

## THE EVOLVING CULTURE OF LARGE-SCALE ASSESSMENTS IN CANADIAN EDUCATION

[Don A. Klinger](#), Christopher DeLuca, and Tess Miller

Faculty of Education

Queen's University

Canada

### Abstract

In recent years, an increase in the number of large-scale assessment programs in Canada has been observed. However, due to the provincial/territorial control of education throughout Canada, the format and purposes of these assessment programs vary. The central purpose of this study was to document the format and explicit purposes of the current large-scale assessment programs in each of Canada's ten provinces and three territories. Through document analysis of publicly accessible policy documents, examination of Ministry websites, and telephone interviews with Ministry of Education officials, specific and general characteristics and purposes of large-scale assessment practices in Canada were obtained and analyzed. This analysis provided an opportunity to examine the commonalities and variations in the frameworks guiding the current large-scale assessment culture that is evolving in Canada. Assessment programs were categorized by their explicit purposes as they related to the functions of accountability, gatekeeping, instructional diagnosis, and monitoring student achievement over time.

Key Words: large-scale assessment, Canadian education (Canada), provincial testing, assessment culture

## **The Current Assessment Culture in Canadian Education**

### **Introduction**

Large-scale educational assessment programs are instruments of public policy (Mazzeo, 2001). To this end, large-scale assessments are increasingly used in jurisdictions not only to measure student achievement but also to hold schools accountable for the educational outcomes of students (Crundwell, 2005; Earl, 1999). Levin (1998) noted “an emphasis on standards, accountability and testing has been a feature of reforms in many countries. Almost everywhere we find more large-scale testing of students and more reporting of the results of these tests than was the case a few years ago” (p. 133). Further, this increased use of large-scale assessments has occurred in jurisdictions having very different educational structures. For example, Britain and Chile, countries with highly centralized educational curriculum commonly developed at the national level, have large-scale testing programs designed either to monitor or certify student achievement in relation to national standards (National Assessment Agency, 2008, Organization for Economic Co-operation and Development, 2004; Qualifications and Curriculum Authority, 2008). The examination program in Chile is also used to implement and evaluate educational interventions and policies, and provide salary incentives to teachers having higher student scores. At the other extreme, the United States represents an educational jurisdiction in which curriculum is commonly developed at the district level, commonly in close alignment with state standards. It was within this context that the No Child Left Behind (2002) act was enacted, resulting in a proliferation of diverse state level assessment programs and an increased focus on systematic measurement of student progress and educational accountability.

The education of children in Canada falls under provincial/territorial jurisdiction. Each province and territory is responsible for the development of curriculum and the assessment of student achievement within its jurisdiction. While systematic differences exist between jurisdictions, education in Canada is best characterized by a set of shared values and goals seeking to meet the changing demands and needs of society over time (McEwen, 1995). Further, large-scale assessment has become a central construct in Canadian education and, in particular, appears to play a critical role in shaping and guiding instruction, curriculum, and policy. McEwen (1995) argues the intended purposes of large-scale assessment programs in Canada are “to monitor student achievement on the programs of study and provide stakeholders with information about how well students are learning” (p. 7). However, the notions

of educational accountability change in response to economic and political movements (Adams & Kirst, 1999). Subsequently, the purposes of testing programs often change or expand, although with no substantive changes to the assessment program presently in place (Rogers & Klinger, 2007).

Nagy (2000) identified three different purposes for large-scale assessments: accountability, gatekeeping, and instructional diagnosis. Accountability relates to public concerns regarding the educational system and student competencies, although Adams and Kirst (1999) provide even further differentiation. These assessments are often associated with program and system examinations intent on providing information about specific schools, school-based initiatives, programs, and teachers. Large-scale assessments impact student-, program-, and system-level decisions under the auspices of accountability to stakeholders. They typically provide data supporting program modifications or reflecting the health of the provincial/territorial educational system. The belief is that such assessment programs effect change in policy, curriculum, and practice, especially in a climate of data driven decision making (Madaus & Kellaghan, 1992).

Gatekeeping is used to grant students privileges such as graduation, admission, or grade promotion. Such assessments are commonly referred to as high-stakes assessments, although there is wide variability in the stakes. At the extreme, results from the large-scale assessments become the sole determiner of a student's success.

Large-scale assessments for instructional diagnosis seek to determine what students actually know with respect to a set of criteria, subject expectations, or learning outcomes (Nagy, 2000). Assessment results are provided to teachers and students in a timely manner in order to support and guide instruction and learning.

