative nature. The students marked their peer’s work as being correct or incorrect rather than providing feedback on how to improve the work. The

formative assessment practiced was nevertheless reported to be successful by the teachers and the students. The promises of formative assessment such as student acquisition of self autonomy, increased motivation, or quantitative measures of increases in student achievement as a result of the formative assessment practices adopted were not reported. Sebba reports that any influences from the lack of external tests was not mentioned by teachers in either supporting or inhibiting their formative assessments practices. Although it is hard to imagine how the teachers could comment on what they have not experienced in terms of the influences of large scale assessments on their teaching practices.

Sebba (2006) states that some schools in Newfoundland were able to implement 'successful' formative assessment initiatives based on the feedback from provincial assessments. She suggests that the provincial assessments, which met with considerable resistance when first implemented, have come to be recognized by school boards, schools, and teachers as providing important data to improve instruction. Sebba suggests that this data became "a key focus of staff development activities and practices, closely linked to evaluation by teachers with individual students" and are in line with formative assessment practices in that they "promote a commitment to a shared understanding of the criteria by which students are assessed" (p. 196). There are few explicit examples offered of formative assessment occurring in the classroom beyond the use of large scale assessments to alter instruction, or how the student the student is involved in the process. Overall, Sebba suggests that there is little evidence in the case studies she examined to suggest that summative assessments limit developments in formative assessment. Of course it has to be pointed out that the definition of formative assessment observed in this study is somewhat removed from the one provided in the theoretical framework.

As interpreted in this work, the teacher-student relationship is vital to the success of formative assessment and one where the student is a key actor in the process.

Sebba's (2006) conclusion is mirrored by Kennedy, Chan, Fok, and Wu (2007) who observed that feedback from the Basic Competency Assessment for Primary 3 and Primary 6 was also being used by teachers to inform instruction in Hong Kong as well as in New South Wales in Australia. Kennedy et al. suggest that in these cases, the summative assessments are not used for ranking schools or creating league tables as in England, but are used to further student learning in altering teaching practices. Although it is important to recall that the uses of large scale assessments in England are also purported to serve as an aid to instruction. Nevertheless, the authors conclude that their studies illustrate how, "summative assessments are used for formative purposes" (p. 203). Unfortunately, the authors provide little evidence on how such feedback has successfully facilitated formative assessment in the classroom or involves individual students.

It is hard to determine how the above 'evidence' on the use of feedback provided by large scale assessments provided by Kennedy et al. and CERI, or standards-based reform in the case of Queensland, have promoted the successful implementation of formative assessment at the classroom level or support Black and Wiliam's (1998) claim that formative assessment has lead to the largest achievement gains of any educational intervention. How CERI can make such claims based on the evidence they provide is a mystery. Interestingly, Brown, Lake and Matters (2009) note that Queensland has recently implemented large scale testing of literacy and numeracy for grades 3, 5, 7 and 9 at the behest of the government and for grades three and six in the New Basics trial. They expect that,

emphases in teaching will shift from nurturing and developmental perspectives towards a more directive, transmission conception of teaching. This shift comes about because of rational responses towards external accountability inspection: whether made public or not, the inspection of teacher work by external agencies triggers a desire to be seen to be doing well. (p. 63)

On a conceptual level there are perhaps potential links between large scale assessments and formative assessment in the classroom. As noted above, for summative assessments to be used as formative assessment in the classroom they need to be based on the same criteria and standards. Yet the potential benefits of large scale formative assessment can be seen as limited in terms of the use of feedback such tests provide. As Wiliam's (2008) points out, the largest learning gain found in the research on formative assessment occurs when feedback is provided within in "minutes, hours, or days" (p. 37) rather than in feedback that is provided in weeks or months later as is the case of large scale assessments.

Buhagair (2007) argues that 'if' large scale tests are being used by teachers to adjust their teaching practices in the future this change can be seen to be formative for the teachers but not for the students taking the tests. He stresses that formative assessment is about addressing how well the teacher is teaching and how effectively students are learning concluding that this can only be done in the classroom in the interaction between the teacher as represented in the theoretical framework of this paper. Through classroom assessment, teachers get continual feedback on how well students are learning what the teacher is trying to achieve. Then students are required, through a variety of classroom assessment exercises, to monitor their learning, to reflect on it, and to take corrective action while there is still time left in the semester to do so.

Similarly Shepard (2008) suggests that within the definition of formative assessment there is the understanding that formative assessment occurs within the classroom. She suggests that, “what makes formative assessment formative is that it is immediately used to make adjustments so as to form new learning” (p. 281). As noted, large scale assessments are not usually available for the immediate use for the classroom teacher to alter their teaching strategies before the conclusion of the teaching term.

Perhaps even more importantly, as the evidence presented thus far in this paper suggests, classroom formative assessment is inhibited by large scale assessments. The dominating influence of large scale assessments and how they inhibit classroom formative assessment becomes clearer when looking at how England’s literacy and numeracy strategies affect instructional strategies within the classroom in context to large scale assessments.

How the Strategies Undermine Formative Assessment

England’s National Literacy Strategy (NLS) describes successful teaching as “discursive, characterised by high quality oral work and interactive, encouraging, expecting and extending pupils’ contributions” (DfEE, 1998, p. 8, as cited in Smith, Wall, & Mroz, 2003, p. 396). Similarly the NLS directs that “high-quality direct teaching is oral, interactive and lively... in which pupils are expected to play an active part by answering questions, contributing points to discussion, and explaining and demonstrating their methods to the class” (DfEE, 1999a, p. 11, as cited in Smith et al. p. 396). Smith et al. (2003) point out that both directives represent a move away from conventional teacher directed lessons to a process of interaction between the students and the teacher. The processes of eliciting student responses and the attention placed on teacher

student interaction that makes the student an active partner in learning are in line with formative assessment practices.

Smith, Wall, and Mroz (2003) conducted a study of 72 primary teachers across England that comprised 35 literacy and 37 numeracy classrooms and examined their teaching practice. Teachers did indeed report that they frequently invited students to elaborate their responses as directed by the Strategies. Yet classroom observations revealed this was not the case. The authors report that more conventional teaching interaction took place with teachers asking more closed ended questions that allowed students limited opportunities for them to elaborate their ideas. They agree with Earl et al. that “changing such pedagogic understanding and practices remain a major challenge in securing the long-term effectiveness of the strategies” (p. 409).

Hardman, Smith and Wall’s (2003) study observed 70 teachers during the literacy and numeracy hour and found, like the study above, that “the NLS is encouraging teachers to use more directive forms of teaching with little opportunities for pupils to explore and elaborate on their ideas” (p. 198). This finding was consistent even when teachers worked with students in small groups. Overall they found that their study supported earlier findings which suggest “that since the introduction of the NLS there has been an increase of whole class teaching which is dominated by teacher-led recitation” (p. 212). They conclude that top-down initiatives such as the NLS has left “deeper levels of pedagogy untouched” (p. 213) due to the processes of teacher adaptation to policy initiatives.

But whether the Strategies actually encourage pedagogical change or simply reinforces conventional practice is open to debate. Formative assessment that is “consistent with the aims of the strategies” as Earl, Watson, and Torrance (2002, p. 49) mentioned earlier in this paper,

reveals a contextual limitation for the implementation of formative assessment practices in context to large scale assessments. For example, the evident contradiction between what the Strategies encourage and what occurs in the classroom may be found in the observations by English, Hargreaves and Hislam (2002). They argue that the NLS offers teachers conflicting advice. They note that although the Strategies, as previously noted, direct for increased pupil interactions they also direct that lessons should be, “well-paced” with “a sense of urgency, driven by the need to make progress” (DfEE, 1998a, p. 8, as cited in English et al. p.12). This contradiction was evident in their classroom observations and case studies where the number of students contributing increased but extended engagement was shortened. The Strategies “urgent” need to “make progress” appears more related to achieving results on standardized tests, which the system uses to mark student progress, rather than encouraging formative assessment. Certainly on a conceptual level formative assessment is not associated in any sense as being executed with urgency.

A massive four-year study on formative assessment was conducted by Britain’s Economic and Social Research Council (ESRC) Learning How to Learn (LHTL) - in Classrooms Schools and Networks Project between 2001 and 2004 (James, Black, McCormick, Pedder, & Wiliam, 2006). This project aimed to examine how formative assessment could be developed as an effective tool to assist students learning in classrooms and across schools by learning more about what effective learning looked like in the classroom (James et al. 2006). The program also attempted to address the rhetoric of lifelong learning as “the phrase ‘learning how to learn’ shifts attention to the learner and provides the ground for the focus on learner autonomy that emerged as a central theme in the research” (p. 104). The study included 40 schools in five local

authorities involving 1,500 teachers and 20,000 students. Data was collected by a series of teacher and student questionnaires and video recordings of lessons.

Drawing on preliminary data from the ESRC-LHTL study Skidmore, Perez-Parent, and Arnfield (2003) used audio recordings of teacher and student interactions in year six classrooms in five schools during the guiding reading part of the literacy hour as directed by the NLS. Their analysis found, in support of the previously noted studies, that teachers dominated the discussion and asked closed ended or test type questions where the teacher already knew the answer. They observe it is the nature of such questions to limit student responses to a few words. Therefore the “space for pupils to articulate and develop their own ideas” on what they have read was limited (p. 52).

Marshall and Drummond’s (2006) analysis of 27 lessons that were also video taped in the ESRC-LHTL study found that almost all the teachers implemented some form formative assessment. However only 20 per cent did so in a way that Marshall and Drummond termed “the spirit” of formative assessment where the underlying principle is based on “a high organization of ideas, promoting student autonomy” (p. 137). Student autonomy in the study is roughly defined as an “expression within the set tasks” (p. 144) rather than genuine student autonomy where students would set their own tasks. This is more clearly illustrated by a common pattern which emerged where ‘the spirit’ of formative assessment was observed. Through a series of sequential activities students framed and negotiated what was to be learned by examining a piece of work and used this information to reassess their own work based on the class discussions. They note, “there is a sense in which, when the pupils reconsider their own work, they are doing so in the light of the collective wisdom of the class” (p. 140). It should be noted however that

successful designs to measure student autonomy quantitatively have yet to be created (Black, McCormick, James, & Peddar, 2006).

The remaining teachers attended to the ‘the letter’ of formative assessment. This is defined as being when teachers follow a rigid set of procedures associated with formative assessment such as implementing group work, increasing wait times between their questions and student responses or the sharing of criteria with students but class discussions were largely controlled and directed by the teacher rather than providing for a free exchange of ideas among the students. The concept of ‘the spirit’ and ‘the letter’ of formative assessment is similar to the notion of ‘convergent’ and ‘divergent’ formative assessment articulated by Torrance and Pryor (2001) noted in the theoretical framework of this paper. The authors, commenting on their earlier work, where they investigated teachers use of formative assessment in primary classroom in England, found that most teachers applied convergent rather than divergent formative assessment.

Wyse and Torrance (2009) suggest that teachers “favour formative, provisional and implicit assessment,” which they argue, is in contrast to the demands of the national Strategies (p. 225). Clearly the national strategies support ‘convergent’ or teaching to ‘the letter’ of formative assessment as indicated in the research noted above. Wyse and Torrance go on to suggest that the demands of the literacy and numeracy strategies distort teachers classroom practices whereby teachers learned very rapidly how to coach and prepare their students for the tests, and that “any benefit to be squeezed from the system...has long been exhausted” (p. 225). They conclude that when teachers coach students to write tests the quality of education the students receive is reduced.

