THE STANDARDIZED TESTING MOVEMENT: EQUITABLE OR EXCESSIVE?

FRANK NEZAVDAL *Brock University*

ABSTRACT. This article addresses the problems associated with increased standardized testing, specifically in the context of the public school system in Ontario. I question the legitimacy of the persistent push for externally mandated comparative data on teaching by taking a holistic standpoint, one that treats teaching and assessing as inseparable parts of a complete process. I begin by situating the debate in the current Ontario context. Next I examine the ideological background to the current government support for standardized tests. In the third section, I discuss in detail the arguments against standardized testing . Finally I suggest non-comparative, qualitative alternatives that are more truly equitable.

LE MOUVEMENT EN FAVEUR DES TESTS STANDARDISÉS : ÉQUITABLE OU EXCESSIF?

RÉSUMÉ. Cet article touche aux problèmes reliés à l'accroissement de l'usage des tests standardisés dans le contexte particulier des écoles publiques de l'Ontario. Je mets en question la légitimité de cette pression constante pour des données comparatives sur l'enseignement en adoptant un point de vue holistique, celui qui considère l'enseignement et l'évaluation comme des parties inséparables d'un tout. Je commence par situer le débat dans le contexte ontarien actuel. Ensuite, j'examine les fondements idéologiques de la position gouvernementale en faveur des tests standardisés. Dans la troisième partie, je présente en détail les arguments qui vont à l'encontre des tests standardisés. Enfin, je suggère d'autres moyens d'évaluation non comparatifs et qualitatifs qui sont beaucoup plus équitables.

Ontario's testing landscape

This article is primarily concerned with the consequences of the current swing of the positivist pendulum as it pertains to assessment in Ontario. More specifically, I examine and question the approaches to standardized tests that are currently being used in Ontario public schools. In 1996 the provincial government of Ontario created the Education Quality and Accountability Office (EQAO). The mandate of this political body is to test achievement of all students in the province in the public school system. At

an estimated cost of $59 million (Ontario Teachers' Federation, 2003), the EQAO has been implementing proprietary performance and criterion-referenced language and mathematics tests since 1997 at the grade 3 and 6 levels, and has since implemented tests in grade 9 and the Ontario Secondary School Literacy Test (OSSLT). Currently the EQAO has plans to bring out 12 new tests from grade 4 through grade 11 by 2007 (Ministry of Education, 2001). There is talk in the media about the possibility of the creation of a new high-school diploma that will be awarded to those students who fail to pass the OSSLT.

Many of the issues that arise in dealing with Ontario's testing context are universal and reoccur wherever similar forms of testing are being employed. However, while debate surrounding the use of standardized tests occurs in all levels of education, the issue becomes more critical when a school system imposes these standardized tests on children and adolescents. The consequences of testing in a K-12 setting are more significant than they may be in an adult or professional learning program because the affected learners in primary and secondary schooling cannot yet be seen as being fully responsible for their own learning. So, while many of the issues discussed herein may apply to post-secondary education, the stakes are different. Here participation in educational programs is not a voluntary act by adults who have taken over much or all of the responsibility for their own educational development, but rather an involuntary requirement imposed on children and adolescents in a mandatory school system.

As matters stand, there is no clear approach or consistent philosophy being applied to assessment practices in Ontario schools. On the one hand, teacher education programs and post-graduate education programs are often exploring an assessment-as-reflective-inquiry paradigm in attempting to create effective teachers, yet on the other hand, policymakers continue to implement policies that further separate the assessment process from the classroom environment. Assessment initiatives are coming from several directions including teacher, school, and school boards, and these sometimes conflict with the goals of the tests being mandated by the EQAO because these tests are autocratically constructed, externally administered, and externally marked.