Recently, there has been an increase in the number and purposes of large-scale educational assessment programs in Canada. However, due to the provincial/territorial control of education throughout Canada, the format and purposes of these assessment programs vary. The central purpose of the present study was to document the format and explicit purposes of the current large-scale assessment programs in each of Canada's ten provinces and three territories. Using document analysis of publicly accessible policy documents, examination of Ministry websites, and telephone interviews with Ministry of Education officials, the specific and general characteristics of the large-scale assessment programs in Canada were obtained. Analyses of the assessment programs provide an opportunity to examine the commonalities and variations in the frameworks guiding the current large-scale assessment culture that is evolving in Canada.

### **A Brief History of Large-scale Assessment in Canada**

To better understand the current assessment culture in Canadian education, it is useful to review the history of large-scale assessment practices in Canada. Formal education began in much of eastern Canada during the 1850's. At the same time, Canadian society began to realize rapid growth in the areas of commerce, government, and technology (Nagy, 2000). As a result of this growth, there was a greater need for educated people. In the early years, emphasis was placed on the development of academic standards and teacher qualification as education was largely directed towards students in the upper social classes. In the mid to late 1800's, the need for secondary level schooling emerged and as a result, so did the need for provincial assessments and, in particular, assessments for gatekeeping. Dr. Egerton Ryerson, Chief Superintendent of Education for Ontario 1846-1876, was instrumental in administering the first entrance examinations in Upper Canada (present-day Ontario) (Putman & Weir, 1925). Ryerson's recommendations for improvement of public instruction in Upper Canada were based on the education practices used in Europe, the British Isles, and the United States (Ryerson, 1868). The first intermediate examinations were held in 1876. The exams had three purposes:

1. Aid both teachers and student regarding the manner in which the work of the school should be directed, providing direction to the teaching, while preventing faulty methods of instruction.
2. Serve as tests for the promotion of pupils, although the data from the assessments would be supplemented by other data for this purpose.
3. Grant certificates that have a qualifying or commercial value, providing a better guarantee of uniformity in standards, rather than relying on teachers or other local authorities to provide such certification. (Millar, 1893, pp. 42-43)

The intermediate inspector, Dr. McLellan, deemed that improvement had followed the adoption of the uniform entrance examinations (McCutcheon, 1941). Subsequently, entrance examinations became more common across the country and especially in the western provinces, with the intention to "ensure a common standard and sense of fairness" (Nagy, 2000, p. 264). Graduation examinations became commonplace and the sole determiner for graduation was based on mandated examination scores. As increased numbers of students from middle and lower classes entered the school system, a structural shift occurred and educational standards became centralized within provinces. Provinces diverged not only in terms of curriculum, but also in the manner in which large-scale assessments were used. The provinces of British Columbia and Ontario illustrate the divergent uses of large-scale assessments in Canada throughout the twentieth century.

*British Columbia.* British Columbia has a long and active history with large-scale assessments and educational accountability. Putman and Weir (1925) surveyed the school system and made recommendations for possible improvements. At the time, an examination system was in use across the province in order to centralize education and unify provincial standards. Their report addressed several concerns and issues in education and focused, in part, on the provincial examination system. Students wrote subject specific examinations in Grades 8, 11, and 12. Putman and Weir identified several faults with this system, including threats to reliability and validity because the tests were highly subjective in both the questioning and marking. The examination program was very expensive to administer and not entirely covered through student fees, resulting in an increasing financial deficit. Teachers' effectiveness was determined based on the number of students passing the examinations rather than on the actual quality of their teaching. Lastly, there was concern the examinations focused too heavily on formal disciplines rather than on content/practical studies. In response to these and other objections, Putman and Weir recommended discarding the Grade 8 examinations and diminishing the influence of the Grade 11 and 12 examinations through the adoption of an "accrediting system" for high schools.

The Putman-Weir Report had an immediate and long lasting effect on the structure and function of schooling in British Columbia. The Department of Education quickly revised its examination program (Johnson, 1964). In 1926, students who failed an examination in one subject were no longer forced to rewrite the entire set of examinations. Rather, students only repeated those exams on which they scored below 50%. In 1931, students became eligible for grade promotion in high school (Grades 9, 10, and 11) based on teacher recommendation rather than by passing final subject specific examinations. By 1937, high school exit examinations were not mandatory for post-secondary admittance and universities considered recommendations by accredited secondary school principals as equivalent. Departmental high school entrance examinations were discontinued in 1939 and public elementary school students could enter their local secondary school based on teacher and inspector approval.

With the exception of the years 1974 to 1983, Grade 12 provincial exit examinations have been administered since the 1920's (Hodgkinson, 1995). These examinations have had two distinct purposes, supporting provincial standards through the inclusion of examination results on the students' academic grades (40%) in specific Grade 12 academic courses, and awarding scholarship money fairly and equitably among high-school graduates.