Yet the amount of distortion that has occurred as a result of the Strategies is open to interpretation. The question that arises is what could explain the differences between teachers who implemented ‘the spirit’ of formative assessment and those who implemented ‘the letter’ of formative assessment. Interestingly Marshall and Drummond (2006) note that data from the LHTL Staff Questionnaire revealed quantitative data that suggested teachers negotiated the expectations and pressures from large scale tests in different ways. Marshall and Drummond suggest that the teachers who attended to ‘the spirit’ of formative assessment held parallel pre-existing beliefs about teaching, learning and student autonomy in their classrooms. The authors note that the teachers who implemented the ‘spirit’ of formative assessment viewed it as an “essentially progressive, rather than fixed, view of what went on in any given lesson” (p. 147). As well they also possessed a greater amount of personal agency and responsibility for the success and failure “in the promotion of pupil autonomy” (p. 133). These previously held beliefs, their pedagogy, influenced the way these teachers implement formative assessment in the classroom. For the other teachers, who were in the majority in the study, it can be argued that their pedagogy was more in line with conventional teacher directed lessons and corresponds with the directives of the Strategies. The issue of teacher pedagogy is an important issue that will be addressed further on in this paper.

It is also important to note as well that focus on criteria and targets identified in the Strategies may lead to instrumentalism. Torrance (2007) notes, in reference to post secondary competence-based assessments, that strict focuses on assessment criteria and setting clear objectives which are too explicit or transparent,

encourages instrumentalism. The clearer the task of how to achieve a grade or award

becomes, and the more detailed the assistance given by tutors, supervisors and assessors, the more likely are candidates to succeed. But transparency of objectives coupled with extensive use of coaching and practice to help learners meet them is in danger of removing the challenge of learning and reducing the quality and validity of outcomes achieved. This might be characterized as a move from assessment of learning, through the currently popular idea of assessment for learning, to assessment as learning, where assessment procedures and practices come completely to dominate the learning experience, and ‘criteria compliance’ comes to replace ‘learning’. (p. 282)

Teaching to ‘the letter’ of formative assessment or ‘convergent’ formative assessment and the dangers of instrumentalism, provide useful heuristics in identifying the problem of when formative assessment is “consistent with the aims of the strategies” argued by Earl et al. (2002) earlier in this paper. Britain’s Literacy and Numeracy Strategies, as the evidence strongly indicates, promotes teacher controlled and target focused instruction to boost marks which is at odds with formative assessment that engages the student as a partner in the learning process.

Modification of large scale assessment in England continues. On October 2008, Secretary of State for Schools Ed Balls suddenly announced that the key stage 3 will be cancelled immediately (Whetton, 2009, 154). However despite this, and the other examples of the reduction of large scale assessments in England, in what can be considered a softening or modification of England’s policy regarding large scale assessments Schools Minister Jim Knight reaffirmed the government’s commitment to standardized tests stating that the assessments, are there to give pupils an understanding of how they're doing nationally, to give parents the opportunity to see how well their child is doing and how well the school is doing, and

for the public to see how well schools generally and how the school system as a whole is performing. (BBC News, 2008)

Nevertheless, perhaps sensing some vulnerability on the part of the government the National Union of Teachers (NUT), which represents most of the teachers in England, and the National Association of Head Teachers (NAHT) are planning to boycott the keystage 2 tests in 2010 (BBC News, 2009). They argue that the test encourages teaching to the test and have nothing to do with learning.

The Lessons From Great Britain

Up to this point, as evidenced in the above discussion on Britain's experience with formative assessment, I have argued that formative assessment has not been able to take root where large scale tests for accountability have been present. Importantly, even 'if' formative assessment is advocated in policy documents the mere presence of large scale tests for accountability is enough, evidently, to discourage its practice by encouraging more transmission based instructional strategies in efforts to raise student scores on such assessments. This is clearly illustrated in the case of England's literacy and numeracy strategies. Ironically, although teachers support formative assessment their response to large scale assessment seems to be a natural adaptation of conventional teacher strategies rather than a distortion.

It appears that Britain's emphasis on standardized tests is experiencing a modification in usage whereby formally mandated yearly tests are disappearing from the assessment landscape. Part of this renewed attention on formative assessment appears to be based on the limits to achievement with standardized tests for accountability. Formative assessment promises, based

largely on Black and Wiliam's (1998) literature review, to produce higher results and can be seen therefore as an incentive for governments in Britain to give it a try. Though large scale assessments may, in some instances, be disappearing from public view they are nevertheless still offered as a tool for teachers within the system to gauge student learning and to provide teachers a means of mediation. It remains to be seen whether or not these new arrangements will encourage 'the spirit' or more 'divergent' forms of formative assessment. Whetton (2009) observes governments in Britain are faced with the dilemma of maintaining the continuing goal for improvement in student achievement as well as maintaining control and accountability. He notes part of the problem stems from the validity of the purpose of the large scale assessments which remains rather obtuse. He concludes that although the winds for change are in the air uncertainty remains over whether any alternative system will necessarily be an improvement and bring with it, its own, unintentional consequences. Such unintentional consequences will be addressed further in this paper.

As I note in the following chapter, Ontario's proposed formative assessment policy raises the same issues and dilemmas regarding large scale tests and accountability that are occurring in Britain. I would suggest though that one notable difference is that Ontario has yet to acknowledge the difficulties in placing the new proposed formative assessment policy within the current student achievement framework which, as I will illustrate, is almost identical to those found in Britain.

Chapter 4

Ontario's Formative Assessment Policy

The purpose for Ontario's proposed formative assessment policy stems from the Ministry's recognition of, "an urgent need to clarify and consolidate, to ensure that policy is aligned, consistent, and clear, and that every student in the system is benefitting from the same high-quality process" (Ontario Ministry of Education, 2008, p. 1). Ontario's proposed formative assessment policy states the,

primary purpose of assessment and evaluation is to improve student learning. Information gathered through assessment helps teachers to determine students' strengths and weaknesses in their achievement of the curriculum expectations in each subject/course in each grade. This information also serves to guide teachers in adapting curriculum and instructional approaches to students' needs and in assessing the overall effectiveness of programs and classroom practices. Assessment is the process of gathering information from a variety of sources (including assignments, day-to-day observations, conversations or conferences, demonstrations, projects, performances, and tests) that accurately reflects how well a student is achieving the curriculum expectations in a subject/course. As part of assessment, teachers provide students with descriptive feedback that guides their efforts towards improvement. Evaluation refers to the process of judging the quality of student work on the basis of established criteria, and assigning a value to represent that quality. (Ontario Ministry of Education, 2008, p. 2)

Included in the proposed policy document is the recognition of the development of students' self-regulatory meta-cognition skills such as self-reflection and self-monitoring where the students become in control of their own learning. Here students come to know when to seek help in understanding what they do not know. Clearly this policy is well aligned to an accepted definition of formative assessment theory and consistent with the theoretical framework outlined in this paper. The document goes on to provide illustrations and examples of formative assessment in the classroom along with guiding principles and recommendations for practice in the classroom.

The Ontario Ministry of Education (2008) further states that;

our policies in Ontario for assessment, evaluation, and reporting are well aligned with the best thinking of experts in this field, and current with international policies and research. Research has shown that the most effective type of assessment for improving student learning is formative assessment, or 'assessment for learning'. (p. 2)

It is puzzling however as to what specific research studies the Ministry is referring to although they do cite Lorna Earl. Earl (2006) primarily relies on Black and Wiliam's (1998) study *Assessment and Classroom Learning* for evidence supporting formative assessment. As illustrated in this paper thus far the research on formative assessment, including international dissemination, is extensive but has not moved much beyond that of 1987 and largely consists of qualitative reports from interested volunteers that may or may not be an accurate reflection of their formative assessment practices. It is also puzzling as to what specific international policies the ministry claims to be "current" with. Certainly, as far as international policies illustrate the jury is still out on the matter with the exception of the claim that standardized tests have

evidently impeded the development of formative assessment in the classroom. Yet it is in this context, that of standardized tests, in which Ontario's proposed formative assessment policy is being placed.

Ontario's Large Scale Assessments

As was the case in Britain debates also emerged in Canada on how well the education system was functioning. Earl (1995) notes that a "climate of concern" about the quality of education began to take hold and that accountability was the watchword in the 1990's, as was the case in the 1980's in Britain (p. 45). The New Democratic Party's (NDP) educational-policy agenda focused on improving equity in Ontario education as well as ensuring economic competitiveness (Paquette, 2001). In 1993, Bob Rae leader of the NDP in Ontario established the Royal Commission on Learning (RCL).

The findings of the RCL (1995) were similar to those articulated by British Prime Minister James Callaghan's speech in 1976, noted earlier in this paper, which cited issues regarding the lack of public confidence and accountability, and to some degree political faith, in the education system. The RCL concluded that the fear that schools were failing in some manner were unfounded stating that, "frankly, we find this fear exaggerated. While the status quo is unquestionably flawed, there is no serious evidence that our schools are failing our kids any more or less than they ever have" (Short Version, What can we expect from our schools? p. 1). Yet the report also acknowledged that there was, "widespread unease that schools have become a kingdom unto themselves, with little need to report to parents or to the world at large what they are doing with our kids, and whether they're doing it successfully" (Short Version, What can we

expect from our schools? p. 1). To address the issue of public accountability the commission noted that,

one of the newfangled theories that's gained some currency in recent years is known as 'outcome-based education'. While somewhat more intricate and controversial than you'd expect, we use it here to mean that schools describe what students are expected to know when they graduate, and then rigorously assess their success in doing so. As one long-time Canadian educator commented, 'This seems so eminently sensible that there must be something wrong with it'. (RCL, Short Version, Assessment: How we know what students learn, p. 1)

Although the commission was introducing and recommending some form of standardized tests, especially in literacy and numeracy they also added that such tests were limited with respect to the quality of learning that occurred noting that, "testing for real understanding, for a student's capacity to think and reason, takes far more sophistication than [*sic*] this; in the trade, it's called authentic assessment, and mostly it must be done by teachers" (RCOL, Short Version, Assessment: How we know what students learn, p. 1).

In response to public concerns over school effectiveness, and on the recommendations of RCL, the Education Quality and Accountability Office (EQAO) was created in 1995 and then supported by different provincial governments in varying degrees. Volante (2007) points out that it was "conceptualized under the tenure of NDP, was created and funded by PC's, and is now operating in partnership with the current Liberal government" (p. 4). The stated objective of the EQAO is, "to ensure greater accountability and contribute to the enhancement of the quality of education in Ontario" (Volante, 2007, p. 4). The implementation of the provincial standard

means that, “it is no longer acceptable for children not to achieve” a level of proficiency mandated by the Education Ministry (Ellis, Lamoureux, Awender, Wessel, & Donohoo, 2008, p. 26).

As was the case in Britain, the implementations of standardized tests have been related to political ideology. It is argued that the focus on large-scale assessments represents a neo-liberal agenda serving the interests of the corporate world that undermines democracy to the detriment of the entire educational system (Ricci, 2004). Moll (2004) argues that large scale provincial assessments in Canada are designed to “catalogue, classify, and rank” the data the assessments provide and do not lead to the improvement of learning (p. 13). She suggests that such assessments represent a shift in power from the classroom teacher to government institutions. Others note that the publication of the EQAO results, particularly by the Fraser Institute, that ranks school performance based entirely on the test results, is really an attempt to provide evidence on the necessity for school privatization (Froese-Germain, 2006).

It should be pointed out here that the EQAO state that their assessment results cannot be used to rank schools considering the different socioeconomic characteristics of each school (EQAO, 2009). The EQAO also provides the Education Quality Indicators Framework which puts student, school and board assessment results into appropriate contexts (Volante, 2006; EQAO, 2009). McAdie and Dawson (2006) argue that teachers do not need the EQAO assessments in order to gather this information to improve teaching. They suggest that teachers have the capacity to use the data they collect in the classroom every day to implement formative assessment. They note that although, “assessment for learning is valuable for teachers” they go on to suggest that, “we do not need large-scale assessments...to be able to learn from

assessments” (p. 32) echoing the arguments noted earlier in Britain.

