Accountability, student assessment, and the need for a comprehensive approach

DR. LOUIS VOLANTE
Brock University

ABSTRACT: Accountability has become synonymous with standardized testing in many Western countries such as Canada, the United States, Great Britain and New Zealand. Schools and districts are increasingly ranked based on their students' performance on standardized tests. Unfortunately, standardized testing measures possess a number of limitations that prevent them from being used as a sole indicator of student performance. This paper proposes a comprehensive approach to student assessment – one focused on improving teacher instruction and student learning. Consisting of seven components, the proposed approach provides an alternative to current accountability systems that distract rather than support the important work of teachers.

Introduction

Holding individual teachers, schools, and districts accountable for student performance continues to be a central feature of educational reform throughout the Western world. Following the lead set by the United States, Great Britain and New Zealand, Canada is increasingly adopting a system that is driven and controlled by standardized testing (Canadian Federation of Teachers, 1999). At the national level, external agencies such as the Fraser Institute publish report cards that rank individual schools according to their performance on provincially administered tests. It has also become increasingly common for newspapers to publish provincial assessment results. Consider the Toronto Star – Canada's largest daily newspaper, which routinely ranks schools within the Greater Toronto Area on Grades 3, 6, 9, and 10 provincial test results. Not surprisingly, schools that score above average use their performance as a measure of the effectiveness of their instruction. Alternatively, below average schools are strongly encouraged by district personnel to improve their performance in successive examinations. This cycle of praise and shame has created a pervasive testing culture within our schools (Cheng & Couture, 2000).

Although standardized testing is not the only assessment option currently operating within Canadian schools, the heightened salience of these measures has come at the expense of classroom assessment approaches that provide essential data for improving teacher instruction and student learning. Thus, assessment of learning (i.e., for accountability purposes) has far outpaced the desire to promote assessment for learning (i.e., for instructional purposes) (Stiggens, 2002). This paper attempts to correct this value imbalance by offering a different vision of educational accountability. Consisting of seven interrelated components, the proposed approach is supported by empirical research and is meant to provoke a dialogue amongst educational leaders such as policy-makers, district personnel, and school administrators. These primary stakeholders are charged with the task of developing accountability systems that not only gauge the quality of education but also support school improvement. In order to better frame the discussion, the next section outlines some of the main strengths and weakness that are commonly associated with standardized testing.

Pros and Cons with Standardized Testing

When designed and administered effectively, standardized tests can offer important benefits to students, teachers, administrators, and policymakers. For students, these measures can provide an external assessment of their knowledge and skills. Such an assessment might motivate them to work harder in school, particularly as they prepare for admittance to post-secondary institutions. For teachers, standardized test scores can be used to help identify areas of strength and weakness within the curriculum. Content not mastered by students provides teachers with valuable information to assess their own pedagogy. Gaps in student learning might encourage some teachers to

pursue specific types of professional development. Lastly, standardized test results provide administrators and policy-makers with evidence to judge the quality of their school programs and provincial policies. Both groups can also use test scores to promote more effective allocation of resources at the school, district, and provincial level.

Despite the previous advantages, current standardized testing programs possess a number of limitations that prevent them from being used as a sole indicator of student performance (Behuniak, 2002). Perhaps their greatest weakness is that they often sample a restricted range of student knowledge and skills. A mandated curriculum often requires students to demonstrate proficiency in a range of subject areas and disciplines. To date, standardized tests have focused almost exclusively on selective aspects of reading, writing, mathematics, and science to a lesser extent. Thus, a number of important curriculum areas are often excluded from these measures. No less worthy in their own right, these excluded areas are an essential component of a student's educational experience. Once more, skills in domains such as visual arts, music, and physical education are vital for a large sector of the workforce.

