UNDERSTANDING THE CONNECTIONS BETWEEN

LARGE-SCALE ASSESSMENT AND SCHOOL IMPROVEMENT PLANNING[1]

Louis Volante and Lorenzo Cherubini, *Brock University*

This study explored how teachers and school administrators connect large-scale assessment results with school improvement planning. Using a semi-structured format, 20 teachers and 17 administrators were interviewed from two school districts in southern Ontario, Canada. The interview protocol contained a range of questions related to teaching and administrative experience, large-scale assessment knowledge, professional development, and instructional planning in response to large-scale assessment results. Analysis of the interviews followed a constant comparison method and suggested few educators, particularly at the secondary level, are using large-scale assessment results in a sophisticated fashion for data-integrated decision-making. The implications of the findings are discussed in relation to professional development, capacity building, and instructional leadership.

Introduction

The utilization of large-scale assessment results for accountability purposes is undeniable within Western educational jurisdictions. Countries such as the United States, England, Canada, Australia, and other European nations such as France and Germany have developed accountability systems that put a strong emphasis on improved test results (Black & Wiliam,

---

2005). In North America alone, every state and province administers external tests which serve as a broad benchmark of district and school effectiveness (Volante & Ben Jaafar, 2008). Some have argued that external testing represents one of the few policy levers that can spur improvements in elementary and secondary schools (Anderson, MacDonald, & Sinnemann, 2004; Barber, 2004). Using widely reported performance data, administrators and teachers are compelled to improve their instructional planning to the benefit of students, schools, and society in general. Yet skepticism exists whether educators possess the requisite skills to use this information in meaningful ways. Some have suggested that large-scale assessment may do more harm than good if it is not carefully considered in relation to other forms of student information (Shirley & Hargreaves, 2006). This study attempts to understand how classroom teachers and school administrators conceptualize their use of large-scale assessment results to inform school improvement planning. The present study was conducted in two school districts in southern Ontario, Canada. As in most Western educational jurisdictions, these districts were situated within a policy context that places a strong emphasis on performance data for accountability purposes.

Large-Scale Assessment and School Improvement

One of the most formidable challenges with the administration and interpretation of large-scale assessment results is how to use this information to spur improvements in schools. At the policy level, the results are meant to hold schools accountable by measuring the degree to which specific standards are meant. The latter is usually accomplished by noting the percentage of students who meet or exceed a specified state or provincial standard in reading, writing, and/or mathematics. These statistics, while helpful in providing a broad metric of student

achievement, offer little to individual schools or teachers in terms of refining their practice. In order to improve pedagogy, educators must disaggregate the data for their student groups, and seek ways to address achievement concerns. If done properly, this type of analysis and corresponding intervention may help close the achievement gap for some of our most vulnerable student populations. Research overwhelmingly supports this relationship between prudent data use and school improvement (see Earl & Torrance, 2000; Heritage & Chen, 2005; Sutherland, 2004; Timperley, 2005; Wohlstetter, Datnow, & Park, 2008).

An emerging body of literature is beginning to document the opportunities and challenges educators confront when trying to make sense of large-scale assessment results (Decker & Bolt, 2008; Ingram, Seashore Louis, & Schroeder, 2004; Marsh, Pane, & Hamilton, 2006). For the most part, this literature has tended to assert the importance of data-driven decision-making as a general characteristic of successful schools. However, we also know that more specific skills such as the capacity for data disaggregation, understanding the degree to which large-scale results align with classroom assessments, and the use of appropriate intervention approaches are fundamental skills for educators (Heritage & Yeagley, 2005; Lachat & Smith, 2005; Mertler, 2007; Ross & Gray, 2008). Essentially, educators must be reflective about their practice (Schon, 1987) and the various forms of data that can be used to refine their teaching (Earl & Katz, 2006; Hayes, & Robnolt, 2007).

The present study attempted to gain a better understanding of the types of analyses administrators and teachers likely engage in when confronted with large-scale assessment results for school improvement planning. The authors were specifically interested in understanding how educators described their use of different forms of data to inform their instructional planning approach. In order to accomplish the latter, a semi-structured interview format was utilized.

Twenty teachers and 17 administrators were individually interviewed and asked a range of questions related to their teaching and administrative experience, large-scale assessment knowledge, professional development, and instructional planning in response to large-scale assessment results. A key objective of the study was to add to the growing literature on this topic but also identify potential gaps in educators' suggested use of data which could inform future professional development and capacity building efforts. The next section provides a brief summary of the context of the study before explaining the theoretical framework and methodology that guided the research.