Like governments in Britain, the position of the EQAO and the Ministry of Education is that the data the EQAO assessments provide will enable teachers to improve instruction (EQAO, 2009). Earl and Torrance (2000) contend that the task of the EQAO was to develop and implement a large-scale assessment system that provided data for accountability and that contributed to improved teaching and learning. The CEO of the EQAO, Joan Green, expressed in the first Provincial Report on Achievement that the,

EQAO believes that large-scale assessment can contribute to positive educational change when it engages educators, parents and students in thought and discussion about what takes place in the classroom. We are testing what children know and can do. We are measuring against provincial standards to obtain information on what we need to do so that all students can learn more and learn better. (EQAO, 1997, p. 3, as cited in Earl & Torrance, 2000, p. 118)

Earl and Torrance (2000) also note that the EQAO assessments would be the basis on which to establish “action plans” (p. 117) and therefore “intentionally focus their efforts on how the assessment could influence teaching and learning” (p. 120). They go on to claim that the, “EQAO had the opportunity to study the experiences and the research from other jurisdictions with a longer history of large-scale assessment before making decisions about the nature of their assessment program” (p. 116). Although Earl and Torrance do not identify the other jurisdictions the reasoning and purpose for the EQAO assessments and the education communities expected response to those assessments are similar to, if not exactly the same as, those found in Britain.

Ontario's Student Achievement Framework

Ontario's formative assessment policy can be seen as part of a broader strategy for improving student achievement supporting established policies such as the revision of the Ontario Curriculum, the EQAO assessments and the Literacy Numeracy Strategy (Levin, 2007). Levin, Glaze, and Fullan (2008) note that Ontario's student improvement initiative, "recognizes schools as ecologies, so it gives attention to building capacity in teachers, improving leadership, involving parents, changing policies, adding resources - all at the same time" (p. 277). The overall structure of school reform in Ontario is described as building on "Fullan's (2006) 'trilevel solution', in which governments, school districts and schools work together on common approaches and strategies. An explicit part of the strategy involves building strong relationships and close connections with boards, schools and other organizations" (Levin, 2007, p. 330). Levin (2007) suggests that Ontario's overall student achievement initiative has been a collaborative approach which has, "in place several new mechanisms for consultation with partners on virtually all programs and policies" (p. 330). For example he points out there is a 'Partnership Table' where the Minister of Education meets with teachers, principals, and school superintendents on a regular basis. Levin also notes that parents and student organizations are also included in the development of policy and its implementation.

Building Teacher Capacity

Levin (2007) points out that the foundation of school improvement rests, "in changing teaching and learning practices in thousands and thousands of classrooms, and this requires focused and sustained effort by all parts of the education system and its partners" (p. 324). The

strategy for change includes, “job-embedded capacity-building led by respected Ontario educators as well as experts from other places,” and the sharing of good practices by the development of learning communities (p. 329). Levin notes that personnel from the Literacy and Numeracy Secretariat work side by side with school district leaders ensuring provincial and board strategies are in alignment. He suggests this is a “collaborative, not a top-down, approach to coherence” (p. 330).

Like Britain, the response to the EQAO assessments, and the heart of the student achievement framework, is Ontario’s Literacy and Numeracy strategy introduced in 2005 (Levin, 2007, p. 332). Levin and Campbell (2009) inform that Ontario’s literacy and numeracy strategy is extensive, comprehensive with an abundance of resources. Klinger et al. (2007) note that elementary schools that have continually low scores on the assessments receive extra funding or resources to increase their scores. Levin (2007) claims such resources provide teachers with “positive pressure” and “increases expectations” (p. 335) and, “furnishes data on an ongoing basis connected to further reform” (p. 335) but avoids comparing schools. As in Britain the strategy is evidently focused on achieving results quickly. Avis Glaze, Ontario’s first Chief Student Achievement Officer and the founding CEO of the literacy and numeracy secretariat, asserts the Strategies are, “committed to taking action with a sense of urgency, on behalf of the children in this province” (2005, p. 38). Levin (2007) also adds that two additional professional development days for teachers have been added and professional development for principals has been expanded as well, noting that the principal’s leadership role is vitally important for improving student achievement at the school and classroom level.

Levin, Glaze, and Fullan (2008) suggest that the framework, while not perfect, represents

a departure from education reforms in other countries since it, “embodies vital principles, grounded in research, that are associated with meaningful and sustainable change” (p. 274). Again, the research Levin et al. are referring to is not identified. Once again though, it is significant to recall that Levin described England’s literacy and numeracy strategy as, “one of the most ambitious and successful examples of large-scale school reform in the world” in 2004 (Leithwood et al. 2004, p. 59). It is also unclear as to what they regard as “meaningful and sustainable change”. Certainly the evidence suggests that scores on large scale assessments eventually taper off. It is therefore somewhat premature to suggest that Ontario’s literacy and numeracy strategies and measures of student success as indicated by increases in EQAO scores will not experience a leveling off in achievement as found in Britain or elsewhere, let alone signify that any real learning has occurred as a result.

What Data Do the EQAO Assessments Provide

There is a continuing debate over the meaning of EQAO assessment scores and the usefulness of the assessments as an instructional aid, as seen in Britain over their large scale assessments. Volante (2007) notes concerns raised by some teachers and others in the education field suggest that differences in yearly scores cannot be used to gauge achievement since they may simply be, “an artifact of variations in test difficulty, scoring procedures, or data analysis procedures” (p. 6). He concludes that serious doubts remain regarding the psychometric parameters of EQAO assessments adding that they need to be scrutinized more carefully “before an argument can be made for authentic student, school, and/or district improvements in the domains of literacy and numeracy” (p. 11).

The EQAO contends that their assessments can in no way be seen as an overall picture of how a student is doing (EQAO Connects, 2008, p. 1), echoing policy texts in Britain noted earlier in this paper. They suggest that the results do however, “provide a snapshot of a child’s achievement, which is an indicator of whether the child has acquired the expected knowledge and skills” (p. 1). The EQAO states that the assessments can be seen as, “an indicator of achievement much as blood pressure is an indicator of health” (p. 1). The EQAO informs that just as a doctor would not offer a complete diagnosis simply based on a patient’s blood pressure, the EQAO assessments cannot provide a complete diagnosis for a student. Whether or not this analogy is appropriate they suggest that the assessments, as an indicator, provide “a good starting point for further analysis” (p. 1). Here it is noted that the assessments together with the student’s school and classroom assessments, teacher observations and student characteristics, provide for a more complete overall interpretation for how a student is doing (EQAO Connects, 2008). As noted in this paper, the EQAO assessments are held up as a feedback device as well as providing accountability (Earl & Torrance, 2000). What needs to be examined therefore is how Ontario’s existing student achievement framework facilitates or prohibits formative assessment practices in the classroom.

Teaching to the Test

As in Britain concern has been voiced that large-scale assessments are narrowing the curriculum since teachers are responsible for teaching the entire curriculum but assessed only on certain portions (McAdie & Dawson, 2006). This encourages, as Ungerleider (2006) suggests, “a resentment in teachers and discourages instruction in those subjects and domains not represented

on the assessment” (p. 876). As a result some teachers have argued that they are being forced to spend “disproportionate amounts of time on tested subjects” (Volante, 2007, p. 8). Levin (2007) stresses that although the foci of the assessments are on literacy and numeracy in elementary schools they are, “complemented by strong support for other curricular areas such as physical activity and the arts, both of which have been expanded in the last three years” adding that this “strategy explicitly rejects narrow views of teaching and curriculum” (p. 329). It is unclear however how this statement provides evidence of an expanded view of teaching and learning.

Levin (2007) notes that, “Ontario teachers and Ontario policy are in agreement that a ‘test preparation’ approach is both inappropriate and ineffective” (p. 331). This of course contradicts the intentions of having the EQAO assessments in the first place. Earl and Torrence (2000) argued that, “if teachers were going to ‘teach to the test’, the assessment units were designed to be worth emulating” (p. 120). They argue that the assessments require complex thinking and applying knowledge in new situations. They also believe the tasks are “curriculum-based and standards-referenced” and include different response formats but with the majority of questions being open ended requiring extended responses. The assessment therefore, they argue, provides teachers with “concrete images of good practice” (p. 21). Similar sentiments are echoed by Levinson (2000) who argues that the practice of ‘teaching to the test’ is filled with dubious connotations but when the assessment is aligned to the curriculum this practice can be seen as a desired strategy. Pervin (2005) also suggests that teaching to the test is totally appropriate. He points out that the transition period from kindergarten to grade seven is of critical importance. He notes that it is here where the transition from ‘learning to read’ to ‘reading for learning’ takes place particularly at the grade three level. He stresses that if students have not made this

transition “then their further learning in all other subjects will be compromised” (p. 39).

Volante (2004) argues that teaching to the test is a major concern for large scale assessments that are norm referenced rather than criterion referenced. In norm referenced assessments teachers may feel obligated to teach to the test which creates conflict over content. In an evident contradiction to previous comments, noted earlier in this paper, Volante (2004) states that, “since its inception in 1997, Ontario’s EQAO has ensured that teachers are principally responsible for the development, administration and scoring” of the standards-based assessments adding that, “other jurisdictions would be wise to follow the lead set by this organization so that standardized tests are more closely aligned with prescribed provincial curriculum” (2004). Ellis, Lamoureux, Awender, Wessel, and Donohoo (2008) suggest that for students to be successful on the EQAO assessments they would need, “ample experience with open-ended questions” adding that, “the thinking processes of the students are evaluated as much as the ‘correct’ answer” (p. 27). The close alignment of the EQAO assessments to the content and achievement expectations outlined in the Ontario curriculum then is argued to suggest that the feedback the EQAO assessments provide may be used to influence teacher instruction, which once again, mirrors the arguments put forth by the British government and their large scale assessments.

Classroom Responses to the EQAO Results

Volante (2006) points out that there is evidence suggesting that the EQAO assessments have been a benefit to elementary teachers for improving planning and classroom practices. He suggests that the EQAO assessments for grade three and six provide schools with the capacity to alter their classroom practices with well-established support structures stating that,

the EQAO provides schools with individual profiles that clearly explaining students' assessment results in relation to provincial standards. These profiles also provide a strategy for teachers to use exemplars to talk to parents and their child about how the assessment information fits with the provincial curriculum expectations as well as other information about the child. Schools and districts are required to prepare reports based on material provided by EQAO along with other material related to important factors such as demographics. These school and district reports include interpretations of the results and an action plan for improving student learning. Collectively, this process seems better able to improve the utilization of large-scale assessment data than simply providing numerical scores on discrete categories. (p. 11)

Earl and Torrence (2000) note that more than 75% of teachers in Ontario had increased their participation in district professional development programs in reading, writing and mathematics that were associated with the grade three and six assessments. Furthermore, more than 98 per cent of the 12,000 teachers who participated in the marking of the EQAO assessments described it as, "among the best professional development experiences of their careers" and influenced their instruction and curriculum planning as a result of that experience (Green, 1998, p. 9).

There are limited studies clearly demonstrating how system wide response to the EQAO assessments unfolds in the Ontario classroom. Volante (2007) notes the research literature on how large scale tests impact students and teachers and the school system in Ontario and Canada is largely an unexplored territory. Wideman (2002) reports on an action research study conducted in 1999/2000 partly funded by the EQAO office. The study included 25 elementary teachers who

volunteered to be involved in the study. The teachers were divided into groups of two and four, to lend support and feedback for each other, from ten schools in two school boards. Since the schools selected were very small the teachers involved in the study represented the total number of teachers for grade three or grade six in each school. Two consultants provided information on *action research* and a one day workshop was presented on *feedback/corrective action*. Action research was defined in this study as one of enquiry where teachers implement a particular strategy and then reflect and discuss how to improve their teaching. Feedback/corrective action was defined as a cyclical process of sharing learning expectations with the students, teaching the expectations and then testing the students. Teachers then analyzed the results and provided feedback to the students on how to improve utilizing strategies such as modeling, the use of rubrics, exemplars, and student conferences. The feedback/corrective action described parallels formative assessment as defined within the theoretical framework of this paper.

