

The Conundrum of Large Scale Standardized Testing: Making Sure Every Student Counts

Mary Jane Harkins

Mount Saint Vincent University

& Sonya Singer

Chignecto—Central Regional School Board

Introduction

I remember well the day the examinations arrived from the Department of Education. I was teaching a class when our secretary piped into my room to tell me that ‘my’ exams had arrived. At lunch time, I wandered down to the main office to pick up my delivery. There, taking up an inordinate amount of space behind our secretary’s desk, were three, fifty-pound boxes filled with English 12 and English Communications 12 provincial exams. Lacking superhuman strength, I knew that I would be unable to carry these monstrosities of boxes, so I commandeered a pilfered shopping cart from the janitors’ supply closet and wheeled the weighty new arrivals down the hall to my classroom.

The following morning, I wheeled the shopping cart to the school’s gymnasium where the examinations were to be administered. Along with the four hundred and eighty examination booklets and a hundred and sixty individualized marking sheets, the shopping cart also held fifty-five scavenged dictionaries which, I had discovered the day before, students were expected to use for all sections of the examinations. The students filed in at 8:00 a.m., and we were underway.

The inveterate dispute between proponents and foes of large scale, standardized testing has been churning since the first homogeneous tests were developed in Europe in the late 1800s. Two hundred years later, the debate still rages in educational circles. This article examines some of the current thinking surrounding the use and implications of standardized assessments, particularly as they pertain to students with

exceptionalities. The article interweaves one teacher's recollections of her first experience with provincial examinations and leads to a broader discussion of critical issues embedded in the construction, administration and scoring of large scale assessments. We hope to generate discussion and raise important questions that contribute to the awareness that every student counts.

There are few issues in education that generate such fractious debate as do matters of assessment; in particular, large scale, standardized assessments. The issues surrounding assessment, particularly standardized tests and the use of such tests, have created a contentious dichotomy. On the one hand there are those who extol the merits of standardized assessments as a cost-efficient, reliable and valid means of determining how well students have succeeded in meeting intended learning outcomes in a particular subject area. Conversely, there are those who argue that the skills, abilities, knowledge and intelligence of individual students cannot be adequately reflected and/or quantified by a one-shot, single test score.

It was our desire to more fully understand the purposes and praxis of standardized testing that led us to begin a critical inquiry into the adoption and implementation of external assessments in one local high school. Although we come from diverse backgrounds, one from English language arts and one from inclusive education, we shared a common interest in how educators can best meet the needs of all students. The critical issue for us was how educational systems can support, assess, and respond to the learning needs of *all* students in meaningful ways.

A Common Vernacular

We began our inquiry by exploring the "language" of standardized testing as defined by practitioners in the field of educational assessment. Standardization of testing refers to the "structuring [of] test materials, administration procedures, scoring methods, and procedures for interpreting results... this helps to ensure accuracy and consistency in measuring progress, determining levels of performance, and comparing performance to others" (Venn, 2004, p. 109). Standardized tests generally include at least some multiple-choice and true-false questions. These can be graded by computer, or by individuals who do not need to understand the material in depth, as long as they have a list of the correct answers. Standardized tests often include written portions as well; these are graded by persons who use rubrics or guidelines to determine the suitability and quality of the written components of the assessment.

Most frequently, tests are administered to people of similar ages

or grades at particular times in development. This practice allows researchers to study correct to incorrect response ratios over time or from region to region. The variety of test-takers in a particular group generates data that is used to establish normative classification for a group. That data can serve as a guide for what that test finds as a normative response for a test item. For that reason, many standardized tests are also called norm referenced assessments which “involves interpreting the performance of individuals and groups in relation to the performance of others” (Venn, 2004, p. 107).

While many standardized assessments are norm-referenced, there are those, like the Nova Scotia provincial examinations, that utilize criterion-referencing. Criterion-referenced tests (CRTs) are intended to measure how well a person performs “in relation to some functional level or criterion” (Venn, 2004, p. 107). In other words, they are used to measure student mastery of instructional objectives or curriculum outcomes, rather than to compare one student with another or to rank students. CRTs are often used as a “benchmark” to identify areas of strength or weakness in a given curriculum, and/or student readiness to move on to a different level of instruction. Typically, raw scores are used to reflect the number of correct responses, the number of completed outcomes, and so forth. Such tests will often use percentages to reflect the level of mastery of a given instructional objective. Raw scores are converted to a percent correct; in other words, students’ scores are compared to a standard criteria or expected outcome (Stiggins, 2005). In this way, validity is understood to be “the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose” (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME], 1999, p.11).