Context of Study

Large-scale assessment in Ontario is conducted under the direction of the Education Quality and Accountability Office (EQAO). Students are tested once per year in grades 3 and 6 in reading, writing and mathematics. High school students are tested in grade 9 mathematics and complete the Ontario Secondary School Literacy Test (OSSLT) in grade 10. Overall, these criterion-referenced assessments are meant to provide a broad metric of student achievement and used to spur improvements in schools, particularly those that are achieving below the provincial standard (level 3 on a 4 point scale). Schools that consistently under-perform are given extra assistance from the Ministry of Education through the Ontario Focused Intervention Partnership (OFIP). OFIP funds are primarily used to deploy student achievement officers across the province and for districts to hire literacy and numeracy coaches and to provide job-embedded professional learning opportunities for their teachers.

It is difficult to categorize these large-scale assessments as low- or high-stakes given the traditional parameters that are used in the literature. For example, only the OSSLT has 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. Nevertheless, large-scale assessment data is highly salient in Ontario with the ranking of schools widely reported in the local media.

School board improvement plans contain a strong emphasis on large-scale assessments as a gauge of educational quality in both elementary and secondary schools (Volante & Ben Jaafar, 2008). In their analysis of 62 Ontario school board improvement plans developed in 2003-2004, van Barneveld, Stienstra, and Stewart (2006) found that only 31% actually made reference to classroom data while all the districts considered EQAO scores of chief importance for guiding instructional and school planning. Ontario's favoritism of large-scale assessment data for driving school improvement appears, like many other jurisdictions in Canada, to be a deeply rooted practice.

Theoretical Framework

Teachers and school administrators' utilization of large-scale assessment data for planning purposes can take various forms. One way to examine this relationship is to examine the level of disaggregation of large-scale assessment data along with the inclusion of classroom assessment data for planning purposes. The importance of disaggregation of large-scale assessment results and the integration between large-scale and classroom-based assessments for school improvement planning are fundamental data literacy skills noted by the provincial assessment office (EQAO, 2005) and also by the broader literature (see Popham, 2005; Wilson,

2004). Using these two critical dimensions, a general taxonomy was developed that considers the lowest level response, one that involves the examination of large-scale results in isolation from other forms of student data (Volante, 2008). Here teachers and administrators make adjustments to teaching and planning on the basis of general test scores in particular subject areas. The second level is similar to the first with the exception that large-scale assessment data is disaggregated for particular student groups (special needs students, English-as-a-Second-Language students, distinct ability groups, etc). The third, and highest, level involves the integration of disaggregated large-scale assessment results with other forms of student assessment information. Educators at the third level make sophisticated teaching/planning decisions based on multiple, and at times, contradictory forms of student assessment information. The third level has been coined data-integrated decision-making (Volante, 2008). This taxonomy provided the overarching theoretical framework that informed the development of interview questions and guided the analysis of findings.

Method

Participants

Participants were selected using a convenience sample method across two school districts in southern Ontario, Canada. The sample consisted of 37 educators; n=17 administrators (11 elementary, 6 secondary) and, n=20 teachers (9 elementary, 11 secondary). Administrative experience ranged between 1 and 20 years, with a mean of 6.0. Teaching experience ranged between 2 and 27 years, with a mean of 11.1. Educators were drawn from 24 schools; 15 elementary and 9 secondary. Sixteen of the participants were male and 21 were female. It is

important to note that many of these participants were recommended by senior district personnel for possessing a range of experiences and interest in student assessment.

Research Site

This study was conducted in two school districts located in the Golden Horseshoe – an area around the western end of Lake Ontario, mainly the south-central region of the province. Half of the population of Ontario lives in or around this area. The student population for both districts was mixed and represented a variety of cultures and socio-economic groups. As with other school districts within the province of Ontario, both districts possessed mandated school improvement plans in relation to provincial large-scale assessments. In general, these districts were selected based on their student diversity, which is a typical feature of schools located within the Greater Toronto Area.

Data Collection

The semi-structured interviews entailed 12 lead questions and lasted approximately 60 minutes. The interview protocol was guided by the work of Rubin and Rubin (1995), and as previously mentioned, contained a range of general questions related to teaching and administrative experience, assessment knowledge, professional development, as well as more specific questions related to their utilization of large-scale assessment data for school improvement planning. Key questions included:

- What does formative assessment mean to you and what does it look like in your classroom/school?