As a starting point the teachers used the EQAO assessment results for their school. Areas in which their students scored lower, as a class, were established as areas of focus. Although it has to be pointed out the teachers were teaching different students than those students which the results were originally based upon. Teachers also reported using a variety of other sources to assess students learning such as student notebook work, classroom tests, assignments and classroom observation, and conversations with students and parents. Teachers reported that feedback provided to students consisted of sharing rubric information as well as using exemplars. Students submitted work to the teacher who provided feedback and then students could resubmit their work. Wideman also notes that teachers used scaffolding to facilitate student progression whereby explicit criteria referenced feedback was gradually reduced as students began to acquire

the desired skills. After a test was given, the teacher would re-teach as necessary and then retest. Many teachers reported that they could see students beginning to take responsibility for their own learning and this necessitated having students know, “where they were; where they were going, and; [sic] what steps they needed to get there” (p. 10).

The classrooms involved in the study experienced dramatic increases on their EQAO results compared to the previous year and considerably higher than schools in the rest of the board in the areas that teachers focused on. This is clarified by Wideman (2002) as he notes that the increase in scores cannot be claimed to have occurred as a result of the intervention strategies of the teachers and not from other unidentified factors. Unfortunately, most of the increases in achievement were attributed to students who were already doing well as opposed to those students were doing less well. Many teachers expressed the view that it was “frustrating” trying to increase learning outcomes in lower achieving students (p. 5). It is not insignificant to remember here that the lower achieving students are precisely the students that should benefit the most from formative assessment (Black & Wiliam, 1998b; Stiggins, 2001, 2005).

At first glance then this study is an excellent example of how the EQAO results may facilitate formative assessment in the classroom. However questions remain as to how teachers implemented feedback/corrective action within the classroom other than the sharing of learning goals. Although the language used to describe what was occurring in the classroom closely resembled the definition and expectations of *corrective/feedback action*, a formula for success emerged; “teach, assess, set goals, reteach, and retest” (Wideman, 2002, p. 9). One can sense here how testing appears to be driving the process. Since the study relied solely on teacher self reports, explicit examples of the feedback offered by teachers and students response to that

feedback is warranted. Widemen notes that teachers at first viewed EQAO results as “unfriendly data” arguing that the results of the assessments were due to socioeconomic backgrounds of the students, flaws in the testing itself, or lack of school resources (p. 5). He suggests however that these concerns were reduced, although not eliminated, at the study’s conclusion with many teachers in the study saying they would continue using the results from the EQAO in the future to influence their instructional focus. Interestingly teachers did not report on how the EQAO assessments influenced their teaching prior to the case study.

Johnson (2005) also provides some insight on how classroom teachers in Ontario have utilized the EQAO assessments. Johnson’s study focused only on schools that were successful on the EQAO assessments regardless of the socioeconomic characteristics of the school. He found that the teachers in grade one to grade three worked closely together as did the teachers in grade four to grade six. Here he notes that the teachers, working as a team, were better able to identify students who needed special skills for improvement on the assessments. A teacher who had participated as an EQAO marker was regarded as invaluable. Working as a team the teachers were better able to identify what resources were needed for preparation for the assessment. Principals noted that it was important to use the results of previous EQAO assessments to identify weaknesses and then modify teaching instruction to improve on areas of need (Johnson, 2005).

The EQAO assessment, evidently, had become the focus of principals and teachers in Johnson’s study, and as such, a school culture was established in terms of organizational teamwork to prepare for the test with administrative support. Wyatt (1996) notes, “one of the most powerful and enduring lessons from all the research on effective schools is that the better

schools are more tightly linked structurally, symbolically, and culturally than the less effective schools” (p. 6). Johnson’s (2005) study appeared to confirm this finding. In this sense the EQAO assessments can be seen to have provided some schools with the capacity and will to work collaboratively, using feedback from the assessments to alter their instructional strategies to increase student achievement on the EQAO assessments. Since the teachers from grade one to three and grades four to six work as a team, it appears that students who wrote the assessments also received altered instruction because of their assessment results when they returned to school in the Fall due to the increased coordination among teachers.

The majority of responses in Johnson’s (2005) study revealed that, “through test preparation, or ‘test wisdom’, closely followed by the importance of strong teamwork” as the reasons given to explain their EQAO results (p. 193). Teachers’ responses to the EQAO assessments were varied. Some teachers did not identify any restrictions on their teaching when preparing for the EQAO assessments and others appreciated the teamwork and support structures within their particular school. Other teachers did, however, express concern that too much time was spent on test preparation and having students focusing on a, “very artificial and limited EQAO vocabulary” (p. 201) which meant that there was little opportunity to “run with a teachable moment” (p. 202). Some teachers also noted that the long term feedback provided by the EQAO assessments was “no substitute for short term feedback from the classroom teacher” (p. 200).

The teaching strategies outlined in the above two studies can be seen as successful examples of how schools and teachers are meant to respond to the EQAO assessments when success is measured in terms of students’ EQAO scores. Since the above studies relied solely on

teacher self reports it is hard to determine whether the formulaic approaches described represents teaching to *the letter* rather than *the spirit* of formative assessment. Given that formulas for success did emerge and the absence of evidence indicating teachers capitalizing on teachable moments or expressions of acquiring increased skills in their questioning techniques, there is the sense the above studies represent more convergent examples of formative assessment rather than the application of divergent formative assessment. As Black and Wiliam (2003) note, facilitating capacity in teachers to implement formative assessment into their classrooms will not be achieved when, “teachers are ‘told’ what to do” and “would not be appropriate” (p. 629). They suggest rather that teachers need support in “developing their own professional practice” (p. 630).

For schools that do not do as well on the assessments the province sends in turnaround teams. These teams are made up of teaching experts, expert principals, and diagnosticians that are shipped into low performing schools if requested by the school. Fullen argues, that the turnaround teams have and do result in increased achievement and that the teachers of these schools have expressed pride in their achievements (Levin, 2005). As one teacher stated, “our discussions now focus on what we can do to improve reading, not on the deficits our students bring to school” (Levin, 2005, p. 40). Levin (2007) concludes then that the EQAO assessments were the subject of considerable controversy when first initiated but are much less so today for the valuable feedback they provide to boards, schools, and teachers.

Teachers’ Acceptance of the EQAO Assessments

However Levin’s (2007) insistence that the EQAO assessments are less controversial

among teachers than when first initiated is not shared, evidently, by the majority of teachers. Overall, elementary teachers in Ontario have not yet accepted the EQAO assessments as a valuable teaching aid. A teacher poll conducted in 2003 by the Ontario College of teachers found that 90 per cent of teachers believed that the EQAO assessments did not aid in student learning and 85 per cent of teachers felt the tests had a demoralizing affect on the students (McAdie & Dawson, 2006). In 2004, the Elementary Teachers Federation found that only 12 percent of elementary teachers felt the EQAO assessments had a positive effect on teaching. In 2005, 71 per cent of teachers felt that standardized tests were the least effective means to improve school achievement. The Elementary Teachers Federation currently argues that student results on the EQAO assessments have turned into an, “obsession driving far to [*sic*] much of what happens in Ontario schools” (Toronto Star, 2009). As in England, Ontario elementary teachers have recently put their support behind the creation of an educational video to illustrate the “negative effects” of the EQAO assessments on students (Toronto Star, 2009).

Volante (2007) concludes that the Ontario system, “is flawed precisely because the province has adopted a myopic view that overemphasizes provincial assessment scores” (p. 16). He suggests that instead of emulating other nation’s reliance of large scale tests Ontario needs to increase the value of teachers’ daily assessments in the classroom concluding that the use of large scale tests in Ontario need to be used as a tool of support rather than as a tool of control in school improvement. For their part teachers in Ontario, noted above, continue to chafe over the province’s emphases on the EQAO assessments. Although some teachers or schools give special consideration to the EQAO assessment to increase student scores on the assessments, the majority of teachers evidently view the assessments with less regard. Admittedly here, beyond

the rhetoric, it remains unclear how the EQAO assessment's influence teacher instructional strategies in Ontario.

Formative Assessment in Ontario: Why Now?

Since Black and Wiliam's literature review has received considerable attention ever since it was first published 11 years ago it is curious why Ontario has only now begun to officially acknowledge formative assessment. It appears that what is now occurring in Ontario is the same situation in England in 2002, when formative assessment first made its appearance on the ministry's website and, coincidentally, at the same time student scores on the large scale assessments began to peak.

Prior to the literacy and numeracy strategy overall scores on the EQAO assessments remained static (Levin, 2005). Levin, Glaze, and Fullan (2008) suggest that Ontario's student achievement framework has been successful since it has lead to sustained increases in student achievement. Although it is too early to claim that the EQAO results have reached their peak, the latest results do not indicate an overall success. Approximately one third of the students writing the grade three and six assessments failed to meet the ministry's provincial standard (Globe & Mail, 2009). Furthermore, the Globe and Mail reports, again based on the latest EQAO results, that one in four students who did not meet ministry expectations on the grade three assessments also "failed to catch up" on the grade six assessments (p. A5). As well, nearly half of the students who failed to meet ministry expectations on the grade six math assessments were also unsuccessful on the grade nine assessments.

Education Minister Kathleen Wynne commented that she was "concerned" over some

aspects of the recent scores but added that 67 per cent of students were meeting or exceeding ministry expectations, adding that there has been a 13 per cent increase in students achieving since 2002 (Globe & Mail, 2009, A5). Of course this fails to acknowledge that this number falls well short of the 75% target set by the government years earlier (Glaze, 2005; Levin, 2007). The Minister added that Ontario is taking steps to improve the results by doing more of the same. She states that the government is adding even more literacy and math coaches to improve students scores on the assessments, or “snapshots of attainment” as termed by the EQAO. What is new is the planned integration of full-day kindergarten. Conspicuously absent from the minister’s comments were any references to the proposed formative assessment policy.

Student scores on the EQAO assessments have not reached the arbitrary target levels set by the government. Ontario is now signaling an intention to add formative assessment to their assessment policies. Add is the key word. As in England, when formative assessment began to appear on the governments radar, the Ontario Ministry of Education’s proposed addition of formative assessment does not, as yet, appear to comprise a new strategy but as noted in this paper being articulated as supporting the existing focus on the EQAO assessments.

Although the EQAO and the Ontario Ministry of Education have acknowledged the importance of teachers’ classroom assessment with the tentative inclusion of the proposed formative assessment policy, the focus clearly remains on elevating student scores on the EQAO assessments. Ontario’s proposed policy can thus be seen as an addition to the existing framework rather than comprising a new strategy. As a consequence, the system response to the data provided by large scale tests is being not so subtly blended within the definition of formative.

Campbell and Levin (2009) state that, “developing ‘assessment for learning’ and use of

data as a systemic strategy also requires balancing capacity building and developmental processes with accountability requirements, sometimes referred to as balancing pressure and support” (p. 50). They add that, “assessment for learning should be considered as a system quality, alongside a range of ‘evidence-based’ sources of information to combine with professional expertise in daily practices to support improved outcomes” (p. 62). Of course this definition of formative assessment is similar to that provided by the OECD and others noted earlier in this paper. It also raises the same concerns voiced in Britain that the government’s emphasis on the data provided by large scale tests results in responses in the classroom focused on raising student scores on those assessments. The continued emphasis on the EQAO assessments, or snapshots of attainment, suggests then that the proposed formative assessment policy may be incorporated as supporting the status quo rather than comprising any new strategy with government focus remaining on raising student scores on the EQAO assessments.

