Equity in Multicultural Student Assessment

LOUIS VOLANTE
Brock University

ABSTRACT: Research on multicultural groups such as Black, Hispanic, First Nation, and English Second Language students suggests that recent standards-based reforms and the large-scale assessment programs that accompany their implementation are unfair. This paper argues that promoting fairness in large-scale assessment will only occur when equity is viewed as a multi-faceted construct. Issues surrounding technical quality, utilization, reporting, and most importantly, educational opportunity, provide an overarching framework for understanding fairness and bias in large-scale assessment. The conclusion suggests comprehensive reforms that strengthen schools, families, and local communities are crucial for closing the achievement gap.

RESUME: Une recherche effectuée sur les minorités multiculturelles telles que les noirs, les hispano-américains, les autochtones, et les étudiants en anglais langue étrangère, sous entend que les récentes réformes s'appuyant sur des méthodes traditionnelles et que les programmes d'évaluation à grande échelle qui accompagnent leur mise en place, ne sont pas équitables. Ce papier soutient que de promouvoir l'impartialité à grande échelle, n'arrive que lorsque l'équité est considérée comme un concept à multi facettes. Les questions concernant la qualité technique, l'utilisation, la couverture et le plus important de tout la pédagogie, apportent une vue globale sur la compréhension de l'équité et le parti pris dans l'évaluation à grande échelle. La conclusion laisse entendre que les réformes d'ensemble qui encouragent l'école, la famille et les communautés locales, sont déterminantes pour réduire l'écart dans les faits.

Large-scale assessment continues to provoke intense debate across North America and much of the Western World. Advocates argue that only through bench-marking student achievement results, can schools become compelled to improve their instructional approach. Armed with

Journal of Educational Thought
Vol. 42, No. 1, 2008, 11-26.

objective assessment data, policy-makers and other educational leaders can make sound decisions for accountability purposes and identify exemplary teaching practices that may be disseminated for system improvement (Covaleskie, 2002; Taylor & Tubianosa, 2001). Critics disagree and argue that over-reliance on large-scale assessment produces an oversimplified vision of educational excellence with predictable winners and losers (Kohn, 2000; McNeil, 2000.) They note a long history of documented abuses and unintended negative consequences associated with high-stakes testing, particularly in multicultural student populations. Current research on Black, Hispanic, First Nation, and English Second Language students continues to support these criticisms, suggesting large-scale assessment programs are often unfair and require significant rethinking. This paper argues that assessment equity is a multi-faceted construct based on four interrelated issues – technical quality, reporting, utilization, and most importantly, educational opportunity. The proposed framework is meant to inform the direction of future assessment-led reforms and ultimately improve educational outcomes for multicultural students.

North American Context

Large-scale achievement testing has been widely implemented because it is perceived to address a broad array of problems, including low standards, weak student motivation, poor curriculum and instruction, inadequate learning, and educational inequality (Darling-Hammond, 2003). Currently, every state, province, and territory within North America (with the exception of Prince Edward Island) administers large-scale assessment measures. These tests continue to exert the most important influence on curricula and pedagogy within North American schools – particularly since they are often used to make high-stakes decisions for students and teachers. In Canada, these measures often comprise a significant portion (i.e., 30% - 50%) of a students' final grade or serve as a graduation requirement (i.e., Ontario Secondary School Literacy Test). In the United States however, these tests also serve high-stakes purposes for teachers and administrators as reflected in the provisions of the federal *No Child Left Behind* (NCLB) act. Essentially, schools that fail to meet "Adequate Yearly Progress" (AYP), as reflected in mandated improvements in test scores, are labeled as "failing," and are eventually taken over by the state. Couple this with recent announcements to introduce merit-based pay incentives that are tied directly to test scores, and it is easy to understand why schools

inevitably devote excessive amounts of time on test content at the expense of other areas of the curriculum. This tendency to teach closely to the test has been reported in virtually every North American jurisdiction that administers large-scale assessment measures.