- What does summative assessment mean to you and what does it look like in your classroom/school?

- How much time do you spend on formative assessment? Summative assessment?

- Please explain your professional development experience in assessment and evaluation.

- How do you/teachers in your school utilize EQAO assessment results for school improvement planning?

- What are the effects of EQAO testing in your school?

- Overall, how would you rate your competence level in assessment?

Each of the questions was tailored for administrators and teachers and was accompanied with a set of probes designed to elicit detailed responses. It is important to note that cohort groups (elementary teachers, elementary administrators, secondary teachers, secondary administrators) were *not* interviewed sequentially – rather individual participants were interviewed on a day/time that fit best with their work schedule. Thus, no order effects can be attributable to the data collection strategy.

Data Analysis

   Analysis of the interviews followed a constant comparison approach (Creswell, 2008). 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 process was repeated for each of the remaining transcripts. Codes from the first transcript were carried over to the second transcript, and so on. This allowed the researchers to note trends across administrators and teachers. Once the initial coding was completed, the researchers examined the alignment of themes with various types of decision-making. The researchers also re-examined divergent responses, with the intent of potentially revising the tri-level conceptual framework.

As previously noted, the lowest level response involved the examination of large-scale results in isolation from other forms of student data. For example, codes that suggested teachers and/or administrators were making adjustments to teaching and planning *solely* on the basis of general test scores were aligned with a level one response theme. Conversely, codes that indicated a progression in thinking by disaggregating assessment results for particular student groups (special needs students, English-as-a-Second-Language students, distinct ability groups, etc) were aligned with a level two responses theme. Only codes that suggested participants were *both* disaggregating large-scale assessment results and integrating the results with other forms of student assessment information when planning were aligned with the third, and highest category. Validity of the findings was determined through triangulation of the data, member check of the transcripts, clarification of the researchers' biases, and the inclusion of discrepant information (Anderson & Arsenault, 2000; Creswell, 2008; Fraenkel & Wallen, 2005).

Results & Analysis

The results of this study identify how classroom teachers and school administrators described their use of large-scale assessment results to facilitate school improvement planning. It is important to point out at the onset that none of the patterns reported could be traced back to a particular teaching and/or training background. That is, educators that offered responses that aligned with a level one, two, or three response, did not come from a particular curriculum and instruction focus (e.g., mathematics, science, language arts, special education, etc) or have a common set of professional development experiences. Rather, the findings suggested patterns of responses somewhat aligned with particular cohort groups (i.e., elementary teachers, elementary administrators, secondary teachers, secondary administrators), irrespective of one's teaching or

prior professional development experiences. It is also important to acknowledge that the findings are based on participants' perspectives, rather than an analysis of actual classroom and/or administrative practice. Nevertheless, the findings provided a window into the overarching schemata – organized pattern of thought – that frame educator responses to large-scale assessment results.

Level One Response

In various contexts all of the Secondary Administrators (SA) indicated that the large-scale assessment data was factored into their school improvement plans. This cohort suggested that the external assessment data constituted one component of the "concrete numbers and data [that] you want to use to see where you need to go" for school planning purposes (SA-1). For the Secondary Administrator cohort, this low-level response to external data is considered a scripted model that distinguishes students' results in isolation of other assessment data and contextual variables. As one participant stated, "We look at the data, see where we have to work, [and] what we need to work at" (SA-2). Their suggested analysis seemed to neglect or account for other contextual variables that may impact upon students' outcomes.

Secondary administrators suggested that the test results identified specific student needs and therefore put the onus on them as school leaders to ensure that student improvement in these areas was the shared responsibility of all staff:

> We have a team of teachers working on things like that [EQAO test results]. We
> have a team of teachers who are looking at the grade 9 practice tests, and they are
> doing the moderated marking. They are going to be seeing where the difficulties
> are showing themselves, and then planning for next year to make sure that the kids
> are developing the skills that they have noticed are weak. When we looked at the
> grade 10 results from last year we were definitely seeing a pattern. Our kids, who
> were not successful, were better writers when they were answering reading

questions. So that is something that we can look at and incorporate into our school growth plan. (SA-6)

Elementary and secondary administrators also acknowledged their responsibility to align the results of the large-scale tests to their annual school improvement plan. One participant, typical of others, explained that "what I do once we get the results [from EQAO] is take them and go to the EQAO website. I pull them up and look at the various aspects of the test [to see] how our school scored" (SA-3). Accordingly, they orchestrated their school resources to suit the emerging student needs as indicated by the assessment scores.

