

**EXAMINING FACTORS THAT INFLUENCE
SCHOOL ADMINISTRATORS' RESPONSES
TO LARGE-SCALE ASSESSMENT***

**Louis Volante, Lorenzo Cherubini, and Susan Drake
*Brock University***

This study examined factors at the school, district, and provincial level that influenced school administrators' responses to large-scale assessment. To understand administrators' perspectives, 5 secondary and 4 elementary administrators from a suburban school district in southern Ontario, Canada were interviewed. Key factors noted by administrators included school improvement planning (elementary versus secondary), departmental structures and teacher resistance within schools, lack of professional development opportunities, inadequate resources and direction from districts, ministry initiative overload, lack of an easily identifiable provincial assessment and evaluation policy document, and pressure to reach provincial achievement targets. The findings underscore the need for greater initial and ongoing professional development for school administrators as well as more thoughtful support from district leaders and provincial policy-makers.

The development and implementation of accountability systems has been one of the most powerful, perhaps the most powerful, trend in educational policy in the last 20 years (Barber, 2004). In the 1980s, Prime Minister Margaret Thatcher's adoption of a standardized test-based accountability system in Great Britain provided a model for

* This research is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC).

proponents of test-driven reform in the United States, Canada, Australia, and other European nations (i.e., Germany, France). This type of test-driven system is often referred to as standards-based reform and still has considerable appeal to policy-makers in much of the English-speaking Western world. For example, in England, the trend is toward total accountability in education with the introduction of the National Curriculum in 1988 to the National Curriculum assessment system that was introduced in 1996 (Whetton, Twist, & Sainsbury, 2000). England currently measures progress against national standards when students reach the ages of 11, 14, and 16 (Holloway, 2003). League tables that summarize the performance of schools are published by local and national newspapers, attracting a considerable amount of political and public attention. The underlying message conveyed to parents continues to be that they should be relatively satisfied with schools that improve their test performance from year to year, and begin to question the quality of instruction in those that have poor performance. Not surprisingly, accountability has become synonymous with external testing.

In the United States, the federal No Child Left Behind Act (NCLB) requires every state to develop standards, standardized tests and accountability systems, and, by mandating the option for students to transfer from schools with low test performance to those with higher performance, NCLB promotes competition between schools (Amrein & Berliner, 2003; Hursh, 2005). Schools that fail to meet Adequate Yearly Progress (AYP), as reflected in mandated improvements in test scores, experience significant stress as a result of the provisions of NCLB (Darling-Hammond, 2004). Couple this with recent announcements to introduce merit-based pay incentives that are tied directly to test scores

and it is easy to understand why the use of large-scale assessment results for accountability purposes continues to provoke fierce debate in the United States.

External testing also attracts widespread attention within Canada. However, unlike England and the United States, where student achievement testing tends to be high-stakes, the consequences of poor test performance are not as profound for Canadian educators and students. For example, in Ontario, testing is conducted under the direction of the Education Quality and Accountability Office (EQAO). Students are tested in grades 3 and 6 in reading, writing and mathematics, grade 9 in mathematics, and 10 in literacy. The results of these assessments are widely disseminated in a manner that invites comparisons across schools and districts. Nevertheless, only the grade 10 Ontario Secondary School Literacy Test (OSSLT) has very important consequences for students, since it serves as a graduation requirement. However, students who have failed or been excluded from writing this test are eligible to take the Ontario Secondary School Literacy Course to fulfill this requirement (Klinger, DeLuca, & Miller, 2008). Moreover, no administrator or teacher is rewarded with merit pay or officially sanctioned based on high or low test scores at any level within the system. Rather, provincial test results are primarily intended to spur improvements within schools. Overall, the Canadian context is somewhat distinct in its attempt to delicately balance accountability and school improvement purposes of large-scale assessment (Volante & Ben Jaafar, 2008).

The chief analytic objective of the present study was to identify factors at the school, district, and provincial level that influenced school administrators' responses to large-scale assessment. By identifying factors at each of the previous levels, the present study provides policy-makers and other educational leaders with starting points for

fostering improved teaching and learning within Ontario's schools. Given the fairly unique context that Ontario provides in terms of different consequences for test performance at various grade levels, the findings should also be of particular interest to those currently engaged in large-scale assessment reform initiatives. Maximizing the intended positive benefits and minimizing the unintended negative consequences of an assessment system remains a formidable challenge for governments across the Western world.

Consequences of Large-Scale Assessment

Proponents of large-scale assessment point to a variety of intended consequences associated with external testing. Some of the most common arguments for external testing are that mandated large-scale assessment programs force students and teachers to work harder (Anderson, 1990), which in turn leads to improved performance on international assessments such as the Trends in International Mathematics and Science Study (TIMSS) (Bishop, 2001). Large-scale assessment improves teachers' assessment and instructional practices, particularly when they are involved in marking these assessments (Gambell & Hunter, 2004; Green, 1998). Unlike classroom assessment which may vary substantially from teacher to teacher, large-scale assessment allows researchers and senior administrators to identify the most successful teaching practices and proficient teachers (Cizek, 2001; Sanders & Horn, 1998). Similarly, the robustness of the test data provides schools with reliable information about the consequences of their past practices and program effectiveness (Anderson & Postl, 2001; Taylor & Tubianosa, 2001). These arguments are often used to intensify and extend large-scale assessment

programs across different elementary and secondary grade levels and core subject areas (i.e., language arts, mathematics, science, social studies).

