ABSTRACT: Renewed calls for greater accountability within schools have led to a rapid expansion of standards-based reforms across the Western world. Establishing and raising standards, and measuring the attainment of those standards, are intended to encourage excellence in our schools. Yet concern is increasing about the fairness of external testing and the undue pressure it places on students and teachers. This paper discusses the impact of recent standards-based reforms and proposes a framework to connect educational leadership with timely assessment reforms at the provincial/state governance, university, district, and school level. The proposed multi-level reforms are grounded in the belief that coherent, well integrated assessment policies are essential for promoting improved teaching practice and student learning.

Introduction

Renewed calls for greater accountability within schools have led to a rapid expansion of standards-based reforms across the Western world. Countries such as the United States, England, Canada, and Australia have all developed educational standards that are primarily measured in relation to large-scale student assessment results. Proponents claim that educational standards coupled with external testing increases accountability and allows for greater instructional consistency (Cizek, 2001; Covaleskie, 2002). Schools are provided with valuable information about the consequences of their past practices and the effectiveness of their programs (Anderson & Postl, 2001; Taylor & Tubiansa, 2001).

Additionally, student performance across schools can be compared with the aim of identifying the most successful teaching practices and proficient teachers (Crundwell, 2005; Sanders & Horn, 1998). The latter is typically accomplished through a value-added assessment approach that measures the impact particular teachers or schools have on student achievement over several years.

Yet concern is increasing about the fairness of external testing and its potential negative impact on public education (Earl, 2003; Jones, 2004; Kohn, 2000). Large-scale achievement testing can result in several unintended consequences for students, teachers, and school systems as a whole. For example, external testing tends to narrow the curriculum by emphasizing basic skills as it deskills teachers who teach to the test (Perkins-Gough, 2004; Popham, 2001). Similarly, large-scale assessment programs tend to detract from authentic forms of teaching and learning since they often do not include performance-based tasks or tap higher-order thinking skills (Neil, 2003; Volante, 2004). Student motivation for learning is also adversely affected and there is a higher dropout rate when high stakes (i.e., graduation) are attached to test results (Amrein & Berliner, 2003; Miller & Tovey, 1996). Lastly, external tests can be culturally biased favoring middle class White students, and thus widen the achievement gap (Lee, 2004; Platt, 2004).

In spite of these concerns, large-scale assessment is a part of a larger picture of educational accountability and school improvement that has become entrenched within North American society. Darling-Hammond (2003) shed light on the complexities of improved student achievement when she argued that standards-based initiatives were expected to spur other reforms. Such reforms included high quality curriculum frameworks, materials and assessments tied to standards, more widely available course offerings reflecting this high quality curriculum, more intensive teacher preparation, professional development guided by related standards of teaching, more equalized resources for schools, and more readily available safety nets for needy students. She noted that at least 47 American states created standards for student learning and many followed a comprehensive approach to enhance multiple aspects of education. Recent research suggests that student achievement improved--but only in states using multiple student measures (Darling-Hammond, Rustique-Forrester, & Pecheone, 2005). This paper discusses the impact of recent standards-based reforms and proposes a framework to connect educational leadership with timely assessment reforms at the provincial/state governance, university, district, and school level. The ultimate objective is
Standards-Based Reform within the Western World

In the United States, large-scale assessment programs are mandated by the federal No Child Left Behind (NCLB) act. This act requires all elementary students between Grades 3 and 8 to be tested annually, with the public release of results. The primary purpose of NCLB is to toughen standards for schools, teachers, and students, and increase public accountability for the education system. Schools that fail to meet Adequate Yearly Progress (AYP), as reflected in mandated improvements in test scores, are labeled as "failing," and may eventually be taken over by the state. Since its inception in 2002, critics have argued that NCLB is a regressive approach to education that fails to consider the complexities of student achievement (Gonzalez, 2002; Hershberg, Simon, & Lea-Kruger, 2004; Kornhaber, 2004).

