Toward Appropriate Preparation for Standardized Achievement Testing

LOUIS VOLANTE
Brock University

ABSTRACT: Standardized achievement testing has become a pervasive feature of schooling in Canada. Unfortunately, the pressure to do well on these tests has facilitated teachers' utilization of inappropriate test preparation practices. This paper distinguishes between appropriate and inappropriate test preparation by examining three key issues: time spent directly on test preparation, content of preparation instruction, and the teaching of test-taking skills. Steps are proposed for test developers as well as individuals at the policy, district, and school level, to help facilitate teachers' utilization of appropriate strategies. The latter are essential if we are to promote authentic learning and strengthen the validity of standardized achievement measures.

RÉSUMÉ: Réussir un examen standardisé est devenu une caractéristique persuasive dans le système scolaire du Canada. Malheureusement, la tension pour réussir à ces examens est telle que les professeurs sont enclins à utiliser des moyens de préparation qui ne sont pas appropriés. Ce rapport fait la distinction entre la préparation aux examens propre et celle qui ne l'est pas. Pour ce fait, il soulève trois questions clé: 1) le temps passé juste pour la préparation de ces examens; 2) le contenu du cours de préparation; 3) l'enseignement des compétences requises pour passer les examens. Ce sont des étapes que l'on propose aux concepteurs d'examens aussi bien qu'aux individuels à un niveau politique, de quartier ou d'école pour aider les professeurs à utiliser facilement les stratégies adéquates. Ces derniers points sont importants pour promouvoir l'authenticité des études et renforcer la valeur de la réalisation de ces mesures réglementées.

Introduction

Standardized achievement testing has become a pervasive feature of schooling in Canada. A quick survey of provincial Ministry of Education websites reveals that every province, with the exception of Prince Edward Island, administers some form of large-scale student

Journal of Educational Thought
Vol. 40, No. 2, 2006, 129-144.

assessment. The approach of individual provinces varies according to the grades tested, sample size, test format, and frequency of administration (MacDonald, 2001). In addition to provincial assessments, some Canadian students are also tested as part of the national School Achievement Indicators Program (SAIP). Operated by the Council of Ministers of Education, the SAIP program assesses 13 and 16 year old students across the country in reading, writing, math, and science. A select number of provinces also participate in international testing programs such as the Third International Mathematics and Science Study (TIMSS), Programme for International Student Assessment (PISA), and/or the Progress in International Reading Literacy Study (PIRLS). Although critics of standardized testing dismiss these measures outright, there is increasing support for large-scale assessment having an impact on schools and ultimately changing education (Chudowsky & James, 2003; Covaleskie, 2002; Earl & Torrance, 2000). The exact level or degree of this impact continues to provoke debate.

Following the lead set by the United States, Great Britain, and New Zealand, Canada is increasingly adopting a system that is controlled by standardized testing (Canadian Federation of Teachers, 1999). These measures play a central role for accountability purposes and often overshadow classroom assessment data (Volante, 2005; Wilson, 2004). Unfortunately, the increased salience of standardized test scores has facilitated teachers' utilization of inappropriate test preparation practices such as relentless drilling on test content, eliminating important curricular content not covered by the test, and providing long practice sessions that incorporate similar items from these measures (Popham, 2000).

In Canada, test-wiseness has spuriously enhanced performance on provincial assessments and remains an ongoing instructional concern (Levison, 2000; Rogers & Bateson, 1991). Teachers in the United States also report that the pressure to do well on standardized tests often results in excessive test preparation (Neil, 2003b; Wright, 2002). In certain instances, this pressure has led to outright cheating where teachers and administrators have given students the answers to reading and mathematics questions (Goodnough, 1999). In England, teachers skew their efforts in the direction of activities that lead to increases in these important scores (Earl, et al., 2003). This assertion is based on the results of a national survey that included more than 4000 primary/junior teachers and resource personnel – suggesting a widespread tendency to

utilize score boasting tactics within this country. Unfortunately, research has indicated that while students' scores will rise when teachers teach closely to a test, learning often does not change (Mehrens & Kaminski, 1989; Shepard, 2000; Smith & Fey, 2000). Overall, the use of inappropriate preparation practices has undermined authentic forms of teaching and learning and eroded the inferences one can draw from these measures.