Intended and Unintended Consequences

The policy intentions and the outcomes of large-scale assessment programs are often incongruent. Although the architects of standards-based reforms sought to correct inequities in the system, the top-heavy testing approach has had virtually the opposite effect and resulted in many unintended negative consequences for multicultural student populations. Research has consistently shown that Black, Hispanic, First Nation, and English Second Language students do not do as well as their White counterparts (Demmert, 2005; Fox, 1999; Froese-Germain, 1999; Gonzalez, 2002; Jones & Ongtooguk, 2002). This achievement gap has not closed, and in many instances, has actually become larger since the introduction of standards-based reforms (McNeil, 2000; Valencia & Villarreal, 2003). The result is that schools serving multicultural students are often compelled to narrow the curriculum even further to boost test scores. When one considers the long-standing criticism that standardized achievement tests do not measure higher-order and critical thinking skills (Ketterlin-Geller, Leanne, McCoy, Twyman, & Tindal, 2003; Yeh, 2001), it seems logical that multicultural students are more likely to be the recipients of a truncated curriculum.

The pressure to raise test scores also extends beyond mere curriculum concerns. In some contexts, including Texas and Massachusetts, there is a sizable bump in the percentage of students who are retained the year before the states' high-stakes exams are given (Kornhaber, 2004). Unfortunately, the extra year that students are given to "catch up" is associated with increased drop-out rates but not with increased learning (Hauser, 2001). High school completion rates have also fallen in Canadian provinces (i.e., Ontario and Alberta) with high-stakes consequences attached to student performance (Volante, 2007). This negative relationship with student retention suggests external testing may actually push students out of school. Similarly, pressure placed on low performing schools which typically operate in the most distressed urban areas, often makes it difficult to recruit and retain creative and talented teachers and administrators (Berlak, 2001). Recent statistics suggest that about two-thirds of Latino, African-American, and Native American eighth-grade math students in the United States have

teachers who do not have an undergraduate degree in mathematics, compared with half of all White students (Haycock, 1998). Thus, multicultural students' grade promotion status and the pedagogical practices they experience are also compromised by current standards-based reform agendas.

Toward a New Framework

Educators and psychologists have strenuously argued for equity in testing and assessment, for "culture free," "culturally fair," and "culturally relevant" or "culturally salient" testing and assessment (Hilliard & Amankwatia, 2003). Unfortunately, these equity oriented concepts do not encompass a broad understanding of the challenges faced by multicultural student populations since they focus primarily on technical quality issues. Although the latter are worthy of attention, the scope of analysis for equitable assessment is much broader and also includes issues related to utilization, reporting, and most importantly, educational opportunity. Thus, assessment equity is a multi-faceted construct that requires a number of inter-related solutions – some of which extend beyond the control of test developers and educators. Indeed, the most pressing problems facing multicultural students require the active engagement of policy-makers since they have a profound influence on the distribution of resources both within and outside of schools. The next section challenges the current value assumptions underpinning each of the previously noted issues as a way to inform future assessment-led reforms.

Technical Quality

Large-scale achievement measures typically reflect the cognitive styles, contexts, and experiences of White middle class students (Purnell, 2000). Assessing and evaluating the progress of students from diverse racial and ethnic groups and social classes is therefore complicated by differences in language, learning styles, and culture (Banks, et al., 2001). Test developers have responded to this challenge by developing sophisticated statistical techniques such as Differential Item Functioning (DIF) and Differential Test Functioning (DTF) that assess of how well two individuals with identical ability but different group membership perform on an item/test (Ellis & Raju, 2003; Gelin & Zumbo, 2003). For example, Banks (2006) analysis of fifth grade reading and language arts data from the Terra Nova test indicated that although

Hispanic and Black students did not differ from matched White students in their cluster scores, Black students were differentially drawn to incorrect options that illustrated various aspects of their culture. The latter suggests that even the design of distracters within multiple-choice items, not just the nature of the question stem, must be closely scrutinized to avoid cultural bias.

Although the previously noted statistical procedures are important components of test and scale development, they merely describe anomalies in test performance. The fact that scores on all commercially produced tests show the same eight to ten percent gap suggests that the gap cannot be fully explained by racial or cultural bias lodged in individual test items (Berlak, 2001). Clearly, more profound structural constraints must be acknowledged before one can faithfully validate particularly assessment approaches (Kane, 2002). Traditional on-demand paper-and-pencil tests seem destined to always favor certain groups since they measure particular skills and types of knowledge. Battiste (1998) refers to this general problem as "cognitive imperialism" and notes the inherent bias of implementing standards from a Western educational frame of reference.