Variations on a Theme

To support the implementation of a new outcome-based curriculum in English Language Arts, Nova Scotia entered the arena of standardized testing in January of 2000. The purpose of these examinations was to assess student achievement in relation to specific curriculum outcomes. In the spring of 1999, grade 12 English Language Arts teachers throughout Nova Scotia were in-serviced on how to administer and score the forthcoming external examinations. As well, information booklets were distributed to all grade 12 students and their parents explaining the rationale and format of these new Atlantic Province Education Foundation external examinations. Additionally, these brochures emphasized

that the writing of these provincial assessments was mandatory for all eligible grade 12 English Language Arts students, including schools where exemptions had formerly been made available to students as rewards for exceptional scholastic achievement and/or excellent attendance.

All eligible grade 12 students enrolled in either English 12 or English Communications [ECM] 12 courses wrote one of two common external assessments known as the Atlantic Provinces Education Foundation [APEF] external examinations. English 12 is considered to be the more academically challenging of the two courses. Both examinations were intended to reflect specific curriculum outcomes prescribed in the Atlantic Canada curriculum documents for senior high English Language Arts (Nova Scotia Department of Education, 1997), and were to account for 30% of the students' overall grade for the course.

Thirteen learning outcome statements, culled from two strands of the English 10-12 curriculum document (i.e., Reading and Viewing, and Writing and Other Ways of Representing), were used as the foundation for the development of the English 12 and English Communications 12 examinations (Nova Scotia Department of Education, 1997, p.18). Although teachers are expected to help students meet outcomes from three primary strands included in the General Curriculum, outcomes that cannot be readily assessed in a structured group forum, such as those contained within the Speaking and Listening strand, were not included in this external assessment. We view this as problematic because it does not acknowledge other ways of knowing, apart from the use of traditional "paper and pencil" assessments.

A second concern is that the learning outcomes for English 12 and English Communications 12 are practically identical, but the scope, emphases, levels of complexity, required degrees of support and instructional adaptations, as well as learning resources for each stream are significantly different. However, there appeared to be few textual or design differences between the English 12 and English Communications 12 examinations when they were first introduced in 2000 (Nova Scotia Department of Education, 2000-2004). This streaming in itself is an interesting practice that produces different types of expectations for students, within the stated provincial outcomes.

As for the items on the test, test constructors write from their own culture and as Campbell (2004, cited in Wink, 2005) stated, "It is difficult for the fish to see the stream" (p. 39). This begs the question, "Who are the test writers and what cultures do they represent?" as these tests typically measure middle or upper class experiences (Christensen, 2000). Connell (1999/1994) has written about this "informal segregation within formally unsegregated institutions" (p. 148). Finn (1999) insightfully

noted that “Savage inequalities persist because a lot of well-meaning people are doing the best they can, but they simply do not understand the mechanisms that stack the cards against so many children” (p. 94).

At 10:00 a.m., the students were permitted to leave the examination area for a one hour break. Most students assembled in the school cafeteria to grab a bite of lunch and to discuss the morning section of the exams. Several expressed concern about the ambiguity of some of the constructed-response questions; others talked about the seeming repetitiveness of some of the responses. One young woman wondered why a scant ten line poem had necessitated the answering of thirteen questions. I had had similar concerns when I glanced through both examinations the previous evening.

In talking with some of my English Communications 12 students, I discovered that many of these learners had experienced significant difficulties in understanding what was being conveyed in the literary prose and informational texts. These students found the vocabulary difficult to understand and the length of these pieces of text to be prohibitive. There was also a discernible level of apprehension about the two writing assignments that would comprise the afternoon section of the exam. There was little I could do to allay their fears about the writing portion of the exam. I knew that these students, most of whom were reluctant readers and writers, would probably experience enormous anxiety when confronted by these two demand-writing components of the English Communications 12 examination. I could only encourage them to try to the best of their abilities, knowing fully well that this would be an exercise in abject frustration for most of them.