Much conjecture exists in school communities regarding the efficacy and usefulness of EQAO tests. Only recently, however, have studies surfaced that reveal that neighbourhood "social capital" (Berthelot, Ross & Tremblay, 2001, p. 25) and other "neighbourhood effects" (Harris & Mercier, 2000, p. 211) have a significant impact on how well students perform on EQAO tests and at school. The Berthelot and Harris studies provide compelling evidence that EQAO test results are strongly linked to local and regional socioeconomic variables. The mass testing of students through imposed

standardized tests does two things: it penalizes students who are socially disadvantaged, and it limits the ability of teachers to help rectify that social inequity.

The OSSLT is now being used to guide the decision to promote students from high school. Teachers' efforts are becoming more constrained in this context because they are no longer the primary instruments used to assess children's literate abilities. It is particularly troubling to know that teachers stand to have no input in the promotion decisions about students who have failed the OSSLT. This loss of input raises inherent questions of teacher independence and responsibility as it pertains to equitable teaching practices. Thus, when teachers and students are being faced with externally mandated tests increasingly disconnected from classroom activities, examining the issues further becomes an urgent task.

Standardized testing: the background.

Assessment is a high stakes social practice occurring in schools and using commonly recognized norms. These norms are not incidentally held but deliberately upheld to stream students – to propel some forward while systematically impeding others. There are two major strains of testing, the comparative and non-comparative methods. Agencies and governments are currently engaged in increasingly frequent and in-depth movements to establish comparative results through "objective" testing.

The development of standardized tests has a long history in the western tradition of schooling, training, and educational programs. Murphy (1997) suggests why: "standardized testing is, perhaps, a prototypical exemplar of this broader desire to control chance" (p. 263). The desire to control chance with large-scale assessments continues to dominate the assessment frameworks in educational institutions in the United States and Canada (Serafini, 2002). The practice of creating and validating tests with the aim of identifying objective knowledge and intelligence has arisen out of the modernist western logico-scientific paradigm. But, as Tindal et al. (2001) demonstrate, "investigation of the reliability and validity of these measures has not kept pace with their development and use" (p. 203). Serafini (2002) describes the traditional view as follows: "'assessment as measurement' is closely aligned with a modernist philosophy and supports the factory model of education" (p. 69). From this perspective, knowledge is broadly seen as existing apart from its users, in that it can be objectively categorized and usefully quantified. Creswell (2002) describes this approach aptly: "just as atoms and molecules are subject to predicable laws and axioms, so also are the patterns of children's behaviors in schools" (p. 45).

The term "mass testing" pertains to two classes of standardized tests: criterion-referenced tests and norm-referenced tests. These types of tests, along

with large-scale performance tests, continue to predominate in the desire to rank and score masses of students. While more will be said about the matter later, the critical point to remember about all standardized tests on which high-stakes decisions are based is that, by definition, mass tests cannot be based on individual students' profiles or needs. Norm-referenced tests compare students to their peers, or a specified percentile-weighted score of similar students. The strength of this type of test is that it can provide a picture of what a student is able to achieve in relation to others. However, gleaning that information is not sufficient; acting upon the information is what helps the struggling student. Criterion-referenced and performance-based tests can provide useful data about a learner's progress. They fail, however, to tell us *why* a student cannot achieve, as the information that these types of tests provide is limited.

The creation and use of standardized tests find varying levels of support in different jurisdictions. They are increasingly being promoted in the name of the "basic skills and accountability" rhetoric of global competitiveness. Historically, there have been numerous "pendulum swings" toward and away from positivist approaches to testing (Linn, 2001, p. 29). And while support for standardized tests has moved in and out of favour, the pendulum has clearly been swinging towards the increased implementation of mass testing in Canada, particularly in Ontario.

The critique of standardized testing

Positions countering the positivist research tradition that supports standardized testing have arisen with the move to post-modern and interpretivist approaches in qualitative research designs. Oddly, these continue to be termed "alternate" approaches (Creswell, 2002. p. 47; Connell, Johnson & White, 1992, p. 209). It is these alternate approaches to assessment that Erickson (1986 cited in Serafini, 2002) asserts have "characteristics [that] align with qualitative research methodologies and with a constructivist perspective towards knowledge" (p. 72).

