

Educational Quality and Accountability in Ontario: Past, Present, and Future

[Louis Volante](#) (Ph.D.)

Assistant Professor
Brock University
Faculty of Education

ABSTRACT

This paper outlines the genesis, limitations, and future directions for the Education Quality and Accountability Office (EQAO) in the province of Ontario. Recent assessment reforms are analyzed and examined in relation to broader Canadian and international literature. Research describing the impact of Ontario's large-scale assessment programs on students, teachers, and the school system is also reported. The discussion outlines measures for strengthening large-scale assessment within the province and proposes a set of three overarching principles to guide future assessment-led reform.

Introduction

Educational accountability is primarily a relationship between three key stakeholders: Taxpayers, elected officials, and teachers. At the most basic level, taxpayers want to know how the education system is performing and expect the government and schools to provide evidence on the value of their investment. In Canada, as in the rest of the Western world, large-scale assessment programs are increasingly being used as the main, and in many cases, sole indicator of system effectiveness. Teachers, administrators, district leaders, and other educational personnel are becoming more and more preoccupied with improving their relative standing on these external tests. In addition to holding education systems accountable for student learning, these achievement tests are also expected to serve a variety of other purposes, including providing useful feedback for instructional decision making, identifying areas for future action, and serving as a fair selection mechanism for grade promotion and/or graduation (Chudowsky & Pellegrino, 2003; Earl, 1999; Taylor & Tubianosa, 2001). Currently, every province and territory, with the exception of Prince Edward Island, administers some form of large-scale student assessment. The approach of individual provinces and territories varies according to the grades tested, sample size, test format, frequency of administration, and most importantly, stakes attached to student performance. This article focuses on the Ontario context and describes the genesis, limitations, and impact of external testing within this province. The discussion focuses on ways to strengthen and re-position the role of large-scale assessment within and outside of the province. The ultimate objective is to move notions of accountability from the realm of simple number crunching to a comprehensive view focused on authentic system improvement. The latter has been sorely lacking in the current mindset that dominates accountability and assessment-led reform.

Genesis of EQAO

With the exception of a few sample assessments of students during the 1970s and 1980s, Ontario had almost no history of large-scale assessment and none with high-stakes for students, schools, and districts (Earl & Torrance, 2000). This situation

changed dramatically with the publication of findings from the Royal Commission on Learning. The Commission held province-wide discussions with educators, policy makers, parents, students and tens of thousands of citizens in what became one of the most extensive public consultations ever undertaken in the history of Canada (Green, 1998). Of the Commission's 167 recommendations to produce sweeping change in the education system, the fifty-first was the creation of an independent, arm's-length testing agency to be called the Office of Learning Assessment and Accountability (Royal Commission on Learning, 1994). The agency would be responsible for the construction, administration, scoring, and reporting of uniform provincial assessments in both elementary and secondary schools. At the elementary level, the Commission recommended the agency develop two assessments for students in Grade 3, one in literacy and one in numeracy, based on specific learner outcomes and standards that are well known to students, teachers, and parents (Recommendation 50). At the secondary level, the Commission recommended the agency develop a literacy test that would serve as a graduation requirement (Recommendation 52). The reports and recommendations of the Office of Learning Assessment and Accountability would go directly to the Minister and the public (Recommendation 55). Additionally, it was recommended that the Ministry of Education develop detailed, multi-year plans for large-scale assessments (program reviews, examination monitoring), which establish the data to be collected, the way implementation would be monitored, how results would be reported publicly, and how educators and the general public should interpret and use the provincial test results (Recommendation 54).

Collectively, the Commission's recommendations provided the impetus for the creation of the Education Quality and Accountability Office (EQAO) in 1995. With the help of classroom teachers, EQAO created large-scale assessment programs in literacy and mathematics for students in Grades 3, 6, 9, and 10. It is not entirely clear why these grades were specifically selected – particularly since the Commission recommended testing for students in grades 3 and 11. Perhaps a more in-depth analysis of trends, which is possible with closer grade testing intervals, was desired. Nevertheless, the domains tested closely parallel similar large-scale assessment programs within Canada and other Western countries. In general, there has been a

noticeable preference to focus testing on the two key areas of literacy and numeracy. This also holds true for national and international assessment programs such as the Pan Canadian Assessment Program (PCAP), Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), and Program for International Student Achievement (PISA). EQAO is also responsible for coordinating Ontario's participation in these assessment programs.