Appropriate Versus Inappropriate Test Preparation

Although the selection and use of appropriate preparation practices is premised on a number of factors, not least of which is a re-examination of the relative importance of standardized test scores, teachers need to clearly understand how to properly distinguish between appropriate and inappropriate strategies. A review of the literature on standardized achievement testing suggests that there are three key issues to consider when preparing students for testing: the amount of class time spent directly on test preparation, the content of preparation instruction, and the teaching of test-taking skills. Although normal classroom instruction prepares students for general test-taking, the first issue involves the amount of class time specifically devoted to a particular standardized measure. This may range from a single lesson to several weeks or even months (Meaghan & Casas, 1995). Conversely, the second issue involves the nature of the preparation instruction and may include routine familiarization with the structure of the test to more intensive mock examinations using cloned items (Popham, 2004). The final issue overlaps with the content of preparation instruction and more narrowly considers the type of test-taking skills that students receive. The latter may include simple suggestions such as skimming all the test pages prior to starting the test to more detailed strategies such as learning how to search a reading passage for clues to the correct multiple-choice answer (Mehrens & Kaminski, 1989). Each of these interrelated issues impacts the quality of instruction within schools and has a profound effect on the psychometric properties of large-scale assessment measures.

Time Spent Directly on Test Preparation

Teachers grapple with the amount of time they should spend preparing their students for standardized achievement testing. They often query whether a period, day, or week of class time devoted to test preparation

is appropriate or excessive. The answer to this question is relatively straightforward. Test preparation time should never come at the expense of non-tested subject matter, even when a significant number of students and parents approve of this instructional shift. A standardized test of literacy and numeracy skills should be addressed within the regular language arts and mathematics instructional schedule. Truncating the instructional time of subjects like music, physical education, or visual arts, in the service of longer test preparation, should never be condoned. Research suggests that narrowing the curriculum to focus instruction solely on tested subjects often alienates a significant number of students whose academic strengths lie outside of those areas (Wei, 2002; Wright, 2002). Administrators in the United States have helped facilitate this negative shift by reassigning staff to tested grades and deleting optional non-tested subjects entirely (Gustafson, 2001).

Standardized testing also tends to constrain the instructional time allotted to mandated curriculum within Canada. In Alberta, time pressures contribute to a narrowing of the curriculum, as available time is focused on the topics that will be assessed on the provincial test (Alberta Teachers Association, 2005). In British Columbia, significant narrowing of the curriculum has also been reported in response to high-stakes provincial examinations (Anderson, 1990; Wideen, O'Shea, Pye, & Ivany, 1997). Not surprisingly, teachers in Ontario also report spending a disproportionate amount of time on tested subjects such as reading, writing, and mathematics (Ontario English Catholic Teachers Association, 2002). A number of teachers within this province have even indicated that they focus much of the second half of the school year on test preparation activities (Meaghan & Casas, 1995). Clearly, there is a need to balance competing interests so that valuable teaching and learning time is not sacrificed in the drive for higher test scores.

Content of Preparation Instruction

The distinction between item-teaching and curriculum-teaching provides a useful dichotomy when considering the dilemma faced by classroom teachers. In item-teaching, teachers narrowly focus their instruction either around the actual items known to be found on the test or a set of look-alike items (Popham, 2001). This practice, which is commonly referred to as "teaching to the test," provides little opportunity for authentic forms of learning. Higher test scores do not necessarily

translate into corresponding levels of student knowledge when teachers employ this technique. As a result, the inferences one can reasonably draw about a students' domain-specific knowledge and skills are severely weakened by item-teaching approaches (Burger & Krueger, 2003). Research has indicated that drilling students on content known to be on the test can make a school look half a year better than a comparable school that did not employ item-teaching preparation (Smith & Fey, 2000). In addition to skewing scores and invalidating a test, this practice also promotes convergent thinking, didactic instruction over discovery learning, involves only short-term memory, and excludes higher-order thinking skills (Kaufhold, 1998). Teachers should never develop or modify curriculum that is based solely on the content or objectives of a particular test (Education Quality and Accountability Office, 1998).