The qualitative paradigm allows for a well-grounded critique of standardized tests. Such tests are ineffective at revealing what a student can do beyond their scope, and they are useless at explaining why a student has not achieved the desired outcomes. Quantitative testing, then, must be seen as having practical limits in usefulness. In the Ontario context, when a teacher is forced to fail a student who has not passed the OSSLT regardless of all other achievements, the test must be seen as being in conflict with the goals of equitable classroom practice. A single test is being used to decide promotion at the end of high-school. Thus, the aims of the EQAO are substantively disconnected from teachers, and their tests offer punitive repercussions for failing students.

Research has been questioning the objective validity of standardized tests since their inception (Resnick, 1980). The questioning of the objectivity of psychometrics has drawn attention to problems of systematic and cultural bias on standardized tests (Chase, 2001; Lewis 2001; Bracey 2001 & Merrow, 2001). Recently, Hoover's (2000) study of the standardized testing movement in Ohio has revealed that socioeconomic factors are consistently the best predictors of standardized test scores in that state. To that end, Kohn (2000a) notes the absurdity of the matter: that it would be far more efficient – with virtually the same results – to promote and demote students according to their household income (p. 7). Hansen (1993) regards the standardized testing movement as being "somewhat analogous to using a tape measure as the treatment to reduce one's waist size" (p. 20). However, the problem with standardized tests has moved beyond validity, long held to refer primarily to whether a test actually measures what it claims to measure. Berlak (1992), in his call for a new science on assessment, concluded that test designers must consider the impact of all tests on schools, teachers, and students. He reminds us that:

> In the real world of schooling, separating means and ends is not possible. All assessment procedures have the power to directly or indirectly shape social relations... moral questions arise in all social relationships, which can either be resolved by the use of direct or indirect power where the values, beliefs and ideologies of those with the ability to impose their will prevail, or by a process wherein conflicts are acknowledged, and mediated recognizing both differences and commonalities in interests and values. (p. 16)

What is happening in Ontario schools is precisely Berlak's (1992) concern: government is imposing its will in a way that shows it is strongly committed to prevailing in this debate.

Assessment as social construct

Connell, Johnston and White (1992) describe assessment as a social technique, "an appraisal of socially-situated practices of reasoning and communication" (p. 209). O'Sullivan (1999) adds that "all observations are 'socially constructed' and are, therefore, systems of interpretation" (p. 91). Educators assess through a necessarily phenomenal world of experience where it is contradictory to affirm the absolute objective validity of intelligence quotient and other forms of standardized tests. That is, teachers must be involved as co-constructors of knowledge and be fully aware of the innate inter-subjectivity of classroom experience. Standardized tests seek to assess individuals, young people who bring a different range of experiences to the classroom, through a most peculiar claim: that all students can learn differently and come from inequitable backgrounds but be evaluated in the same way, at the same time, by the same test – designed, of course, by those who

are the social power-holders. Such a notion of objective testing is typically upheld by the classically conservative from the 'assessment-as-measurement' movement, who, likely, by the vanity generated in their own successes on such 'objective' tests, believe that the unbiased standardized test is not only possible, but useful and desirable. Adding to the inappropriateness of standardized tests are the limited factors that can possibly be tested: recall and interpretation that require students to have the same social, economic and language-based learning experiences to make similar sense of the questions. How students reason for their answers on such tests is never known, only the singular selected or written response. Assessing learning requires that we assess the reasoning. Assessing reasoning demands that we genuinely know and understand our students and their needs – those necessary precursors to effective teaching.

The main problem with standardized tests is that they gauge pluralistic knowledge against one standard set of knowledge. Unlike teachers, tests cannot strive to include and bridge the gaps between dominant knowledge and the students' knowledge. One test cannot possibly assess a classroom of individual instincts and pluralistic powers any more than we could singularly assess a nation. Stereotyping is problematic for the same reason: it reveals some limited aspects of the target groups, then it accentuates their differences simply because they are not part of the dominant norm.