In all cases the secondary administrator participant cohort underscored the need to prepare students in advance for the range of skills and competencies that the grade 10 literacy test requires. In varying degrees of formal implementation, participants stated the importance of "kids writ[ing] a mock test in grade 9...For four or five days we free up a teacher to work around literacy in preparation for the literacy test" (SA-5). Another participant justified the investment of curricular time to prepare for the external assessments by explaining, "You know the students will face these questions in EQAO and you want to give them the best shot they can, so you have to give them opportunities" (SA-4). This cohort emphasized the importance of addressing the specific competencies inherent in the literacy test well in advance of its actual administration.

The Secondary Teacher (ST) cohort's use of external data was also predominantly indicative of a first-level response. Secondary teachers reviewed students' baseline profiles as they were reported by the test results. One participant's response was typical of the others from this cohort: "We go to the data [and identify the specific students] who were unsuccessful" (ST-6). These participants used the external assessment data to create a synopsis of students' needs. In this context the literacy test "pinpoints areas of the testing where our students have problems.... For example, our students have trouble making inferences and with simple things

like multiple-choice questions" (ST-8). Many of these participants are involved with school

literacy teams or school improvement committees that deconstruct the test questions as per the

students' responses: "We have actually taken the grade 9 math results and looked at each

question specifically as to what students had difficulty with, and how we can improve" (ST-10).

The same individual admitted that "it was a great activity" in which to be involved, but was not

useful for the other teachers since "the grade 9 math teachers were not there" (ST-10). Consider

as well this statement from another secondary teacher: "I am on the student success team so the

EQAO assessment results are definitely something that are brought up at the meeting and are

shown, broken down, and analyzed a little bit" (ST-11). Secondary teachers recognized the

utility of analyzing students' specific responses on the test itself.

Ultimately, the vast majority of this participant cohort reported that they are "not aware"

of the processes in the school that connect external assessment data to other forms of student data

(ST-8). For the most part, secondary teachers' knowledge of data disaggregation and integration

is a product of information shared at professional development and staff meetings. One

participant confessed,

> I will be honest with you. The only experience that I have had with EQAO [and]
> connections to my teaching are from the school EQAO committee…and the ideas
> they give. Other than that it is only when we have our staff meetings where they
> tell us how many students passed and how many did not. That is really the extent
> of it. (ST-4)

In numerous instances, therefore, secondary teachers were passive recipients of test result data as

their awareness of student outcomes was generally determined by the extent to which such

information was disseminated during faculty meetings.

Secondary teachers' level one response to large-scale assessment data was influenced by

their skeptical approach to testing in the first place, and how it adversely effects curricular time

in schools. One participant's reflection captured the sentiment of the others: "I truly believe that the school and I do not necessarily know if it is this school or if it is other schools, just uses the results to aid you in teaching for the test the following year" (ST-5). In many instances participants shared their experiences with teaching to the test practices: "A week before the literacy test we do a really intensive sort of road show where we are going through and teaching all of the areas that we are wanting [students] to improve…I have never personally analyzed the results myself" (ST-7).

Secondary teachers also expressed reluctance to engage in curricular practices that funneled student' competencies towards external assessments. One participant was candid in stating, "With EQAO results, there is a lot of teaching to the test which in itself defies all common assessment practices. But that is what is going on. Some schools are getting tremendous results in their EQAO scores and therefore it must be a good school. But in reality?" (ST-9. The previous response underscored the underlying suspicion that many secondary teachers shared regarding the reliability of provincial assessment results in lieu of test preparation practices.

In only one instance did an elementary administrator suggest that individual item responses were principally used for school improvement target-setting. However, this participant qualified that she is a new administrator in the school and is committed to analyzing "the trends in [the data] and to understanding how to hook it back to curriculum" (EA-1). This participant, however, has the intention of incorporating a more sophisticated response to data analysis in the school: "We understand that these kids are not giving up. When we start to look at their answers, we compared it to what they are doing in class" (EA-1). In fact, she has begun to implement level two-type response initiatives that are more in line with her cohort.