Despite the fact that teachers of grade 12 English in all four Atlantic provinces had received in-servicing on how to administer and mark these examinations, there was a general sense of stunned disbelief at the length and complexity of these new external assessments. A three and a half hour time frame; three examination booklets per student; fifty-six selected and constructed-response questions; a literary prose essay and a piece of transactional writing; and a value of thirty percent of the year’s final mark in grade 12 English. This created a backlash of discontent from teachers, students, and parents about the perceived lack of fairness in the format and content of the ECM 12 exam.

This situation demonstrates a shift from teacher-designed assessments to external, standardized assessments that alters the power relationships of teachers. This shift has implications for teachers’ classroom teaching and for students, in the larger socio-economical arena. Assessments are no longer valuing the uniqueness and individual strengths of all students. Historically, it has been found that “children from working-class, poor, and minority ethnic families continued to do

worse than children from rich and middle-class families on tests and examinations, were more likely to be held back in grade, to drop out of school earlier, and were much less likely to enter college or university” (Curtis, Livingstone & Smaller, 1992; and Davis, 1948, as cited in Connell, 1999/1994, p. 148).

The focus of standardized assessments now appears to be on functional knowledge and skills believed to be of value in society. Kozol (1985, as cited in Lankshear, 1993) referred to this as a reductionist approach to learning as it is “reducing the individual to the status of mere objects and means, rather than confirming and exalting them as ends in themselves” (p. 91). We agree that the reliance on this type of student assessment has the potential to perpetuate social inequities in schools.

Most of my English 12 students were able to complete the two writing tasks in approximately seventy minutes without any apparent difficulties. On the other hand, all of my English Communications students were finished well before the first hour had elapsed. I knew that that was not a good sign. My worst fears were confirmed that evening when I began marking the English Communications 12 examinations. The results for the constructed and selected-response questions in Booklet 2 were better than I had expected, but the marks for Booklet 3, which contained the two writing tasks, were dismal. Of the twelve students who wrote the English Communications 12 exam that day, only three students attempted both writing tasks. Seven students attempted the transactional writing activity only, and the other two students left Booklet 3 blank; they did not attempt either task. By doing so, they had forfeited 50% of their final examination mark in English Communications 12.

I ran into these two students in the school cafeteria a couple of days later. When I asked them why they had not attempted either writing task in Booklet 3, one of the young men said something that still haunts me to this day. He said, “I was tired; I was frustrated ... this exam just proved to me how stupid I really am. Maybe that was their point, Miss, to show ECM kids that they don’t really belong in a grade 12 classroom; that they don’t deserve to graduate like the smart kids.”

Christensen (2000) wrote of how her students blamed themselves for their low test scores on standardized tests: “As far as they [her students] were concerned, there was no need to go to college because this test confirmed their stupidity” (p. 60). Christensen stated that many of her African-American males who were misplaced in special education classes had “amazing dexterity with verbal language and astute social/political insights, but their literacy skills were underdeveloped” (p. 172). These strengths are not going to be valued in large scale standardized tests, yet the results of test scores can play an important role in the selection process for post-secondary education and employment in the workplace.

Corbett (1999) asserted that “Everything I know as an educator and learner militates against the notion of standardization, and a uniform set of outcomes. Learning is not about assimilating standard material, it is about engagement with ideas and materials which are useful and important, here and now” (p. 171). Using this approach to learning, it appears that “the ‘core’ values almost always come out ‘right’ and marginal children almost always come out ‘wrong’” (Corbett, 1999, p. 174). As Dewey (1938) informed us:

Every theory which assumes that importance can be attached to these objective factors only at the expense of imposing external control and of limiting the freedom of individuals rests finally upon the notion that experience is truly experience only when objective conditions are subordinated to what goes on within the individuals having the experience. (p. 41)

Revising the Conceptual Framework

In August of 2004, the Nova Scotia Department of Education responded to public and professional concerns. Evaluation Services initiated a major overhaul of both the English 12 and English Communications 12 external examinations (Nova Scotia Department of Education: Evaluation Services, 2004). In reviewing the changes to both examinations, one can see clearly that much more consideration was given to altering the presentation and response format of the “non-academic” exam. Assiduous attention was paid to the English Communications 12 exam during these review sessions.