It is important to note that even within tested disciplines, standardized tests often sample a restricted range of knowledge and skills. Test developers routinely acknowledge the challenges of developing exhaustive measures in particular areas. For example, judging the literacy levels of students requires data related to reading, writing, speaking, and listening. Not surprisingly, standardized tests focus on the first two to the exclusion of speaking and listening. Thus, test results are of limited value when considering the breadth of curricula in particular domains. Unfortunately, teachers are often pressured to change their instructional approach based on a single measure. The latter may lead to unwise pedagogical refinements. Clearly, knowledge concerning the types of conclusions that can and cannot be drawn from particular test batteries is essential for program review and modification.

Even meticulously developed standardized tests are prone to measurement error, given the brief nature of their administration. Classifying a student based on his or her performance at one point is analogous to a doctor diagnosing a patient with high blood pressure based on one reading. Just as a patient's blood pressure will fluctuate over time, so too will a student's performance. Factors such as test anxiety adversely affect student performance (Hardy, 2003). Reviews of the literature suggest that the influence of test anxiety on the number of children who pass or fail high-stakes, standardized exams is potentially considerable (McDonald, 2001).

Lastly, the adoption of a test-driven educational system creates unhealthy competition between schools. This competition has led to inappropriate preparation strategies such as spending excessive amounts of time on tested content to the exclusion of mandated curricula. Some teachers also teach to the test, that is, they train students by using actual or look-alike items from previous assessments (Popham, 2001). Research has shown that this practice of teaching to the test has little effect on student learning (Neil, 2003). Collectively, the preoccupation with high-stakes, standardized testing has narrowed the curriculum and distracted teachers from a thoughtful discourse on the utility of diverse assessment approaches (Linn, 2000). The next section outlines seven key components within a comprehensive approach.

Comprehensive Approach

A comprehensive approach to student assessment is one that: (1) expands the assessment repertoire to include sound classroom-based assessment data; (2) supplements classroom assessment data with appropriate standardized measures; (3) sets realistic targets and standards for students, teachers, and schools; (4) downplays cross-school comparisons in favor of improvement for individual students; (5) utilizes a value-added criterion to interpret student performance in context; (6) provides teachers and administrators with professional development aimed at enhancing assessment literacy; (7) monitors and reviews the assessment system regularly. These seven components were frequently noted within the broader assessment, accountability, and educational reform literature. Nevertheless, the decision to incorporate a component was ultimately guided by the desire to develop a robust framework that promotes assessment for accountability as well as instructional purposes.

Classroom Assessment

Consider the variety of methods that teachers use to understand and learn about their students. They assess students by observing them in the classroom, evaluating their day-to-day class work, grading their homework

assignments, keeping logs on student development, and administering tests. These forms of student assessment have often been shunned for their lack of objectivity. Policy-makers typically argue that only standardized testing decreases the level of subjectivity in decision-making while helping in the evaluation of learning programs and systems (Burger & Krueger, 2003). Although rarely used, classroom work produced during the course of instruction (often referred to as "curriculum-embedded" assessment) can also serve policy purposes of assessment while maintaining a high degree of objectivity (Chudowsky & James, 2003).

One way teachers can collect credible evidence of their own effectiveness is through the use of a pretest/posttest design in which they give identical classroom assessments at the start of a semester and again at its conclusion (Popham, 2003). Teachers can tie test content directly to previous and current instructional material. Even more importantly, this form of assessment provides a high degree of contextualization that can be connected to students' personal educational experiences (Chudowsky & James, 2003). The latter alleviates the content bias that is often associated with large-scale assessment. Curriculum-embedded assessment also provides a strong basis for further instruction since pretest data can identify gaps in student knowledge and skills. By addressing these deficiencies, teachers can demonstrate individual student growth in particular domains over the course of an academic year. Thus, the pretest/posttest design can serve the interests of teachers for instructional purposes and also policy-makers who want valid data for accountability reasons.