In England, the Standards and Effectiveness Unit (SEU) of the Department for Education and Skills (DfES) has developed standards for students in kindergarten to Grade 9. These standards are tied to the National Literacy and Numeracy Strategies (NLS and NNS), and require annual assessments for students in each of these grades. Nevertheless, the pivotal measures that receive the most widespread attention are the public rankings of schools in relation to achievement on Key Stage 1 (end of Grade 3), Key Stage 2 (end of Grade 6), and Key Stage 3 (end of Grade 9) tests. DfES has argued that the literacy and numeracy standards along with national targets provide an important lever to improve classroom practice and pupil learning in schools across England. Not surprisingly, the high profile of the national targets led some educators to adopt counter-productive teaching activities such as narrowing of the curriculum that were directed at increasing these highly publicized scores (Earl, Levin, Leithwood, Fullan, Watson, Torrance, et al., 2003). In line with Darling-Hammond's (2003) research, Earl et al. (2003) argued for a shift toward consolidated targets that encompass more than large-scale literacy and numeracy assessment scores.

Unlike the United States and England, which have federal departments of education, Canada's education system is the responsibility of provincial and territorial governments. Despite this important difference, every province and territory, with the exception of Prince Edward Island, has developed provincial standards that are tied to large-scale assessment measures. The approach of individual provinces and territories varies according to the grades tested, sample size, test format, and frequency of administration. Although the impact of these large-scale testing programs is mixed, research within Canada has highlighted similar concerns to those expressed in the United States and England. Research within this country has also noted significant narrowing of the curriculum in response to provincial assessments (Cheng & Couture, 2000; Levinson, 2000; Wideen, O'Shea, Pye, & Ivany, 1997). The latter is in spite of the fact that provincial testing tends to carry more important consequences for Canadian students than it does for teachers. For example, provincial testing is currently used as a graduation requirement (i.e., Ontario) or forms a significant percentage of a senior high school students' final grade (i.e., 50% in Alberta, Newfoundland, and Quebec) (Volante, 2006). Nevertheless, test scores have not been used to determine teachers' progression through the ranks or serve as the basis for merit pay increments as in parts of the United States. Thus, even low-stakes testing contexts for teachers can still produce similar curriculum implementation concerns.

Similar to Canada, Australia's education system is the responsibility of the state or territory. Each of its six states (Queensland, New South Wales, Victoria, South Australia, Western Australia, and Tasmania) and two territories (Australian Capital Territory and Northern Territory) has developed their own respective curriculum frameworks, standards, and assessment procedures. Unlike the preceding countries, high-stakes assessment and reporting typically includes classroom-based assessments (also known as curriculum-embedded assessment) in addition to centralized external tests. Nevertheless, the increased push for more vigorous forms of accountability recently motivated the development of national literacy and numeracy benchmark assessment procedures for students in Grades 3, 5, and 7. Although the nature of these tests vary across the country, raw data for the different states and territories is equated through statistical and qualitative judgment procedures to enable comparisons of performance against the national benchmark standards (Cumming & Maxwell, 2004). As in England, the results of these assessments are reported in relation to the percentage of students that achieve the expected level.

Multi-Level Assessment Reform

Given the public's desire for accountability of school systems, standard-based reforms and external testing seem here to stay. Instead of engaging in a debate of the value of standardized testing, we argue that we need to maximize the potential for positive outcomes emerging from both classroom and large-scale assessment. For example, Black, Harrison, Lee, Marshall, and William (2004) report on "assessment for learning" in the classroom. Through a series of projects, they have shown that teachers can teach well and also produce good test results from their students. They emphasize such things as questioning techniques, feedback without grades, peer assessment, self-assessment, and
the formative use of summative tests as instructional strategies. Earl (2003) and Earl and Katz (2004) also describes the relative importance of assessment of learning (summative), assessment for learning (formative), and assessment as learning (the assessment is not graded but acts as a learning tool). Stiggins (2004) calls for new ways to think about assessment since high-stakes tests without supportive environments harm struggling students. For him, the answer is a balance between classroom and large-scale assessment in a synergistic system. Common in all these visions is teachers' view of assessment as pedagogy that is readily integrated into their best instructional strategies. They recognize different purposes for assessment and use them accordingly.