Although research suggests the inclusion of performance-based assessment components in public examinations presents some important challenges (Elwood, 1999), there are compelling reasons to view this approach as a more viable and equitable alternative for the future. Soodak (2000) articulates three key reasons why performance-based assessment should be adopted for accountability purposes. First, performance assessment may accommodate differences in students' ways of knowing by permitting, and even encouraging, multiple presentation and response formats. Second, because performance assessments are curriculum-embedded, they may minimize opportunities for biases resulting from a student's lack of familiarity with the language, context, and format of the test. Third, students and parents are more likely to feel that they are involved in the assessment process when performance measures are employed. Educators can also improve assessment by promoting the inclusion of oratory skills to balance reading skills, by ensuring culturally relevant curriculum, and by factoring in students' language and experience when judging their abilities (Fox, 1999).

Utilization

It is important to note that changes in the basic form of an assessment, whether it be traditional or authentic in nature, are unlikely to enhance equity unless we also change the ways in which assessments are used (Darling-Hammond, 2003). Large-scale assessment programs that are reliable, valid, and include multiple measures that tap different learning styles and cultures, are still unfair when they are utilized for high-stakes decisions that outstretch their abilities. For example, denying a student their earned diploma based on a single test score is not fair nor is it an effective method for raising academic achievement within a school system (Johnson, Boyden, & Pittz, 2001). It is equally unacceptable to use test results to classify a student as disabled when he or she did not have a disability or to track a student into an inferior or segregated educational placement (Soodak, 2000). Sadly, multicultural students such as Black, Hispanic, First Nation, and ESL students are less likely to pass exit exams and are more likely to be placed into special education classes than their White counterparts (Case & Taylor, 2005; Handy, 1999; Zhang & Katsiyannis, 2002). In addition to limiting students' life chances, large-scale assessments may also reinforce biases teachers may have toward particular groups of students. For example, research suggests that multicultural student populations and those at-risk for dropping out report more bias from their teachers than do White students and those who are more academically advanced (Wayman, 2002). The cumulative effect is that multicultural students – particularly those who perform poorly on external tests – are more susceptible to becoming disenchanted and disengaged from school.

Using assessment data to classify a school is equally problematic and also leads to a host of negative consequences. The customary practice of ranking schools in local newspapers has precipitated documented cases of inappropriate teacher and administrator behaviour (Simner, 2000). Encouraging low-performing students to stay home the day of the test, offering subtly cues to correct responses, and even providing correct answers to multiple-choice questions become more common when relative school performance becomes highly publicized. Ranking also tends to drive highly qualified and competent teachers out of schools serving vulnerable student populations such as those within inner city environments (Delphi, 1998; Hargreaves & Fink, 2006). Thus, it seems clear that the practice of using assessment results to compare and rank schools creates unhealthy competition and stifles the more desirable objective of fostering a community of learners (Volante, 2005). The

alternative is to utilize assessment results primarily as vehicles for developing action plans to improve instruction, staff development, school organization, resource and community engagement – as ways of improving students' lives and ensuring that all children receive a high quality education (Earl, 1999).

Reporting

Research across North America has consistently demonstrated that between 40 and 50 per cent of students' scores on a standardized achievement test is attributable to socio-economic status (Crislip & Heck, 2001; Lytton & Pyryt, 1998; Sacks, 2000). Indeed, research from the three major international assessments: Program in International Student Assessment (PISA); Progress in International Reading and Literacy Study (PIRLS); and Third International Mathematics and Science Study (TIMSS); suggests that the link between socioeconomic status and achievement is a pervasive feature around the world (see Marks, Cresswell, & Ainley, 2006; Myrberg & Rosen, 2006; Nash, 2003; Papanastasiou, 2000; Thorpe, 2006; Willms, 2006). Unfortunately, parents and the public are often not provided with information that allows them to clearly understand the differences between groups of students or the limitations that exist with particular measures. This is compounded by the fact that the differences that do exist between groups of students are often exaggerated by analysis and reporting techniques. Berlak (2001) illustrates this point within the California context where the gap between low and high performance on a basic literacy test is 3.2 multiple-choice test items on a 50 item test. Although this is a relatively insignificant difference from an educational point of view, he notes how the norming of the test and the manner in which cut scores are set grossly inflates this difference for persons of color. For example, African-American test-takers are 3.5 times more likely to fail the test than Whites, Latino/Hispanics more than twice as likely and Asian Americans more than 1.5 times more likely than Whites.