There were a number of significant amendments to the January 2005 version of the English Communications 12 provincial examination. For the first time, the two writing tasks, the completion of text-specific selected and constructed-response questions, as well as the reading of the selected texts, were divided between the morning and afternoon sessions of the examination, in four separate booklets. In the morning, students were required to prepare a 250 word literary prose essay and complete the short-answer questions that were based on the selected pieces of text found in the examination booklet.

The readings included one piece of literary prose and two pieces of informational text. Students were given a maximum of two and a half hours to complete this first section of the provincial assessment. It should be noted here that while students were required to read more pieces of written text on this exam, the text length was significantly shorter than in previous years. This would help to ensure that English Communications 12 students would have adequate time to complete all

reading tasks. As well, texts were strategically selected so as to maximize student engagement with the readings; that is to say, particular attention was given to choosing texts that reflected the interests and life experiences of typical grade 12 students.

The prompts for the literary prose essay were simplified in an attempt to aid students in formulating coherent thesis statements. Students were asked to choose one of three questions or statements that pertained to the piece of literary prose text included in the examination booklet. Additionally, these students were required to complete a skeleton outline of their ideas before actually writing the essays. To complement this change to the composition section, the provincial examination advisory committee produced a new marking guideline for the English Communications 12 literary prose essay. Teachers now evaluate this piece of demand-writing using the following criteria for assessment: purpose, supporting details, organization and matters of correctness. Previously, student essays were assessed on thought and detail, organization, matters of correctness and matters of choice; the same four criteria by which English 12 students' literary prose essays were (and continue to be) assessed.

The afternoon section of the revised English Communications 12 examination involves the distribution of two additional booklets to students. The text booklet contains a poem and two pieces of media text. Again, the types of poetic and visual texts to which these students will be exposed have been expanded to include free verse poems, historical or contemporary songs and/or song lyrics, and advertisements, editorial cartoons and/or promotional brochures. The student workbook contains a section that includes instructions for the 250 word transactional writing activity, space for writing the draft letter, accompanying questions and spaces for students' responses to these questions. Students are given a maximum of two hours to complete this section of the Nova Scotia English Communications 12 examination. As with the literary prose essay, a new scoring rubric has been developed to assist teachers in assessing the transactional writing activity. The changes appeared to more accurately reflect the intended learning outcomes for this program and new assessment rubrics were developed. However, the underlying questions remain: how do you judge the standard of the work produced by students with such diverse interests, needs and abilities with a traditional "paper and pencil" test? What is a fair assessment?

Special Provisions

Though the terminology varies from province to province, such as adaptations, modifications and/or accommodations, there is general

agreement that special provisions may be needed to support the various needs of individual students during the writing of standardized assessments. These provisions generally fall into four discrete categories: presentation format, response format, setting of the test and timing of the test (Thurlow, House, Scott, & Ysseldyke, 2000). For the purpose of this paper, we will use the term “accommodations” when referring to any or all categories of provided support. Accommodations are defined as “supports and services that enable students with special needs to demonstrate their competencies in the skills being measured by the test” (Ontario Department of Education, Education Quality and Accountability Office, 2004, p.1).

Within these categories, there are several types of accommodations that are in common usage throughout Canada. These include: (a) providing alternate settings to minimize external environmental stimuli; (b) permitting time extensions and/or scheduling flexibility to allow for short but frequent breaks in order to maintain on-task behaviors; (c) alternate presentation formats that acknowledge that some students may have difficulty with interpretation and/or retention of textual material and directions; and (d) response modifications, as many students tend to have substantially stronger expressive verbal abilities than written alacrity. More specifically, these provisions might include the full text version of the test on audiotapes, in Braille, or large print; use of sign language or oral interpretation; verbatim readers and/or scribes; and the use of word processors as well as assistive technology. Glaring inconsistencies are evident between and among provinces, with respect to the types of permitted interventions and details provided, such as the type and amount of support provided by scribes, readers and/or interactive technology (see reference list for provincial and territorial Department of Education websites).

If the intended purpose of using accommodations during standardized tests, is “to provide accurate and comparable measurement for everyone, and unfair advantage for no one” (AERA et al., 1999, p.61), then several interrelated issues become problematic. In particular, educators need to be mindful of the following:

What kinds of accommodations can be used during assessment?

What do testing accommodations have to do with classroom instruction?

Who should be involved in making decisions about assessment accommodations?