In order to achieve an effective balance across these multiple purposes and modes of assessment, educational leadership must be appropriately exercised at the provincial/state governance, university, district, and school level. Multi-level reforms that promote coherent, well integrated assessment policies offer the best prospect for realizing improved teaching practices and student learning within our schools. This relationship between educational leadership and policy coherence is pivotal for the successful implementation of large-scale reforms (Fullan, 2005; Leithwood, Jantzi, Earl, Watson, Levin, & Fullan, 2004; Leithwood, Seashore Louis, Anderson, & Wahlstrom, 2004; Spillane, 1998).

Provincial/State Governance Leadership

Policy-makers at the provincial/state governance level play a critical role in fostering multi-level assessment reform. They have the power to significantly influence the current Western zeitgeist that scorns classroom assessment data and reduce the deleterious effects that accompany the over-reliance on large-scale assessment results. By integrating a range of curriculum-embedded assessment measures for accountability purposes, state superintendents and provincial ministers of education can help focus attention on where it is needed most—improving the reliability and validity of classroom assessment data. This type of assessment data is the most timely and relevant to changing instruction and improving student learning (Stiggins, 2002). Another advantage of integrating classroom assessment data for accountability purposes is that it invites teachers to become partners in the reform agenda by re-affirming the importance of what they do on a day-to-day basis. Sending teachers to work for a school year and giving disproportional importance to an external test given on one day is a recipe for failure and disillusionment with the entire testing process. Although the latter is not indicative of every jurisdiction in North America, the testing requirements of NCLB and the central importance of many provincial assessment scores (i.e., Alberta, British Columbia, and Ontario) strongly suggests that classroom assessment data is often overshadowed for the purposes of maintaining public confidence and educational accountability in general.

Of course, classroom assessment practices must be strengthened for the public to have sufficient trust in the robustness of the data. For practicing teachers, this means that professional development must be targeted toward improving their assessment literacy. Research suggests that when governments have made these types of significant investments, improvements in student performance and teachers' self-efficacy levels have been realized - particularly in the neediest school districts (Volante & Melahn, 2005). New teachers also require teacher induction programs that help build their assessment competence. These beginning teachers deserve special attention since they initially grapple with challenges such as lesson planning, classroom management, and generally becoming socialized within the profession. For them, assessment is often an afterthought rather than the beginning point of instruction.

University Leadership

Despite the central importance of assessment and evaluation, there are many teacher education programs that do not require completion of a separate course in classroom assessment. Many pre-service programs have chosen to address this topic within their curriculum and instruction classes which results in an uneven knowledge base for recent graduates. What becomes the responsibility of many faculty members may get the attention of none of them. Thus, it is not surprising that assessment training at the pre-service level has been regarded as woefully inadequate (Popham, 2004). Research also supports this claim as pre-service candidates have identified this as an area of special concern (Campbell & Evans, 2000). Given this knowledge gap, many pre-service programs should be re-designed to give more careful attention to assessment and evaluation issues. Indeed, there are many faculty members who themselves would benefit from more intensive assessment and evaluation training.

Universities can also facilitate improvements in classroom assessment and evaluation by offering in-service teachers opportunities to sharpen their skills in this area. Ironically, both authors currently work in a province whose teacher accreditation body fails to sanction assessment and evaluation courses within continuing education departments. Since the demand for non-accredited courses is low, universities have not developed/offered sections in this area. This unfortunate cycle could easily be rectified by having accreditation bodies mandate the delivery of in-service assessment and evaluation courses at all levels-national, state, provincial, and territorial. Providing these types of course offerings for practicing teachers also imparts the message that developing assessment knowledge and skills is an integral part of becoming an effective educator. Such skills are no less worthy of attention than taking an
Teachers may also assume team-leadership roles to support literacy and numeracy initiatives on a school-wide basis.