The Elementary Teacher (ET) cohort's responses were also characteristic of a level one response to assessment data. As one participant stated, "We look through those EQAO results and take them to see what areas we did poorly in and then we build goals from there" (ET-1). Reminiscent of the other comments, one participant suggested,

> When we get the results we sit down as a division [a division might include primary: grades 1-3, junior: grades 4-6, or intermediate: grades 7-10] and have a look at where we have been and where we are going…. And that is where you can draw the conclusions and base our school growth plan on...that is pretty much a one-shot deal. You bring it in, you look at it…because it really is a snapshot that gives you more of a general direction. (ET-4)

Although the elementary teacher cohort reported their practice of evaluating EQAO scores on different fronts, they unanimously suggested that they did so in isolation of other forms of data.

Some elementary teachers explained that data interpretation was the responsibility of school improvement committees who then provided direction for the classroom teachers in their respective division:

> We have a school improvement team that has looked at the data from EQAO and were able to come up with SMART goals…. So we have looked at it as individual teachers, divisions and from a school wide [perspective]. We know where we are strong and try not to give up any of the areas that we are doing well but try to strengthen and improve upon the areas that tend to be weaker. (ET-2)

It is important to note that school improvement teams are made up of teachers from different grades, not just those from grades 3 and 6.

Overall, the elementary teacher cohort provided responses to large-scale assessment that were geared towards practical information and were used mainly to identify gaps in their curriculum and student learning. In line with the others, one individual stated, "we found that some of the grade 3 students were low in comprehension which meant that we needed to reinforce those strategies in the lower grades" (ET-8). A different participant expressed the same type of response: "We meet and look at the results – we set goals and that is our plan for the

year. The following year we see where we need to improve – say if we did better in math then next year we will work more on the language" (ET-9). The data represented a means to pinpoint the curricular concerns that influenced students' results.

Level Two Response

The results suggested little evidence that secondary administrators were disaggregating external assessment data for specific student groups. For a few of the secondary administrators, test preparation practices lent themselves to a slightly more sophisticated response to large-scale assessment data. These participants were approaching a level two response since they factored their "school board's focus on the identified students" in terms of "using assisted technologies" to support specific students, and subsequently "focused on applied students'[community college/vocational stream]" results to identify gaps in teacher's instruction:

> We really hammered the grade 10 applied teachers saying, 'look, you guys have to pick it up. The academic students [university stream] are naturally able to do this, but the applied have to be taught. You cannot expect them to be able to do a news report unless they have had practice. (SA-4)

Another individual from this cohort seemed to approach a level two response, indicating that their school has "a plan for our academic kids, a plan for our applied level kids [and] a plan for the kids who we have identified in grade 9 as struggling" (SA-5). Similarly, another secondary administrator stated: "We are data-driven. We are looking at the data and then we are putting specific focus on how to build success with these kids" (SA-6). The focus on at-risk students was a distinguishing factor for many school administrators' responses.

Like the secondary administrator cohort, few secondary teachers were approaching a level two response. The two individuals who were at least conceptualizing a more advanced level response explained that the data is analyzed in his department by grade, subject area, and level:

> Whether a student is in the applied level, whether he is academic, [or in] a special needs program… [they ask about] the accommodations being provided for these students…. Within the department we look at what instructional strategies are working, what is the best practice, what works for this group of students…and within groups of students. (ST-1)

These participants expressed a greater tendency to focus on students' skill development in the context of their external test results and their classroom performance. They did not over-simplify the test data in their explanations of student needs and instructional practice. Instead, they accounted for the data by examining their practices relative to external and internal outcomes.

Secondary teachers perceived their responses to the data as a gainful employment of their time given the contributions they could make to student learning. As one secondary teacher stated,

> We turned ourselves inside out to make sure we were using data and then turning that into an effective preparation that could be measured in better results…. To take their results, map the data, give them [teachers] back materials they can use to change instruction…we'll also identify the kids [who are struggling] specifically by name. (ST-2)

Similar to their administrative colleagues, some secondary teachers' responses also indicated a concern for at-risk students.

While elementary teachers' responses were not genuinely indicative of level two or three responses to data analysis, the elementary administrator cohort did identify various level two practices. This cohort identified a regimented process of data analysis that included aligning Individualized Item Reports (IIR) [reports from EQAO that provide an itemized analysis of test scores] to "making connections" with, as one participant described, individual students (EA-2). Students, thus, "have a hard time making these real life deep connections… it is tough because… particular groups of kids may not be able to make that connection" (EA-2). One administrator stated that the staff's inexperience with examining data in the context of specific student groups

left her to "disaggregate the data myself" (EA-8). In all level two response cases, elementary administrators expressed their preference to examine "different ways [of] looking at data" particularly for those populations that have "a large special needs population" (EA-10). In similar contexts, therefore, elementary administrators' recognized the potential impact of "varying student populations [and] what they have been exposed to outside school" upon both the external large-scale assessment results and teachers' evaluation outcomes.