Currently, EQAO administers tests to Grades 3 and 6 students in the areas of reading, writing, and mathematics. Grade 9 students are tested in mathematics, while Grade 10 students complete the Ontario Secondary School Literacy Test (OSSLT). The latter is considered a high-stakes test for students, since it serves as a graduation requirement. Interestingly, all of the major political parties in Ontario (i.e., Liberal, New Democratic Party, Progressive Conservative) continue to support the overall mandate of EQAO, and have played an important role in its inception or ongoing development. For example, EQAO was initially conceptualized under the tenure of the NDP, was created and funded by the PC's, and is now operating in partnership with the current Liberal government. Thus, despite the rhetoric, the continuities of the PC education policy are more striking than its discontinuities, and mimic broader trends across North America (Gidney, 1999).

The main objectives of these tests are to provide data for both accountability purposes and improved teaching and learning (EQAO, 1998). More specifically, EQAO currently describes their mandate as follows:

EQAO will ensure greater accountability and contribute to the enhancement of the quality of education in Ontario. This will be done through assessments and reviews based on objective, reliable and relevant information, and the timely public release of that information along with recommendations for system improvement (<http://www.eqao.com/AboutEQAO/01about.aspx?Lang=E>).

Wolfe, Childs, and Elgie (2004) noted three main objectives for the provincial assessments: 1) report on results of the test(s); 2) report of the quality and effectiveness of education; and 3) report to accountability boards. Not surprisingly, these three functions closely parallel the purpose and scope of other provincial and territorial assessment systems within Canada.

In order to facilitate improved teaching and learning, EQAO provides teachers and administrators with individual reports that present a profile of a students' performance and a strategy to use the exemplars when talking to parents. These reports are discussed in relation to broader curriculum expectations and other information that is presently available about the child. EQAO also requires districts and schools to prepare their own reports and school improvement action plans based on assessment findings and other information which is likely to affect student learning (e.g., demographics, program descriptions). EQAO asserts that improvement planning is a strategy that brings about educational change by increasing district school boards' and schools' capacity to design and manage change that will improve student outcomes. Collectively, these procedures were enacted to boost the utilization of the provincial assessment data and ultimately spur macro level improvements within the system. Nevertheless, the ability of any assessment program to act as a catalyst for system improvement is heavily dependent on the psychometric properties of the assessment measures.

Psychometric Limitations

At the simplest level, test reliability refers to the consistency of scores while validity refers to the appropriateness of the inferences stemming from the assessment. In Ontario, the current assessment materials and practices suffer from a number of reliability and validity concerns. In their examination of inter-rater reliability, Wolfe, Wiley, and Traub (1999) found a 70% to 80% probability that a students' performance would be marked correctly. Although this result is relatively high in comparison with other large-scale assessments, it does suggest that teachers may be receiving incorrect information for a quarter of their students' responses. Similarly, Wolfe, Childs and Elgie (2004) examined the impact of the number of test items within the assessments and concluded that the testing programs biggest difficulty is their limited number of items. Increasing the number of test items would undoubtedly improve the reliability of the inferences stemming from the assessments. While these studies provide important information for the general public, other forms of reliability such as test-retest reliability,

reliability estimates for various subgroups of the population, and data regarding measurement error are lacking (Crudwell, 2005).

Given that EQAO administers criterion-referenced assessments that are closely aligned with the provincial curriculum; these tests likely have acceptable levels of content validity. However, other forms of validity, such as consequential validity, which examine the impact of external tests on students and teachers, have not been provided nor systematically examined. Not surprisingly, Wolfe, Childs, and Elgie (2004) argued for the introduction and on-going sustainability of an active program of validity research. This is clearly a pressing concern and seems to be supported by voices from the field. For example, board statisticians and assessment personnel have been complaining that EQAO does not publish detailed technical reports to accompany the assessment results. As a result, it is difficult to determine whether difference in the scores from year to year constitute a real difference or are merely an artifact of variations in test difficulty, scoring procedures, or data analysis procedures (Ontario English Catholic Teachers Association, 2002). Given the previous concerns, it is not surprising that many in the education community are questioning the authenticity of the steady gains in test scores over the last five years (see http://www.eqao.com/pdf_E/06/06P034E.pdf for provincial trends since 2001). Clearly, the level of precision in test scores must be determined before any government can boldly assert that student learning is improving. Not surprisingly, those in the measurement community, such as the Joint Committee on Testing Practices (2005), have argued that the level of precision in test scores is the first consideration for developing and selecting appropriate large-scale assessment measures. To date, there is no conclusive evidence to suggest EQAO has satisfied this basic requirement.