**Substantive and Engaging Learning Experiences for Students**

Teachers can integrate processes of improvement rooted in developmental needs of their students (Cherubini, 2006). Teachers can be empowered to be proactive in devising what they deem to be appropriate action plans best suited for the diverse student population that are not achieving to standard. School leaders can determine strategic programming that includes not only intensive student capacity development in the areas of literacy and numeracy, but an embedded awareness of students' fundamental knowledge and learning needs. In this case, as with other measures of standardized assessments that identify common standards of student progression (for example, the Developmental Reading Assessment and Reading Running Records in the primary grades) the data can be instrumental in exposing the amorphous components of a school's curriculum while identifying intervention programs, resources, and in-school supports to sustain student improvement initiatives. Such a process is instrumental in complementing the developmental phases of students' academic and social growth.

**Large-Scale Assessment Practices**

As previously noted, assessment tends to drive the curriculum, particularly in high-stakes testing contexts (Kohn, 2000; Popham, 2001; Stiggins, 2002; Volante, 2004). In this light, large-scale assessment data can be a useful point of departure as a diagnostic piece for school administrators. Test scores can help identify trends and segments of the student population that are not achieving to standard. School leaders can determine strategic programming that includes not only intensive student capacity development in the areas of literacy and numeracy, but an embedded awareness of students' fundamental knowledge and learning needs. In this case, as with other measures of standardized assessments that identify common standards of student progression (for example, the Developmental Reading Assessment and Reading Running Records in the primary grades) the data can be instrumental in exposing the amorphous components of a school's curriculum while identifying intervention programs, resources, and in-school supports to sustain student improvement initiatives. Such a process is instrumental in complementing the developmental phases of students' academic and social growth.

**District Leadership**

The most pressing challenge for senior school board administrators is to constitute a positive alliance against the overriding public misperception that distorts the results of large-scale assessments. Among the objectives for district leaders is to facilitate a progressive discourse between educational stakeholders that details the goals and outcomes of standardized testing practices in relation to not only improving scores, but to enhancing overall student learning. In a collaborative effort with schools, district leaders can be catalytic in mobilizing a paradigm shift from discussing large-scale assessment practices as isolated and insulated events towards acknowledging their role as one means (and not the only means) of examining trends in student learning and performance. It is fair to deduce that school personnel, and not district officials, are in the eye of the storm as far as public accountability is concerned. District leaders can be the filters of interpretation so that large-scale assessments do not assume a disproportionate importance in the district. They are also pivotal in supporting principals to gauge the political complexity that results from the dubious process of publicly ranking schools.

In more concrete terms the onus rests on district leaders to provide sufficient resources, human capital, and adequate time for administrators and teachers to legitimize their work to, in turn, address student learning and development. This translates into timely and relevant in-services where teachers are relieved of their classroom duties and given opportunities to collaborate on what constitutes best practice. Beginning teachers who are assigned to teach in tested grades that are particularly high-stakes (i.e., Grades 3, 6, 9 in England) may especially deserve extended in-services to discuss establishing a non-threatening and participatory approach in their classrooms during the formal external assessment days in particular (Olebe, 2005). It means targeting specific schools that rank below not only regional, but school board standards, and assigning trained personnel to work with teachers to reaffirm their integrative capabilities and enhance their proficiency to apply proven strategies that improve student learning. Further, district leaders can support school administrators in developing formal school improvement plans that address the entire curriculum, including non-tested areas, as a primary objective. Superintendents can assist principals in utilizing the data from the externally imposed assessments to account for the school interventions that best address under-achieving students.