Provincial testing at other grades has been much more variable. The Provincial Learning Assessment Program (PLAP), established in 1976, provided a broad means of assessing

learning outcomes in the absence of the Provincial examination program, although they were never used to determine students' marks. Annual assessments were given to samples of students in Grades 4, 7, and 10, focusing on specific subjects (most commonly language arts, mathematics and science) each year. The purposes of the PLAP were to inform professionals and the public about the strengths and weaknesses of the education system and assist the ministry, school districts, and schools in making decisions linked to curricula, allocation of resources, while also monitoring student learning over time (British Columbia Ministry of Education and Ministry Responsible for Multiculturalism and Human Rights, 1992, p. xviii). The PLAP remained after the reintroduction of the provincial examination program in 1984 but was replaced in the mid 1990's by the Foundations Skills Assessment (FSA). The FSAs are linked to the provincial curriculum and requiring students to apply skills and concepts to complete complex, realistic tasks (British Columbia Ministry of Education, 2007b). The FSAs were initially given annually to all Grades 4, 7, and 10 students and assessed reading, writing, and mathematics.

Currently, many of the Grade 12 examinations have become optional for students and new provincial examinations have been introduced at the Grade 10 and Grade 11 levels, albeit their contribution to the students' marks is 20% of the final course mark. This was coupled with the removal of the FSA at the Grade 10 level. Further changes to the FSA program have also been made for the 2007-2008 school year, including an online multiple-choice component, an earlier administration date and local scoring of the open-ended responses.

*Ontario.* While Ontario was among one of the first provinces to administer secondary school entrance exams, these were discontinued in the early 1930's (Nagy, 2000). For much of the 20<sup>th</sup> century, Ontario did not emphasize large-scale assessments in its educational system (Royal Commission on Learning, 1994). In response to public concerns, standardized exit examinations (departmental exams) in Grade 13 were administered in all core subject areas in the 1950's and 1960's (Royal Commission on Learning). The results from these exams were initially the sole basis for university admission. However, this requirement changed in the mid-1960's and students were accepted to postsecondary education based on a combination of departmental examination scores and teacher-derived marks.

The Hall-Denis Report (Ontario Department of Education, 1968) appraised the social and educational issues affecting the Ontario school system. The report's central recommendation was to "establish the responsibility of every school authority to provide a child-centred learning continuum inviting learning by individual discovery and inquiry" (p. 179). Standardized assessments of achievement were contrary to the basic tenets of this philosophy

(Raphael, 1993), thus departmental Grade 13 examinations were discontinued in the late 1960's.

Ontario did not administer any provincial assessments in the 1970's and 1980's. Instead, the focus was on teacher assessment of student achievement. The Ontario Schools Intermediate and Senior Division 1984 document supported a teacher directed assessment program:

For the most part, it is recognized that the most effective form of evaluation is the application of the teacher's professional judgment to a wide range of information gathered through observation and assessment. In order to help teachers evaluate student achievement, curriculum guidelines will describe appropriate evaluation techniques.

(Royal Commission on Learning, 1994, p. 134)

In the absence of a provincial assessment program, the Ontario Assessment Instrument Pool (OAIP) was created. OAIP was an assessment bank for teachers to improve the quality, diversity, and use of their classroom assessment instruments. It was not until Ontario's most recent educational reforms that large-scale assessments were reinstated in the province. These reforms, made in response to the Royal Commission on Learning (1994), included the formation of the Education Quality and Accountability Office (EQAO), and the introduction of province wide literacy and numeracy testing at Grades 3 and 6, and mathematics in Grade 9. Since their inception, these assessments have undergone several modifications, including a large reduction and length and the use of centralized marking procedures.

In 2002, the Ontario Secondary School Literacy Test (OSSLT) was first administered to Grade 10 students, with successful completion being a requirement for graduation. However, students who do not successfully complete the OSSLT can now complete the Ontario Secondary School Literacy Course (OSSLC) to satisfy this graduation requirement.

The ongoing shifts in provincial assessment practices are evident even from this brief historical account of assessment changes and practices in British Columbia and Ontario. Canada's large-scale assessment programs have reflected the changing pedagogical, political, and social perspectives of each province/territory. And they continue today. Over the last 20 years, the number and types of provincial large-scale assessments in Canada have increased dramatically. Currently, every province and territory has at least one, and more commonly, multiple provincially administered large-scale assessment programs. Given the variety of such examination programs throughout Canada, it is important to understand the various structures, purposes, and uses of these programs.

## **Theoretical Framework**

In order to examine the provincial/territorial assessment programs, a categorical framework for examining large-scale assessments was established. Based on previous research (McEwan, 1995; Nagy, 2000; Taylor & Tubianosa, 2001), the frameworks of accountability, gatekeeping, instructional diagnosis, and monitoring student achievement over time were used to classify the purposes and uses of each assessment program. Nagy's definition for accountability was used in this framework, including monitoring student, school, district, or province level performance with respect to an established provincial standard. A relationship between accountability and reporting exists where the strategy of reporting either confirms or contests the intended accountability purpose of the assessment. Thus if an assessment has the explicit purpose to support the provincial commitment that students are achieving the expected standard, and the results are publicly reported, the purpose is considered to be one of accountability.