Critics counter the previous arguments by citing a large body of research documenting the unintended consequences of testing on students and teachers. Given the nature of these traditional paper and pencil measures, large-scale assessment may undermine the quality of education by penalizing divergent thinking, creativity, and intellectual work in general (Hess & Brigham, 2000; Wideen, O'Shea, Pye, & Ivany, 1997). The elevated importance of individual test scores routinely narrows and distorts the curriculum by encouraging "teaching to the test" techniques which take valuable time away from non-tested subjects, particularly when high-stakes are attached to results (Kohn, 2002; Neil, 2003; Smith & Fey, 2000). Research also suggests that emphasizing large-scale assessment widens the gap between minorities and white students by ignoring key factors known to affect student performance such as socio-economic status, language of origin, and the students' physical and/or emotional health (McNeil, 2000; Sacks, 2000; Valencia & Villarreal, 2003). Lastly, large-scale assessment for accountability purposes supports policy-decisions that have not been adequately scrutinized and are often used as a lever to hold teachers responsible for results with inequitable resources (American Educational Research Association, 2000; Darling-Hammond, 2004; Kane, 2002).

Collectively, arguments for and against large-scale assessment suggests a variety of contextual factors may lead particular testing programs to have a positive and/or negative impact on students, teachers, and the school system in general. Understanding these factors, particularly as they are identified by school leaders, appears to be of utmost importance given the growing trend of standards-based reform across many English

speaking Western nations. A large body of research has noted the relationship between leadership and school effectiveness (see Leithwood & Jantzi, 2006; Leithwood, Seashore Louis, Anderson, & Wahlstrom, 2004; Ross & Gray, 2006). Nevertheless, literature examining factors that influence administrative responses to large-scale assessment and the implications for school improvement planning in a Canadian context remains relatively sparse.

Assessment Leadership

Research consistently demonstrates that school leadership has a measured impact on teacher beliefs and student achievement (Leithwood & Jantzi 2006; Nettles & Herrington, 2007; O'Donnell & White, 2005; Waters, Marzano, & McNulty, 2003). Thus, leadership in utilizing student assessment and evaluation data, particularly as it is exercised within schools, is increasingly being viewed as a pressing concern within North America (Noonan & Renihan, 2006; Popham, 2005; Sheppard, 2000; Volante & Cherubini, 2007). Despite this growing recognition, very little formal training exists for the development of assessment literacy in classroom teachers and school administrators (Stiggins, 1999, 2000; Volante & Melahn, 2005). Assessment literate educators possess 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 (Gallagher & Ratzlaff, 2008). Unfortunately, school administrators are not required to complete a course in assessment and evaluation as part of their training in the vast majority of North American jurisdictions (Arter, Stiggins,

Duke, & Sagor, 1993; Lukin, 2004). The previous statement also applies to the Ontario context where elementary and secondary administrators are required to complete a Principal Qualification Course that provides very little attention in the area of student assessment and evaluation.

Finding ways to strengthen administrators' leadership skills, particularly in response to large-scale assessment programs, requires an analysis of key issues at the school-, district-, and provincial-level. Factors at each of the previous levels may present opportunities and serious challenges to bringing about positive changes in response to large-scale assessment. As previously suggested, few empirical studies exist that examine the perspectives and responses of school administrators to mandated large-scale assessment within a Canadian context. Many of the existing studies also lack attention to important factors outside of individual schools, such as those at the district and provincial level.

Conceptual Framework

Leadership with student assessment and evaluation has often been examined in relation to the knowledge and beliefs of individual teachers (Lukin, 2004; Ross & Gray, 2006), and/or the role of school administrators in promoting data-driven decision making in schools (Johnson, 2000; Marsh, Pane, Hamilton, 2006; Mertler, 2007). Much of this research has focused on school level issues such as the types of summative assessment data commonly utilized for school improvement planning. Little attention has been directed at understanding leadership with student assessment and evaluation as it is effected by district and provincial level factors. The present study uses the conceptual

framework of Volante and Cherubini (2007) to gain greater understanding of the perspectives of a sample of elementary and secondary school administrators receipt of large-scale assessment in relation to school-, district-, and provincial-level factors. Implicit in the framework is the belief that educational leadership must be effectively exercised at the school level if in fact assessment is to contribute to sustainable improvements within schools. In this context, assessment drives school curriculum. Large-scale assessment data, in turn, can serve as a useful diagnostic measure for school administrators. This tri-level framework, therefore, provided a fertile conceptual ground for a re-consideration of the opportunities and constraints of large-scale assessment that embraces and intersects with the leadership capacities within contemporary elementary and secondary schools.