Currently, large scale assessments in Britain are undergoing modification at the same time as emphasis on formative assessment is being increased. It remains to be seen if Ontario will continue to follow developments in Britain and modify the terms and conditions of the EQAO assessments in the coming years. It must be recalled though that formative assessment has not yet flourished as a result of any policy initiatives, in Britain, or elsewhere. To a large degree the number one barrier outlined thus far in this paper is that formative assessment has been that the influence of large scale tests prohibits its practice. Although this is certainly an important consideration when preparing the soil for formative assessment to take root it is by no means, as will be discussed in the next chapter, the only factor.

Chapter 5

Beyond Large Scale Assessments

Formative Assessment and Teachers' Assessment Skills

Stiggins (2001) argues, based on his own studies and classroom observations, and on a rather extensive literature review, that teachers' assessment capabilities are seriously wanting. Particularly he notes that teachers tend to base their assessments on recall of facts, completion of work over quality, with little attention to the application of knowledge, and had difficulty in "recognizing assessment of higher ordered reasoning" (p. 9). Although teachers reported using a wide variety of assessment practices, Stiggins speculates that "teachers were being given responsibility for assessing achievement targets that they themselves had not mastered, and did not understand" (p. 9). Harlen (2005) also expresses concerns on the reliability of teacher made tests noting that "there is evidence of unreliability and bias in teachers' assessment" (p. 212). Tierney (2006) notes that teachers in Ontario have expressed "frustration or concern about their lack of assessment expertise" (p. 253). This evidence suggests then that teacher assessment competence is questionable. Importantly, teacher assessment techniques incorporate little, if any, formative assessment practices as defined within the theoretical framework of this paper.

An explanation for why teachers do not incorporate formative assessment into daily classroom activities may be, as Stobart (2006) observes, "what the research evidence makes clear...is just how complex the process of feedback in learning is" (p. 141). Stobart notes that research studies conducted on classroom assessment suggest that conventional teacher feedback

usually consists of praise, the assignment of marks or grades, or is of an evaluative nature that signals approval or disapproval to the student. Such methods, suggests Stobart, only serve to interfere with increases in learning. Stobart argues that fundamentally feedback has to be task focused. However, he observes that assessment research, “has shown that even with expert teachers relatively little of this... ‘task focused’ descriptive feedback takes place” (p. 142). Evidently then, formative assessment is not only complex but involves processes that are a departure from how teachers assess their students, as defined the theoretical framework of this paper.

Principals’ Assessment Skills

The importance of the principal’s role in sustaining school assessment reform is recognized in Ontario’s student achievement framework as previously noted in this paper. Volante, Cherubini, & Drake (2008) note that the research in this area suggests that it is becoming necessary for principles to possess an “assessment literacy” in order to be active and informed agents in promoting assessment reforms. They define “assessment literacy” as having knowledge of,

a variety of skills and know various assessment methods and their purposes, the rudiments of technical assessment quality, how to embed assessment in instruction and curriculum, how to interpret various forms of assessment data, and how to use data to adjust curriculum and instruction. (p. 6)

Noonan and Renihan (2006) stress that principals need to be as knowledgeable as their teachers regarding new assessments methods occurring at their schools or experience the

frustration of having “power without authority” (p. 13). However, Volante, Cherubini, and Drake (2008) maintain that, despite a recognition of the importance of principals and teachers possessing assessment literacy to effect long-lasting change in the classroom, “very little formal training exists for the development of assessment literacy in classroom teachers and school administrators” (p. 7). They point out that the Ontario Principal Qualification Course, “provides very little attention in the area of student assessment and evaluation” (p. 7). However, Teirney (2007) observes that simply addressing assessment literacy through professional development courses has not been sufficient to alter instructional pedagogies.

The limited assessment expertise then among teachers and principals is of concern especially when considering the complexities involved in undertaking a program of formative assessment. Grave doubts arise as to whether simply providing more professional development courses or in-service training will prove effective in altering teacher assessment practices. This raises questions then over whether Ontario’s existing ‘ecological’ framework will prove sufficient in building the necessary, and rather specific teacher capacities required for formative assessment. This may help explain why a number of studies, outside of ‘clinical experiments’ have observed that teachers have been largely unsuccessful in adopting formative assessment practices despite receiving extensive training.

Full Disclosure: Formative Assessment in the Real Classroom

Black and Wiliam (1998b) argued that the evidence for formative assessment for raising standards presents a strong “prima facie case” for formative assessment (p. 148). Though they acknowledge that the question is not “Does it work? but How do we get it to happen?” (Black &

William, 2003, p. 629). Studies conducted in a 'real' classroom reveal how difficult formative assessment actually is.

As noted earlier, teacher pedagogy is an important variable for the implementation of formative assessment. Yet evidence suggests that altering teacher pedagogy is quite difficult to achieve. Munby and Lock (2000) conducted a qualitative analysis on a teacher who was interested in adopting student centered-assessment practices into his classroom. The study undertaken attempted to shed some light on how teachers evaluated their students since it is an area, the authors note, that "is relatively unexamined" (p. 2). The teacher, referred to as George in the study, taught a grade seven and eight split class comprising 27 middle class students. He had 18 years teaching experience and expressed an interest in establishing, "goal-setting, assessment portfolios, and student-led conferencing into his classroom" (p. 2). The project was a collaboration between the teacher and the researchers and was conducted for more than a year.

Munby and Lock (2000) found that George was unable to incorporate student centered assessment practices into his classroom. George believed that the teacher's main role and responsibility rested in "the transmission of information" (p. 4). As such George emphasized "factual knowledge in his instructional practices" and regarded student learning as an "acquisition of factual information" (p. 4). George's view of teaching and learning conflicted directly with student centered assessment in that he did not consider how his instruction would need to be altered to accommodate the new assessment. Therefore the new assessment practices were not incorporated into his teaching, but were added onto his regular teacher centered practices. As such, George abandoned the assessment practices when he felt he would not be able to cover the curriculum in time.

This study is important in that it identifies two barriers that may prove hindrances for teachers to effectively implement formative assessment into the classroom; teachers' beliefs and the context of the classroom. Munby and Lock (2000) note, "George's attempts to implement new assessment practices were constrained by the context of the teaching environment and by the beliefs he held about teaching and learning" (p. 7). They also question whether new assessment practices can ever be mandated with any real chance of success. Despite George's desire to implement new assessment practices, and despite professional support and resources when working collaboratively with the research team, he failed.

Dixon and Williams (2003) note that the *Report of the Literacy Task Force* recognized literacy as the focus of governmental policy in New Zealand. The report stressed that the success of the policy rests on the formative assessment capabilities of the teachers and therefore conducted a study on teachers' formative assessment capabilities. They found that teachers had a solid understanding of formative assessment theory and were quite able to explain the key aspects of formative assessment. However, paradoxically, when the teachers were asked to describe their own teaching practices, "there were noticeable gaps and confusions in their articulated understanding of formative assessment" (p. 4). For example teachers stressed that it was important to work closely with children but they did not make references to sharing learning goals or providing students with feedback. As noted in this paper, feedback and the sharing of learning goals is a key component to formative assessment. It remained unclear to the researchers why "important facets of formative assessment appear to be absent from teachers' descriptions of their own practice" (p. 9). The researchers note that the teachers' descriptions of their teaching practices need to be confirmed through observation. They suggest however that in order to

increase teachers' formative assessment capacities attention needs to be placed on not only "the theoretical and conceptual notions underpinning formative assessment" but on the deconstruction of "their current pedagogical practices" as well (p. 9).

Inspired by Black and Wiliam's (1998) literature review researchers at the Stanford Education Assessment Laboratory and the Curriculum Research & Development Group (CRDG) Stanford embarked "with considerable euphoria," on a project "to provide embedded formative assessments in a nationally used curriculum developed at CRDG: Foundational Approaches in Science Teaching" (Shavelson, 2008, p. 294). The idea behind the project was to, "embed formative assessments in a nationally used science curriculum" (p. 294). These embedded assessments were designed to, "elicit students' explanations-formally within a unit, as part of a lesson plan, or as 'on-the-fly' teachable moments occur-teachers would take this opportunity to close the gap in student understanding" (p. 293). The following is a review of their studies.

Yin, Shavelson, Ayala, Ruiz-Primo, Brandon, and Furtak (2008) embedded formative assessments within an enquiry science unit (middle school) in six classrooms teaching the same unit to compare with a control group of six classrooms, also teaching the same unit. The embedded formative assessment did not have an impact on student motivation, motivation or conceptual change. Yin et al. (2008) suggest that this study did not provide evidence for the effectiveness of formative assessment but did highlight "the difficulty and importance of effectively implementing formative assessment" (p. 354).

Part of the problem Yin et al. (2008) suggest was the "inadequate implementation of the treatment by the experimental teachers," even after training, that may have been due to, ironically, inadequate formative assessment in terms of providing feedback to the teachers during the experiment (p. 355). The experiment successfully elicited student responses but the teachers

were not able to capitalize on them. Yin et al. also suggest the lack of success in their study may have been due to other uncontrollable factors such as community effects, the school environment, students' family backgrounds, and academic preparation. Including these variables is puzzling if the authors are suggesting that a student's background is an indicator of limited achievement in terms of formative assessment. Again it should be pointed out that variables existing outside of the classroom, particularly a student's family background or socioeconomic status, are the kinds of variables that formative assessment, as argued by formative assessment advocates, will have the most influence upon (Black & Wiliam, 1998b; Stiggins, 2001, 2005). Yin et al. (2008) conclude that "simply embedding formative assessment in the curriculum does not guarantee improved learning and teaching" (p. 356).

Furtak, Ruiz-Primo, Shemwell, Ayala, Brandon, Shavelson, and Yin (2008) further analyzed the results of Yin et al. (2008) by examining the *fidelity of implementation*, which is a way of, "determining the alignment between the implementation of a treatment and its original design" (p. 362). Furtak et al. (2008) examined the videos of the classroom instruction during the original experiment to determine what teachers actually did in the classroom. They differentiated between what the experiment intended, and what was enacted in the classroom and the learning achieved. Furtak et al. concluded the study, "raises important questions about the feasibility of formative assessment in general" (p. 387). Here they note that, "embedded formative assessments may be too restrictive and time-consuming for teachers" and teachers may well lack the necessary skills to lead whole class discussions, especially if, "they have weak backgrounds in science" (p. 387). They suggest that the most important lesson to be derived from this study was the uncertainty in the teachers' ability to successfully implement formative assessment and the researchers ability to facilitate it. They note that "that despite" their "best

intentions to design a treatment and instruments to measure it” the study was limited by the extent to which they were able to influence teacher capacity to successfully capitalize on the teachable moments using formative assessment (p. 388).

Concerns in the Subsystem

If formative assessment can be interpreted as a major innovation, Van der Vegt, Smyth, and Vandenberghe (2001) note that such innovations can have significant effects on the strategic subsystem of which the teacher is apart. The reason for this is that policy implementations upset “in varying degrees, the ‘steady state’ of the organization” (p. 12). They suggest that policy implementations may require teachers to “re-articulate their subject domains” (p. 19). This involves establishing new collaborations and a “re-designing of the school’s operational subsystem” that “comes close to the bone” for teachers (p. 19). Here the authors note that teachers may have to renegotiate their present influence or identity, invest new efforts, and addresses their professional competencies, within the new collaborative or reorganized structure. Questions may emerge such as, “Will my ideas about professional work still be supported in this new school?; Do I really want to be a loyal part of this new organization?” (p. 16). These subsystem responses will have direct influences upon the successful implementation of the policy.