Impact of Testing

It should be noted from the outset that recent assessment-led reforms have not been widely embraced by the majority of Ontario's teachers nor their unions. Many educators within the province view provincial assessment with a suspect eye and dispute the taken-for-granted assumption that external testing will lead to system improvement. Teachers point to data which shows that Ontario's provincial assessment results reflect

regional, linguistic, and socio-economic disparities rather than differences in the quality of teaching (Allington, 2000; People for Education, 2002). For them, the millions spent on large-scale assessment programs should be re-invested directly into classrooms, where it would have a more profound and lasting impact (English Teachers Federation of Ontario, 2001; Ontario English Catholic Teachers Association, 2002). Unfortunately, research literature on the impact of large-scale assessment for students, teachers, and the school system is relatively scarce for the Ontario and Canadian context. The ensuing discussion pieces together research from within and outside Ontario as a way to inform our understanding of this important topic.

In general, testing can produce two general types of emotional reactions in students. For one group, testing may cause a hyper-motivation to succeed and provide the necessary impetus to get 'serious' about school. The latter is obviously a desirable objective that proponents view as a positive consequence of large-scale assessment (Cizek, 2001; Covalieskie, 2002). For other students, testing may lead to apathy or lack of a genuine effort, particularly for students who experience significant anxiety and feel they will not be successful (Burger & Krueger, 2003). For these students, not trying serves as a defense mechanism since their poor performance can be attributable to lack of effort, not their low ability. The Ontario context provides an interesting place to examine these two reactions and their effect on student performance; particularly since some assessment results carry significant consequences for students (i.e., graduation requirement) while others do not.

The distinction between low- and high-stakes testing is a key consideration when evaluating the impact of provincial testing. Both academics and practitioners have argued that students are placed at increased risk of educational failure and dropping out when external testing carries high-stakes consequences (American Educational Research Association, 2000; Canadian Federation of Teachers, 1999). Research tends to support this concern. For example, Kane (2002) analysis found that low achieving students are 25% more likely to drop out of school in states that employ graduation tests versus non-tested states. Recent announcements by the Ontario government suggest that the province may be experiencing a similar trend. For example, the high school completion rate was steady in the mid 1990's to 2001 at 78 per cent, but

dropped sharply in 2001 to 71 per cent, and has remained relatively unchanged (People for Education, 2006). The 2001 date is significant since the OSSLT was introduced as a graduation requirement during the 2000/2001 school year. King's (2002) comprehensive study, which included a sample of 49,796 students from 133 schools in 58 districts, provides an important caution for Ontario and other contexts utilizing high-stakes tests for graduation purposes. Namely, he asserted that the high failure rate of 30% on the OSSLT creates an additional burden for 'at-risk' students, effectively stripping away their motivation.

These trends in Ontario are not a surprising finding given that other provinces like Alberta have similar concerns. Despite having one of the countries most advanced assessment systems, Alberta boasts the lowest percentage of high school students entering postsecondary institutions in Canada. The Alberta Teachers Association (2005) has argued that the latter is an unintended result of their accountability system's continuing over-emphasis on high test scores. Clearly, the provincial Ministry of Education needs to re-examine or remove a required 'Pass' on the OSSLT as a requirement for graduation (Ryan & Joong, 2005). Even the recent creation of the Ontario Secondary School Literacy Course – an alternative route for students who repeatedly fail the OSSLT – does little to change the prospect of creating a two-tiered class of graduating students.

As researchers the world over have found, external testing can strongly influence how teachers educate students (Black & William, 1998; Webb, 2005; Wideen, O'Shea, Pye, & Ivany, 1997). Subjects that typically get assessed (i.e., language arts, mathematics, and science) assume greater importance than non-assessed subjects (i.e., music, visual arts, and physical education) or facets of the curriculum (i.e., reading and writing versus speaking and listening). Schools and districts skew their teaching to reflect this value imbalance by narrowly focusing instruction on simulated test activities and content, particularly in cases where the results are made public (Popham, 2001). Thus, even high performing students are robbed of a well-balanced educational experience that promotes a diverse range of knowledge and skills. Teachers in Ontario have not been immune to the previous forces and have reported spending a disproportionate amount of time on tested subjects (Ontario English Catholic Teachers

Association, 2002). In some instances, teachers within this province have indicated they focus much of the second half of the school year on test preparation activities (Meaghan & Casas, 1995). Collectively, the excessive focus on test scores and unhealthy competition between teachers and schools often impedes forms of professional collegiality such as the sharing of resources and best practices (Volante, 2005). Hargreaves and Fink (2006) also provide examples of how this type of competition between schools has a ripple effect in the system so that low-achieving schools often fail to attract and/or lose their most experienced educators.