The dilemma exists that some of what is collectively valued and perpetuated as desirable knowledge within the dominant norm is destructive and unsustainable, yet this model of using one standard for valuing and identifying knowledge and learning remains dogmatically prominent. O'Sullivan (1999) raises a compelling point regarding the construction of human communities, that "one of the perennial problems that human communities have faced in the past, and still experience in the present, is a state of solidarity with the community as an in-group while excluding and denigrating an out-group" (p. 246). In the communities of schooling, high-stakes tests do precisely that: they create a state of solidarity within the group that passes, and exclude the out-group by virtue of their "failure". When Western society at large believes it is the global in-group, it is no wonder that we tacitly perpetuate this and overtly test for adherence in the classroom. Comparative standardized testing remains an accepted form of ethnocentricity, perpetuating marginalization.

The political dimension

The methodology of evaluating student performance has a widely debated history in Canada because the top policy-makers are rarely educators or students of education. In Ontario, the provincial government's creation of EQAO has created a situation for the provinces' stakeholders that commits students to achieve satisfactory outcomes on standardized high-stakes tests.

The provincial government's agenda ignores the possibility that standardized testing may function as a form of educational ethnic cleansing (Kohn, 2000b). Tests become high-stakes when they are utilized to assess students in a manner that has broadly-based and far-reaching implications on students' future achievement potential. Such tests create additional pressure on schools and teachers to ensure that students perform well on single tests, while teacher and school performance is increasingly being assessed by the results of students' achievement on such tests. Clearly, there is nothing progressive about the current inability to reconcile testing approaches with political aims. The teacher, the best fit to teach, is forced to fail a student who cannot pass one literacy test, heedless of why, or to teach only to the test to give the struggling student a chance of passing.

Standardized tests have been called into question as being either tautological, non-inclusive, or both (see Kohn 2000b; Chase, 2001; Lewis 2001; Bracey 2001 & Merrow, 2001). To date, the EQAO has not demonstrated enough about their implementation, their methodology, and the research basis of their tests to counter the critique that there are major problems with the tests themselves. My attempts to acquire empirical support for the technical adequacy of EQAO tests have proven futile as there are currently no provisions for public accessibility to any EQAO data beyond the test results that are annually released to the schools and media. It should be noted that an absence of empirical support is not unique to EQAO tests for, as Tindal et al. (2001) report, "inspection of 12 commercially criterion-referenced tests revealed that only 4 test manuals addressed reliability and validity" (p. 203). While these are problems of some commercially available tests that may apply to EQAO tests, testing practices in Ontario begin to move some distance further from rationality, if Allingham's (2000) accusations are correct:

> [There is] administrative ineptitude in the marking process. Considering that a panel of 'experts' . . . arbitrarily set the passing standard at about 70% . . . the writers had no stake in the process [and] many markers were not even teachers (p. 69)

If, indeed, non-teachers mark these tests, then it is clear how substantially disconnected from the classroom this process is becoming. This should be critically troubling because "marking has considerable consequences for educational and societal opportunities and is arguably one of the key instruments for selection and control" (Verkuyten, 2000, p. 453). When non-teachers are the markers, the test results are stripped further of any beneficial involvement in the classroom. The validity of the testing instrument is no longer the whole issue. There are also questions of authenticity of the testing environment. While the Ontario Curriculum documents specify outcomes for students to utilize dictionaries, peer and teacher feedback for their writing, the EQAO "'on demand' series of writing tasks did not give

students access to these customary resources" (Allingham, 2000, p. 69). This potentially renders the results on these tests ambiguous. The second crucial problem with these tests is more broad: one rooted in a problematic situation of dichotomous approaches and unreconciled philosophical understandings of educational praxis at the varied levels of educational policymaking and implementation in Ontario. In other words, politicians, touting the virtues of standardized "objective" tests and "accountability" are bulldozing their tests into the classroom, seemingly unaware of the potentially disastrous consequences of high stakes testing.