Our definition of gatekeeping expands on that proposed by Nagy (2000). In addition to granting student privileges such as graduation, admission, or grade promotion, gatekeeping also encompasses any contribution of assessment scores to students' final course marks. Therefore, gatekeeping can impact students' educational decisions to varying degrees. For example, an assessment where the score contributes 10% to a student's final course mark will impact decision-making to a lesser extent than an assessment where the score contributes 50% to the student's final grade, which in turn will count less than a score that is the sole determiner of a student's grade.

Instructional diagnosis includes the assessment of students' strengths and weaknesses so that appropriate programming and instruction can be provided to individual students. The outcomes of data collected from instructional diagnostic assessments should have direct and immediate impact on the student's education. In contrast to accountability, the results are only reported to those stakeholders directly involved with programming and instruction.

Assessments monitoring student achievement provide information regarding student performance over time. Typically, such assessments provide data that can be used for temporal cohort analyses and/or vertical scaling, resulting in large-scale program or system modifications. Therefore, assessments monitoring student achievement over time document changes in students' knowledge and skills, changes in students' strengths and weaknesses at a more global level. These results typically do not directly impact the education of those students writing the assessments.

In choosing this framework, we deliberately did not differentiate examination programs in terms of low- and high-stakes. This decision was based on two factors. First, high-stakes assessments are defined in terms of consequences and importance, for example, marks, grade promotion, graduation, or admission (Santrock, Woloshyn, Gallagher, Di Petta, & Marini, 2004). Second, although low-stakes assessments do not directly impact individual students' educational decisions, it is not clear students make these distinctions. This is further complicated by those policies encouraging or allowing teachers to include provincial assessment results in determining the students' school grades.

### **Methodology**

Information regarding provincial and territorial assessment programs was collected through a document analysis of publicly accessible (electronic and print) policy documents issued by the Ministries of Education. Specific details regarding the explicit purposes of assessment programs, assessment development and administration (including responsibility, invigilation, and scoring), exclusion guidelines, and reporting methods were obtained. Telephone contact was made with representatives from each Ministry (Department) of Education to confirm the accuracy of information or supplement the information we found in the documents.

Using the explicitly stated purposes from each province/territory, assessments were classified in terms of the classification of their purposes and uses. Purposes that used the term "accountability" or associated terms (e.g., monitoring, ongoing reviews) and publicly reported teacher, school or board results were placed within the accountability category. Assessments in which the results were used to support student promotion, provide access to further education, or provided a part or sole determination of a grade were included under the framework of gatekeeping. Assessment programs considered under instructional diagnosis included those describing such a role, with the release of results allowing sufficient timing to make changes to instruction based on those students' strengths and weaknesses. Finally, assessments were classified as monitoring student achievement over time if they provided information about student progress or enabled comparisons of student performance across grades.

### **Characteristics of Provincial/Territorial Assessment Programs in Canada**

Specific information regarding individual practices per province or territory is detailed below (Provincial/Territorial Assessment Program Profiles) and is also summarized in Tables 1 through 3. Through a comparison of provincial/territorial assessment programs, several general characteristics can be deduced regarding large-scale assessment practices in Canada. For example, with the exception of Nunavut, formal provincial exclusion policies have been

established in each of the provinces and territories. In documenting the large-scale assessment programs across Canada, the results highlight the wide variety of assessment programs, policies, and procedures in place. An overview of general characteristics of large-scale assessments in Canada, their explicitly-stated purposes, and the evolving culture of large-scale educational assessments in Canada follows from these analyses.

### **Provincial/Territorial Assessment Program Profiles**

**Alberta.** Two assessment programs are administered in Alberta on a yearly basis: the Achievement Testing Program (ATP) and the Alberta Diploma Examinations Program (ADEP, Alberta Education, 2007). All of the provincial tests are developed by Alberta Education, who, along with school authority personnel are responsible for ensuring that high-quality education is provided to all students in the province. The test items are initially developed by trained classroom teachers, reviewed and revised by assessment specialists in the Ministry, pilot tested, and reviewed and revised prior to use. The tests include multiple-choice, numeric- and constructed-response items. The items and tasks included in these tests are referenced to the learning outcomes identified in the corresponding Programs of Study.

The ATP is conducted annually in all elementary and secondary schools across the province. Alberta Education states the purposes of the Achievement Testing Program are to (a) determine if students are learning what they are expected to learn; (b) report to Albertans how well students have achieved provincial standards at given points in their schooling; and (c) to assist schools, jurisdictions, and the province in monitoring and improving student learning. In addition, improvement of student learning and establishing accountability are also cited as purposes of the Achievement Testing Program. Students in Grades 3, 6, and 9 are administered tests in mathematics, Français, French language arts, and English language arts (reading and writing). In addition, Grade 6 and 9 students complete tests in science and social studies. Tests are administered by the classroom teacher. Although formal marking of the tests is completed by selected teachers at a central location, teachers are encouraged, but not required, to mark their students' tests and use the mark as part of their students' grades.