What criteria should guide decisions about the use of testing

accommodations in district and state [provincial and territorial] assessments? (Thurlow, Elliot, & Ysseldyke, 2003, p. 29)

Those who oppose the use of accommodations during large scale assessments argue that these accommodations give an unfair advantage to the students using them. This, they believe, affects the validity of the test results. Goh (2004) maintains, however, that “the purpose of using accommodations is to increase the validity of test results, so that the results may accurately reflect the construct (e.g., aptitude or achievement) measured by the test, rather than the disabilities or language barriers of the students” (p.15). Thurlow, et al., 2003, asserted that, “a system is accountable for *all* students when it makes sure that all students *count* in the *evaluation program* of the education system” (p.3).

These divergent viewpoints underscore the complexity of selecting and implementing appropriate accommodations during standardized testing. Furthermore, recent studies indicate that additional empirical research is necessary to determine the effects of accommodations on the reliability and validity of test scores (AERA, et al, 1999; Goh, 2004; Thurlow, Elliott, & Ysseldyke, 2003). Westwood (2005/2003) has asked the question: “Is it fair to judge the standard of work produced by a student with a mild intellectual disability, or a student that is deaf, against the standard applying to students of average or good ability in the class?” (p. 155).

Another pivotal question arises whenever one considers the use of accommodations during standardized testing: who has the authority and the expertise to decide if a student participates and under what conditions? To be considered for accommodations, in most provinces, students must have written documentation that identifies their need for intervention and a rationale that supports intended accommodations. Provincial policies and/or procedures in Canada cover the gamut from school-based, generic provisions, based on instructional accommodations to detailed procedures requiring protracted documentation substantiated by medical and/or psycho-educational reports. Within the scope of curriculum theory, there is an inextricable relationship between and among individual students, instructional practice and assessment; furthermore, we contend this is most salient at the classroom level.

In general, classroom teachers have become adept at supporting diverse learners and providing opportunities for them to demonstrate their knowledge and skills in various mediums that allow for multiple ways of knowing. However, from our experiences working with secondary school teachers in Nova Scotia, it would seem that many educators lack confidence and proficiency in understanding the rudimentary relation-

ship that exists between instructional accommodations and the use of such accommodations during standardized testing. If this simple but essential juxtaposition of selecting and aligning appropriate accommodations for particular students to particular test constructs is missing, then the validity of the whole process becomes compromised. Validity is, therefore, a shared and necessary responsibility between the test developer and those administering the test.

Interestingly enough, it appears that there were no provisions for alternate types of assessments or activities in place for those students who were unable to participate in provincially mandated standardized testing. There was, and continues to be, a general provision that allows for exemptions for those students who have Individualized Program Plans, and accommodations for those children who have documented instructional adaptations (Nova Scotia Department of Education, 2005b). This finding raises troubling questions about the likelihood of marginalization of some students and total disregard of others, within a learning environment that purportedly espouses an inclusive philosophy of education. In Canada, inclusive education is guaranteed under the *Charter of Rights and Freedoms* in section 15(1): “Every individual is equal before and under the law and has a right to the equal protection and equal benefit of the law without discrimination based on race, national or ethnic origin, colour, religion, sex, age, or mental or physical disability” (Government of Canada, 1982). We would suggest that a system is accountable only when it can *demonstrate* that *all* students *count* in the assessment procedures used in *all* its testing protocols.

Future Directions

This “modern” version of standardized assessment is a relatively new phenomenon in Nova Scotia’s high schools. In the years since its inception, there have been occasional missteps, intermittent success stories and underlying discontent. In other words, Nova Scotia’s experience with standardized testing has been fairly consistent with the growing pains experienced by other jurisdictions within Canada that have adopted large-scale, external examinations. Although education is under provincial and territorial jurisdiction in Canada, it would be beneficial for all members of educational directorates to try to agree on a solid foundation on which to build a common vision of best practices as they pertain to student assessment.