In Canada, differences between schools and districts are also reported in a manner that invites misunderstanding and possible incorrect interpretations. For example, many schools are often compared based on raw score differences with little, if any, attention paid to measurement error. This is despite the fact that when uncertainty is taken into account, many – sometimes most – differences in raw scores between schools and districts disappear (Earl, 1999). Thus, it seems

clear that those in the education community need to model effective dissemination of assessment information, whether at the international, national, state, or provincial level (McEwen, 1999). Failure to do so inevitably leads to the perception that multicultural students and the schools that serve them are inferior.

Educational Opportunity

Schools can be thought of as a collection of opportunities to learn with fairness of opportunity determined by comparing the learning opportunities students have within, across, and outside of schools. Opportunities to learn within a school are related to factors such as a) funding (i.e., small class sizes, quality resources/technology, extracurricular activities); b) teacher quality (i.e., experience, preparation to teach the content, participation in high-quality professional development, verbal ability, and opportunity to receive teacher rewards and incentives); c) rigor of the curriculum (i.e., breadth and depth of content coverage, time engaged in learning); and d) school organization (i.e., safe and orderly learning environment, effective parent involvement strategies, avoidance of rigid grouping practices) (Darling-Hammond, 1997; Gonzlez, 2002). Research suggests that all of these factors are below standard for multicultural student groups (Dreeban & Gamoran, 1986; Lee, 2004). For example, in terms of teacher quality, the least qualified teachers typically end up teaching the least advantaged students – a situation exasperated by inadequate supports for beginning teachers and the poor working conditions within many schools, especially those in urban settings serving multicultural students (Darling-Hammond, 2001).

Similarly, the relationship between funding and curriculum necessitates that schools possess adequate resources to deliver quality programs. As one 16-year-old student from Oakland, California commented:

> We lack resources, we have no money, and other schools when you look at their high [test scores] you wonder why. It's because they have money, they have resources. They have all that which we need over here at the bottom and people don't even bother to think maybe that's the problem. And yet you want to blame it on the youth. (Johnson, Boyden, & Pittz, 2006, p. 9)

This relationship between resources and student achievement can also be extended to broader relationships. For example, Willms (2005) suggests that Alberta and Ontario's high performance relative to other

Canadian provinces on PISA is partly attributable to the level of school resources within those educational jurisdictions.

Research is also emerging that should inform future school organization reforms. Consider a recent analysis of student achievement in 30 countries by Marks, Cresswell, and Ainley (2006). These authors found that schools with tracking and other forms of differentiation tended to magnify socioeconomic inequalities in education. The latter is a significant finding since tracking fosters a negative self-image and creates reduced expectations for a student's future (Froese-Germain, 1999). Although the previous authors do not suggest completely abandoning such practices, they argue for policies that would delay the allocation of students to different tracks and for greater movement between tracks. Similarly, research suggests that the creation of small schools is vital for improving student achievement (Fine & Powell, 2001). These findings are especially important for multicultural students since they are consistently over-represented in less rigorous academic programs and are taught within large urban schools. Clearly, schools must be organized to promote contexts of learning in which all students, particularly those from at-risk multicultural groups, are encouraged to form high aspirations for achievement within a supportive learning environment.

Equally important is the recognition that many home-related factors, which are beyond a school's control, have a profound impact on student achievement. For example, student achievement correlates highly with parent education levels and family income. Since ethnic minorities are disproportionately represented in the low-income sector, inequalities in opportunities to learn will invariably contribute to the achievement gap of multicultural students (American Educational Research Association – AERA, 1999; Hilliard & Amankwatia, 2003). Indeed, the presence of material resources such as those classified as *wealth measures* (i.e., room for the student, link to the internet, and number of computers, televisions, mobile phones) along with *educational resource measures* (i.e., quiet place to study, possession of a dictionary, textbooks, calculators) have a significant impact on PISA reading, science, and mathematics achievement scores – particularly in countries with more income inequality such as the United States, Mexico, Brazil, Portugal (Marks, Cresswell, & Ainley, 2006). Similar findings have arisen from analyses of PIRLS (Myrberg & Rosen, 2006) and TIMSS (Papanastasiou, 2000) data. Thus, making educational opportunities equal within schools

is necessary but insufficient for many multicultural students who have socioeconomically disadvantaged family backgrounds (Lee, 2004).