In contrast, curriculum-teaching requires teachers to direct their instruction toward a specific body of content knowledge or a specific set of cognitive skills represented by the test (Popham, 2001). When prepared with such an approach, students should be able to transfer what they have learned into a novel situation. The latter suggests that authentic learning has been achieved. Ideally, strong convergence exists between provincial curricula and the content within standardized measures. This allows teachers to accomplish the dual goal of test preparation as well as the teaching of mandated curriculum. A recent evaluation of Ontario's provincial assessment program noted that many assessment questions did not match the curriculum and those that did sometimes did not address the full expectation (Wolfe, Childs, & Elgie, 2004). This finding underscores the importance of selecting standardized tests that are more closely aligned with provincial curricula. Such tests support teachers' instructional efforts and promote student learning (Chudowsky & Pellegrino, 2003).

Teaching Test-Taking Skills

Critics of standardized testing typically argue that the teaching of test-taking skills is an illegitimate practice. The Canadian Federation of Teachers (1999) refers to these types of skills as "tricks" to be avoided. Although we should share the Federation's concern with the teaching of unethical test-taking skills, a thoughtful discourse on this subject needs to carefully delineate among various test-taking competencies. Certainly, dubious approaches should be eliminated. For example, there are cases where children have been taught to read by learning to look at the answer options to multiple-choice questions and then scan the short

passage to find the clue for the proper response (Neil, 2003a). Not surprisingly, independent evaluators have found that these children are often unable to explain what they have just read. The implication is that they actually lack the reading comprehension skills that a correct answer would suggest they possess.

Nevertheless, one should not confuse or associate all test-taking skill efforts with shortcuts to success. Perhaps the simplest test-taking skill – reading the question properly – is one that eludes many students. Studying various types of questions can be a useful endeavor for future test takers. Calkins, Montegomery, and Santman, (1999) point out that questions such as "how was the central problem resolved?" or "which statement is NOT true about the narrator?" are typically not the types of questions children are asking themselves when they read. Students should also have the opportunity to take different types of multiple-choice item formats that assess different levels of cognition (e.g., comprehension, application, analysis, synthesis, etc. – Miyasaka, 2000). Awareness of the structure of common standardized questions and the diversity of item formats helps students understand precisely what is being asked and assessed.

Students also need to be made aware of the test's general format including the variety of assessment approaches (e.g., multiple-choice, short answer, extended response, performance tasks) they will encounter (Miyasaka, 2000). As a former Marking Strategy Leader with the Education Quality and Accountability Office (EQAO) in the province of Ontario, I was amazed at the number of students that misinterpreted the length requirements for open-ended writing questions. Students would routinely cram their answers below the question, failing to realize that they could use both sides of the page for their response. In some cases, students seemed to miss sections of the test entirely, particularly those at the end of test booklets. Familiarizing students with these types of formatting issues does not compromise the integrity of individual items. These skills however, are essential if we are to avoid underestimating students' abilities.

Perhaps the most important skill students need to master before undertaking any standardized test is the ability to control their anxiety level. Reviews of the literature suggest that test anxiety has a significant influence on the number of children who pass or fail high-stakes tests (MacDonald, 2001). Programs such as Testbusters have been successful in reducing test anxiety and social-evaluative concerns in school-aged children (Beidel, Turner, & Taylor-Ferreira, 1999). Test familiarization

and anxiety reduction can actually improve validity since scores that were invalidly low because of anxiety might become validly higher (Messick, 1996). Thus, addressing children's fear of academic failure is not only essential for getting an accurate picture of student strengths and weaknesses; it is also an obligation of every conscientious educator. In many respects, the development of the previously noted test-taking skills ensures that students are test-ready. This is in stark contrast to test-wise students who lack the requisite knowledge for which they seemingly have demonstrated through their inflated scores.