Method

Participants

Participants included 9 administrators: 3 elementary principals, 1 elementary vice-principal, 2 secondary principals, and 3 secondary vice-principals. Teaching experience ranged between 8 and 26 years, with a mean of 15.8 years. Administrative experience ranged between 1.5 and 8 years, with a mean of 5.6 years. Collectively, the administrators had between 16 and 32 years of educational experience and represented approximately one-quarter of the schools within one municipality of the school district. Five of the administrators were males and four were female. The purposeful sampling method (Creswell, 2008) ensured participants represented a range of administrative experience, and were drawn from both genders and school levels (i.e., elementary vs.

secondary) fairly equally. All of the administrators were certified educators with the Ontario College of Teachers.

It should be noted that the majority of participants had a broad understanding of large-scale assessment. Apart from one elementary and one secondary principal, the remaining administrators did not reference other large-scale assessments beyond the Ontario context; instead, their experience in using and subsequently interpreting large-scale assessment was limited to provincial test scores that were the direct outcome of the students in their schools. The interpretive assessments of the test scores respective of their individual schools was supported by the school board research officer who provided statistical and comparative analyses of student achievement in light of previous years and other student cohorts across the district. The two administrators who made reference to large-scale assessments beyond Ontario did so from different perspectives, although neither actually cited specific examples such as the Programme for International Student Assessment (PISA) or the Pan-Canadian Assessment Program (PCAP). In the case of the elementary school principal, Canadian and international student test scores were not necessarily a priority since the School Improvement Plan (referring to a mandatory document prepared by school principals on an annual basis and presented to the school board) does not require principals to juxtapose their school's test scores with other large-scale assessments. From the secondary school administrator's perspective, national and international large-scale assessments represented a further layer of accountability for student performance.

Research Site

The present study was conducted in a suburban school district in southern Ontario, Canada. The student population for this district was mixed and represented a variety of cultures and socio-economic groups. As with other school districts within the province of Ontario, this district possessed mandated school improvement plans in relation to EQAO test results. A noteworthy feature of this district is that it did not possess an official student assessment and evaluation policy at the time of this study, as do some, but not all of the other school districts in the province.

Data Collection

The semi-structured interviews of approximately 60 minutes involved a set of scripted questions. The interview protocol was guided by the work of Rubin and Rubin (1995) and asked a range of general questions related to teaching and administrative experience, large-scale assessment knowledge, professional development, as well as more specific questions related to their responses to large-scale assessment programs. Sample questions included:

- Can you provide a brief description of your current student population?
- Have you had any professional development in the area of assessment and evaluation? Explain.
- How would you rate your confidence level in assessment?
- What factors do you consider when interpreting EQAO assessment results?
- How do you utilize EQAO assessment results for assessment planning and/or instruction?
- What are the positive effects, if any, of EQAO testing in your school?
- What are the negative effects, if any, of EQAO testing in your school?
- What suggestions do you have for improving EQAO testing?

Each of the main questions was also accompanied with a set of follow-up questions designed to elicit more detailed responses (see Appendix A). Nevertheless, there was enough flexibility for participants to go beyond the questions as they chose.

Data Analysis

The data analysis method included a meticulous line-by-line coding of each transcript before identifying the labels and codes that best represented emerging themes. Codes were assigned to each line directly in the margins of the transcripts. Entries with codes having similar meanings were merged into a new category. This constant comparative process was repeated for each of the nine transcripts (Bogdan & Biklen, 2003). Codes from the first transcript were carried over to the second transcript, and so on. This allowed the researchers to note trends across participants. Once the initial coding was completed, the alignment of themes across various levels within the system (including school, district, and province) were examined using domain analysis (Spradley, 1979).

Results and Discussion

This section attempts to capture the most salient findings that emerged from the present study. For ease of comprehension, the results and discussion are organized around the three interrelated levels: school-, district-, and provincial factors.

School-Level Factors

School Improvement Planning. Results suggested a noticeable difference between how elementary and secondary administrators were utilizing large-scale assessment data to promote school improvement. In the majority of cases, however, participants indicated that they had relatively little professional development in the area of assessment and evaluation. In the case of the five secondary school administrators, their most significant professional development occurred during Secondary School Reform in Ontario (late 1990's) when they were in the role of department chairs. As a result, all participants, including the elementary administrators, rated their confidence level in assessment as average citing their inexperience as classroom teachers with the new Ontario curriculum in the post-reform era. Typically, elementary administrators used results to target intervention for particular groups of students. Consider this response:

It's [EQAO results] used for communication for divisional meetings. How to move the kids forward, the ones that aren't at level 3 [provincial achievement standard] ... a lot of our school improvement plan is based on the areas of concern that come from the EQAO item analysis... We use our assessment data to see which kids are at risk ... For example, in the genuine student profile they look at the accommodations checklist or language student reading survey or running records, recording checklist, rubrics, comprehension questions ... and then they store the data on the student profile, almost like the data wall, but it's on the individual basis. So the teacher does the class profile, this is the data wall. (Elementary Principal)

The findings suggested that elementary principals were utilizing large-scale assessment results within an integrated data-driven decision-making framework that considers a variety of different types of assessment data for program planning.