By recognizing the importance of both within and outside school factors on student achievement we can avoid the defeatist myth that schools make no difference while avoiding the pitfalls of the other extreme, that they make all the difference (Rothstein, 2002). One way to recognize this balance is to utilize a value-added assessment approach for interpreting large-scale assessment results. Value-added assessment considers family background and emphasizes the degree of progress in students before making judgments about appropriate levels of achievement (Hershberg, Simon, & Lea-Kruger, 2004; Wickstrom, 1999). Essentially, this approach permits an examination of variables schools have control over – instructional approach –– with those factors they can not control – school demographics (Crudwell, 2005). Although there are challenges associated with value-added assessment, it does bring to the forefront one essential truism: not all children begin the achievement race at the same start line. Assessment-led reforms that fail to acknowledge this important fact ensure many children and schools will continually be classified as failures – even when they jump from 10% to 50% achievement in one academic year.

Conclusion

Diversity in schools is both an opportunity and a challenge to understand the fundamental differences that influence student learning and achievement. We must find ways to respect multicultural students and help create a collective ethic for which all children feel welcomed and appreciated. An important starting point for this discussion is the recognition that over-reliance on traditional large-scale assessment measures has often marginalized our most vulnerable student populations. Policy-makers need to begin to examine how various forms of data and multiple representations of student work can be used to present a more accurate and robust picture of student achievement. Essentially, performance-based measures and other forms of curriculum-embedded assessment must be incorporated into existing accountability frameworks. However, no meaningful dialogue is possible if we continue to endorse a market-placed solution to education that minimizes differences in opportunities to learn.

Fortunately, the preceding discussion suggested some targeted reforms can be utilized to address differences in opportunities to learn, and in doing so, help close the achievement gap. Some of the suggested

changes relate to improving the quality of the teaching/learning process within schools (i.e., high-quality teachers, less tracking, smaller class sizes and schools) while other address more complex home-related factors (i.e., parent education levels, community resources, greater income equity) that require the active engagement of senior policymakers. In truth, many of these factors indicate that sound educational policies cannot be divorced from the economic and social policies that characterize particular governments and countries. To do so, treats schools as islands in an ocean, with many multicultural students relegated to contexts that are academically impoverished or essentially barren,

Thus, the ultimate question is not *is this test fair?* but *what is the context for this assessment?* (Cline, 2000). Simply designing large-scale assessment measures with culturally neutral test items will not eliminate persistent problems with the achievement gap. Indeed, characterizing assessment equity within the narrow bounds of technical quality confines the scope of solutions solely to test developers and schools. Concerns with assessment design, administration, analysis, and reporting must also be informed by the inherent inequities that many multicultural students face within and outside of schools. Policy-makers must acknowledge the multi-faceted nature of assessment equity and promote comprehensive reforms that strengthen schools, families, and local communities. The latter would suggest an authentic desire to improve the educational outcomes of multicultural student groups. Clearly, standards-based reforms and the convention blame-and-shame approach with the public ranking of schools have not delivered on their promises. It's time to explore other options.

REFERENCES

American Educational Research Association. (1999). *Standards for educational and psychological testing.* Washington, DC: AERA.

Banks, K. (2006). A comprehensive framework for evaluating hypotheses about cultural bias in educational testing. *Applied Measurement in Education, 19*(2), 115-32.

Banks, J.A., Cookson, P., Gay, G., Hawley, W.D., Irvine, J.J., Nieto, S., Schofield, J.W., Stephan, W.G. (2001). Diversity within unity: Essential principles for teaching and learning in a multicultural society. *Phi Delta Kappan, 83*(3), 196-98.

Battiste, M. (1998). Enabling the autumn seed: Toward a decolonized approach to Aboriginal knowledge, language, and education. *Canadian Journal of Native Education, 22*(1), 16-27.

Berlak, H. (2001). *Academic achievement, race, and reform: Six essays on understanding assessment policy, standardized achievement tests, and anti-racist alternatives.* California: United States Department of Education.

Case, R.E. & Taylor, S.S. (2005). Language difference or learning disability? Answers from a linguistic perspective. *Clearing House, 78*(3), 127-30.

Cline, T. (2000, September). *Principles and practice of fair assessment.* Paper presented at the Minorities in a Pluralistic Society at the Turn of the Millennium, Brno, Czech Republic.