Facilitating Appropriate Test Preparation

Teachers need to be supported in their efforts to adopt and utilize appropriate test preparation practices. In order to facilitate this process, a series of steps are required from test developers, policymakers, district leaders, administrators, and classroom teachers. The next sections outline the key actions required by each of these stakeholders.

Test Developers

A test can only be useful if the educational objectives are clearly outlined in advance, if the curriculum is well designed around those objectives, and if the test is aligned with the curriculum (Rubenstein, 2003). In instances where there is significant disparity between the tests' content and mandated curriculum, teachers may abandon the latter in order to more closely mirror the content and skills to be assessed. Greater test-curriculum alignment seems absolutely essential if teachers are to move toward appropriate test preparation practices. The latter is particularly difficult with norm-referenced tests that typically compare student performance to a representative sample of students at the national level (Gronlund, 2003). Thus, test developers should more carefully focus their efforts on developing sound criterion referenced tests since they are capable of measuring student performance against clear standards tied to mandated provincial or state curricula (Covaleskie, 2002). One major caveat to consider is that the tests' standards not represent the most minimalist competencies in students. The latter would dumb down instruction and provide the impetus for lowering student expectations. Ontario, Alberta, and British Columbia are three large provinces that currently use criterion-referenced tests to assess student performance.

Another important dimension for test developers to consider is the test's construct validity – the extent to which the test examines the

construct under investigation. Construct under-representation occurs when a test is too narrow and fails to include important facets of the curriculum. If constructs or aspects of the curriculum are under-represented on the test, teachers may overemphasize those areas that are well-represented and downplay those that are not (Messick, 1996). Thus, broad curriculum sampling is essential to minimize the teaching of a narrow curriculum and strengthen the validity of the inferences derived from test scores. In cases where the test is not fully representative of the curriculum standards, multiple forms should be used to sample every outcome (Kane, 2002). Although standardized achievement tests are rarely designed to be a completely faithful exemplar of criterion behaviours, test developers have an ethical obligation to address these limitations and provide strong evidence on the validity of their measures (Austin & Mahlman, 2002; Messick, 1996).

Test developers can also minimize the use of inappropriate test preparation strategies by designing measures that increasingly emphasizing critical thinking over lower-order thinking skills. Well-implemented instruction that focuses on critical thinking is more likely to motivate and foster meaningful learning of relevant domain-specific skills and knowledge (Yeh, 2001). The latter is in contrast to lower-order skills that tend to facilitate the use of worksheets, drills, and other rote practices for test preparation purposes. Although tests measuring critical-thinking are difficult to develop they provide a deeper understanding of a students' learning profile. Collectively, the lack of test/curriculum alignment, broad curriculum sampling, and critical thinking tasks, facilitate the use of inappropriate preparation practices.

Policy Level

Policymakers also play a pivotal role in moving teachers toward the adoption of appropriate preparation practices. These individuals must resist the temptation to overemphasize standardized test scores as indicators of educational quality. Although standardized measures do provide useful information concerning student learning and performance, they often do not permit an extensive sampling of a child's strengths and weaknesses. If a test actually covered all of the knowledge and skills in every content domain, multiple assessments would have to be given during the course of the school year (Kane, 2001). This type of approach requires a significant investment both in terms of test administration and validation. Nevertheless, Messick (1996) reminds us

that even a validated test can be subverted by coaching or test preparation practices emphasizing test-wiseness strategies.