The ADEP for Grade 12 students includes 14 examinations for selected core academic subjects in English, French, social studies, mathematics, science, biology, chemistry, and physics (Alberta Education, 2007). The Grade 12 ADEP has three explicit purposes including: (a) to certify the level of individual achievement in selected Grade 12 courses; (b) to ensure that province-wide standards of achievement are maintained; and (c) to report individual and group results. The examination mark counts for 50% of a student's final mark. To obtain credit in a diploma examination course, students must achieve a final composite score (school mark and

examination mark) of 50% or higher. The tests are administered by school personnel (although not the specific subject teacher) and are centrally marked by teachers who have been nominated by school superintendents and selected according to criteria identified by Alberta Education.

Summary results for all of the tests are reported to students, parents, schools, and the public annually. Detailed reports are provided to teachers, principals, and district administrators. The detailed reports can be used by teachers and administrators in planning and in the delivery of effective instruction.

**British Columbia.** Two provincial assessment programs are administered in British Columbia: the Foundations Skills Assessment (FSA) and the Graduation Program Provincial Examinations (GPPE; British Columbia Ministry of Education, 2007a, 2007b). All of the assessments are developed for the Ministry of Education with item development support from experienced classroom teachers. Many assessments include both multiple-choice and constructed response items, and are referenced to the learning outcomes outlined in the corresponding provincial curriculum.

The British Columbia's Foundation Skills Assessment is an annual province-wide assessment for Grades 4 and 7 students. The purposes of the FSAs are to determine how well students are achieving the basic skills and to help schools in their planning to improve student achievement. Reading comprehension, writing, and numeracy are assessed in each of the two grades. The assessments are administered by classroom teachers. Given the purposes of this assessment program, the administration and marking of these assessments have undergone a series of changes over the past three years. For example, beginning with the 2007-2008 school year, the assessments are administered in early to mid February. Scoring of the open-ended items occurs at the district level with a sample of assessments scored centrally in a summer monitoring session.

As part of the Graduation Program Provincial Examinations, secondary students are required to take five course-based provincial examinations (Language Arts 10 and 12, Science 10, Mathematics 10, and Social Studies 11/12). The purpose of these examinations is to ensure graduating students have consistent levels of minimum competency. Students may elect to take additional provincial graduation examinations related to specific Grade 12 academic courses. Grades 10 and 11 examinations account for 20% of students' final course marks with Grade 12 examination scores accounting for 40%, when they are written. In the case of the provincial examination program, the open-ended portion of the Social Studies 11, Civic Studies 11, BC First Nations Studies 12, Grade 10 English and Français Langue Première 10 exams are scored

locally at the school or district level. All other provincial examinations with open-ended portions are scored in central marking locations by teachers selected by the Ministry of Education.

**Manitoba.** Manitoba's Provincial Assessment Program (PAP) includes 3 testing programs conducted on an annual basis, the Grade 3/4 Assessment, the Middle Years Assessment (MYA) and the Grade 12 Standards tests. The main purpose of the PAP is to "provide feedback to students, teachers and parents about student learning: a) informing instructional planning and helping to determine the need for changes or student specific interventions; b) providing system-wide information that assists in identifying trends and making decisions about resources and support; and 3) providing the public with general information about student achievement to sustain confidence in the education system." (Manitoba Department of Education, Citizenship and Youth, 2007).

Grade 3 students complete tests in the areas of reading and numeracy. Students in French Immersion complete the numeracy test in Grade 3 and the reading test in Grade 4. The tests are administered by classroom teachers at the beginning of the school year. The early administration enables the results to be used to inform parents and support instruction at the student, classroom, school, school division and provincial levels.

The Middle Years Assessment is a unique assessment program in Canada. The assessments are classroom based formative assessments, focusing on Grade 7 students' level of engagement with school and certain competencies in mathematics and Grade 8 students' skills in reading (comprehension) and writing (expository). Based on the most recent, stable evidence of student achievement obtained during teaching, student reports are created during the last two weeks of January showing each student's level of attainment on each of the identified key competencies. The reported competencies are based on mid-year criteria provided by the Department based on curricular, grade-level learning outcomes. The Department also provides assessment criteria supported by rubrics, continua, and student examples, so that reported performance results are as reliable and as valid as possible.