**School Leadership**

As previously noted, assessment tends to drive the curriculum, particularly in high-stakes testing contexts (Kohn, 2000; Popham, 2001; Stiggins, 2002; Volante, 2004). In this light, large-scale assessment data can be a useful point of departure as a diagnostic piece for school administrators. Test scores can help identify trends and segments of the student population that are not achieving to standard. School leaders can determine strategic programming that includes not only intensive student capacity development in the areas of literacy and numeracy, but an embedded awareness of students' fundamental knowledge and learning needs. In this case, as with other measures of standardized assessments that identify common standards of student progression (for example, the Developmental Reading Assessment and Reading Running Records in the primary grades) the data can be instrumental in exposing the amorphous components of a school's curriculum while identifying intervention programs, resources, and in-school supports to sustain student improvement initiatives. Such a process is instrumental in complementing the developmental phases of students' academic and social growth.

Also, school administrators can focus the attention of their teachers on current assessment and evaluation practices. They can then review these practices and their bearings in juxtaposition to any negative and adverse implications of external assessment protocol. By discussing large-scale assessment measures and teacher assessment objectives within their school community, principals can seek input from parents and school councils on strategies to support learning and achievement on the entire curriculum in the context of situated learning. In so doing, parents can be reassured that the school's effort towards improving student learning in tested areas is not merely fattening the proverbial prize pig at the cost of starving the other farm animals.

Further, in this era of standards-based reform it is essential that teachers also assume incremental leadership roles by processing their students’ test scores, making specific interpretations based on their classroom observations and evaluations, and by advocating on behalf of their students for the necessary interventions at the school level (Zependa, Mayers, & Benson, 2003). To shadow the research, both beginning and experienced teachers can be empowered to be proactive in devising what they deem to be appropriate action plans best suited for the diverse developmental needs of their students (Cherubini, 2006). Teachers can integrate processes of improvement rooted in substantive and engaging learning experiences for their students (Merideth, 2000).

Teachers may also assume team-leadership roles to support literacy and numeracy initiatives on a school-wide basis.
School steering committees can amalgamate their professional expertise and collective experiences to align pedagogical practices to both classroom and large-scale assessments. Such collaborative efforts lend themselves to establishing fluent lines of communication between the school and the parent community, particularly concerning the distinct but interrelated niches of imposed large-scale assessment and the classroom assessment instruments teachers employ to collect evidence of student learning. Reminiscent of Hargreaves and Fink's (2006) notion regarding the breadth of sustainable leadership, teachers can impact upon reforming their own classroom assessment and pedagogical practices instead of resigning themselves to the mandates of a presiding structure and top-down imposed targets.

Supporting Multi-Level Assessment Reform

One way to support multi-level reform is to provide policy-makers who have direct control over provincial/state governance a strong justification for adopting a range of large-scale and curriculum-embedded assessment approaches for accountability purposes. Certainly, assessment methods and procedures around the world are constantly changing and permit an examination of different assessment models for accountability. Research is beginning to emerge that has shown how both traditional and alternative classroom assessment data has been successfully integrated for standards-based accountability purposes in pockets of the United States, England, and Australia (Wilson, 2004). By studying the interactions between large-scale and curriculum-embedded assessment, new models of assessment could be made more helpful (Black & Wiliam, 1998). Similarly, the refinement and dissemination of these important innovations may compel more policy-makers at the provincial/state governance level to adopt broader notions of what counts as worthy data for educational decision-making purposes. Essentially, curriculum-embedded assessment should be an integral part of an overall accountability framework to facilitate the types of changes proposed at the university, district, and school level. Maintaining the status quo approach ensures a myopic vision of educational excellence and the continual replication of inauthentic forms of teaching and learning that do little to promote high-order thinking skills or the application of new skills to novel tasks.