In many respects, the utility of a standardized achievement test is premised on a careful balancing act. If the assessment measure becomes too important or high-stake, teachers will skew their teaching in the direction of inappropriate preparation practices likely to produce elevated scores. Unfortunately, research suggests that this type of teaching discourages inquiry and active student learning (Wideen, et al., 1997). Conversely, if a test becomes irrelevant for educational decision-making, students will likely view the process as a trivial exercise. Thus, "middle-stakes" assessment seems essential if we are to trust the implications that follow from these measures. Policymakers can facilitate this shift toward middle-stakes assessment by utilizing large-scale test scores *only* in conjunction with curriculum-embedded assessment (i.e., classroom assessment data). This approach ensures that both forms of assessment are equally valuable and worthy of attention when deliberating the status of individual schools. Only by using different approaches to assessment can a complete picture emerge as to what a student understands and can do (McMillan, 2000; Wilson, 2004). Not surprisingly, the most recent code of fair testing practices prepared by the Joint Committee on Testing Practices (2005) affirms the importance of multiple data sources, arguing against the use of single test scores as sole determinants for making decisions.

In line with the shift toward middle-stakes assessment and multiple measures comes the recognition that additional factors, aside from schooling, influence student performance. Clearly, policymakers should consider how student outcomes are affected by extraneous factors such as socio-economic status and the support structures in place for families (Earl, et al., 2003). This type of analysis is typically referred to as value-added assessment and has helped many policymakers understand the degree of progress in the teaching-learning process (Vaughan, 2002). Far from providing excuses for vulnerable student populations, the value-added approach considers important contextual information that is often lacking from the examination of student performance.

District Level

The importance of test/curriculum alignment was previously noted. Sadly, the average overlap between standardized test content and the texts teachers rely on for their instruction hovers somewhere around 50% (Hilliard, 2000). Although the preceding analysis comes from the United States, this test/curriculum mismatch underscores the importance of choosing appropriate large-scale achievement measures. Forcing teachers to chose between provincial curricula and standardized test content that requires different knowledge and skills places them in an awkward position. Those who focus their efforts on mandated curriculum will have students who attempt standardized test items on topics they may have never been taught. Conversely, teachers who direct their instruction toward a specific body of content knowledge or a specific set of cognitive skills represented by the standardized test may take valuable instructional time away from other facets of the mandated provincial curriculum. District leaders need to clearly articulate these concerns to test developers and government leaders who typically impose mandated provincial testing.

District-level staff, particularly those who hold senior administrative positions, can also facilitate teacher's utilization of appropriate preparation practices by downplaying cross-school comparisons. Ranking schools based on single measures produces unhealthy competition with predictable winners and losers. Not surprisingly, "good teachers" will seek out "good schools" to avoid the scrutiny that results from relatively poor performance (Delphi, 1998). In this respect, low performing schools are increasingly marginalized when ranking occurs. High performing schools that over-emphasis test scores are also disadvantaged. Studies in Chicago, Milwaukee, and Philadelphia have shown that these types of schools tend to take attention away from changes that produce school improvement such as effective leadership, high expectations for all students, and cohesive staff with a clearly articulated vision and knowledge of best practices (Duke, 1999). Thus, district-level staff should disaggregate data to inform the provision of supports to particular groups of students and schools. This approach underscores the importance of being improvement-oriented rather than results-oriented.

Districts also need to allocate resources so that teachers understand what appropriate standardized preparation looks like. Too often, testing is imposed from provincial governments, with little, if any thought of the professional development needs of classroom teachers. A stark example of this occurred in Ontario during the late 1990s when the government

forced school districts to eliminate virtually all of their professional activity days for teachers. This drastic cut-back in professional development occurred at the same time new curricula and provincial testing measures were being introduced. Not surprisingly, this lack of professional development led to a great deal of misunderstanding about the assessment process. Ironically, providing teachers with in-service opportunities ensures a certain degree of uniformity in how teachers understand and approach large-scale assessment measures. In this sense, districts can ensure a test is truly standardized both in terms of its administration as well as the approaches teachers have employed for preparation purposes. This type of uniformity strengthens the inferences one can reasonably draw from large-scale assessment data.

School Level

A conscientious principal discourages subject matter drills and practice solely for the test (Power, 1999). The former is widely recognized as an unethical and educationally indefensible approach to test preparation (Mehrens & Kaminski, 1989; Miyasaka, 2000; Popham, 2004). When such cases occur within a school, the principal has an ethical obligation to intervene and provide proper counseling. This can be accomplished through an individual one-on-one conference or through an inclusive staff meeting where administration addresses assessment related concerns and clarifies the utility of different preparation practices. These teachable moments provide a useful way to promote assessment literacy in practicing teachers. The latter has been related to a number of positive outcomes, including enhanced teaching practices and student learning (Stiggins, 2002).