In Grade 12, Standards Tests occur in language arts and mathematics (Manitoba Education and Training, 2007). Tests are offered in both English and French and results account for 30% of the students' final course mark. These assessments are developed by the Ministry of Education with input from teacher committees. The tests are marked locally by a team of teachers supervised by a local marking coordinator who has received training and support from the Department. A sample of test booklets are sent to the Department of Education to measure and report the level of agreement between the local and central marking process.

Assessment results from each of the three testing programs are only reported to the student, parents, and teachers. There are also supporting documents for teacher use. A summary provincial report is produced for the Grades 3 and 4 assessment results.

**New Brunswick.** Currently, New Brunswick has the most comprehensive examination program in Canada with annual provincial assessments at several grade levels and additional provincial assessments being piloted. Further, as Canada's only officially bilingual province, there are separate Anglophone and Francophone sectors within the Department of Education. Each sector administers its own provincial assessment program although the purposes of both are the same. "The Department of Education administers a comprehensive Provincial Evaluation Program to monitor overall student achievement at particular points in the system. This provides important feedback at provincial and local levels about students' knowledge and skills" (New Brunswick Department of Education, 2007a). These assessment and evaluation programs are intended to improve student achievement and to keep parents informed about their child's progress. All provincial assessments are developed by the Department of Education with teacher input. Assessments include both selected and constructed response items that are based on the provincial curriculum.

Anglophone students in Grades 2, 4, 7, and 9 complete a literacy assessment focusing on reading comprehension and writing. The purpose of these assessments is to support the provincial commitment above and ensure students are able to read, comprehend, and write at an expected level by the end of the specified grades. The Grade 9 Literacy Assessment is also a graduation requirement for students. "Successful completion of the assessment will be achieved by the student who performs at an 'appropriate' or 'strong' level on both components (Part One – Reading Comprehension and Part Two – Writing) which comprise the assessment. Successful completion of the English Language Proficiency Assessment is required to graduate with a New Brunswick diploma" (New Brunswick Department of Education, 2007b, p. 5). Anglophone students in Grades 5 and 8 write assessments in mathematics. Currently, a Grade 6 science assessment is being piloted and a Grade 3 mathematics assessment is under development. Lastly, the province offers French certification for Anglophone students who successfully complete the French Language Proficiency Examination.

The structure of the assessment program in the Francophone system is similar at the elementary level although the actual examinations do vary. Provincial assessments for Francophone students begin in Pre-kindergarten (Évaluation de la petite enfance – appréciation directe) and Kindergarten (Évaluation de la petite enfance – appréciation de l'enseignante) for diagnostic purposes. Grade 2 Francophone students complete a literacy assessment focusing

on silent and oral reading. Grade 3 students complete a mathematics assessment (Mathématiques) (New Brunswick Department of Education, 2006). Beginning in September 2008, Grade 4 students will write a literacy (reading) (Littératie) assessment to provide Grade 4 teachers with a current report of students' reading abilities reflecting reading loss during the summer months. In Grades 5 and 8, Francophone students complete provincial examinations in science (Sciences de la Nature), mathematics (Mathématiques), and French reading and writing (Français). Lastly, Grade 10 Francophone students complete an assessment in science (Sciences de la Nature) and Grade 11 students complete an assessment in history (Histoire). Rather than a single proficiency examination as used for the Grade 9 assessment in the Anglophone system, Francophone students write provincial examinations in Français 11 and Mathématiques 11. These examinations account for 40% of each student's final course grade. There is an English oral proficiency examination for all students in Grade 10 but it is an oral examination only.

Scoring generally occurs in centralized scoring locations. Districts and schools are provided provincial, district and school level information, and students receive an individual report summarizing their results.

**Newfoundland and Labrador.** Two assessment programs are administered annually in Newfoundland and Labrador, the Criterion Referenced Testing (CRT) program and the Public Examination (PE) program (Newfoundland and Labrador Department of Education, 2007). CRTs and the Public Examinations contain both selected and constructed response items which are developed by practicing teachers, and test development specialists at the Department of Education. Assessments are administered by classroom teachers. Marking of provincial assessments is completed centrally with teachers from various school districts participating.

The CRT in English language arts and mathematics are administered to all students in Grades 3, 6, and 9, focusing on specific strands within the respective curriculum. For example, the Grades 3 and 6 Language Arts assessment focus on listening and speaking, reading and viewing, and writing and other ways of representing. The Department randomly selects a sample of students to complete the speaking portion of the assessment. The purpose of the CRT program is to (a) allow useful comparisons of performance between students in Newfoundland/Labrador and those in other schools nationally; (b) provide an established standard on which to assess student achievement over time; and (c) provide the Department of Education with the information needed to make important decisions regarding students' strengths and weaknesses. CRT results are reported to schools, school districts, parents, and students at the beginning of the next school year. There is also an Elementary Science

Assessment Resource support document focusing on scientific literacy and process. The document is designed for teachers to use to help students acquire the necessary levels of scientific knowledge and skills.