The significance of the above is especially relevant to formative assessment policy. It is important to note here that formative assessment will require teachers to have a high level of competency in their subject domains (Black & Wiliam, 1998a). Similarly, Dixon and William (2003) note that, “unless teachers have sufficient content knowledge they will have difficulty noticing gaps and contradictions in children’s learning, and will be unable to utilize children’s

existing knowledge to promote new learning” (p. 5). Hattie and Timperley (2007) note that “deep understandings of the subject matter” are necessary to provide feedback that can link ideas and relationships (p. 103). Furtak et al. (2008) suggest that teachers may lack the necessary skills in their subject domains to effectively implement formative assessment leading them to question the “feasibility of formative assessment in general” (p. 387).

It remains questionable then as to whether formative assessment can be effectively implemented in an elementary or middle school classroom where teachers may not have the background or knowledge to adequately facilitate students’ self-assessment capacities even though such capacity, notes Black and Wiliam (2008a), is “essential” (p. 30). Formative assessment can then be seen to challenge the autonomy and professional identity of teachers in that it may require an examination into their level of basic competence as educators. This of course can be seen to be a significant barrier to teacher acceptance of formative assessment. As with standardized testing, current reforms which emphasize formative assessment may be interpreted as an assault upon teachers’ professional competence and personal integrity.

Steinburg (2008) notes that before the accountability era teachers had an escape valve in that poor results could be attributed to the students themselves, school management, or the social economic characteristics of the community. She argues that accountability “closes off these avenues of emotional relief” (p. 55). As one teacher noted in Johnson’s (2005) study commented, after moving from an affluent school district to a less affluent one, “I went from being one of the best teachers in the region to being one of the worst teachers in the region in one year” (p. 165). Although the addition of the EQAO’s Education Quality Indicators Framework, once again, can be seen to open the emotional release valves as results are once again put into various contexts. Yet Steinburg also notes that the emotional release valves may also be sealed with formative

assessment since, “formative assessment disallows emotional distance from student failure by arguing that teachers need to ‘re-form’ their teaching in response to student misconceptions” (p. 60). It must be recalled that although formative assessment as interpreted in this essay, is understood as a partnership between the student and the teacher both of whom are responsible for learning, it is a partnership where the teacher provides the student with the initial means to move forward. This is especially so for low performing students who, the research suggests, are the ones expected to benefit most from formative assessment. Steinburg (2008) notes that formative assessment may therefore present emotional challenges for teachers if students do not progress since such failure may represent teacher inadequacy or even incompetence which may lead teachers to directly confront the “limits of their self efficacy” (p. 59). The blame for student failure is not so easily passed off on the student if the teacher is not able to effectively execute formative assessment since failure implies a change in tactics is required by the teacher that will lead to the student’s success.

Unintended Consequences

Cohen, Moffitt, and Goldin (2007) observe that, “policy makers who define problems and devise remedies are rarely the ultimate problem solvers” adding that, “they depend on the very people and organizations that have or are the problem to solve it” (p. 522). They argue that policy instruments used to influence teaching practices can be conceptualized as intending to build capacity, creating the capability in teachers to affect the new practice. They note that the effectiveness of certain policy instruments to affect practice depends to a large degree on the differences or gaps between conventional practice and the new practices proposed by the policy. To address this gap policy instruments require that the knowledge is conveyed directly to the

teachers or “indirectly enables practitioners to acquire knowledge themselves” (p. 536). Here however, they note, there lies an inherent danger. Policies that aim to change practice, “often create incompetence by requiring practitioners to do things that they do not know how to do” (p. 538).

Such dangers have been noted in formative assessment research. Smith and Gorard (2005) observed the results of a pilot project conducted in a middle school in Wales that implemented formative assessment practices by changing the feedback provided to students. Here, summative testing and student marks were removed and replaced with enhanced formative assessment feedback. The teachers had intensive training in providing effective formative assessment feedback. The results however were rather disappointing. Smith and Gorard report that in the three subjects, Math, Science, and Welsh the treatment was ineffective. However for English the results proved to be “actually harmful” (p. 28) with reduced summative scores being reported. Smith and Gorard conclude, “there are important lessons to be learned for schools and researchers about how research-based designs fare in the ‘real world’ and when control is passed from the designers to the practitioners alone” (p. 36). Despite receiving training the teachers provided feedback in a conventional manner with brief, evaluative comments rather than providing explicit criterion references. Despite being informed on the process the students responded with frustration in not knowing how they were doing in the class since they did not receive any marks.

Similar results were found by Niven and Meyer (2007). However in this study teacher capacity to implement formative assessment did not appear to be an issue. The research professors applied formative assessment in an Academic Literacy course at the University of KwaZulu-Natal in South Africa. The Academic Literacy course makes “extensive use of process

writing methodologies and formative assessment” (p. 14). Niven and Meyer, who taught the course, prepared extensively for the use of formative assessment and they were confident that they had “significantly raised” their “standards of formative assessment” particularly in written feedback (p. 16). However, they were “surprised” that the “effort” they put into the extensive constructive written feedback did not result in increased achievement or increases in student motivation (p. 17). Rather, in some cases the feedback was internalized by the students as “criticism and failure” (p. 17). Furthermore, in spite of the fact that Niven and Meyer believed that the feedback provided was “deliberately kind and helpful” it “humiliated and angered the students” (p. 19).

Niven and Meyer (2007) suggest that the students’ lack of experience with extensive and constructive feedback may have been too ‘alien’ to them. The authors note that prior to each assignment the students were provided clear assessment criteria. However, Niven and Meyer suggest that this, “was not enough to challenge the students’ underlying notion that assessment is intrinsically summative and that the comments that lecturers make on their writing mean that they have failed in some way” (p. 19). Furthermore Niven and Meyer suggest that “content-focused feedback makes unexpectedly difficult cognitive demands,” on some students, which prevents them from acquiring “independent self-reflection or meta-cognition” which in turn does not produce increases in motivation (p. 16). Relatedly, Nitko (1995, as cited in Sebatane, 1998, p. 5) also suggests that students who are having problems in specific areas may “lack the meta-cognitive skills needed to evaluate the quality of their own progress” and may not even be cognizant of what their problem may be. Despite their best efforts and extensive knowledge of formative assessment Niven and Meyer conclude, “it is certainly more difficult to reproduce Black and Wiliam’s findings, that strengthening the practice of formative assessment enhances

learning” (p. 20).

From the above two studies it can be suggested that students also need to be informed to some degree on the principles and purposes of formative assessment in order for it to have any chance of success. Interestingly, as well, in one study teachers could be seen as at fault, in the other study, the students. Such noted differences only add to the complexities involved in adapting formative assessment to the classroom.

Chapter 6

Implications

From the studies and observations presented in this paper I have identified limitations in Ontario's formative assessment policy in the context of Ontario's student achievement framework. The issue here is not whether or not the EQAO assessments are any good or are an accurate measurement of student achievement of curriculum expectations. Rather, the issue is how large scale assessments influence teachers formative assessment practices. I suggest that the evidence I have presented in this paper clearly suggests that the presence of standardized tests and emphasis on increasing student scores on such tests inhibit practices that are consistent with formative assessment. This appears to be the case regardless of whether formative assessment is acknowledged and promoted. Although the EQAO assessments are being soft pedaled to some extent, being described as snapshots of achievement by both the Ministry of Education and the EQAO, and although both the Ministry of Education and the EQAO acknowledge that teacher assessments provide for a more complete overall picture of how a student is performing, the emphasis clearly remains on raising EQAO scores. In this regard the EQAO assessments and Ontario's literacy and numeracy strategy do not provide the fertile soil necessary for the development of formative assessment. This then highlights the problem of implementing formative assessment in context to Ontario's student achievement framework. Ontario's student achievement framework is structured to respond to the EQAO assessments and will, as the evidence presented in this paper suggests, undermine developments in formative assessment.

Research literature on formative assessment suggests that formative assessment is not

easily practiced by teachers even with extensive training that in many respects goes well beyond the 'limited' training outlined in Ontario's student achievement framework. Providing teachers with professional training opportunities in an attempt to build teachers formative assessment capacities does not ensure their effective implementation in the classroom. Wiliam (2008) suggests that, "the necessary changes in classroom practice, although often apparently quite modest, are actually difficult to achieve" (p. 38). As an example to illustrate this point he notes a teacher, with 25 years of teaching experience, who wanted to ensure that she allowed every student a chance to answer questions in class began to use popsicle sticks with the students' names on them to keep track of whom she called. However, after picking up a popsicle stick with a particular student's name on it she often began her questions with, "Does anyone know...?" (p. 38). He also notes that most teachers are aware of the findings such as increasing wait times by a three or four seconds after asking a question to allow students to consider a response increases student achievement. Nevertheless research still shows that teachers still allow approximately one second (Wiliam, 2008).

The changes required for successful implementation of an effective formative assessment regime is a complex affair involving more than small pieces of wood or adding a three-second pause. Sebatane (1998) notes that the changes required for formative assessment involves "radical changes in teacher pedagogy" (p. 2). Certainly, after consideration is given to the complexities involved in effectively and successfully implementing formative assessment in the classroom the changes appear to be directly the opposite of modest. This is why empirical research illustrating the importance of feedback for increasing student achievement and providing techniques to enact formative assessment in curriculum documents is simply not enough. As Tierney (2007) observes, "changing assessment practices is not simply a matter of

increasing teachers' assessment literacy through professional development workshops, but a more comprehensive process that requires a conceptual shift for all stakeholders" (p. 259).

Elwood (2006) argues that changing teacher pedagogy requires allowances to be made for the increased role of the student in the teacher-student relationship in the learning process.

Elwood questions how issues of equality will be addressed in this relationship. She notes that an important limitation in formative assessment research is that:

considerations of the social and cultural aspects and settings of classrooms seem more rhetorical than actual and that while lip-service is paid to the notion that formative assessment can deliver equality of outcomes, the work reported is not substantial enough to support these claims and more research is needed to understand the fundamental relationship between formative assessment and equity. (p. 228)

She notes that research on gender leads to significant questions on how gender will influence the teacher-student dialectic during the process of formative assessment. To be sure, there remain a number of important variables that formative assessment research has yet to answer and Ontario's Student Achievement Framework has yet to address.

For effective formative assessment to occur in the classroom there may be a need to address the issue of teacher qualifications regarding subject content since a high degree of foundational knowledge is necessary to facilitate higher order thinking in context to subject content. This is especially true in the elementary classroom where teachers are responsible for teaching several subjects, being perhaps expert in none. Yet as Navarro (2008) observes establishing what content knowledge is sufficient is difficult since there are few instruments

available to “assess that knowledge” (p. 254). Teacher unions have rejected suggestions regarding teacher qualification testing in the past (Majhanovich, 2002).

Certainly though some measure of assurance that teachers are competent in the subject matter they are teaching as outlined in the curriculum documents is warranted. The question here then is what depth is necessary beyond the prescribed content? What separates the teacher from the person on the street when both may be fully able to read and understand the course text? What depth of understanding is necessary to provide the feedback necessary to move the student forward in a way that provides genuine learning? How can this depth of content knowledge be measured accurately?

On the other hand having a high level of content knowledge does not guarantee that a teacher possesses “an understanding about how to teach it” (Navarro, 2008, p. 254). There are few mechanisms currently in place to ensure that teachers have an adequate understanding of how to teach what they know particularly in regard to formative assessment. Knowledge and an ability to teach and assess students are needed but how to provide a mechanism that will ensure teachers are capable in these realms remains unclear. Have Faculties of Education ensured that their graduates have acquired the skills necessary to be effective teachers? Is it even possible that they can presume to do so? What indeed makes a teacher successful? Fullan has suggested, in his reference to what is needed for successful school reform, that the key “is as simple and complex” as getting teacher support for the reform (Fullan, 1991, as cited in Van der Vegt, Smyth, & Vandenberghe, 2001, p. 9). Yet the above evidence on formative assessment suggests that developing successful formative assessment capacity involves much more than simply acquiring teacher support.