Despite the previous concerns, limited research in elementary schools has documented some positive effects on teachers. For example, Wideman (2002) found that schools were able to use EQAO data to improve student learning by developing action research projects that were tied to the grades 3 and 6 results. Similarly, Earl and Torrance (2000) found that over 75% of teachers increased their participation in staff development in reading, writing, and mathematics and took advantage of district staff development programs linked to the grade 3 assessment. Overall, their findings suggested that the grade 3 assessment process and recommendations had a noticeable effect on improvement planning and practices in Ontario schools. Green (1998) also reported that over 98% of teachers viewed participation in grade 3 marking as one of the best professional development experiences of their careers. This result was based on a large sample of over 12,000 educators and confirms earlier findings which suggested teachers were changing programs and instruction as a result of their marking experience (EQAO, 1997). Collectively, these findings suggest that positive consequences can result from provincial assessments. Nevertheless, the lack of corresponding research from high schools suggest more work is required, particularly since the OSSLT is used as a high-stakes graduation requirement. Indeed, the previously noted benefits to elementary teachers suggest EQAO must significantly bolster the participation rate of active high school teachers marking the OSSLT.

Reporting Challenges

Perhaps the most insidious challenge facing Ontario's large-scale assessment programs is that their results are typically reported in a manner that far outstretches

their abilities. Not all aspects of student learning may be assessed through on-demand paper-and-pencil tests. Consider the four parameters that comprise literacy: reading, writing, speaking, and listening. Although EQAO assessments do a fairly good job of assessing reading and writing, they are not designed to examine speaking and listening components. This inability to assess many performance-based skills such as speaking clearly, designing a class project, or working effectively in a group, are important limitations that should shape public understanding. Designing more authentic situations for capturing the complexity of cognition and learning requires breaking out of the current paradigm to explore alternative approaches to large-scale assessment for all Canadian provinces and territories (Chudowsky & Pellegrino, 2003).

It seems imperative that the use of test results be well scrutinized and the reasons for testing and communication strategies incorporate the limitations of the results being reported (Burger & Krueger, 2003). In Ontario, one may access district data from EQAO's website. Test results are also widely reported in local newspapers with schools ranked from highest to lowest. This lack of interest in the complexities that shape student performance by the media has led the general public to draw many inappropriate conclusions (Cheng & Couture, 2000). This is despite the fact that research has frequently demonstrated that ranking schools can lead to teacher and administrator abuses, such as cheating (Simner, 2000). Sadly, the Ministry of Education in Ontario mandates the release of data in a manner that encourages such comparisons (Crudwell, 2004). This is despite EQAO's stated opposition to using data to rank schools. Educators have a responsibility to become assessment literate so that they can draw appropriate conclusions and inform the public of misguided and misleading information (Popham, 2004; Stiggins, 2002).

Interestingly, Ontario recently developed the Education Quality Indicators Framework to report on a range of factors impacting student achievement. EQAO (2004) argued that the framework is intended to provide: 1) demographic and other education-related environmental information that will help teachers, administrators, and the public interpret student achievement scores in the context of the school, board and province; and 2) information that can be used by decision-makers at the provincial, board, and school levels for improvement planning as they create the best possible

learning environment for students. The data are derived annually from student, teacher, and principal questionnaires, assessments, and school board student information systems. The Education Quality Indicators Framework data is reported annually, as part of the school, board, and provincial assessment results.

Thus, the Education Quality Indicators Framework provides important information for interpreting provincial assessment results in relation to contextual variables such as socio-economic status and linguistic background. Clearly, more than numerical scores on assessment measures are required if the public is to understand and evaluate the quality of education in the province (EQAO, 2004). Namely, a comprehensive picture of the unique and complex characters of schools, boards, and the province is pivotal. Unfortunately this message may be going unheeded, particularly since some important stakeholders (i.e., parents) tend to be underutilizing the detailed information provided by the provincial testing agency. For example, in their analysis of parental knowledge of large-scale assessment within the province, only 13.5% of parents visited EQAO's website (Mu & Childs, 2005). These authors suggested that in lieu of possible inaccessibility to the Internet, EQAO should make sure information reaches parents in other ways. This information should help clarify appropriate uses and limitations of provincial assessment results, and in doing so, protect students against important decisions based on single test scores.