If classroom test data is to stand up to public scrutiny, a certain degree of standardization is also required. For example, consistent rules must be applied in the administration and grading of pretest/posttest designs. Tests must also cover a range of material in particular subject areas. Despite these stipulations, curriculum-embedded assessment still provides teachers with the flexibility to make adjustments that are often impossible with on-demand standardized testing procedures. For example, teachers can administer multiple tests with different formats at different intervals throughout the school year.

Policy-makers would also be wise to consider the strengths of alternative assessment approaches, such as performance assessment. Consider the range of skills required to make a science project or write a research report. Both types of tasks do not readily lend themselves to standardized paper-and-pencil measures. Nevertheless, both are essential for success in school, particularly as a student progresses into higher grades. Compiled into portfolios, collections of student work can demonstrate the student's progress over time in a variety of subject areas and subdisciplines. In order to provide objectivity, teachers could develop analytic rubrics for the grading of students' portfolios. Research suggests that rubrics have been successfully utilized in large-scale writing assessment (Herman, 1992). The Education Quality Accountability Office (EQAO) in the province of Ontario currently uses analytic rubrics to assess writing on the Grade 10 Ontario Secondary School Literacy Test (OSSLT). A similar approach could also be adopted to assess students' portfolio collections.

Incorporating classroom assessment data in educational decision-making both values teachers' work and provides them with a basis for instructional modification. Collectively, classroom testing and performance assessment provide an important alternative to the sole reliance on standardized testing. Research has shown that both traditional and alternative classroom assessment data has been successfully integrated for decision-making purposes in the United States, England, and Australia (Wilson, 2004). The utilization of classroom assessment data also seems essential for strengthening the validity of standardized test results that policy-makers hold in such high regard. The increased importance of such data should lead to a corresponding decrease in inappropriate standardized test preparation, such as teaching-to-the-test techniques. The next section discusses important changes needed to increase the utilization of standardized test data.

Standardized Testing

There is general agreement that large-scale assessment should have an impact on schools and on changing education (Earl & Torrance, 2000). Rather than dismissing the use of these tests outright, educators should advocate for important changes that would permit their integration in meaningful ways. Distinguishing between norm-referenced and criterion-referenced tests is a logical starting point. Whereas norm-referenced tests compare student performance against statistical norms so that one-half of students are typically classified as "below grade level" (i.e., less than 50th percentile), criterion-referenced tests measure student performance against clear standards so that all

students can be successful (Covaleskie, 2002). Perhaps the most promising aspect of criterion-referenced testing is that teachers have been able to use these results to improve their instruction (Wideman, 2002). Thus, administrators and teachers should advocate for the use of criterion-referenced testing procedures within their schools.

Another essential step for the utilization of standardized testing is to ensure that districts select measures which have sufficient alignment between a test's content and the mandated curricula. Jurisdictions with a large degree of incongruity between these two areas facilitate the use of maladaptive preparation strategies. In an attempt to raise the performance of their students, teachers often devote excessive amounts of time teaching test content and administering mock examinations. This preparation time often comes at the expense of non-tested disciplines. Clearly, accountability systems need to work to support, not distract teacher's instructional efforts.

The influence standardized testing has (or not) on teachers and teaching depends on how teachers interpret the nature of the test and use results to guide their actions (Cimbricz, 2002). Unfortunately, most standardized testing regimes do not provide teachers with professional development related to the testing process or precise feedback on the performance of their students. This has led to a lack of utilization of test results. One must query how any teacher could tailor instruction to address a gap in student learning without such feedback and support. If the ultimate goal of standardized testing is to improve instruction and learning, an equal, if not greater, effort must be placed on feedback and reporting. The status quo approach will only exacerbate teachers' frustration with standardized testing, and lead to disillusionment with the entire testing process.