Administrators and teachers also need to share their assessment knowledge with parents and other primary stakeholders. Their first responsibility is to dispel the notion that achievement scores in themselves are adequate and sufficient indicators of school performance (Earl, 1998). Parents need to understand that standardized achievement tests are one kind of assessment tool among many that need to be employed in accurately measuring student learning (Marshak, 2003). Helping parents understand the limitations of standardized test scores may go a long way to reducing the public pressure experienced by district and school personnel. This type of active dialogue with parents will also inspire a greater sense of loyalty to the school (Lewington, 1999).

Conclusion

Teachers work in an environment increasingly dominated by external accountability and a pervasive culture of testing (Cheng & Couture, 2000). Unfortunately, the growing salience of high-stakes test scores for decision-making purposes has facilitated the use of inappropriate test preparation practices. This article encourages teachers and administrators to carefully consider three key issues: the amount of time spent directly on test preparation, the content of preparation instruction, and the teaching of test-taking skills, when contemplating preparation practices within their classrooms and schools. Understanding these issues is essential for teachers to move toward appropriate standardized test preparation.

Nevertheless, teachers' preparation practices do not occur in a vacuum, free from the stresses of external accountability systems and public scrutiny. In line with this comes the recognition that additional factors, aside from the will of individual teachers, are instrumental in promoting appropriate preparation strategies. Indeed, the preceding discussion called into question the utility of high-stakes testing and argued for a more balanced assessment approach. Utilizing multiple measures along with a shift toward middle-stakes testing are two important elements within a balanced system. Collectively, test developers along with individuals at the policy, district, and school levels, all play a role in helping teachers utilize appropriate standardized test preparation practices. Failing to do so undermines authentic forms of teaching and learning and compromises the validity of large-scale assessment measures.

REFERENCES

Alberta Teachers Association (2005). *Testing and accountability in education*. Available from http://www.teachers.ab.ca

Anderson, J.O. (1990). *The impact of provincial examinations on education in British Columbia: General report*. Victoria, BC: British Columbia Department of Education.

Austin, J.T., & Mahlman, R.A. (2002). *High-stakes testing: Implications for career and technical education.* (Report No. NDCCTE7). Washington, DC: Office of Vocational and Adult Education.

Beidel, D.C., Turner, S.M., & Taylor-Ferreira, J.C. (1999). Teaching study skills and test-taking strategies to elementary school students. *Behavior Modification, 23*(4), 30-46.

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). Available from http://www.ucalgary.ca/~iejll

Calkins, L., Montegomery, K., & Santman, D. (1999). Helping children master the tricks and avoid the traps of standardized tests. *Practical Assessment, Research & Evaluation*, *6*(8). Available from http://pareonline.net/getvn.asp?v=6&n=8

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

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

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

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

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

Duke, D.E. (1999). Real learning: Do tests help? *Action for Better Schools*, *6*(2), 1-12.

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

Earl, L., Levin, B., Leithwood, K., Fullan, 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, UK.

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.

Education Quality and Accountability Office. (1998). *Educators handbook*. Toronto, ON: Queen's Printer of Ontario.

Goodnough, A. (1999, December 9). New York City teachers nabbed in school-test cheating scandal. *National Post*, p. B1.

Gronlund, N.E. (2003). *Assessment of student achievement* (7th ed.). Boston: Allyn and Bacon.

Gustafson, R.W. (2001). The future of Canadian public education: Canadian or American? *Education Canada*, *42*(1), 40-43.

Hilliard, A.G. (2000). Excellence in education versus high-stakes standardized testing. *Journal of Teacher Education, 51*(4), 293-304.

Joint Committee on Testing Practices. (2005). Code of fair testing practices in education (revised). *Educational Measurement: Issues and Practice*, *24*(1), 23-26.