In contrast, secondary administrators primarily utilized the data as a broad benchmark of school performance, and often promoted “teaching to the test” techniques to boost poor performance, particularly in low-achieving schools.

Our school always has the lowest score in the district on the OSSLT, always has, always will. It's a solid 45% lower. We've never been able to get out of that rut. Our scores go up, but everyone else's goes up higher. So as a result it is constantly something that we have to work on. One of our OSSLT projects is what we call the practice test, where we take kids out of class and give them questions from the last year's test and from the past four or five years. (Secondary Principal)

Further probing revealed that the importance of the OSSLT, coupled with the pressure from the publication of school results, compelled many secondary administrators to adopt these types of test preparation practices. Participants recognized that these practices were not necessarily the most prudent use of curriculum time, and as a result were a negative effect of EQAO testing in their schools, yet remained committed to their implementation. According to Popham (2001), such test preparations are not educationally defensible since they often do not lead to transferable knowledge and skills, which is an essential feature of authentic student learning.

It is also worth noting that secondary administrators did not discuss the previous types of test preparation practices in relation to the grade 9 mathematics test. Indeed, administrators made little mention of the grade 9 mathematics test when answering any of the questions related to EQAO testing. Perhaps this is not surprising given that the grade 9 test does not serve as a graduation requirement as does the grade 10 OSSLT. Nor does the numeracy test receive as much attention as the literacy test does in the media. Essentially, secondary school contexts seemed more likely to foster “teaching to the test” techniques in relation to the more salient literacy test versus the low-stakes testing in

grade 9. These findings suggest administrative and teacher responses to large-scale assessment are strongly influenced by the implications attached to test results and the scope of test reporting.

A fairly recent content analysis of 62 Ontario school board improvement plans revealed that only 42% were combining large-scale provincial assessments results with other types of achievement data to better understand and validate the measures of student achievement (van Barneveld, Stienstra, & Stewart, 2006). Although data-driven decision-making is encouraged by EQAO and has been linked to improved student learning and achievement (see Johnson, 2000; Marsh, Pane, & Hamilton, 2006; Mertler, 2007), the present findings suggested that elementary principals were more likely to utilize large-scale assessment results in conjunction with other sources of student achievement information than their secondary counterparts. Perhaps greater professional development must be directed at the secondary level so that these administrators are better able to contend with the pressures of external testing and more thoughtfully integrate large-scale assessment results with other forms of student assessment data.

Departmental Structure & Teacher Resistance. The findings also suggested that the departmental organization of high schools made secondary teachers particularly resistant to change. Consider the following response to a question related to assessment and evaluation practices:

The general challenge in a high school environment ... you're dealing with teachers that really like to take ownership of their own subject area and tend to kind of hang on to it, they're a little resistant to change. (Secondary Vice-Principal)

Other high school administrators made similar comments and used the term “silo effect” to discuss how secondary teachers generally responded to best practice initiatives around assessment and evaluation. Consider the following responses:

You are dealing with the challenge of the silo effect, I'm the teacher, it's my classroom. Yes I have the course outline, I'm going to follow it, but I don't want to be doing it in lockstep with you over there. (Secondary Principal)

The problem is that in education people walk in, teach, and I'm talking about secondary school, they close the door, and they have their own jungle for 75 minutes. So this whole silo effect, pretty much kills any intent to dialogue ... they might just give lip service that they need to improve, and in reality they're not doing that ... it's an attitudinal problem. (Secondary Vice-Principal)

The previous responses suggest that secondary school administrators have to address the implications of teacher resistance and departmental structure by challenging traditional pedagogical models and instructional formats to generate more integrated learning experiences for students. In this sense, educational leadership requires that school administrators exercise both prudent and sensitive judgment. To challenge traditional and arguably embedded institutional structure and practice has genuine implications for teachers above and beyond measures of external assessment. It requires the courage to position student development and achievement at the forefront of school programming at the expense of challenging “what teachers are used to doing.” As an example, school leaders must commission all teachers to share responsibility for promoting literacy skills whether they teach English, social science, mathematics, science, or other subjects. Essentially, responsibility for student achievement must be evenly distributed across the teaching staff.

The previous paradigm shift requires that school leaders embody a commitment that student learning is most successful in contexts that interpret and integrate different types of assessments, including those from different subject areas, to the needs and interests of the students themselves. Classroom-based assessment and evaluation practices are, then, directly connected to this multi-dimensional model of teaching and learning. By establishing best practice initiatives around classroom assessment, students, parents and other stakeholders may be better situated to integrate the different perspectives that classroom and large-scale assessment results provide as complementary measures of student learning. In this vein, both instructional leadership (Brower & Balch, 2005; Ruff & Shoho, 2005; Southworth, 2002) and transformational leadership (Leithwood & Jantzi, 2006; Leithwood & Riehl, 2003), are especially salient for enacting change in contemporary secondary schools. Perhaps the integration of both leadership paradigms (Marks & Printy, 2004) represents an opportunity to facilitate more positive responses to large-scale assessment programs in the face of significant challenges.