Targets & Standards

Jurisdictions that set unrealistic targets and standards do little to mobilize and motivate their teachers. Research suggests that imposing unrealistic, high targets undermines the credibility of the target-setting exercise, decreasing rather than increasing teachers' efforts to improve (Earl et al, 2003). Similarly, the teaching of a narrow curriculum that is often associated with higher standardized test targets is likely to alienate a large portion of students whose academic strengths lie outside of commonly tested subjects (Volante, 2004). Thus, targets are only effective when they are attainable and include a broad range of data sources. Current levels of performance and past gains provide a context for judging future gains and long-range targets of performance (Linn, 2000).

In order for assessment to connect to school improvement in a meaningful way, teachers must understand in advance of teaching the achievement targets that their students are to hit (Stiggens, 2002). Deciding acceptable achievement targets should be a collective responsibility. Jones (2004) suggests that teachers, parents, and community members review data about student performance and make decisions about promotion, placement, graduation, and so on. Incorporating diverse perspectives on student targets and standards provides a basis for thoughtful discourse on the knowledge, skills, and dispositions we seek in our children. It also serves as an avenue for discussing the legitimacy and effectiveness of diverse assessment approaches. Those assessment approaches that improve teacher instruction and student learning should remain the focus of any accountability system.

Cross-School Comparisons

Accountability systems designed to help schools improve and students learn will downplay cross-school comparisons in favor of improvement for individual students. Compelling evidence from the United States indicates an increase in incidents of cheating and test misuse when schools are ranked publicly (MacDonald, 2001). In Canada, testing agencies such as EQAO have argued against ranking since it is misleading and does not contribute to the well-being of students (EQAO, 1998). Clearly, the ranking of schools serves as an unnecessary distraction for practicing teachers.

In many respects, a comprehensive approach is improvement-oriented rather than results-oriented. Improvement-oriented approaches focus on individual student learning as the ultimate measure of success. Such approaches break down or disaggregate data to identify and learn from pockets of success (Gratz, 2000). Thus, every school can be a winner and contribute to the well-being of all students. Conversely, results-oriented approaches focus on school performance by aggregating data that has little, if any, opportunity for informing student instruction. The excessive focus on higher test results pits teacher against teacher and school against school so that a community of learners is

discouraged. School results are seen as relative so that there will always be winners and losers within the system. Teachers within losing schools eventually contemplate moving to a better performing school or leave the profession entirely (Delhi, 1998).

Value-Added Criterion

Educators, parents, politicians, and the public are all responsible for contributions to the quality of schools, and none of them can be held responsible for things over which they have no control (Earl, 1998). A value-added criterion seems essential so that teachers do not unnecessarily shoulder the blame for poor student performance. Essentially, the value-added approach suggests that student success is achieved when students progress as much as, or more than, might be expected due to prior attainment and background. The appropriate degree of progress is judged against factors such as family background and the knowledge and skill level students bring to a task (Wickstrom, 1999).

Critics have charged that the value-added approach makes excuses for poor student performance in particular classrooms, schools, and districts. This critique is not well-founded when one considers the range of factors that are likely to influence student performance. In fact, the relationship between socio-economic status and educational achievement continues to be recognized as one of the most stable relationships in educational research (Earl et al., 2003). A primary teacher can hardly take credit for the strong showing of his students on a test when all of them come from affluent households and were able to read, write, and perform simple arithmetic before they entered the classroom. Conversely, a teacher within an inner city school should not be held accountable for poor student performance when her school lacks basic resources, and all her students come from difficult home environments (Wright, 2002). The value-added approach considers these factors, and emphasizes the degree of progress in students when making judgments about appropriate levels of achievement.

In the absence of a value-added approach, teachers will seek out "bright-students" in "good schools", since it will be easier to demonstrate their effectiveness. High staff turnover in low performing schools makes it particularly difficult to maintain continuity in instruction or forge enduring professional communities. Not surprisingly, teachers within poor performing schools are continually marginalized within accountability systems that do not consider the unique nature of their student population. These teachers are more likely to report additional job stress or leave the teaching profession when they are compared to higher socioeconomic schools with better standardized test scores (Wright, 2002). When one considers the significant investment needed to train teachers, the value-added approach seems essential for retention purposes.