Kane, M. (2001, April). *The role of policy assumptions in validating high-stakes testing programs*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, *21*(1), 31-41.

Kaufhold, J. (1998). What's wrong with teaching to the test? *School Administrator*, *11*(55), 14-16.

Levinson, C.Y. (2000). Student assessment in eight countries. *Educational Leadership*, *57*(5), 58-61.

Lewington, J. (1999). Accountability: Reality or pretense? *Education Canada*, *39*(3), 36-37.

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

Marshak, D. (2003). No Child Left Behind: A foolish race into the past. *Phi Delta Kappan*, *85*(3), 229-231.

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). Available from http://pareonline.net/getvn.asp?v=7&n=8

Meaghan, D.E., & Casas, F.R. (1995). On standardized achievement testing: Response to Freedman and Wilson and a last word. *Interchange*, *26*(1), 81-96.

Mehrens, W.A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practices*, *8*(1), 14-22.

Messick, S. (1996). *Validity and washback in language testing.* (Report No. ETS-RR-96-17). Princeton, NJ: Educational Testing Service.

Miyasaka, J.R. (2000, April). *A framework for evaluating the validity of test preparation practices.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

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

Neil, M. (2003b). The dangers of testing. *Educational Leadership*, *60*(5), 43-46.

Ontario English Catholic Teachers Association. (2002). *Weighing in: A discussion paper of provincial assessment policy.* Available from http://www.oecta.on.ca/pdfs/weighingin.pdf

Popham, W.J. (2000). The mismeasurement of educational quality. *School Administrator*, *57*(11), 12-15.

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

Popham, W.J. (2004). *Classroom assessment: What teachers need to know* (4th ed.). Boston: Prentice Hall.

Power, M.A. (1999). *Ethical standards in testing: Test preparation and administration*. Vancouver, Washington: Washington Educational Research Association.

Rogers, W.T., & Bateson, D.J. (1991). The influence of test-wiseness on performance of high school seniors on school leaving examinations. *Applied Measurement in Education, 4*(2), 159-83.

Rubenstein, J. (2003). *Test preparation: What makes it effective?* (Report No. ED 480-063). North Carolina: U.S. Department of Education.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Smith, M.L. & Fey, P. (2000). Validity and accountability of high-stakes testing. *Journal of Teacher Education, 51*(5), 334-344.

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

Vaughan, A.C. (2002). Standards, accountability, and the determination of school success. *Educational Forum, 66*(3), 206-13.

Volante, L. (2005). Accountability, student assessment, and the need for a comprehensive approach. *International Electronic Journal for Leadership in Learning, 9*(6). Available from http://www.ucalgary.ca/~iejll

Wei, H.H. (2002, April). *Teachers' responses to policy implementation: Interactions of new accountability policies and culturally relevant pedagogy in urban school reform.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Wideen, M.R., O'Shea, T., Pye, I., & Ivany, G. (1997). High-stakes testing and the teaching of science. *Canadian Journal of Education, 22*(4), 428-44.

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

Wolfe, R., Childs, R., & Elgie, S. (2004). *Final Report of the External Evaluation of EQAO's Assessment Processes.* Toronto, ON: Ontario Institute for Studies in Education of the University of Toronto. Available from http://www.eqao.com/pdf_e/04/04p014e.pdf

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). Available from http://cie.ed.asu.edu/volume5/number5

Yeh, S.S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher, 30*(9), 12-17.

Louis Volante obtained his Ph.D. at the Ontario Institute for Studies in Education at the University of Toronto. Prior to joining Brock University, he held faculty appointments at Concordia University and the University of Hawaii. His research focuses on assessment for learning, fairness/bias in assessment, assessment reform, and educational change. He currently teaches assessment and evaluation at both the pre-service and graduate level.

Author's Address:
Brock University
Faculty of Education
1842 King Street East
Hamilton, Ontario
CANADA L8K 1V7
EMAIL: Louis.Volante@Brocku.ca