Level Three Response

While none of the secondary administrators identified a level three response, one secondary teacher provided responses that suggested a high level of data analysis. This participant expressed an awareness of the impact to weave assessment data with "a whole lot of other areas in which we gather data and evaluate data, assess it, and take it into the classroom. That is the key – taking the numbers and taking the data we have collected and putting it into practice in the classroom" (ST-1). This individual recognized the potential of the cross-departmental dialogue of different types of data analysis within the school. This process of integrating disaggregated test results with other student data would, according to this participant, illuminate various intersections that would be critical for school improvement: "Building collaboration and understanding about what is taking place, not only in our own classrooms but in the classrooms of our colleagues, and bringing that all together as qualitative and quantitative data is what is important" (ST-1).

Although representing a relatively small number of the total population of this study, three elementary administrators reported a level three response to large-scale assessment data. In these instances the elementary administrators factored curriculum benchmarks, national large-

scale assessments, and standardized reading assessments into their analysis of the results of

large-scale assessments:

> We do PM benchmarks in grade 3 and CASIE in grades 4 to 8 [standardized
> literacy tests focusing on reading and comprehension]. We put them [the results]
> up on our data wall in the learning resource teacher's room. From there we can
> gather data from all of the tasks that have been given that have been either teacher
> administered like the spelling inventory and various pieces of writing…and any of
> the other pieces such as the EQAO…in the early Fall we can develop profiles for
> a class of students and for individual students. We include the report card marks
> as well. (EA-4)

This type of response represents a clearly defined notion of data management. It also represents

an assessment paradigm whereby various forms of student data are not mutually exclusive and

instead contribute to a rich data set.

This sample of elementary administrators made a concerted effort to process the data in

teams or as a staff. "We go through it and I say is there anything we can take from this and we do

it as divisions…we incorporate that into something that we are already doing" (EA-5). Another

individual credited the talents of the school planning team that lead her to "value more than just

the EQAO data" (EA-6). This individual uses the EQAO data as a premise for asking her staff

"some focus questions" before asking them to engage in more profound topics "that are deeper

about where we need to go" (EA-6). Working from this complex conceptual perspective allows

this administrator to "open up" discussion amongst her staff and account for "other information

we need to have in order to make good decisions" (EA-6). Good decision-making included, for

these participants, an analysis of the assessment data in light of the school curriculum. As one

Elementary Administrator suggested, "you should see the parallels between EQAO and the

curriculum…it lets us [school staff] talk about the different things that we can do" (EA-7). By

accounting for large-scale results and other forms of student data, some teachers "woke up to the

fact that they have to be flexible, reflective, and individualized to help meet student needs" (EA-

7). According to this participant, therefore, assessment data and external tests forced some teachers to be reflective and re-evaluate their pedagogical practices.

Discussion

The results of this study suggested that few educators conceptualized the use of large-scale assessment results in a manner consistent with data-integrated decision-making. Participant responses suggested that external data is typically not disaggregated for particular student groups or examined in relation to other forms of student data. The main finding is consistent with the earlier work of van Barneveld, Stienstra, and Stewart (2006) that noted the dominance of EQAO tests scores for school planning in 2003-2004. Thus, the current policy context (i.e., salience of EQAO) and organizational context (i.e., role of school improvement planning committees) likely acted as a powerful mediator of participants discourse during our study. It is difficult to say whether our participants' responses closely aligned with their actual classroom practice. This relationship between perspective and practice represents an important area for future study. Despite the previous limitation, the present study does suggest the need for more focused professional development so that educators, particularly at the secondary level, can make better use of large-scale assessment data.

The overall pattern from the secondary panel may be partly due to the larger size and departmental organization of high schools. Namely, there may be a greater diffusion of responsibility when educators are not instructing students in tested areas versus smaller elementary schools that tend to share responsibility for student success within grade level and division teams. Thus, it is important to acknowledge that secondary schools are fundamentally different structures from their elementary school counterparts and that departmental

19

specialization is a key element in understanding differences with respect to instructional and assessment expertise (Sisken, 1990). Perhaps, secondary schools may require a different professional development model to enact meaningful changes in response to large-scale assessment results.