District-Level Factors

Lack of Professional Development. Elementary and secondary administrators identified a variety of factors that were pivotal to promoting effective responses to large-scale assessment and school planning. Lack of release time for teachers and limited professional development opportunities for themselves and their teachers were noted as particularly important. Consider the following response from an elementary principal:

I don't know if we've been in-serviced exactly on assessment ... There's not necessarily a program as part of the principalship or anything like that that's directed specifically at assessment. (Elementary Principal)

Similarly, consider the following response from the secondary panel:

Professional development, experience and assessment. "Not much", but probably in comparison to the general population of teachers, probably had more. I know that I'm not anything close to an expert. I guess I have an understanding of the issues but I don't think I have any answers or any clear personal view of where it should go. (Secondary Vice-Principal)

Interestingly, there appeared to be an inverse relationship between years of administrative experience and self-reported assessment knowledge and competence. That is, more experienced administrators explicitly noted that because they had been out of the classroom for a long time, they did not have up-to-date knowledge and skills in this area.

Consider too that the role of school principal has become increasingly complex. In the absence of any specific or targeted district professional development, district or ministry assessment policy, principals are called upon to demonstrate competencies across various tiers of academic, program, and resource outcomes while still thoughtfully operating a school. As a result, they are potentially default-embedded into having a profound understanding of assessment and evaluation (especially in the public eye), while the responsibilities of best practices for assessment and evaluation has essentially fallen more on the shoulders of subject-specific department chairs and divisional heads. This is somewhat troubling given the instructional leadership role that administrators are typically ascribed within schools. The implication is that in the absence of professional development, more experienced administrators may become progressively more out of date with recent developments in the field of assessment and evaluation and may not be able to effectively integrate different types of assessment such as classroom and provincial test data.

Inadequate Resources and Direction. Administrators pointed to a lack of resources and board direction in their attempts to build assessment capacity within their respective schools. Consider the following responses from the secondary panel:

You should be conscious of the fact that our board does not have a written assessment and evaluation policy – I've been waiting for one for four year at least, as I've been principal. I was interested in seeing one also when I was a vice-principal and also in previous years. (Secondary Principal)

I don't see that happening right now. I see teachers doing different things... a lot of different kinds of assessment. I'm comfortable with that, that's good. But I don't think they take the next step. ... What feedback am I getting from the kids about this experience? I think in their [teachers] minds I'm being told by the board to do it this way, so who cares what the kids think. I'm doing it the way I was told. Which goes back to the old thing ... I'm delivering curriculum. I'm not so concerned about checking for understanding. I did my job. I taught you. And I taught you the way the board said to teach you. (Secondary Principal)

The previous responses suggest that district leaders need to provide more targeted direction to school-based leaders that extend beyond mere mechanical and pragmatic considerations to help teachers teach for deep learning and mindfully consider how classroom and large-scale assessment and evaluation practices can inform planning. By moving their attention beyond the glitter-value of achieving test scores that exceed provincial averages, school districts can model for school leaders a focus on large-scale assessment as it contributes to school improvement and not necessarily on the impact of school comparisons within their respective school jurisdictions. The inclusion of a board assessment and evaluation policy document may help promote vertical and horizontal capacity building in relation to student assessment (Volante & Cherubini, 2007). In the absence of more thoughtful district support, capacity building in the prudent use of assessment data will remain a formidable challenge (Cizek, 1995; McMillan, 2000; Trevisan, 1999; Volante, 2007).

Provincial-Level Factors

Ministry Initiative Overload. Responses suggested that the volume of new initiatives coming from the ministry of education was contributing to administrator stress and apathy. Consider the following responses from both panels, when asked how the ministry could assist them to develop best practices around assessment and evaluation:

Reduce the number of initiatives. Give teachers/staffs and opportunity to digest, implement, perfect one or two at a time before introducing another ... perhaps say “no” to some of the Ministry initiatives even if there is a monetary amount attached. The concern should be sustainability of the initiative. (Elementary Principal)

There's so many initiatives coming down the pipe and you only have so much time. I think that's one of the things about the secondary school that you can't rely on. On the administration side there's probably three of you, you've got department heads ... Not everybody is keen on jumping into these initiatives. (Secondary Vice-Principal)

Collectively, administrators in both panels echoed the concern that too many ministry initiatives was resulting in shallow attention. Even more troubling was the suggestion that buy-in to any particular initiative was compromised by the belief that many initiatives would not be sustained over an extended period of time.