Covaleskie, J.F. (2002 April). Two cheers for standardized testing. *International Electronic Journal for Leadership in Learning, 6*(2). Retrieved June 5, 2004, from http://www.ucalgary.ca/~iejll/volume6/covaleskie.html

Crislip, M.A. & Heck, R.H. (2001). *Accountability, writing assessment, and equity: Testing in a multilevel model.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Crudwell, R.M. (2005). Alternative strategies for large-scale student assessment in Canada: Is value-added assessment one possible answer? *Canadian Journal of Educational Administration and Policy, 41.* Retrieved April 4, 2006, from http://www.umanitoba.ca/publications/cjeap/articles/crundwell.html

Darling-Hammond, L. (1997). *The right to learn.* San Francisco: Jossey-Bass.

Darling-Hammond, L. (2001). Apartheid in American education: How opportunity is rationed to children of color in the United States. In Johnson, T., Boyden, J.E., & Pittz, W.J. (Eds.), *Racial profiling and punishment in U.S. public schools: How zero tolerance policies and high stakes testing subvert academic excellence and racial equity* (pp. 39-44). Oakland, CA: Applied Research Center.

Darling-Hammond, L. (2003). *Standards and assessments: Where we are and what we need.* Retrieved January 13, 2004, from http://www.tcrecord.org

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

Demmert, W.G. (2005). The influences of culture on learning and assessment among Native American students. *Learning Disabilities Research and Practice, 20*(1), 16-23.

Dreeban, R. & Gamoran, A. (1986). Race, instruction, and learning. *American Sociological Review, 51,* 660-69.

Earl, L. (1999). Assessment and accountability in education: Improvement or surveillance? *Education Canada, 39*(3), 4-6.

Ellis, B.B. & Raju, N.S. (2003). *Test and item bias: What they are, what they aren't, and how to detect them.* North Carolina: United States Department of Education.

Elwood, J. (1999). Equity issues in performance assessment: The contribution of teacher-assessed coursework to gender-related differences in examination performance. *Educational Research and Evaluation, 5*(4), 321-44.

Fine, M. & Powell, L. (2001). Small schools: An anti-racist intervention in urban America. In Johnson, T., Boyden, J.E., & Pittz, W.J. (Eds.), *Racial profiling and punishment in U.S. public schools: How zero tolerance policies and high stakes testing subvert academic excellence and racial equity* (pp. 45-49). Oakland, CA: Applied Research Center.

Fox, S.J. (1999). *Student assessment in Indian education or what is a roach?* Washington, DC: United States Department of Education.

Froese-Germain, B. (1999). *Standardized testing: Undermining equity in education.* Ottawa, ON: Canadian Teachers Federation.

Gelin, M.N. & Zumbo, B.D. (2003 April). *Differential domain functioning on the numeracy component of the Foundation Skills Assessment: Bringing the context into picture by investigating sociological/community moderated test and item bias.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Gonzalez, R. (2002). *The No Child Left Behind Act: Implications for local educators and advocates for Latino students, families, and communities.* Washington, DC: National Council of La Raza.

Handy, A.J. (1999). *Ethnocentrism and black students with disabilities: Bridging the cultural gap,* Volume I. New York, NY: Vantage Press.

Hargreaves, A. & Fink, D. (2006). The ripple effect. *Educational Leadership, 63*(8), 16-21.

Hauser, R. (2001). Should we end social promotion? Truth and consequences. In G. Orfield & M.L. Kornhaber (Eds.), *Raising standards and raising barriers? Inequality and high-stakes testing in public education* (pp. 151-78). New York, NY: Century Foundation.

Haycock, K. (1998). *Thinking K-16: Good teaching matters a lot.* Washington, DC: The Education Trust.

Hershberg, T., Simon, V.A., & Lea-Kruger, B. (2004). The revelations of Value-Added: An assessment model that measures student growth in ways that NCLB fails to do. *School Administrator, 61*(11), 10-12.

Hillard, A.G. & Amankwatia, B. (2003 April). *Assessment equity in a multicultural society: Assessment and instructional validity in a culturally plural world.* Paper presented at the National Council on Measurement in Education, Chicago, IL.

Johnson, T., Boyden, J.E., Pittz, W.J. (2001). *Racial profiling and punishment in U.S. public schools: How zero tolerance policies and high stakes testing subvert academic excellence and racial equity.* Oakland, CA: Applied Research Center.

Jones, K., & Ongtooguk, P. (2002). Equity for Alaska natives: Can high-stakes testing bridge the chasm between ideals and realities? *Phi Delta Kappan, 83*(7), 499-550.

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

Ketterlin-Geller, L., Leanne, R., McCoy, J.D., Twyman, T., & Tindal, G. (2003). How do critical thinking measures fit within standards-based reform? *Assessment for Effective Intervention, 28*(3-4), 37-48.