There are issues concerning students' acceptance of formative assessment practices. Even if teachers can implement formative assessment students may not respond as predicted by formative assessment theory. Even if students are prepared for learning under formative assessment through capacity enabling implements that build upon their understanding of the process they may still not be able to participate in the manner formative assessment theory predicts due to meta-cognitive issues thereby undermining the entire process. Without student cognizance of their own learning the feedback loop cannot be completed. Hattie and Temperly note that, "only when students are grounded in and committed to the goals of learning" will feedback be effective (p. 103). How to establish this grounding and commitment remains unclear.

Conclusion

The lessons of history teach that the walls of the classroom are formidable barriers that have easily withstood the arsenals of policies that were aimed to alter teachers' teaching practices (Cuban, 2006). Tierney (2006) notes that "policy is often named as the impetus for action, especially in relation to large-scale assessment and professional development" (p. 252). Large scale assessments however, as the evidence I have presented indicates, have breached the walls of the classroom not as an innovation but rather as a reinforcement of conventional teaching practices. I have therefore argued that Ontario's proposed formative assessment policy has been placed in an existing framework that was designed primarily for different purposes namely, to respond to and support the EQAO's assessments. I have acknowledged the argument made by others that the EQAO assessments also serve to promote more effective teacher instructional

strategies to boost scores on the EQAO assessments but I have pointed out that such instructional strategies have not been established as supporting formative assessment practices in context of accountability. Clearly more than an alignment between the EQAO assessments and the Ontario curriculum is needed for the successful implementation of formative assessment in the classroom.

Formative assessment is an incredibly difficult and complex assessment technique that requires vital preconditions to be in place for its effective implementation. The preponderance of empirical evidence I have presented overwhelmingly suggests that the first precondition for effecting successful formative assessment is the absence of standardized tests. The unanswered question is whether or not the two can exist simultaneously? Theoretically the answer is yes but the relevance of theory to the peculiarities of the classroom is open to debate. As of today, the empirical research suggests formative assessment and standardized tests for accountability have not yet been able to achieve a harmonious co-existence. Articulations of system wide formative assessment only serve to hide the tensions between formative assessment and large scale assessments for accountability and prevents the full potential of formative assessment to be realized in the classroom. I have also suggested that the implements or practices outlined within Ontario's existing student achievement framework are insufficient to build or alter principal, teacher and student capacity for the effective execution of formative assessment in the classroom.

As of today there remains a paucity of quantitative measures of increases in student achievement as a result of formative assessment in a 'real classroom'. Elwood (2006) claims that the question as to, "whether whole-scale changes to the use of formative assessment practices yield the overall biggest gains is still a moot point" since there is no evidence of sustained

achievement and the, “continued improvement of low-achieving students through comprehensive change to formative assessment” (p. 227). Therefore, studies on assessment, and particularly formative assessment, have come to be dominated by qualitative studies (Tierney, 2006) which, Elwood (2006) suggests, must be “viewed with caution” (p. 227). This does not mean that formative assessment should not be promoted. However, it does raise concerns outlined in this paper regarding the accountability agenda, standardized tests and formative assessment.

I think it is important for both teachers and governments to be familiar with just how complex formative assessment actually is in order to facilitate the conditions necessary to develop its practice within the classroom. Volante (2007) suggests that “by adopting a collaborative approach that is informed by recent advances in the field, Ontario could develop an accountability framework that appropriately re-positions large-scale assessment to support, not control school improvement” (p. 14). However, as I have noted in this paper, regardless of how unobtrusively large scale tests are articulated within a system of accountability they have not, as yet, been able to facilitate formative assessment. On the contrary, as the evidence I have presented suggests, they only serve to impede the implementation of formative assessment.

Campbell and Levin (2009) suggest Ontario’s student achievement framework is promoting evidence-based practices yet the evidence clearly indicates that formative assessment is inhibited by large scale tests for accountability. Certainly more research is required on how teachers respond to the EQAO assessments and the literacy and numeracy strategy in the Ontario classroom in regard to influences on formative assessment. Formative assessment promises higher learning gains but, so far, verification of learning gains in the ‘real’ classroom primarily comes from qualitative teacher reports. To verify learning gains current methods of system

accountability requires quantitative data from large scale assessments which the empirical research clearly indicates inhibits the practice of formative assessment. Therefore addressing accountability with formative assessment is problematic and remains a Catch-22. Whatever public trust exists in the education system will need to be maintained in terms of accountability lest schools once again run the risk of being perceived as 'kingdoms unto themselves'. How to provide a mechanism for accountability that would facilitate formative assessment without impeding its successful implementation remains, as yet, unresolved. Successful implementation of formative assessment may indeed require a renegotiating by all stakeholders in terms of what the purposes of education are and how they will be achieved.

There may be some readers who will find fault with my concept of formative assessment in that it is too narrow or restrictive. Some may even suggest that this was intentional in design in order to purposely conflict with large scale assessments for accountability. To this criticism I would point out that my concept of formative assessment is similar to the one adopted by most of the literature presented in this paper and indeed, the field. It needs to be pointed out that the Ministry has drawn upon this very same literature to support the proposed policy. Such an objection then is unfounded.

An underlining current that exists in this paper, and one that I did not have room to explore in more depth, concerns the limitations of educational research. As Hargreaves notes, "there is no vast body of research which, if only it were disseminated and acted upon by teachers, would yield huge benefits in the quality of teaching and learning" (1996, as cited in Hammersley, 2007, p. 9). This appears to be an accurate statement thus far for both standardized tests for accountability and formative assessment. I suggest therefore that at the very least

government policy initiatives that are purported to be evidence-based need to be clarified in terms of who's evidence, and the context of that evidence, and that the limitations of that evidence be fully acknowledged. Anything less would be disingenuous.

References

- Allal, L. & Lopez, L.M. (2005). *Formative assessment of learning: a review of publications in French*. Retrieved 9 April, 2009 from http://www.oecd.org/document/28/0,2340,en_2649_35845581_35338140_1_1_1_1,00.html.
- Assessment Reform Group. (2006). The role of teachers in the assessment for learning. Retrieved September 2, 2009 at <http://www.assessment-reform-group.org/publications.html>
- BBC News. (2008). MPs Condemn Tests. Retrieved September 02, 2009 at http://news.bbc.co.uk/2/hi/uk_news/education/7396125.stm.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Educational Assessment, Principals, Policy and Practices*, 5(1), 7-74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappa*, 80(2), 139-148.
- Black, P., & Wiliam, D. (2003). "In praise of educational research": Formative assessment. *British Educational Research Journal*, 29(5), 623-637.
- Black, P., McCormick, R., James, M., & Peddar, D. (2006). Learning how to learn and assessment for learning: a theoretical inquiry. *Research Papers in Education*, 21(2), 119-132.
- BMG Research Report. (2008). *Evaluation of the developing thinking and assessment for learning development programme*. Retrieved 4 August, 2009 at <http://bing.search.sympatico.ca/?q=welsh%20assembly%20government&mkt=en-ca&set>

Lang=en-CA.

- Brookheart, M. (2007). Expanding views about formative classroom assessment, A review of the literature. In J. McMillan (Ed.), *Formative Classroom Assessment* (pp. 43-62). NY: Teachers College Press.
- Brown, G., Lake, R., and Matters, G. (2009). Assessment policy and practice effects on New Zealand and Queensland teachers' conceptions of teaching. *Journal of Education for Teaching*, 35(1), 61-75.
- Buhagair, M. (2007). Classroom assessment within the alternative assessment paradigm: revisiting the territory. *The Curriculum Journal*, 18(1), 39-56.
- Campbell, C. and Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability*, 21, 47-65.
- CERI OECD Policy Brief (2005). Formative assessment: improving learning in secondary classrooms. Retrieved 9 April, 2009 at <http://www.oecd.org/dataoecd/19/31/35661078.pdf>
- Cohen, D., Moffitt, S., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education*, 113, 515-548.
- Cuban, L. (2006). Getting past futile pedagogical wars. *Phi Delta Kappan*, 87(10), 793-795.
- Daugherty, R., & Ecclestone, K. (2006). Constructing assessment for learning in the UK policy environment. In J. Gardner (Ed.), *Assessment and learning*, (pp. 149-167). London: Sage.
- Daugherty, R. (2008). Reviewing national curriculum assessment in Wales: How can evidence inform the development of policy? *Cambridge Journal of Education*, 38(1), 73-87.
- Daugherty, R. (2009). National curriculum assessment in Wales: adaptations and divergence

- Educational Research*, 5, 247-250.
- Davies, J., & Ecclestone, K. (2009). 'Straitjacket' or 'springboard for sustainable learning'? The implications of formative assessment practices in vocational learning cultures. *The Curriculum Journal*, 19(2), 71-86.
- Dixon, H., & Williams, R. (2003). Teachers' understanding and use of formative assessment in literacy learning. *New Zealand Annual Review of Education*, 12, 1-12.
- Earl, L. (1995). Assessment and accountability in education in Ontario. *Canadian Journal of Education*, 20(1), 45-55.
- Earl, L., & Torrance, N. (2000). Embedding accountability and improvement into large-scale assessment: What difference does it make? *Peabody Journal of Education*, 75(4), 114-141.
- Earl, L., & Torrance, N. (2002). Front row seats: what we've learned from the national literacy and numeracy strategies in England. *Journal of Educational Change*, 3, 35-53.
- Eason, G. (2009). Teachers back boycott of stats. *BBC News*, Retrieved 5 August 30 July, 2009 at http://news.bbc.co.uk/2/hi/uk_news/education/7994882.stm
- Education Quality and Accountability Office. (2009). *EQAO school, board and provincial results 2007-2008*, Retrieved 18 April 2009, from <http://www.eqao.com/categories/home.aspx?Lang=E>
- Ellis, J., Lamoureux, G., Awender, T., Wessel, D., & Donohoo, J. (2008). Of class, culture, and accountability. *The International Journal of Learning*, 15(2), 24-33.
- Elwood, J. (2006). Formative assessment: possibilities, boundaries and limitations. *Assessment in Education*, 13(2), 215-232.

- Elwood, J. (2009). The English national curriculum assessment system: a commentary from Northern Ireland. *Educational Research, 51*(2), 251-254.
- English, E., Hargreaves, L., & Hislam, J. (2002). Pedagogical dilemmas in the national literacy strategy: primary teachers perceptions, reflections and classroom behaviour. *Cambridge Journal of Education, 32*(1), 9-26.
- EQAO Connects. (2008). Retrieved 28 February, 2009 from <http://www.eqao.com/eMagazine/2008/05/askUs.aspx?Lang=E&ArticleID=09&ItemID=24>.
- Froese-Germain, B. (2004). In M. Moll (Ed.) *Passing the test, the false promises of standardized testing* (pp. 182-186). Ottawa: Canadian Center for Policy Alternatives.
- Furtak, E. (2003). *Formative assessment in K-8 science education: A conceptual review*. Retrieved Jan. 30, 2009, from http://www7.nationalacademies.org/bose/Furtak_Commissioned_Paper.pdf.
- Furtak, E., Ruiz-Primo, M., Shemwell, J., Ayala, C., Brandon, P., Shavelson, R. & Yin, Y. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education, 21*, 360-314.
- Gareis, C. (2007). Reclaiming an important teacher competency: The lost art of formative assessment. *J Pers Eval Educ, 20*, 17-20.
- Glaze, A. (2005). Literacy and numeracy: An educational imperative. *Orbit, 35*(3), 37-38.
- Green, J. M. (1998). Authentic assessment: Constructing the way forward for all students. *Education Canada, 38*(3), 8-12.
- Growing Success. (2008). Retrieved March 2, 2008, from

http://www.ocup.org/resources/documents/EDU_GS_binder_010708_BMv2.pdf

Hall, K. (2000). A conceptual evaluation of primary assessment policy and the education policy process in the Republic of Ireland. *Compare*, 30(1), 85-101.