Professional Development

When assessment is integrated with instruction, it informs teachers about what activities and assignments will be most useful, what level of teaching is most appropriate, and how assessment provides diagnostic information (McMillan, 2000). Unfortunately, most teachers have not been adequately trained in assessment, and need substantial and ongoing professional development to create valid and reliable tasks and build effective classroom assessment repertoires (Jones, 2004). Assessment-literate teachers understand how to transform their expectations into assessment exercises and scoring procedures that accurately reflect student achievement (Stiggens, 2002). These teachers are better able to make their students' progress much more transparent and visible to parents, administrators, and policy-makers. In the absence of such training, skeptics will continue to dismiss classroom assessment data as "self-interested home cooking" (Popham, 2003).

Assessment literacy training should be provided to both practicing as well as aspiring teachers. Thus, faculties of education must take the lead and ensure that their training programs are graduating teacher candidates who utilize sound assessment practices. The significant investment in preservice education provides a solid foundation for tomorrow's teachers. These new teachers can also use their assessment knowledge to help improve any accountability system that relies on unsuitable standardized tests (Popham, 2004).

System Review

Lastly, a comprehensive approach monitors the implementation and effectiveness of the assessment system on a regular basis. This type of system "check-up" should be conducted on an annual basis and include the perspectives of a variety of stakeholders. Parents, teachers, administrators, policy-makers, and community members, as well as elected officials, should have a voice in shaping and revamping any particular approach. Representatives from each of these groups offer a unique perspective that provides important checks and balances within the system. Local councils must be created that have real power to effect school change (Jones, 2004). Clearly, no single group, such as politicians, should be able to impose top-down reforms without approval from other vested parties, particularly classroom teachers. As Fullan (2003) reminds us, lasting educational change results from an appropriate balance of top-down and bottom-up input.

The litmus test for any system review is whether the adopted approach is having a beneficial impact on teacher instruction and student learning. Are students progressing as much as might be expected, given their prior attainment and background? If not, what are the impediments to such progress, and what type of corrective steps can be taken? These steps may include a reanalysis of the relative weighting of particular sources of assessment data, an adjustment in educational targets and standards, or continued professional development at the preservice or inservice level.

Conclusion

A comprehensive approach to student assessment requires a fundamental shift in our thinking about accountability, namely, that teachers can produce valid and reliable sources of assessment data within their own classrooms which can serve in planning and decision-making. The objective standard that is often used to dismiss classroom assessment is flawed when one considers the alternatives previously discussed. Collectively, different forms of curriculum-embedded assessment such as pretest/posttest designs and portfolios provide an important mechanism to measure student performance. More importantly, these approaches offer teachers vital information needed to improve their classroom instruction.

Accountability means taking information and using it to make judgments – about quality, about how good is good enough and, most importantly, about how to make changes that will enhance and extend student learning (Earl, 1998). Relying on single flawed standardized test measures does little to produce the desired improvements we seek in education. The previous approach underscored the need for educational leaders to broaden their definitions of what counts as evidence of success. By utilizing classroom assessment data for educational decision-making purposes, widespread improvements in teacher instruction and student learning can be achieved (Wilson, 2004).

References

Behuniak, P. (2002). Consumer-referenced testing. Phi Delta Kappan, 84(3), 199-207.

Burger, J. M., & Krueger, M. (2003). A balanced approach to high-stakes achievement testing: An analysis of the literature with policy implications. International Electronic Journal for Leadership in Learning, 7(4). Online at http://www.ucalgary.ca/~iejll.

Canadian Federation of Teachers. (1999). Province-wide assessment programs. Online at http://www.ctf-fce.ca/e/what/other/assessment/testing-main.htm

Cheng, L., & Couture, J. (2000). Teachers' work in the global culture of performance. The *Alberta Journal of Educational Research*, 46(1), 65-74.