Provincial Assessment Policy. Responses suggested that a lack of an easily identifiable provincial assessment policy was a pivotal concern for school administrators. Consider the following response from administrators in the elementary panel:

What would really be helpful is if the ministry came out with an assessment document. Just like we have curriculum policy document, we should have an assessment policy document. It's come out with bits and pieces. (Elementary Principal)

We'd love to have a ministry document on assessment for the elementary level. A one stop shop. (Elementary Principal)

Interesting, the Ministry of Education in Ontario recently initiated a process to develop a comprehensive assessment policy framework for student assessment, evaluation, and reporting. The initial document "Growing Success" was released in January of 2008 with the final assessment, evaluation, and reporting policy framework to be released in September of 2009. As noted within the document, "there is an urgent need to clarify and consolidate, to ensure policy is aligned, consistent, and clear, and that every student in the system is benefiting from the same high-quality process" (Ontario Ministry of Education, 2008, p.2).

Provincial Achievement Targets. Administrators also noted that the pressure of reaching provincial achievement targets coupled with the publication of disaggregated test results as an area of special attention. Consider the following responses from elementary and secondary administrators:

I am concerned that it [provincial assessments] focuses so much on grades 3 and 6 ... a lot of teachers still think, "that's not my problem, it's for the grades 3 and 6 teachers." We need to move away from that... The school is more than just the results of a test, and we have to be careful with the message we're giving to parents. I know some parents that go on the Internet and look at EQAO scores and say "I want to go to that school," and yet it doesn't speak to the rest of the school. (Elementary Vice-Principal)

The provincial targets as stated are a little unrealistic given our clientele ... The literacy team, which I am part of is extremely proactive this year, to put in place those aspects we hoped would mitigate the problem. We have our students in grade 10 take a practice test before they actually write it. We are also going to *have* them do a mock OSSLT test in the cafeteria. So what we are trying to do there is duplicate. (Secondary Vice-Principal)

Responses such as the ones above suggest there are competing tensions between the policy intentions of large-scale assessment and administrators' difficulties in promoting

appropriate test preparation practices and communities of practice that share responsibility for student learning and achievement.

The pressure of achievement targets and the challenges of test preparation have genuine implications on school programming. Pre-tests make a demand on two of the most important commodities related to teaching and learning – time and human energy. To implement a practice test entails in-servicing faculty, informing parents of the practice and justifying its implementation as a prudent use of curricular time, administering the pre-tests during school hours, and evaluating its results. At some point as well, validity considerations have to be accounted for to determine the correlation between pre- and actual test scores. Schools that choose to utilize pre-test practices have to consider the extent to which they skew the accuracy of authentically assessing students' knowledge and skills (see Haladyna, Nolen, & Haas, 1991; Smith & Fey, 2000). District leaders and provincial policy-makers must be proactive in providing the necessary professional development to ensure schools are utilizing educational defensible approaches to standardized test preparation. Failure to do so may undermine that predictive validity of the tests they hold in such high regard and provide the public with a false sense of authentic improvement in student learning and achievement (Volante, 2004). Thus, the impact of traditional reporting techniques for large-scale assessment results need to be accounted for in district wide professional development.

It is difficult, perhaps impossible due to freedom of information laws, to prevent the media from publishing crude rankings of schools. Nevertheless, the Ministry should take the lead in trying to minimize the misunderstanding of large-scale assessment results that inevitably follows the publication of raw test scores. By publicly acknowledging the

existence of measurement error, limitations in single test scores, as well as the socio-economic factors that strongly influence student achievement, the general public may become more assessment literate and better equipped to make informed decisions on the quality of education within individual schools, districts, and the province. Concern for increased test scores should also be tempered by the realization that traditional paper-and-pencil measures are unlikely to tap the richness of skills required for the current knowledge economy.

Conclusion

Mandated large-scale assessment programs are often promoted as a way to improve schools. The present study indicated that a variety of interrelated factors at the school-, district- and provincial-level present significant challenges in bringing this goal to complete fruition. Key factors noted by administrators included school improvement planning (elementary versus secondary), departmental structures and teacher resistance within schools, lack of professional development opportunities, inadequate resources and direction from districts, ministry initiative overload, lack of an easily identifiable provincial assessment and evaluation policy document, and pressure to reach provincial achievement targets. Although some of these challenges are already being addressed (i.e., development of a provincial assessment, evaluation, and reporting policy framework) or likely require fairly straightforward solutions (for example, greater initial and ongoing professional development for administrators and teachers); others require more careful deliberation (i.e., restructuring of secondary departments/schools and making achievement results a collective rather than individual responsibility). By modeling and

promoting effective instructional practices through a shared vision, administrators at all levels of the school system have the potential to foster positive responses to large-scale assessment so that student learning and achievement is improved.