Hardman, F., Smith, F., & Wall, K. (2003). 'Interactive whole class teaching' in the national literacy strategy. *Cambridge Journal of Education*, 33(2), 197-215.

Hargreaves, D. (2007). Teaching as a research-based profession: Possibilities and prospects (The teacher training agency lecture 1996). In M. Hammersley (Ed.), *Educational research and evidence-based practice* (pp. 3-17). CA: Sage.

Harlen, W. & Crick, R. (2003). Testing and motivation for learning. *Assessment in Education*, 10(2), 169-207.

Hattie, J. And Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.

Haycock, K. (2008). Improving achievement and closing gaps between groups: The role of assessment in helping us move further, faster. In C. Dyer (Ed.) *The Future of Assessment, Shaping Teaching and Learning* (pp. 305-320). NY: Taylor Francis Group.

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.

Hayward, E. (2007). Curriculum, pedagogies and assessment in Scotland: 'Ah kent yir faither'. *Assessment in Education*, 14(2), 25-268.

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140-145.

- Hutchinson, C. and Hayward, L.(2005). The journey so far: assessment for learning in Scotland. *The Curriculum Journal*, 16(2), 225-248.
- James, M. (2000). Measured lives, the rise of assessment as the engine of change in English schools. *The Curriculum Journal*, 11(3), 334-364.
- James, M., Black, P., McCormick, R., Peddar, D., and Wiliam, D. (2006). Learning how to learn, in classrooms, schools and networks: aims, design and analysis. *Research Papers in Education*, 21(2), 101-118.
- Johnson, D. (2005). *Signposts of success: interpreting Ontario's elementary school test scores*. ON: C.D. Howe Institute.
- Kennedy, K., Chan, J., Fok, P., & Yu, W. (2008). Forms of assessment and their potential for enhancing learning: conceptual and cultural issues. *Education Research Policy Practice*, 7, 197-207.
- Kirton, A., Hallam. S., Peffers, J., Robertson, P., & Stobart, G. (2007). Revolution, evolution or a Trojan horse? Piloting assessment for learning in some Scottish primary schools. *British Educational Research Journal*, 33(4), 605–627.
- Klinger, D., Deluca, C., & Miller, T. (2008). The evolution of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, 76, 1-34.
- Koller, O. (2005). Formative assessment in classrooms: a review of the empirical German literature, Retrieved 9 April, 2009 from http://www.oecd.org/document/28/0,2340,en_2649_35845581_35338140_1_1_1_1,00.html

- Leithwood, K., Jantzi, D., Earl, L., Watson, N., Levin, B. & Fullan, M. (2004). Strategic leadership for large-scale reform: the case of England's National Literacy and Numeracy Strategy. *School Leadership & Management*, 24(1), 57-79.
- Leung, C., & Rea-Dickens, P. (2007). Teacher assessment as policy instrument: Contradictions and capacities. *Annual Review of Applied Linguistics*, 29, 64-79.
- Leung, C., & Scott, C. (2009). Formative assessment in language education policies. *Annual Review of Applied Linguistics*, 29, 64-79.
- Levin, B. (2005). Turnaround teams [Ontario's new literacy and numeracy strategy]. *Orbit*, 35(3) 40.
- Levin, B. (2007). Sustainable, large-scale education renewal. *Journal of Educational Change*, 8, 323-336.
- Levin, B., Glaze, A., & Fullan, M. (2008). Results without rancor or rankings, Ontario's success story. *Phi Delta Kappan*, Dec. 273-280.
- Marshall, B., & Drummond, M. (2006). How teachers engage with assessment for learning: lessons from the classroom. *Research Papers in Education*, 21(2), 133-149.
- Majhanovich, S. (2002). Conflicting visions, competing expectations: Control and de-skilling of education. *McGill Journal of Education*, 37(2) 159-175.
- McAdie, P., & Dawson, R. (2006). Standardized testing, classroom assessment, teachers, and teacher unions. *Orbit*, 36(2) 30-33.
- Moll, M. (2007). Large scale educational assessment, the new face of assessment. In M. Moll (Ed.) *Passing The Test, The False Promises of Standardized Testing* (pp. 11-16). Ottawa: Canadian Center for Policy Alternatives.

- Munby, H., & Lock, C. (2000). Changing assessment practices in the classroom: a study of one teacher's challenge. *Alberta Journal of Educational Research*, 46(3), 267-276.
- National Curriculum Task Group on Assessment and Testing. (1988). *A report*, London: Department of Education and Science.
- Navarro, M. (2008). Assessment and school reform, lessons from 15 years in the field. In C. Dwyer (Ed.) *The future of assessment, shaping teaching and learning* (pp. 245-278). New York: Lawrence Erlbaum Associates.
- Niven, P., & Meyer, B. (2007). Understanding the impact that principled formative assessment feedback has on first year student's writing: Is it useable nor not? An action research project. *The International Journal of Learning*, 14(8), 13-22.
- Noonan, B., & Renihan, P. (2006). Demystifying Assessment Leadership. *Canadian Journal of Administration and Policy*, 56, 1-20.
- OECD Policy Brief (2005). *Formative assessment - Improving learning in secondary classrooms*. Retrieved 9 April, 2009 at <http://www.oecd.org/dataoecd/19/31/35661078.pdf>
- Ontario Ministry of Education (2008). *Growing success: Assessment, evaluation, and reporting – Improving Student Learning*. Toronto, ON: Queen's Printer for Ontario. Retrieved March, 11, 2009, from http://www.ocup.org/resources/documents/EDU_GS_binder_010708_BMv2.pdf
- Paquette, J. (2001). Cross-purposes and crossed wires in education policy-making on equity: the Ontario experiences, 1990-1995. *Curriculum Studies*, 33(1), 89-112.
- Pervin, B. (2005). Turning around student achievement in literacy: Ontario's new literacy and

- numeracy strategy. *Orbit*, 35(3) 39-40.
- Priestly, M. and Sime, D. (2005). Formative assessment for all: a whole-school approach to pedagogic change. *The Curriculum Journal*, 16(4), 475-492.
- Ricci, C. (2004). The case against standardized testing and the call for the revitalization of democracy. *The Review of Education, Pedagogy, and Cultural Studies*, 26(4), 339-361.
- Royal Commission on Learning, Ontario. (1995). *For the love of learning: Report of the Royal Commission on Learning* (Vols. 1-4, Short Version and CD-ROM). Toronto: Author.
Retrieved August 14, 2009 at
<http://www.edu.gov.on.ca/eng/general/abcs/rcom/short/short.html>
- Sebatane, E. M. (1998). Assessment and classroom learning: A response to Black & Wiliam. *Assessment in Education: Principles, Policy & Practice*, 5(1), 123.
- Sebba, J. (2006). Policy and practice in assessment for learning: The experience of selected OECD countries. In J. Gardner (Ed.), *Assessment and learning* (pp. 169-183). London: Sage.
- Shavelson, R. (2008). Guest editor's introduction. *Applied Measurement in Education*, 21, 293-294.
- Shavelson, R. J., P. J Black, W. Dylan, and J. Coffey (2004). On linking formative and summative functions in the design of large-scale assessment systems, submitted to *Educational Evaluation and Policy Analysis*.
- Shavelson, R., Young, D., Ayala, C., Brandon, P. Furtak, E., Ruiz-Primo, et al. (2008). On the impact of curriculum-embedded formative assessment on learning: a collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21,

295-314.

- Sheppard, L. (2008). Formative Assessment, Caveat Emptor. In C. Dyer (Ed.) *The future of assessment, shaping teaching and learning* (pp. 279-303). New York: Taylor Francis Group.
- 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.
- Smith, F., Hardman, F., Wall, K., & Mroz, M. (2003). Interactive whole class teaching in the national literacy and numeracy strategies. *Cambridge Journal of Education*, 33(3), 395-411.
- Steinberg, C. (2008). Assessment as an "emotional practice". *English Teaching: Practice and Critique*, 7(3), 42-64.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5.
- Stiggins, R.J. (2005). From formative assessment to assessment for learning: a path to success in standards-based schools. *Phi Delta Kappan*, (Dec.) 324-327.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 169-183). London: Sage.
- Taras, M. (2007a). Machinations of assessment: Metaphors, myths and realities. *Pedagogy, Culture & Society*, 15(1), 55-69.
- Taras, M. (2007b). Assessment for learning: Understanding theory to improve practice. *Journal of Further Learning and Higher Education*, 31(4) 363-371.
- Tierney, R. (2006). Changing practices: influences on classroom assessment. *Assessment in*

- Education*, 13(3), 239-264.
- Torrance, H. (1993). Formative assessment: Some theoretical problems and empirical questions. *Cambridge Journal of Education*, 23(3), 333-343.
- Torrance, H. and 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.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education*, 14(3), 281-294.
- Troman, G. (1989) Testing tensions: the politics of educational assessment. *British Educational Research*, 15(3), 279-295.
- Ungerleider, C. (2006). Reflections on the use of large-scale student assessment for improving student success. *Canadian Journal of Education*, 29(3) 873-883.
- Van der Vegt, R., Smyth, L., & Vandenberghe, R. (2001). Implementing educational policy at the school level: Organization dynamics and teacher concerns. *Journal of Educational Administration*, 39(1), 8-23.
- Volante, L. (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, 35, 1-19.
- Volante, L. (2005). Alternate strategies for large scale student assessment in Canada: Is value-added assessment one possible answer. *Canadian Journal of Educational Administration and Policy*, 41, 1-21.
- Volante, L. (2006). An alternate vision for large-scale assessment in Canada. *Journal of*

- Teaching and Learning*, 4(1), 1-14.
- Volante, L. (2007). Educational quality and accountability in Ontario: Past, present, and future. *Canadian Journal of Educational Administration & Policy*, 58, 1-21.
- Volante, L., Cherubini, L., & Drake, S. (2008). Examining factors that influence school administrators' responses to large-scale assessment. *Canadian Journal of Educational Administration and Policy*, 84, 1-30.
- Welsh Government Assembly Website (2009). *Developing thinking and assessment for learning: development programme*. Retrieved on August 4, 2009 at <http://wales.gov.uk/topics/educationandskills/curriculumassessment/thinkingandassessmentforlearning/?lang=en>.
- Whetton, C. (2009) A brief history of a testing time: national curriculum assessment in England 1989-2008. *Educational Research*, 51(2), 137-159.
- Wideman, R. (2002). Using action research and provincial test results to improve student learning. *International Electronic Journal For Leadership in Learning*, 6(20),
- William, D. (2006). Assessment for learning: Why no profile in US policy. In J. Gardner (Ed.), *Assessment and learning* (pp. 169-183). London: Sage.
- William, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3&4), 283-289.
- William, D., & Thompson, M. (2008) Integrating assessment with learning, What will it take to make it work? In C. Dyer (Ed.) *The future of assessment, shaping teaching and learning* (pp. 53-82). New York: Taylor Francis Group.
- Wolf, P. (2007). Academic improvement through regular assessment. *Peabody Journal of*

Education, 82(4), 690-702.

Wyatt, T. (1996). School effectiveness research: dead end, damp squib or smouldering fuse?

Issues in Educational Research, 6(1), 79-112.

Wyse, D. and Torrance, H. (2009). The development and consequences of national curriculum assessment for primary education in England. *Educational Researcher*, 51(2), 213-228.

Yin, Y., Shavelson, R., Ayala, C., Ruiz-Primo, M. Brandon, P., & Furtak, E. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21, 335-359.