Comprehensive Framework

Over-reliance on large-scale assessment for accountability has been fraught with flawed assumptions, oversimplified understandings of school realities, undemocratic concentration of power, undermining of the teaching profession, and predictable disastrous consequences for our most vulnerable students (Jones, 2004; Kohn, 2000). This narrow view of educational quality often leads teachers to adopt inappropriate test preparation strategies that produce spurious improvements in test scores that do not reflect authentic student learning (Smith & Fey, 2000). Clearly, if large-scale assessment is to act as a positive force for improved teaching and learning, accountability must be based on comprehensive notions of educational quality. In line with this truism, three overarching principles must be respected when designing and

implementing a provincial/territorial assessment and accountability framework. Namely, educational accountability must be conceptualized as a multifaceted concept, examined in relation to important contextual factors, and negotiated with a range of stakeholders. These principles provide the foundation for meaningful assessment-led reform.

Conceptualizing Educational Quality

Large-scale assessment data is part of an accountability system; it is not the entire system itself (Darling-Hammond, 2004). These measures must be used in conjunction with other data sources if one is to understand the complex nature of our schools. There may even be instances when a district and/or province consider lower assessment scores acceptable in light of improvements in other areas. For example, a significant improvement in the high school completion rate will lead to a larger sample of students writing a particular test. This broader sample will undoubtedly include students who are at the lower end of the achievement scale. Which objective is more worthy: Higher test scores with a restricted sample or lower test scores with a higher school retention rate? Educational leaders need to make sure they see both the forest and the trees when conceptualizing educational quality.

In line with a shift towards broader notions of educational quality, must be recognition that classroom assessment, often referred to as curriculum-embedded assessment, also has an important role to play in shaping views of educational quality. Policymakers who shun classroom assessment data position schools to promote inauthentic forms of learning that do little to equip our students for the challenges of a knowledge economy. What are needed are leaders willing to restore the value imbalance that has often existed between classroom and large-scale assessment (Volante, 2006). These complementary forms of assessment can be utilized to promote meaningful change within a comprehensive accountability framework. Fortunately, research is emerging in pockets of the United States, England, and Australia where both large-scale and curriculum-embedded assessment have been successfully integrated for accountability purposes (Wilson, 2004).

Contextual Factors

Educators, parents, politicians, and the public are all responsible for contributions to the quality of schools, and none of them can be held responsible for things over

which they have little or no control (Earl, 1998). For example, a teacher can hardly take credit for the strong showing of her students on the OSSLT when most of them are gifted and come from affluent households. Conversely, a teacher working in an inner city school with numerous English as Second Language (ESL) students should not be held accountable for poor student performance when her students lack basic resources and fundamental English skills. This is precisely what is occurring in Ontario, as schools with high ESL student populations are consistently ranked the lowest within the province. EQAO assessment results are showing that this gap between ESL and non-ESL students is increasing (People for Education, 2002). This lack of consideration for extraneous variables is occurring despite the recognition that students are ineligible for ESL support after having been in Canada for 3 years, regardless of their ability to communicate.

Crudwell (2005) has argued that a value-added criterion provides the best way to understand these contextual factors when evaluating student performance data. The value-added approach considers these factors, and emphasizes the degree of progress in students when making judgments about appropriate levels of achievement. Essentially, this approach permits an examination of variables schools have control over (i.e., instructional approach) with those they can not control (i.e., school demographics). Thus, the effects of confounding variables are greatly diminished when academic progress is examined through a value-added approach. Although this approach is a more powerful means to improve education, value-added assessment is not without limitations. For example, the requirement for multiple testing points during a school year easily doubles or triples the costs associated with provincial testing programs. These increased costs will undoubtedly create further resistance from teacher unions and even advocates who are concerned about fiscal constraints within the overall education budget. Perhaps one way to circumvent this challenge is to test a smaller sample of students at multiple times during a school year. For example, one-third of the provinces school's could be tested at the beginning, middle, and end of each academic year. This carefully selected sample, that accurately represents each school district, still allows researchers to identify best practices that can be disseminated for the benefit of the entire education community. Similarly, a third of the student population tested at three

critical periods would not lead to prohibitive testing costs. The latter is also in keeping with the general philosophy of using large-scale assessment to support, not control school and system improvement.