Student needs and ethical concerns

The goal of streaming students via mass comparative assessments works efficiently, but to what end is this streaming? Why do we presume that, because historically only the "brighter" students have excelled, this is a desirable outcome? To the contrary, society is maximally benefited by maximal possible achievements for all students. Further, the issue here is not simply a problem with the idea of streaming students into vocational and academic directions, but rather, that there is a significant potential of erroneously streaming students based on antecedent conditions that may take many years to correct. Educating students is about maximizing learning by meeting needs, by propelling passions and by nurturing human curiosity, not closing doors forever because of one test. Socially fair and just teaching must begin from the proposition that all students are always making something positive of themselves.

Students need basic skills and students need to become literate. While everyone involved claims to be interested in serving the needs of students, the disagreement about how this can be best achieved rests in whether the purpose of any testing is to help, or to hinder, students. The main question, then, is whether it is ethical to use high stakes tests to prevent students from proceeding if they are not achieving the basic skills. If assessment is used to identify the needs of students because it is a necessary precursor to learning optimally, we can then aim at extending learning. Holding students back is not needs-based unless we are prepared to do something more for them. If we are simply preventing students from getting a high school diploma at the end of their schooling, what is the point? Any testing that is not aimed at identifying what a student needs to learn is, by definition, not needs-based. However, the issue is complicated by a shroud of simplistic externally-originating assumptions about assessing needs. Comparative tests not based in serving needs – those with implications that stand to prevent entry into other educational institutions without educative recourse, like the OSSLT – are unethical, especially when we remember what groups are being denied.

If we claim, as the Education Quality and Accountability Office (2002) does, that we "value the well-being of learners above all other interests"

(para. 1), then what follows is that only needs-based assessment is ethically sound because it is the only assessment type that reflects this stated concern for the well-being of students. When the purpose of assessment is to decide promotion, or when the stakes of a test require students to pass in order to continue on the path of formal education, clearly, the focus has shifted away from the needs of students. Students who failed the test cannot graduate, and are denied entry into further schooling when this is precisely what these students need. There should be no illusions about these consequences: the end result is to exclude those who have, for complex reasons, failed to meet the requirements of one test. And stunningly, nearly one-third of secondary school students failed the OSSLT in 2000 (Canadian Press, 2001). By giving tests of exclusivity, we are ignoring most those who need help.

The alternative: the possibilities of qualitative assessment.

Qualitative assessment is an approach that aims to assess student learning across multiple modes of accomplishment, that is, through interviews, journaling, anecdotal records or portfolios, to name a few. The links between qualitative research and qualitative assessment practices are strong. A holistic approach to assessment chooses a method of assessing that is best fit for the specified purpose in the specified context. The chosen method of assessing must be intrinsic to classroom experience rather than external to it. Carefully-developed diagnostic tools and classroom-integrated methods are essential. We need assessments that participate intimately in the classroom, as Serafini (2002) writes:

> Teachers, and in many instances students, collect evidence of student's learning and use this information to guide curricular decisions. In this way, the curriculum is responsive to the assessment process, and the assessments we choose are responsive to the experiences we provide children. (p. 71)

It is an impossible aim for mass testing to be responsive to students' classroom experiences, especially when conducted after-the-fact.

Non-comparative methods of assessment aim to assess students in an unbiased and just way. They involve the student in the process; they can be implemented by taking the standpoint of those who are most disadvantaged; and they are able to drive and participate in what is being taught. Additionally, non-comparative assessments can avoid potential problems of student embarrassment and test anxiety, thereby making the learning experience pleasurable and more effective.

Impediments to a qualitative approach

The impediments to abolishing standardized tests are, however, many. They may include teachers who foresee a greater workload; they may include a public that distrusts teachers; but they clearly include politicians who sup-

port standardized tests as a simple solution for the perceived failings of the educational system. And there may be those "on the top" who prefer to uphold the status quo for obvious reasons. What cannot be overlooked in these viewpoints, however, is that there is no logical connection between standardized testing and promoting genuine educational achievement. Rather, the evidence is to the contrary.