Kohn, A. (2000). *The case against standardized testing: Raising scores, ruining the schools.* Portsmouth, NH: Heineman.

Kornhaber, M.L. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Educational Policy, 18*(1), 45-70.

Lee, J. (2004). Multiple facets of inequity in racial and ethnic achievement gaps. *Peabody Journal of Education, 79*(2), 51-73.

Lytton, H. & Pyryt, M.C. (1998). Predictors of achievement in basic skills: A Canadian effective school study. *Canadian Journal of Education, 23*(3), 281-301.

Marks, G.N., Cresswell, J., & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational Research and Evaluation, 12*(2), 105-28.

McEwen, N. (1999). Using international test results to improve student performance. *Education Canada, 39*(3), 44-46.

McNeil, L. (2000). Creating new inequalities: Contradictions of reform. *Phi Delta Kappan, 81*(10), 729-34.

Myrberg, E. & Rosen, M. (2006). Reading achievement and social selection in independent schools in Sweden: Results from IEA PIRLS 2001. *Scandinavian Journal of Educational Research, 50*(2), 185-205.

Nash, R. (2003). Is the school composition effect real? A discussion with evidence from the UK PISA Data. *School Effectiveness and School Improvement, 14*(4), 441-57.

Papanastasiou, C. (2000). Internal and external factors affecting achievement in mathematics: Some findings from TIMSS. *Studies in Educational Evaluation, 26*, 1-7.

Purnell, R.B. (2000). *Making dry bones live: The role of testing in equity and access.* Paper presented at the National Council of Teachers of English, Milwaukee, WI.

Rothstein, R. (2002). *Out of balance: Our understanding of how schools affect society and how society affects schools.* Chicago, IL: The Spencer Foundation.

Sacks, P. (2000). *Standardized minds.* Cambridge, MA: Perseus Books.

Simner, M.L. (2000). *A joint position statement by the Canadian Psychological Association and the Canadian Association of School Psychologist on the Canadian press coverage of the province-wide achievement test results.* Retrieved March 4, 2004, from http://www.cpa.ca/documents/joint_position.html

Soodak, L.C. (2000). Performance assessment and students with learning problems: Promising practice or reform rhetoric? *Reading and Writing Quarterly, 16,* 257-80.

Taylor, A.R. & Tubianosa, T. (2001). *Student assessment in Canada: Improving the learning environment through effective evaluation.* Kelowna, BC: Society for the Advancement of Excellence in Education.

Thorpe, G. (2006). Multilevel analysis of PISA 2000 reading results for the United Kingdom using pupil scale variables. *School Effectiveness and School Improvement, 17*(1), 33-62.

Valencia, R.R. & Villarreal, B.J. (2003). Improving students' reading performance via standards-based reform: A critique. *The Reading Teacher, 56*(7), 612-21.

Volante, L. (2005). Accountability, student assessment, and the need for a comprehensive approach. *International Electronic Journal for Leadership in Learning, 9*(6). Full text available from http://www.ucalgary.ca/~iejll/volume9/volante.htm.

Volante, L. (2007). Standards-based reform: Can we do better? *Education Canada, 47*(1), 54-56.

Wayman, J.C. (2002). *Student perceptions of teacher ethnic bias: Implications for teacher preparation and staff development.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

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

Willms, J.D. (2006). Variations in socioeconomic gradients among cantons in French- and Italian-speaking Switzerland: Findings from the OECD PISA. *Educational Research and Evaluation, 12*(2), 129-54.

Willms, J.D. (2005). Why the difference? *Education Canada, 45*(1), 54-55.

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

Zhang, D. & Katsiyannis, A. (2002). Minority representation in special education: A persistent challenge. *Remedial and Special Education, 23*(3), 180-87.

Louis Volante is an Assistant Professor in the Faculty of Education, Brock University, teaching assessment and evaluation courses at both the pre-service and graduate level. Prior to joining Brock University, he held faculty appointments at Concordia University and the University of Hawaii, and also lectured at the Ontario Institute for Studies in Education of the University of Toronto, where he obtained his Ph.D. His research interests include classroom and large-scale assessment, international comparisons of assessment reform, educational program evaluation, teacher education and development, 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.

Author's Address:
Brock University
Faculty of Education
1842 King Street East
Hamilton, Ontario, L8K 1V7
EMAIL: louis.volante@brocku.ca