Educational leaders, particularly those at the school and district levels, should also continually lobby for well-designed large-scale measures that are closely aligned with the prescribed curriculum. Misalignment with content standards causes the information yielded from large-scale measures to be essentially irrelevant in the determination of school effectiveness (Anderson, 2005). Misalignment also forces teachers and administrators to abandon the prescribed curriculum and focus their instruction on test content and trivial preparation practices as a form of self-preservation. Research has documented how these teaching to the test techniques can improve school performance, sometimes making them look half a school year better than a comparable school that did not employ such strategies (Smith & Fey, 2000). The latter suggests elevated test scores in high-stakes accountability contexts (Carnoy & Loeb, 2002) may be more an artifact of test preparation than the power of threats/sanctions to spur authentic improvements in classroom practice and student learning. Clearly, external tests should closely mirror the state, provincial, or territorial curricula and include higher-order thinking tasks that are essential within our knowledge economy. Ironically, when broad curriculum coverage and critical thinking skills figure more prominently within external tests, teaching to the test actually becomes a desirable objective. It is quality, not ease of development, administration, and scoring, that should always be the first consideration in the selection of any large-scale assessment measure.

Conclusion

The reality remains that classroom and large-scale assessment occur in a variety of contexts in contemporary schooling. Our position is that reform initiatives, respective to multi-level reforms already discussed, conceptualize their broadening impact as being unified by the common objective of improving student performance and learning, and in the process remedying the disconnect that often occurs between the two in standards-based reform contexts. For this to occur, assessment leadership and multi-level reform should be outcome oriented and married to a student-centered approach to teaching and learning. Ultimately, it is to recognize a compelling and intuitive interpretation of assessment and leadership that embraces reform as a transition from compliance to effectiveness.

However, effectiveness should be understood as a product of authentic teaching and learning that focuses on the demonstration of higher-order thinking skills and knowledge transfer. Where assessment and student scores are concerned, leading educators, irrespective of their role as provincial/state policy-makers, program directors, district personnel, or school faculty, can create environments that recognize the various skills and talents within our increasingly diverse student population and more importantly, implement assessment frameworks able to capture such complexity. Although the logistics may at first appear formidable, the mandate to support these reform initiatives calls for collaborative and sustained professional dialogue that promotes synergy between classroom and large-scale assessment to strengthen our individualized attention to the whole student. Failure to balance our assessment methods, as previously discussed, leads to predictable negative consequences for students, teachers, and the school system in general.
Note

References


Authors

**Louis Volante (PhD)**
Assistant Professor
Brock University
Faculty of Education
1842 King Street East
Hamilton, Ontario, L8K 1V7
Email: Louis.Volante@Brocku.ca
Phone: (905) 547-3555 ext. 3621
Fax: (905) 547-9500

Professor Volante has worked as an elementary and secondary school teacher, project coordinator, and most recently as a consultant for governments across Canada, England and the United States. He has held faculty appointments at Concordia University and the University of Hawaii and has also lectured at the Ontario Institute for Studies in Education of the University of Toronto, where he obtained his Ph.D. Currently, he is an assistant professor in the Faculty of Education, Brock University, teaching assessment and evaluation courses at both the pre-service and graduate level.

Professor Volante's research interests include classroom and large-scale assessment, fairness/bias in assessment, professional development, organizational leadership, and educational change. He is the 2005 recipient of the R.W.B. Jackson Award for the most outstanding English language journal article by the Canadian Educational Researchers'Association.

**Lorenzo Cherubini (EdD)**
Assistant Professor
Brock University
Faculty of Education
1842 King Street East
Hamilton, Ontario, L8K 1V7
Email: Lorenzo.Cherubini@Brocku.ca
Phone: (905) 547-3555 ext. 3603
Fax: (905) 547-9500

Lorenzo Cherubini earned an Ed.D. from The University of Southern Queensland, Australia, an M.A. (American
Literature) and B.A. at McMaster University, Ontario, Canada, and a B.Ed. at Brock University, Ontario. His specialty is in Teacher Induction with research focus on post-industrial influences on organizational leadership concepts and school culture. Over the course of the last decade, Professor Cherubini has employed Grounded Theory methodology in various local and international collaborative research projects. He has taught at the secondary school level, served in the roles of guidance counselor, department chair of English and Modern Languages, and school administrator in both the elementary and Secondary school panels. Currently, he is an assistant professor in the Faculty Of Education, Brock University.