Chudowsky, N. P. & James, W. (2003). Large-scale assessment that supports learning: *What will it take? Theory into Practice,* 42(1), 75-83.

Cimbricz, S. (2002). State-mandated testing and teachers' beliefs and practice. Educational Policy Analysis Archives,

10(2), Online at http://epaa.asu.edu/epaa/v10n2.html

Covaleskie, J. F. (2002). Two cheers for standardized testing. International Electronic Journal for Leadership in Learning, 6(2). Online at http://www.ucalgary.ca/~iejll/volume6/covaleskie.html

Delhi, K. (1998). Shopping for Schools. *Orbit,* 29(1), 29-33.

Earl, L. (1998). Developing indicators: The call for accountability. *Policy Options*, 6, 20-25.

Earl, L., & Torrance, N. (2000). Embedding accountability and improvement into large-scale assessment: What difference does it make? *Peabody Journal of Education*, 75(4), 114-41.

Earl, L., Levin, B., Leithwood, K., Fullan, M., Watson, N., Torrance, N., Jantzi, et al. (2003) *England's National Literacy and Numeracy Strategies: Final report of the external evaluation of the implementation of the strategies*. Department of Education and Employment, England.

EQAO. (1998). *Educators handbook*. Toronto, Ontario: Queen's Printer of Ontario.

Fullan, M. (2003). *Change forces with a vengeance*. New York: Falmer.

Gratz, D. B. (2000). High standards for whom? *Phi Delta Kappan*, 81(9), 681-87.

Hardy, L. (2003). Overburdened, overwhelmed. *American School Board Journal*, 190(4), 18-23.

Herman, J. L. (1992). What research tells us about good assessment. *Educational Leadership*, 49(8), 74-78.

Jones, K. (2004). A balanced school accountability model: An alternative to high-stakes testing. *Phi Delta Kappan*, 85(8), 584-90.

Linn, R. (2000). Assessment: Responsibility and reasonable expectations. *Educational Researcher,* 32(7), 3-13.

MacDonald, T. (2001). *To test or not to test: A question of accountability in Canadian Schools.* Online at http://policy.ca/archive/20010622.php3.

McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology: An International Journal of Experimental Educational Psychology,* 21(1), 89-101.

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. Practical Assessment, Research & Evaluation, 7(8). Online at http://pareonline.net/getvn.asp?v=7&n=8

Neil, M. (2003). High stakes, high risk: The dangerous consequences of high-stakes testing. *American School Board Journal,* 190(2), 18-21.

Popham, W. J. (2004). All about accountability / Why assessment illiteracy is professional suicide. *Educational Leadership,* 62(1), 82-83.

Popham, W. J. (2003). The seductive allure of data. *Educational Leadership,* 60(5), 48-51.

Popham, W. J. (2001). Teaching to the test. *Educational Leadership,* 58(6), 16-20.

Stiggens, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan,* 83(10), 758-765.

Volante, L. (2004). Teaching to the test: What every educator and policy maker should know. Canadian Journal of Educational Administration and Policy, 35. Online at http://www.umanitoba.ca/publications/cjeap/articles/volante.html

Wickstrom, R. A. (1999). Accountability: Toward an appropriate approach. *Education Canada,* 39(3), 16-19.

Wideman, R. (2002). Using action research and provincial test results to improve student learning. I*nternational Electronic Journal for Leadership in Learning,* 6(20). Online at http://www.ucalgary.ca/~iejll/volume6/wideman.html

Wilson, M. (Ed.). (2004). *Towards coherence between classroom assessment and accountability: 103rd yearbook of the National Society for the Study of Education, Part II.* Chicago: University of Chicago Press.

Wright, W. E. (2002). The effects of high stakes testing in an inner city elementary school: The curriculum, the teachers, and the English language learners. *Current Issues in Education,* 5(5). Online at http://cie.ed.asu.edu/volume5/number5