It is our hope that as educators, we will learn from our experiences and find ways to increase the degree and regularity of our students’ success stories as we continue to engage in meaningful discourse about the

use and implications of standardized assessments. Standardized assessments need to be re-evaluated if this “500-pound gorilla” (Kohn, 2002 as cited in Wink, 2005, p. 39) that has “transformed the joy of learning into the terror of testing for students, teachers, administrators and families” (Wink, 2005, p. 39) is to be removed from the backs of teachers. Teachers’ stories can offer critical perspectives on the impact of standardized assessments on the daily realities of teaching and student learning for as Neilsen (1999) asserted, “The time for counternarratives of teachers’ work has arrived” (p. 14). Central to the process of needed school reform will be our willingness to acknowledge and value the diverse ways in which students learn and represent their knowledge and skills.

What does one say to a hulking, seventeen year old man-child standing there before you with tears welling up in his eyes? At the time, I could not think of a single thing to say to him. I simply wrapped my arms around him and cursed silently at the examination development committee that had destroyed the last vestiges of this young man’s tenuous self-esteem. Unfortunately, his was not the only heart wrenching story I heard over the next few days. Many of my students sought me out to talk about the examinations. Several spoke about the physical and mental fatigue they experienced during the writing of the exams. A number of them expressed concerns about the lack of clarity in task instructions and the ambiguity of some of the possible answers to the constructed-response questions. Many of my students were right to be concerned: on average, the final exam marks that year were ten to fifteen points lower than their term marks in English.

“If we consider that all of us are diminished by a school system and an educational process that does not ‘honour the history, culture, social realities, abilities, and diversity of each of us’ (Lewis, 1993, p. 194), then we must begin to envision how our differences can function as a rallying point for change, rather than as the ‘normalizing categories’ (Butler, 1991, p. 308) that fling and confine students to the margins” (Singer, 1997, p. 127). In this rapidly spinning cycle of educational reform, we must never lose sight of the fact that *every student counts*.

References

- American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Alberta Ministry of Education. (2005). *Provincial achievement testing program*. Retrieved May 9, 2005, from http://www.education.gov.ab.ca/k_12/testing/
- British Columbia Ministry of Education. (2004). *Provincial examinations: About*

- exam adjudication*. Retrieved May 9, 2005, from <http://www.bced.gov.bc.ca/exams/adjudication/>
- Butler, J. (1991). Imitation and gender subordination. In H. Abelove, M. A. Barale, & D. Halperin (Eds.), *The lesbian and gay studies reader*. London, UK: Routledge.
- Campbell, D. (2004). *Choosing democracy: A practical guide to multicultural education*. Upper Saddle River, NJ: Prentice Hall.
- Christensen, L. (2000). *Reading, writing, and rising up: Teaching about social justice and the power of the written word*. Milwaukee, WI: Rethinking Schools.
- Connell, R.W. (1999/1994). Poverty and education. In E. Mintz & J. T. Yun (Eds.), *The complex world of teaching: Perspectives from theory and practice* (pp. 145-167). Cambridge, MA: Harvard. (Reprinted from *Harvard Educational Review*, 64(2), 125-149)
- Corbett, M. (1999). From here to uncertainty: A preoccupation with code. In A. R. Neilsen (Ed.), *Daily meaning: Counternarratives of teachers' work* (pp. 165-182). Mill Bay, BC: Bendall Books.
- Curtis, B., Livingstone, D. W., & Smaller, H. (1992). *Stacking the deck: The streaming of working-class kids in Ontario schools*. Toronto: Our Schools/Our Selves Education Foundation.
- Davis, A. (1948). *Social-class influences upon learning*. Cambridge, MA: Harvard University Press.
- Dewey, J. (1938). *Experience and education*. New York: Collier Books.
- Éducation, Loisir et Sport, Québec. (2005). *Ministerial examination of college English*. Retrieved May 8, 2005, from <http://www.mels.gouv.qc.ca/>
- Finn, P. J. (1999). *Literacy with an attitude: Educating working-class children in their own self-interest*. Albany, NY: State University of New York Press.
- Goh, D. S. (2004). *Assessment accommodations for diverse learners*. Boston, MA: Pearson Education.
- Government of Canada. (1982). *Charter of rights and freedoms, section 15 (1)*. Retrieved May 8, 2005, from <http://laws.justice.gc.ca/en/charter/index.html>
- Kohn, A. (2002). The 500-pound gorilla. *Phil Delta Kappan*, 84(2), 112-119.
- Kozol, J. (1985). *Illiterate America*. New York: Anchor & Doubleday.
- Lankshear, C. (1993). Functional literacy from a Freirean point of view. In P. McLaren, & P. Leonard (Eds.), *Paulo Freire: A critical encounter* (pp. 90-118). New York: Routledge.
- Lewis, M. (1993). *Without a word: Teaching beyond women's silence*. New York: Routledge.
- Manitoba Education, Citizenship and Youth. (n.d.). *Assessment and evaluation: Overview of the provincial assessment program*. Retrieved May 9, 2005, from <http://www.edu.gov.mb.ca/ks4/assess/index.html>
- Neilsen, A. R. (Ed.). (1999). *Daily meaning: Counternarratives of teachers' work*. Mill Bay, BC: Bendall Books.
- New Brunswick Department of Education. (2003). *Provincial examination program: Guidelines for accommodations and exemptions*. Retrieved May 8, 2005, from <http://www/gnb.ca/0000/publications/eval/guidelin.pdf>
- Newfoundland Department of Education. (2003). *High school certification manual*.