Stakeholder Involvement

As Fullan (2003) reminds us, lasting educational change results from an appropriate balance of top-down and bottom-up input. Thus, an effective accountability framework requires an inclusive process that values the perspectives of a diverse range of stakeholders. Too often, top-down reforms are implemented from policymakers that have little, if any, understanding of the daily challenges faced by students, parents, and teachers within our schools. A comprehensive approach brings these stakeholders into the fold and provides both formal and informal mechanisms to hear their concerns. This approach ensures that the indicators of educational quality, which define the system, are widely embraced and valued by those directly affected in practice. Research in England suggests shifting from test targets to consolidated targets that encompass challenges faced by schools is pivotal for sustaining large-scale reform (Earl, Levin, Leithwood, Fullan, Watson, Torrance, Jantzi, Mascall, & Volante, 2003). It seems logical that the nature and scope of these broader objectives must be informed by those directly affected in practice. Although the reforms suggested by the Royal Commission of Learning were initially embraced, the direction and scope of EQAO's mandate continues to provoke resistance from many educators. Thus, ongoing consultation with primary stakeholders is vital for maintaining system stability.

Ongoing collaboration allows us to not only discuss the direction we want our schools to take, but more importantly, examine how we are going to get there. Talking with students, parents, educators, and other primary stakeholders may reveal important factors that stand in the way of academic excellence. Although these conversations will likely produce some predictable suggestions such as improvements in classroom resources, smaller class sizes, and more rigorous forms of community support, others may reveal more novel strategies such as assessment literacy training that may be overlooked by policymakers. Recent research has suggested that such training is a key ingredient to large-scale reform, and has led to improved self-efficacy and instruction amongst teachers (Volante & Melahn, 2005). Thus, a thoughtful dialogue about

ameliorating some of these barriers is an essential aspect of any educational reform agenda.

Future Considerations

To date, EQAO has not adequately documented the lived experiences of teachers and students directly affected by provincial assessment programs. For example, how has provincial assessments affected instruction in tested and non-tested subject areas? How many administrators and teachers presently possess the statistical and assessment literacy to make prudent use of provincial test results? How has testing affected administrator and teacher retention/burnout? What effect has testing had on student learning, particularly for low achieving students? In general, what are the intended and unintended consequences of testing within the Ontario and broader Canadian context? All are worthy questions that need to be addressed more systematically and underscore the importance of examining the consequential validity of the province's assessment programs. Essentially, the assumptions built into provincial assessment systems need to be supported if a strong case is to be made for the validity of their proposed interpretation and use (Kane, 2002).

For the immediate future what is also needed is to study the interactions between large-scale assessment and curriculum-embedded assessment to see how models of assessment that external tests can provide could be made more helpful (Black & Wiliam, 1998). Longitudinal research can make it possible to isolate those aspects of an assessment system that are pivotal for sustaining improvements in teaching and learning and providing accountability information for the public. Similarly, research on other jurisdictions may shed light on how large-scale assessment results have been effectively reported in a manner that is consistent with their limitations. As the preceding discussion suggested, the current practice of ranking schools based on mean scores is unacceptable. As Earl (1999) reminds us, when uncertainty is taken into account, many – sometimes most – differences in raw scores between schools and districts disappear.

Conclusion

Establishing and raising standards, and measuring the attainment of those standards are intended to encourage excellence in education and provide the public with a means for holding our teachers, administrators, and school system accountable. Yet, the preceding discussion suggested that the current basis for judging educational quality and accountability in Ontario is flawed precisely because the province has adopted a myopic view that overemphasizes provincial assessment scores. This is despite the fact that many forms of test reliability and validity have yet to be examined within the provincial assessment system. Clearly, the psychometric properties of the provinces' various assessment programs must be researched more rigorously before an argument can be made for authentic student, school, and/or district improvements in the domains of literacy and numeracy.

Rather than emulate other jurisdictions which rely heavily on large-scale assessment results, Ontario and Canada need to adopt a more comprehensive framework for judging educational quality. Such an approach values teacher's day-to-day classroom work by incorporating curriculum-embedded assessment into our decisions of acceptable student achievement. This type of approach provides policymakers with a more robust analysis of student achievement that is able to consider various performance-based skills essential for future success. The nature and specific details of this synergistic assessment approach must be based on a collective process that values the opinions of diverse stakeholders. 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. The stakes associated with maintaining a top-heavy testing approach are too high – particularly for students who are at-risk *and* those interested in developing the requisite skills to become future leaders within the knowledge economy.