Public Examinations are held in the high-school exit level courses for all core academic subject areas and the results account for 50% of the student's final course marks. Final marks for public examination courses are mailed to students in July and are reported to schools and school districts during the same time period.

A comprehensive report containing three year trends is produced for the CRTs and Public Examinations. These reports are distributed to schools and school districts and are also posted on the Department of Education's website.

**Northwest Territories.** The Northwest Territories is currently using Alberta's Achievement Testing Program (ATP) and Alberta's Diploma Examination Program (ADEP) in the same way the assessments and examinations are used in Alberta (Northwest Territories Department of Education, 2007). The Northwest Territories Department of Education website links directly to the Alberta Learning site for Alberta Achievement Testing and Diploma Examination information.

**Nova Scotia.** Nova Scotia administers two assessment programs on an annual basis, the Program of Learning Assessment for Nova Scotia (PLANS), and the Nova Scotia Examinations (NSE) (Nova Scotia Department of Education, 2007). PLAN tracks both individual achievement and system performance. Assessments are provided in English, French Immersion, and French First Language. The purposes of the assessment program are to provide "information to improve the quality of educational decision making" and "identify the needs of students so that they can be supported." Hence students who do not meet expectations receive support each year until they reach the next assessment level. The provincial assessments in Nova Scotia are developed by the Ministry of Education, teachers, and advisory committee groups. The assessments include both selected and constructed response items and are administered by classroom teachers. In the case of the Grades 3, 6, and 9 assessments, scoring is completed centrally with the exception of the Early Language Literacy Assessment which is scored regionally by teachers under the direction of department professional staff. Marking of the NSEs is completed by the classroom teacher.

Grade 3 students write the Early Language Literacy Assessment in September and the Early Elementary Mathematical Literacy Assessment in June. Students in Grade 6 write the Elementary Literacy Assessment in October and those in Grade 9 write the Junior High Literacy Assessment in May. The literacy assessments focus on reading and writing skills. There are

also plans to add a mathematics examination for Grades 6 and 9. The Department of Education describes these assessments as “assessments for learning.” The Nova Scotia Examinations (NSE) are administered to Grade 12 students in core subject areas (Nova Scotia Department of Education, 2007). NSE results account for 30% of the students’ final course marks.

Representative samples of examinations are marked by the Department of Education to provide provincial and board results. Students receive their individual results while schools receive the school, school board, and provincial results.

**Nunavut.** Two assessment programs are administered annually in Nunavut. Currently, there is a Grade 3 assessment program and a Grade 12 examination program that uses the Alberta Diploma Examinations in the same manner as it is used in Alberta (Nunavut Department of Education, 2007).

Grade 3 students are assessed in mathematics in all four official languages (English, French, and the eastern and western dialects). A Grade 7 and Grade 9 assessment in reading is currently under development with plans of implementing a Grade 7 reading assessment in 2009.

**Ontario.** Unlike the other provinces, the provincial assessment program in Ontario is the responsibility of the Educational Quality and Accountability Office (EQAO), which is at arms-length to the Ministry of Education. EQAO has its own Board of Directors, and the Chief Executive Officer reports directly to the Board and to the Minister of Education. EQAO annually administers three province wide testing programs, the Grade 3 and Grade 6 Literacy and Numeracy Assessment, the Grade 9 Numeracy Assessment, and the Ontario Secondary School Literacy Test (OSSLT, Educational Quality and Accountability Office, 2007). All assessments are developed by EQAO with support from teachers and private contractors. All assessments include a combination of selected response and constructed response items linked to the specific learning expectations in the provincial curriculum documents. The Grade 3 and 6 assessments are administered in May and the OSSLT is administered in March. The Grade 9 mathematics assessment is administered at the end of each semester. All assessments are supervised by classroom teachers either in the classroom or a central location within the school. Marking of the assessments is completed centrally using teachers from across the province.

The purposes of the Grades 3 and 6 Literacy and Numeracy Assessments are to (a) assess students and report yearly data on the level at which students are meeting curriculum expectations in reading, writing, and mathematics; (b) provide data to assist schools in improvement planning and target setting; and (c) provide support for the implementation of the curriculum. These scores do not contribute to students’ final grades. The purpose of the Grade

9 Numeracy Assessment is to determine whether or not students have met the mathematics expectations in the Ontario Curriculum and to identify areas of weakness so that appropriate remediation can take place. Teachers commonly include a portion of the provincial test in calculating students' final grades, to a maximum of 10%.

The OSSLT is administered to Grade 10 students and is a graduation requirement. However, students eligible to write the test twice but who have failed or been excluded are eligible to take the Ontario Secondary School Literacy Course (Education Quality and Accountability Office, 2007). Successful completion of this course satisfies the graduation requirement. The purpose of the OSSLT is to determine whether a student has the literacy (reading and writing) skills required to meet the standard for understanding reading selections and communicating in a variety of writing forms expected by the Ontario curriculum across all subjects up to the end of Grade 9.