Kachur (2000) suggests the current problems are somewhat out of teachers' hands because there has been a "failure of a critical intellectual community to provide a viable alternative to the neoliberal and neoconservative agenda" (p. 65). Although Kachur may be right in pointing to a failure in our system, we do not need more alternative approaches. We need a new mainstream approach.

The role of education is to create a better society by using socially just and equitable approaches. The often brutal inequities of the whole should never be mimicked in the classroom. The classroom ought to be the microcosm of the social ideal: a safe, genuine and equitable place. When the statistics of our societal successes and failures are used as indicators for where we need improvement, our vision of education can stand up to meet this failure – as a reality check against a dynamic and changing social order to which we need to respond with deliberation. Hutchinson (1999) states the purpose of education as follows:

> Education, in its richest sense is based in the assumption that there is something about the human spirit that can rise above its social enculturation. If this were not the case, we would simply never escape forms of domination and discrimination because we would simply accept our society's message regarding the correct form of social relations. (p. 90)

We know that standardized testing and social conditions are inequitable and biased. But we have the methods and awareness to assess justly. Since the tension between policymakers and practitioners remains a part of our educative landscape, the question remains: What more can we do? Kohn (2000a) suggests that community and teachers overtly resist standardized testing, but practically, there is no reasonable way for teachers to resist implementing what has become a necessary part of keeping their position. And teacher resistance to such testing is politically advertised as teachers not wanting to be "accountable." This is an easily-sold claim when politicians have a voice that teachers lack. Moreover, the value of teacher perceptions of assessment is being undermined by policymakers through their locating the shortcomings on EQAO tests onto teachers; they are doing this by concurrently introducing teacher-testing in the province. Regrettably, the increasing pressure on schools and teachers that are not performing well according to this political movement may continue into the foreseeable future. So, most importantly, we must remain aware and show that it is the case that political policy has entered too far into the classroom.

Shaker (2000) argues that the increasing stakes proposed by teacher testing is a further symptom of the same problematic issues in the system: that policy is overstepping its bounds as "the hunger for simple answers has overwhelmed good judgement" (p. 3). Sadly, the research of Anders and Richardson (1994) shows that some teachers have internalized the notion that "tests developed elsewhere were objective, and could therefore be trusted" (p. 399). In this regard, teachers, who are politically encouraged to feel lack of confidence about their own judgments, may have already internalized their own professional subordination. Murphy (1997) suggests that forcing teachers to implement standardized testing places teachers "in the unenviable position of explicitly denying their own judgement" (para. 30). Yet, the need to utilize qualitative non-comparative assessments and maintain a commitment to equitable practices has never been stronger.

The government that now forces accountability into the classroom has ironically forgotten that they are the same people who approve teacher education programs. And as Issler (2001) points out, "the process-product orientation to teaching presumes too much of a causal relation between teaching and learning and, thus, requires a greater degree of accountability than is realistic and necessary" (p. 342). It is a case of undermining teachers' professionalism when we have politicians mandate their testing into the classroom. Public support of the policy keeps it in place for now, but the reach of the public opinion in this debate should be questioned, for if we believed completely in the views of the majority, we would not need judges in courts of law. Citizens could simply decide cases by clicking buttons online. Through this example, the underlying condition is obvious: we trust that judges are the best fit to make legal determinations. Teachers are considered less professional than judges because the public and the policymakers do not trust teachers' determinations about students. We need to develop teacher education and assessment processes that recognize that a healthy educational system is not as simple as effectively transmitting information to students whereby teachers, administrators, and school districts are directly or indirectly evaluated according to student test scores. Pearson et al. (2001) describe the current accountability trends as destructive because increasingly "reputations, jobs, salaries, and resources are at stake" (p. 177). And the greater the consequences of test score results, the more "that pressure compels educators to use any means necessary to raise test scores" (Pearson et al, 2001, p. 177).