Further, the results of the study underscore for school administrators that provincial test results can be a significant tool in their strategic programming apart from the formalities of School Improvement Plans, School Council Reports, and school-based statistics. More specifically, school administrators can readily distinguish segments of the student population that are not achieving to standard. By focusing on this population, school administrators garner the collective capacities of their staff to build comprehensive student portfolios in the areas of literacy and numeracy, while heightening teachers' appreciation of how these students learn differently. The large scale assessment scores can also be considered with other standards of student progress (including classroom assessment data or other standardized measures such as the Developmental Reading Assessment) to further situate each student's unique achievement, skills, and capacities. The combination of these scores, along with the statistical analyses provided by the board research officer, can be significant measures in identifying intervention programs, resources, and in-school supports specific to each student's unique needs.

The present study also highlighted the impact of the policy context on the prescribed curriculum and teachers' pedagogical practices. Publicly accessible scores and the resulting ranking of schools by the media promoted "teaching to the test techniques" that are unlikely to lead to transferable skills. In the absence of new approaches to large-scale assessment reporting, policy-makers and district leaders need to carefully consider

how to best support administrators and teachers within schools so that they can effectively contend with these pressures. Support itself in many ways represents the cornerstone of school improvement (Leithwood et al., 2004). In this context, administrators and teachers need to be included as critical partners in the process of examining patterns that emerge in classroom and large-scale assessment data within an integrated framework. As a core action team, individuals from all levels of education can identify together the essential features, respective of their students and schools, to promote school effectiveness and improvement. It is ultimately districts and provinces that build capacity in the effective utilization of different forms of student assessment and evaluation that will be properly positioned to realize authentic improvements in teaching and learning.

References

- AERA. (2000). *Cautions issued about high-stakes tests*. Retrieved October 17, 2008, from http://www.educationworld.com/a_issues/issues110.shtml.
- Amrein, A.L., & Berliner, D. (2003). The effects of high stakes testing on student motivation and learning. *Educational Leadership*, 60(5), 32-37.
- Anderson, J. O. (1990). *The impact of provincial examinations on education in British Columbia: General report*. Victoria, BC: British Columbia Department of Education.
- Anderson, B., & Postl, B. (2001, April). *Using large data sets as a basis for school improvement*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Arter, J. A., Stiggins, R., Duke, D., & Sagor, R. (1993). Promoting assessment literacy among principals. *NASSP Bulletin*, 77(556), 1-7.
- Barber, M. (2004). The virtue of accountability: System redesign, inspection, and incentives in the era of informed professionalism. *Journal of Education*, 185, 7-38.
- Bishop, J. H. (2001). A steeper, better road to graduation. *Education Next*, 1(4), 56-61.
- Bogdan, R. C., & Biklen, S. N. (2003). *Qualitative research for education* (4th ed.). Toronto: Allyn & Bacon.
- Brower, R. E., & Balch, B. V. (2005). *Transformational leadership and decision making in schools*. Thousand Oaks, CA: Corwin.
- Catano, N., & Stronge, J. H. (2006). What are the principals expected to do? Congruence between principal evaluation and performance standards. *NASSP Bulletin*, 90(3), 221-237.
- Cizek, G. J. (1995). The big picture in assessment and who ought to have it. *Phi Delta Kappan*, 77, 246-249.
- Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27.
- Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106, 1047-1085.
- Gallagher, C. W., & Ratzlaff, S. The road less travelled. *Educational Leadership*, 65(4), 48-53.

- Gambell, T., & Hunter, D. (2004). Teacher scoring of large-scale assessment: Professional development or debilitation? *Journal of Curriculum Studies*, 36, 697-724.
- Green, J. M. (1998). Authentic assessment: Constructing the way forward for all students. *Education Canada*, 38(3), 8-12.
- Haladyna, T.M., Nolen, S.B., & Haas, N.S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.
- Hess, F. M., & Brigham, F. (2000). None of the above: The promise and peril of high-stakes testing. *American School Board Journal*, 187(1), 26-29.
- Holloway, J. H. (2003). A global perspective on student accountability. *Educational Leadership*, 60(5), 74-76.
- Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline of educational equality. *British Educational Research Journal*, 31, 605-622.
- Johnson, J. J. (2000). Data-driven school improvement. *Journal of School Improvement*, 1(1), 16-19.
- Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, 21(1), 31-41.
- Klinger, DeLuca, & Miller, (2008). The evolving culture of large-scale assessments in Canadian Education. *Canadian Journal of Educational Administration and Policy*, 76. Retrieved September 1, 2008, from <http://www.umanitoba.ca/publications/cjeap/articles/klinger.html>.
- Leithwood, K., & Jantzi, D. (2006). Transformational school leadership and large-scale reform: Effects on students, teachers, and their classroom practices. *School Effectiveness and School Improvement*, 17, 201-227.
- Leithwood, K. A., & Riehl, C. (2003). *What we know about successful school leadership*. United Kingdom: National College for School Leadership.
- Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning: Review of research*. New York: Wallace Foundation.
- Lukin, L. E. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 23(2), 26-32.
- Marks, H. M., & Printy, S. M. (2004). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, 39, 370-397.

- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. Santa Monica, CA: Rand Education.
- McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, 7(11). Retrieved August 14, 2007, from <http://pareonline.net/getvn.asp?v=7&n=8>.
- McNeil, L. (2000). Creating new inequalities: Contradictions of reform. *Phi Delta Kappan*, 81, 729-734.
- Mertler, C. A. (2007). *Interpreting standardized test scores: Strategies for data-driven instructional decision making*. Thousand Oaks, CA: Sage.
- Neil, M. (2003). High stakes, high risk: The dangerous consequences of high-stakes testing. *American School Board Journal*, 190(2), 18-21.
- Nettles, S. M., & Harrington, C. (2007). Revisiting the importance of the direct effects of school leadership on student achievement: The implications for school improvement policy. *Peabody Journal of Education*, 82, 724-736.
- Noonan, B., & Renihan, P. (2006). Demystifying assessment leadership. *Canadian Journal of Educational Administration and Policy*, 56. Retrieved March 14, 2007, from <http://www.umanitoba.ca/publications/cjeap/articles/noonan.html>.
- O'Donnell, R. J., & White, G. P. (2005). Within the accountability era: Principals' instructional leadership behaviours and student achievement. *NASSP Bulletin*, 89(645), 56-71.
- Ontario Ministry of Education (2008). *Growing success: Assessment, evaluation, and reporting – Improving Student Learning*. Toronto, ON: Queen's Printer for Ontario. Retrieved March, 11, 2008, from http://www.ocup.org/resources/documents/EDU_GS_binder_010708_BMv2.pdf.
- Popham, W. (2001). Teaching to the test. *Educational Leadership*, 58(6), 16-20.
- Popham, W. (2005). *Assessment for educational leaders*. Boston: Allyn & Bacon.
- Ross, J. A., & Gray, P. (2006). School leadership and student achievement: The mediating effects of teacher beliefs. *Canadian Journal of Education*, 29, 798-822.
- Rubin H. J., & Rubin, I. S. (1995). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage.
- Ruff, W. G., & Shoho, A. R. (2005). Understanding instructional leadership through the mental models of three elementary school principals. *Educational Administration Quarterly*, 41, 554-577.

- Sacks, P. (2000). Predictable losers in testing schemes. *School Administrator*, 57(11), 6-9.
- Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12, 247-255.
- Sheppard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Smith, M. L., & Fey, P. (2000). Validity and accountability of high-stakes testing. *Journal of Teacher Education*, 51, 334-344.
- Southworth, G. (2002). Instructional leadership in schools: Reflections and empirical evidence. *School Leadership and Management*, 22(1), 73-91.
- Spradley, J.P. (1979). *The Ethnographic Interview*. New York: Holt, Rinehart, & Winston.
- Stiggins, R. (2000). Learning teams for assessment literacy. *Orbit*, 30(4), 5-7.
- Stiggins, R. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, 81, 191-198.
- 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.
- Trevisan, M. S. (1999). Administrator certification requirements for student assessment competence. *Applied Measurement in Education*, 12(1), 1-11.
- Valencia, R. R., & Villarreal, B. J. (2003). Improving students' reading performance via standards-based reform: A critique. *The Reading Teacher*, 56, 612-621.
- van Barneveld, C., Stienstra, W., & Stewart, S. (2006). School improvement plans in relation to the AIP model of educational accountability: A content analysis. *Canadian Journal of Education*, 29, 839-854.
- Volante, L. (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, 35. Retrieved August 24, 2008 from <http://www.umanitoba.ca/publications/cjeap/articles/volante.html>.
- Volante, L. (2007). Educational quality and accountability in Ontario: Past, present, and future. *Canadian Journal of Educational Administration and Policy*, 58. Retrieved November 21, 2007, from http://umanitoba.ca/publications/cjeap/articles/volante_educational%20quality.html.

- Volante, L., & Ben Jaafar, S. (2008). Profiles of educational assessment systems worldwide: Educational assessment in Canada. *Assessment in Education, 15*, 201-210.
- Volante, L., & Cherubini, L. (2007). Connecting educational leadership with multi-level assessment reform. *International Electronic Journal for Leadership in Learning, 11*(12). Retrieved December 6, 2007, from: <http://www.ucalgary.ca/~iejll/>.
- 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.
- Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Aurora, CO: Mid-Continent Research for Education and Learning.
- Whetton, C., Twist, E., & Sainsbury, M. (2000 April). *National Tests and Target Setting: Maintaining Consistent Standards*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Wideen, M. R., O'Shea, T., Pye, I., & Ivany, G. (1997). High-stakes testing and the teaching of science. *Canadian Journal of Education, 22*, 428-444.

Appendix A

Sample follow-up questions were designed to elicit more detailed responses from the participants. These included:

- To what extent do you feel comfortable delivering professional development to your staff in the area of assessment and evaluation?
- Do you feel that the statistical output of your school's test scores by the school board research officer is sufficient in meeting your purposes?
- How willing are teachers to be involved in interpreting EQAO assessment results?
- How would you describe the interest level of the members of your School Council in the school's EQAO results?
- Have you noticed a change in their level of interest since the first year of EQAO testing in Ontario?