Despite some of the differences across cohorts in our study, the present findings suggest modeling data-integrated planning may be helpful to many teachers. This is particularly the case since many of them tend to have idiosyncratic assessment practices and conceptions of teaching that are often resistant to change (Brown, 2004; Ingram, Seashore Louis, & Schroeder, 2004). It seems, as well, that for teachers and the majority of educators, large-scale external assessment has historically been perceived as fundamentally disconnected from their classroom practices. Stated differently, the realities associated with standardized testing for Ontario educators certainly have some implications in terms of rationalizing school and classroom interventions, but are not necessarily pivotal considerations for how teachers and administrators cultivate their planning and pedagogy. This is not to suggest that current practitioners would deny the inherent value of using data to inform their instruction; however, eliciting this kind of response can be very difficult if such practices are not readily known to educators.

School administrators are commissioned to support teachers' professional development, and also keep them accountable to the purposeful integration of different forms of student data. For this to occur, teachers and administrators have to appreciate the value of systematic, informed, and multi-faceted data analysis as the experiences of the relatively few elementary administrators responses attest to. For these participants, their sophisticated responses are indicative of advanced data literacy. These educators recognize how the authenticity of student achievement data generated in their classrooms can in fact be advanced by other forms of student

assessment to develop a clearer understanding of students' strengths and areas of concern. As the results suggested, finding ways to promote data literacy remains a formidable challenge within contemporary schools.

Although targeted professional development initiatives in this particular context may seem too financially costly given the size of Ontario, the Ministry of Education and the various school districts within this province already have much of the infrastructure in place with literacy and numeracy coaches in individual schools, assessment literacy coordinators in school districts, and student achievement officers deployed by the province. Using this broad range of expertise to develop more sophisticated responses to large-scale assessment would be a prudent investment given the current assessment capacity within Ontario's schools. This type of professional development could easily be woven into existing in-services that teachers routinely experience and supported in the provincial assessment policy document *Growing Success: Assessment, Evaluation, and Reporting – Improving Student Learning* that was initially released by the Ontario Ministry of Education in 2008 and subsequently revised in 2010.

The available research strongly suggests that school leaders with a strong background in instructional design and assessment are pivotal for school success (Copland, 2003; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Leithwood & Riehl, 2003; Southworth, 2002). Unfortunately, the majority of administrator responses in the current study were indistinguishable from teachers. Nevertheless, the findings are also somewhat encouraging since they identified three elementary administrators, who appear to be using a reflective team analysis process, to promote data-integrated decision-making. Findings from this relatively small group of administrators underscore the importance of promoting shared ownership of improvement targets.

It is reasonable to assume that most educators will naturally default to a simplistic planning approach given the salience of large-scale assessment data and the diffusion of responsibility via school improvement planning committees. Nevertheless, the small group of elementary administrators suggested conceptualizing data in a more sophisticated fashion is still possible in such policy contexts. Collectively, these administrators, along with other teachers and administrators in other school districts, represent cases worthy of future inquiry. Understanding how these individuals developed their planning capacities should inform the design of future professional development efforts in this area. These "best practice cases" also have the added advantage of being context specific, which is common problem when sharing professional development and/or school improvement strategies from significantly different school districts.

Conclusion

Faced with increasing accountability, schools and districts are implementing a variety of methods for gathering, storing, analyzing, and reporting different forms of data, but they are moving forward with paltry amounts of guidance (Wayman, 2005; Wayman & Springfield, 2006). The present study affirms this criticism and suggested direction must be provided to enhance educators' use of large-scale assessment data. Overall, few educators, particularly at the secondary level, were conceptualizing large-scale assessment results in manner consistent with data-integrated decision-making. Some elementary administrators were employing an advanced response by integrating the disaggregated large-scale assessment results with other student assessment data and used this information for instructional planning. Developing the instructional leadership skills of school administrators and the overall assessment capacity of teachers is vital if schools are to make prudent use of assessment data for school improvement

planning. Given the millions of dollars that are spent every year on large-scale assessments across North America and much of the industrialized world, greater attention to effective planning is warranted.

References

Anderson, G., & Arsenault, N. (2000). *Fundamentals of educational research* (2nd ed.). London: Falmer.

Anderson, B., MacDonald, D. S., & Sinnemann, C. (2004). Testing and school improvement: Can measurement of results help improve the performance of schools? *Phi Delta Kappan, 85*, 735-739.