Conclusion

The implementation of holistic assessment requires both political and practical components. Reed (1992) proposes multicultural and invitational approaches to education that will need to "shape significant changes in both structure and delivery systems of schools" (p. 73). While we must continue

to improve our systems, we must now utilize equitable approaches in our classrooms. Reed states: "this challenge [the accountability of words and deeds by those in an educational setting] will . . . benefit children who are culturally, economically, linguistically, and racially different or otherwise disinvited" (p. 73). One cannot become a social agent of change through merely saying politically correct catch-phrases of social equality. It will take the deed, to assess inclusively, qualitatively, and holistically – deliberately. We cannot assume that teachers will automatically do this, just as we cannot assume that policymakers will automatically see a need to do it – but we must raise social awareness of the issues. We must collectively strive to bridge the gap that is widening between teachers and policymakers as the push for comparative data impinges on the professionalism of teachers and on the environment of the classroom.

High stakes comparative testing is not benign. It is inconsistent with the pursuit of equitable and authentic teaching. Teachers have a primary duty to provide a just and equitable classroom, and they have an obligation to know and state the matter when policy is infringing beyond its healthy bounds. The school community at large needs to become aware of the control agenda of the positivist, autocratic, and politically driven ideology that continues to mandate standardized tests as a punitive control gate at the exit door of public schooling. We must strive towards a reconciliation of the split between teaching, teacher education, and policy-making. Ultimately, if teachers are destined to flourish as reflective, knowledgeable professionals, they will need support from government in the removal of mass-testing schemes. Teachers must decidedly become those best fit to teach and assess students in an equitable way. Let us continue to develop the competencies of teachers as assessors and let us hope we manage to get mass-testing out of the classroom. As for the mass-testing movement itself, Eric Hoffer's (1951) description of the function of mass movements is fitting: "the technique of a mass movement aims to infect people with a malady and then offer the movement as a cure" (p. 42).

REFERENCES

Allingham, P.V. (2000). The Ontario secondary school literacy test: Mr Harris's high-stakes version of BC's foundation skills assessment. *English Quarterly*, 32(3/4), 68-69.

Anders, P., & Richardson, V. (1994). Teachers as game-show host, bookkeeper, or judge? Challenges, contradictions and consequences of accountability. *Teachers College Journal*, 94, 382-395.

Berthelot, J., Ross, N., & Tremblay, S. (2001). Factors affecting grade 3 student performance in Ontario: A multilevel analysis. *Education Quarterly Review*, 7(4), 25-36.

Berlak, H. (1992). The need for a new science of assessment. In H. Berlak, F., M. Newmann, E., Adams, D., A, Archbald, T., Burgess, R. Raven, & T.A. Romberg (Eds.). *Toward the development of a new science of educational testing and assessment* (pp. 1-21). Albany: State University of New York Press.

Bracey, G.W. (2001). The condition of public education. *Phi Delta Kappan*, 83(2), 157-170.

Canadian Press (March 8, 2001). Interest in private tutors on the rise in wake of Ontario literacy shock. *Canadian Press Newswire*. Retrieved from CBCA Fulltext Education Index

Chase, B. (2001). Challenging the almighty test. *National Education Association Today*, 20(2), 5.

Clark, A. M. (1998). The qualitative-quantitative debate: moving from positivism and confrontation to post-positivism and reconciliation. *Journal of Advanced Nursing, 27*, 1242-1250.

Connell, R., Johnston, K., & White, V. (1992). Measuring up: Assessment, evaluation and educational disadvantage. Belconnen, ACT, Australia: Australian Curriculum Studies Association

Creswell, J. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, New Jersey: Merrill Prentice Hall.

Education Quality and Accountability Office (2002). *EQAO Mandate*. Retrieved February 20, 2002, from http://www.eqao.com/eqao/home_page/01e/1_2e.html

Hansen, J. B. (1993). Is educational reform through mandated accountability an oxymoron? *Measurement & Evaluation in Counseling & Development*, 26(1), 11-22.