References

- American Educational Research Association (2000). *Cautions issued about high-stakes tests*. Retrieved October 30, 2006, from http://www.education-world.com/a_issues/issues110.shtml.
- Alberta Teachers Association (2005). *Accountability in Education*. Retrieved October 30, 2006, from <http://www.teachers.ab.ca>.
- Allingham, P. V. (2000). The Ontario Secondary School Literacy Test: Mr. Harris's high-stakes version of B.C.'s Foundation Skills Assessment. *English Quarterly*, 32(3-4), 68-69.
- Black, P. & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-48.
- Burger, J. M., & Krueger, M. (2003). A balanced approach to high-stakes achievement testing: An analysis of the literature with policy implications. *International Electronic Journal for Leadership in Learning*, 7(4). Retrieved October 30, 2006, from <http://www.ucalgary.ca/~iejll/volume7/burger.html>.
- Canadian Federation of Teachers. (1999). *Province-wide assessment programs*. Retrieved October 30, 2003, from <http://www.ctf-fce.ca/e/what/other/assessment/testing-main.htm>.
- Cheng, L., & Couture, J. (2000). Teachers' work in the global culture of performance. *Alberta Journal of Educational Research*, 46(1), 65-74.
- Chudowsky, N., & Pellegrino, J. W. (2003). Large-scale assessment that supports learning: What will it take? *Theory into Practice*, 42(1), 75-83.
- Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27.
- Covaleskie, J. F. (2002). Two cheers for standardized testing. *International Electronic Journal for Leadership in Learning*, 6(2). Retrieved October 30, 2006, from <http://www.ucalgary.ca/~iejll/volume6/covaleskie.html>.
- Crundwell, R. M. (2005). Alternative strategies for large-scale student assessment in Canada: Is value-added assessment one possible answer? *Canadian Journal of Educational Administration and Policy*, 41. Retrieved October 30, 2006, from <http://www.umanitoba.ca/publications/cjeap/articles/crundwell.html>
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106(6), 1047-85.

- Earl, L. (1999). Assessment and accountability in education: Improvement or surveillance? *Education Canada*, 39(3), 4-6.
- Earl, L. (1998). Developing indicators: The call for accountability. *Policy Options*, 6, 20-25. Retrieved October 30, 2006, from <http://www.irpp.org/po/archive/jul98/earl.pdf>.
- Earl, L., Levin, B., Leithwood, K., Fullan, M., Watson, N., Torrance, N., Jantzi, D., Mascall, B., & Volante, L. (2003). *England's National Literacy and Numeracy Strategies: Final Report of the External Evaluation of the Implementation of the Strategies*. England: Department for Education and Skills.
- Earl, L., & Torrance, N. (2000). Embedding accountability and improvement into large-scale assessment: What difference does it make? *Peabody Journal of Education*, 75(4), 114-41.
- Education Quality and Accountability Office (2004). *Completing the picture: The education quality indicators framework*. Toronto, Ontario: Author.
- Education Quality and Accountability Office (1998). *Parents Handbook*. Toronto, Ontario: Queen's Printer of Ontario.
- Education Quality and Accountability Office (1997). *Provincial report on achievement; English language schools*. Toronto, Ontario: Queen's Printer of Ontario.
- English Teachers Federation of Ontario (2001). *Adjusting the optics: Assessment, evaluation and reporting*. Retrieved October 30, 2006, from <http://www.efo.on.ca/attachments/adjustingtheoptics.pdf>.
- Fullan, M. (2003). *Change forces with a vengeance*. New York: Falmer.
- Gidney, R. D. (1999). *From Hope to Harris: The Reshaping of Ontario's Schools*. Toronto, ON: University of Toronto Press.
- Green, J. M. (1998). Authentic assessment: Constructing the way forward for all students. *Education Canada*, 38(3), 8-12.
- Hargreaves, A., & Fink, D. (2006). The ripple effect. *Educational Leadership*, 63(8), 16-20.
- Joint Committee on Testing Practices. (2005). Code of fair testing practices in education (revised). *Educational Measurement: Issues and Practice*, 24(1), 23-26.
- Jones, K. (2004). A balanced school accountability model: An alternative to high-stakes testing. *Phi Delta Kappan*, 85(8), 584-90.

- Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, 21(1), 31-41.
- King, A. (2002). *Double Cohort Study – Year 2 Report*. Ministry of Education and Training, Toronto.
- Kohn, A. (2000). *The case against standardized testing: Raising scores, ruining the schools*. Portsmouth, NH: Heineman.
- Meaghan, D. E., & Casas, F. R. (1995). On standardized achievement testing: Response to Freedman and Wilson and a last word. *Interchange*, 26(1), 81-96.
- Mu, M., & Childs, R. (2005). What parents know and believe about large-scale assessments. *Canadian Journal of Educational Administration and Policy*, 37. Retrieved October 30, 2006, from <http://www.umanitoba.ca/publications/cjeap/articles/childs.html>.
- Ontario English Catholic Teachers Association (2002). *Weighing in: A discussion paper of provincial assessment policy*. Retrieved October 30, 2006, from <http://www.oecta.on.ca/pdfs/weighingin.pdf>.
- People for Education (2006). *The Annual Report on Ontario's Public Schools 2006*. Toronto, ON: Author.
- People for Education (2002). *English as a second language in Ontario Schools*. Toronto, ON: Author.
- Popham, W. J. (2004). All about accountability / Why assessment illiteracy is professional suicide. *Educational Leadership*, 62(1), 82-83.
- Popham, W. J. (2001). Teaching to the test. *Educational Leadership*, 58(6), 16-20.
- Royal Commission on Learning (1994). *For the love of learning: Report of the Royal Commission on Learning*. Toronto, ON: Queens Printer for Ontario. Available [online]: <http://www.edu.gov.on.ca/eng/general/abcs/rcom/main.html>.
- Ryan, T. G., & Joong, P. (2005). Teachers' and students' perception of the nature and impact of large-scale reforms. *Canadian Journal of Educational Administration and Policy*, 38. Retrieved October 30, 2006, from http://www.umanitoba.ca/publications/cjeap/articles/ryan_joong.html.

- Simner, M. L. (2000). *A joint position statement by the Canadian Psychological Association and the Canadian Association of School Psychologists on the Canadian press coverage of the province-wide achievement test results*. Retrieved October 30, 2006, from http://www.cpa.ca/documents/joint_position.html.
- Smith, M. L., & Fey, P. (2000). Validity and accountability of high-stakes testing. *Journal of Teacher Education*, 51(5), 334-44.
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-65.
- Taylor, A. R., & Tubianosa, T. (2001). *Student assessment in Canada: Improving the learning environment through effective evaluation*. Kelowna, BC: Society for the Advancement of Excellence in Education.
- Volante, L. (2006). Principles for effective classroom assessment. *Brock Education Journal*, 15(2), 134-147.
- Volante, L. (2005). Accountability, student assessment, and the need for a comprehensive approach. *International Electronic Journal for Leadership in Learning*, 9(6). Full text [online]: <http://www.ucalgary.ca/~iejll/volume9/volume9.html>.
- Volante, L., & Melahn, C. (2005). Promoting assessment literacy in teachers: Lessons from the Hawaii School Assessment Liaison Program. *Pacific Educational Research Journal*, 13, 19-34.
- Webb, T. P. (2005). The anatomy of accountability. *Journal of Education Policy*, 20(2), 189-208.
- Wideen, M. R., O'Shea, T., Pye, I., & Ivany, G. (1997). High-stakes testing and the teaching of science. *Canadian Journal of Education*, 22(4), 428-44.
- Wideman, R. (2002). Using action research and provincial test results to improve student learning. *International Electronic Journal for Leadership in Learning*, 6(20). Retrieved October 30, 2006, from <http://www.ucalgary.ca/~iejll/volume6/wideman.html>.
- Wilson, M. (Ed.). (2004). *Towards coherence between classroom assessment and accountability: 103rd yearbook of the National Society for the Study of Education, Part II*. Chicago: University of Chicago Press.

Wolfe, R., Childs, R., & Elgie, S. (2004). *Final Report of the External Evaluation of EQAO's Assessment Processes*. Toronto, ON: Ontario Institute for Studies in Education of the University of Toronto.

Wolfe, R., Wiley, D., & Traub, R. (1999). *Psychometric perspectives for EQAO: Generalizability theory and applications*. Toronto, ON: Ontario Institute for Studies in Education of the University of Toronto.