- Retrieved May 8, 2005, from <http://www.ed.gov.nl.ca/edu/hsc/hsc.htm>
- Northwest Territories Education, Culture and Employment. (2005). *Education, culture and employment*. Retrieved May 9, 2005, from <http://www.ece.gov.nt.ca/>
- Nova Scotia Department of Education. (1997). *Atlantic Canada English language arts curriculum guide: Grades 10-12*. Retrieved May 6, 2005, from http://www.ednet.ns.ca/pdfdocs/curriculum/ela10_12web.pdf
- Nova Scotia Department of Education. (2005a). *About "PLANS."* Retrieved May 10, 2005, from <http://plans.ednet.ns.ca/about.shtml>
- Nova Scotia Department of Education. (2005b). *Adaptations: Strategies and resources*. Retrieved November 15, 2005, from <http://www.ednet.ns.ca/pdfdocs/studentsvcs/adaptations.pdf>
- Nova Scotia Department of Education: Evaluation Services. (2000-2004). *Nova Scotia examinations: English 12 and English / Communication 12*. Halifax, Nova Scotia: Nova Scotia Department of Education.
- Nova Scotia Department of Education: Evaluation Services. (2004). *Nova Scotia examinations English 12 and English / Communications 12 Teacher Information Guide*. Halifax, Nova Scotia: Nova Scotia Department of Education.
- Nunavut Department of Education. (2005). *Kavamat Elihaktoliginikot Havakviat*. Retrieved May 9, 2005, from <http://www.gov.nu.ca/education/index.htm>
- Ontario Department of Education, Education Quality and Accountability Office. (2004). *Guide for accommodations, special provisions, deferrals and exemptions*. Retrieved May 8, 2005, from <http://www.edu.gov.on.ca/extra/eng/ppm/127.html>
- Prince Edward Island Department of Education. (2005). *Education*. Retrieved May 9, 2005, from <http://www.gov.pe.ca/educ/>
- Saskatchewan Department of Education. (2004). *Special provision policy*. Retrieved May 8, 2005, from <http://www.sasklearning.gov.sk.ca/>
- Singer, S.E. (1997). *Voices from the margins: Lesbian teachers in Nova Scotia's schools*. Unpublished master's thesis, Mount Saint Vincent University, Halifax, Nova Scotia, Canada.
- Stiggins, R. J. (2005). *Student-involved assessment for learning*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Thurlow, M. L., Elliot, J. L., & Ysseldyke, J. E. (2003). *Testing students with disabilities: Practical strategies for complying with district and state requirements*. Thousand Oaks, CA: Corwin Press.
- Thurlow, M. L., House, A. L., Scott, D. L., & Ysseldyke, J. E. (2000). Students with disabilities in large-scale assessments: State participation and accommodation policies. *Journal of Special Education, 34*(3), 154-163.
- Venn, J. J. (2004). *Assessing students with special needs*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Westwood, P. (2005/2003). Adapting curriculum and instruction. In K. Topping, & S. Maloney (Eds.), *The RoutledgeFalmer reader in inclusive education* (pp. 145-159). London, UK: RoutledgeFalmer.
- Wink, J. (2005). *Critical pedagogy: Notes from the real world* (3rd ed.). Boston: Pearson.
- Yukon Department of Education. (2005). *Yukon Education*. Retrieved May 9, 2005, from <http://www.education.gov.yk.ca/>