Harris, R., Mercier, M. (2000). A test for geographers: the geography of educational achievement in Toronto and Hamilton. *Canadian Geographer*, *44*(3), 210-227.

Hoffer, E. (1951). *The true believer*. New York: Perennial Library.

Hutchinson, J. (1999). *Students on the margins - Education, stories, dignity*. Albany: State University of New York Press.

Hoover, R. (2000). *Forces and factors affecting Ohio proficiency test performance: A study of 593 Ohio school districts*. Retrieved February 10, 2002, from http://www.cc.ysu.edu/~rlhoover/OPTISM/3_primaryfindings.html

Issler, K. (2001). A conception of excellence in teaching. *Education*, *103*(4), 338-344.

Kachur, J. (2000). Northern dreams: Schooling and society in Canada. In K., Mazurak, M. Winzer, & C. Majorek (Eds.) *Education in a global society: A comparative perspective* (pp. 53-66). Needham Heights, MA: Allyn & Bacon

Kohn, A. (2000a). *A case against standardized testing: Raising the scores, ruining the schools*. Westport, CT: Heinemann.

Kohn, A. (2000b). High stakes testing as educational ethnic cleansing [Electronic Version]. *Education Digest*, 66(4), 13-19

Lewis, A. C. (2001). Heads in the sand. *Phi Delta Kappan*, 83(1), 3-5.

Linn, R. L. (2001). A century of standardized testing: Controversies and pendulum swings. *Educational Assessment 7*(1), 29-39.

Merrow, J. (2001).Undermining standards. *Phi Delta Kappan*, *82*(9), 652-660.

Ministry of Education (June 22, 2001). *Ontario expands province-wide testing*. Ontario Ministry of Education News Release. Retrieved December 20, 2002 from http://www.edu.gov.on.ca/eng/document/nr/01.06/nr0622.html

Murphy, S. (1997). Literacy assessment and the politics of identities [Electronic Version]. *Reading & Writing Quarterly*, *13*(3), 261-278.

O'Sullivan, E. (1999). *Transformative learning: Educational vision for the 21st century*. Toronto: University of Toronto Press.

Ontario Teachers' Federation (2002). *Realizing the Promise of Public Education*. Retrieved December 20, 2002 from http://www.talkschools.on.ca/provincial/testing/15.htm

Pearson, P. D., Vyas, S., Sensale, L. M., Kim,Y. (2001). Making Our Way through the Assessment and Accountability Maze. *Clearing House*, *74*(4), 175-183.

Reed, C. (1992). Invitational and multicultural perspectives: What do they have in common? In J. Novak (Ed.), *Advancing Invitational Thinking* (pp. 47-75). San Francisco, CA: Caddo Gap Press.

Resnick, D. P. (1980). Minimum competency testing historically considered. In D. C. Berliner (Ed.), *Review of research in education* (Vol. 8, pp. 3-29). Washington, DC: American Educational Research Association.

Serafini, F.W. (2002). Dismantling the factory model of assessment. *Reading & Writing Quarterly*, *18*(1), 67-86.

Shaker, P. (2000). *Teacher testing: A symptom.* (ERIC Document Reproduction Service No. ED450071)

Tindal, G., Fuchs, L. S., Fuchs, D., Shinn M. R., Deno S. L, Germann, G. (1985). Empirical validation of criterion-referenced tests. *Journal of Educational Research* 4(3), 203-209.

Verkuyten, M. (2000). School marks and teachers' accountability to colleagues. *Discourse Studies*, *2*(4), 452-473.

FRANK NEZAVDAL is a Master of Education (Integrated Studies) student at Brock University, St. Catharines, Ontario. He can be reached at fnezavda@ed.brocku.ca.

FRANK NEZAVDAL est étudiant à la maîtrise (éducation intégrée) à l'Université Brock à St. Catharines, Ontario. On peut le joindre à l'adresse électronique suivante : fnezavda@ed.brocku.ca.