Barber, M. (2004). The virtue of accountability: System redesign, inspection, and incentives in the era of informed professionalism. *Journal of Education, 185*, 7-38.

Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal, 16*, 249-261.

Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education, 11*, 301-318.

Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis, 25*, 375-395.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Decker, D. M., & Bolt, S. E. (2008). Challenges and opportunities for promoting student achievement through large-scale assessment results. *Assessment for Effective Intervention*, *34*, 43-51.

Earl, L., & Katz, S. (2006). *Leading schools in a data rich world*. Thousand Oaks, CA: Corwin.

Earl, L., & Torrance, N. (2000). Embedding accountability and improvement into large-scale assessment: What difference does it make? *Peabody Journal of Education, 75*(4), 114-141.

EQAO. (2005). *Using data to promote student success (Grades 3 & 6 assessments of reading, writing and mathematics).* Toronto: Queen's Printer of Ontario.

Fraenkel, J. R., & Wallen, N. E. (2005). *How to design and evaluate research in education*. Princeton, NJ: McGraw-Hill.

Hayes, L. L., & Robnolt, V. J. (2007). Data-driven professional development: The professional development plan for a reading excellence act school. *Reading Research and Instruction, 46*, 95-119.

Heritage, M., & Chen, E. (2005). Why data skills matter in school improvement. *Phi Delta Kappan, 86*, 707-710.

Heritage, M., & Yeagley, R. (2005). Data use and school improvement challenges and prospects. *Yearbook of the National Society for the Study of Education*, *104*, 320-339.

Ingram, D., Seashore Louis, K., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*, 1258-1287.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, *112*, 496-520.

Klinger, D., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian Education. *Canadian Journal of Educational Administration and Policy*, *76*. Retrieved from http://www.umanitoba.ca/publications/cjeap/articles/klinger.html

Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk, 10*, 333-349.

Leithwood, K. A., & Riehl, C. (2003). *What we know about successful school leadership*. United Kingdom: National College for School Leadership.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. Santa Monica, CA: Rand Education.

Mertler, C. A. (2007). *Interpreting standardized test scores: Strategies for data-driven instructional decision making*. Thousand Oaks, CA: Sage.

Ontario Ministry of Education. (2008). *Growing success: Assessment, evaluation, and reporting – improving student learning*. Toronto: Queen's Printer for Ontario. Retrieved from http://www.ocup.org/resources

Ontario Ministry of Education. (2010). *Growing success: Assessment, evaluation, and reporting in Ontario's schools, first edition, covering grades 1-12*. Toronto: Queen's Printer for Ontario. Retrieved from http://www.edu.gov.on.ca/eng/policyfunding/growSuccess.pdf

Popham, W. (2005). *Assessment for educational leaders*. Boston: Allyn & Bacon.

Ross, J. A., & Gray, P. (2008). Alignment of scores on large-scale assessments and report-card grades. *Alberta Journal of Educational Research*, *54*, 327-341.

Rubin H. J., & Rubin, I. S. (1995). *Qualitative interviewing: The art of hearing data.* Thousand Oaks, CA: Sage.

Schon, D. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass.

Shirley, D., & Hargreaves, A. (2006). Data driven to distraction. *Education Week, 26*(6), 32-33.

Sisken, L. S. (1990). *Different worlds: The department as context for high school teachers* (Report No. pp.90-126). Washington, DC: Office of Research and Improvement. (ERIC Document Reproduction Services No. ED338592)

Southworth, G. (2002). Instructional leadership in schools: Reflections and empirical evidence. *School Leadership and Management*, *22*, 73-91.

Sutherland, S. (2004). Creating a culture of data use for continuous improvement: A case study of an Edison Project School. *American Journal of Evaluation, 25*, 277-293.

Timperley, H. S. (2005). Instructional leadership challenges: The case of using student achievement information for instructional improvement. *Leadership and Policy in Schools, 4,* 3-22.

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. (2008, August). Testing achievement. *Education Today, 8*, 22-23.

Volante, L., & Ben Jaafar, S. (2008). Profiles of education assessment systems worldwide: Educational assessment in Canada. *Assessment in Education*, *15*, 201-210.

Wayman, J. C. (2005). Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk, 10*, 295-308.

Wayman, J. C., & Stringfield, S. (2006). Data use for school improvement: School practices and research perspectives. *American Journal of Education, 112,* 463-468.

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.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness & School Improvement*, *19*, 239-259.