Group level results for all of the assessments are reported publicly at the provincial and school levels. Individual student results are reported to students and parents, and schools and school boards receive detailed results at the student, item, and strand levels.

**Prince Edward Island.** With the completion of its task force on student achievement (Kurial, 2005), Prince Edward Island became the last educational jurisdiction in Canada to implement province wide large-scale educational assessments. Based on the recommendations of the task force, annual provincial assessments are being implemented in language arts and mathematics for all Grades 3, 6, and 9 students and designated subjects at the senior high-school level. The assessments are linked to the provincial curriculum (Prince Edward Island Department of Education, 2007). Teachers are involved in all aspects of the assessments. Item development is completed by teachers working in collaboration with the Department of Education and school boards. These assessments consist of selected response and constructed response items that are based on specific curriculum outcomes. Administrations of the assessments occur in May and June at the school under teacher supervision. Assessments are scored by teachers during central marking sessions and results are presented to students, schools, and school boards.

The Grade 3 Literacy Assessment (reading and writing) was first administered in the 2006/2007 school year, while the Grade 6 Language Arts Assessment was introduced in 2007/2008. The purposes of these assessments are not to rank or compare students or schools but rather to provide accurate achievement information to inform parents, teachers, and students; help improve teaching and learning; and guide professional development. The results are also used to help districts/boards plan resources and support, and help the Department of

Education monitor student learning, with the goal of targeting areas for improvement through curriculum redesign or program initiatives (Prince Edward Island Department of Education, 2007). The Grades 3 and 6 assessments are not used to determine students' marks.

The primary purpose of the Grade 9 Mathematics Assessment is to monitor the mathematics skills and knowledge of students. Further, the assessment contributes 10% to students' mathematics grade in their Mathematics 9 course. Grade 9 students receive an overall score and subtopic scores in each of the seven strands assessed.

Although results were not released to the public at the inception of the provincial assessment program, school, district, and provincial results are now available to the public on the Ministry website. This change was made in response to requests based on the Freedom of Information and Protection of Privacy Act.

**Québec.** Two large-scale annual assessment programs are administered in Québec: the Learning and Evaluation Situation (LES) Compulsory Cycle 3 Examinations and the Certification Examinations (Ministère de l'Éducation, du Loisir et du Sport, 2007a). The Ministère de l'Éducation, du Loisir et du Sport (Québec Ministry of Education, Recreation, and Sport) also develops optional examinations for teacher use. All assessments are developed by the Ministère de l'Éducation, du Loisir et du Sport. The provincial assessment program is under transition as the new curriculum becomes fully implemented over the next three years. All assessments use constructed response and there is an increased use of performance assessment items considered to represent more realistic and complex tasks. Administrations of the compulsory and Certification Examinations are completed at the schools and marking is completed by teachers locally in a marking centre. Random samples of students' examinations are scored during a summer marking session conducted by the Ministère.

The Cycle 3 examinations are given to students at the end of Cycle 3 (equivalent to Grade 6) between April and June. The subjects included are language (French and English reading, writing, and listening) and mathematics. The goals of these assessments are to describe the competency levels attained by students at the end of elementary school. The LES also "offers guidance to teachers who seek to inform themselves about the effectiveness of their classroom practices" (Ministère de l'Éducation, du Loisir et du Sport, 2007b, p. 3). Hence the examinations are designed to support teachers' practices within an assessment for learning context, albeit considering the teacher as the learner. Results for the Grade 6 examinations are not included in the students' course marks.

Certification Examinations are administered to students in Secondary 4 (Grade 10) and Secondary 5 (Grade 11) in core subject areas (i.e., language, mathematics, history, science,

etc.), and are used to certify competency in specific subject areas. The results from these assessments contribute up to 50% of students' final course marks. The Certification Examinations in language arts and mathematics do not contain any selected response items and there is a desire for future examinations in the other disciplines to also reduce the number of selected response items on their certification examinations.

The Ministère is responsible for publishing the results of all of the examinations. Official documents are published with individual and school level results for students and schools. The Ministère also publishes an annual report intended for educational institutions and the general public (Ministère de l'Éducation, du Loisir et du Sport, 2007b).

**Saskatchewan.** Two provincial assessment programs are given annually in Saskatchewan, the Assessment for Learning Program (AFL) and the Grade 12 Departmental Examinations (Saskatchewan Ministry of Education, 2007a). The Ministry is responsible for the development of student assessments under the provincial Continuous Improvement Framework. The assessments are developed by teachers under the guidance of the Ministry and items are based on the current provincial curricula. All AFL assessments and language arts Departmental Examinations include a combination of selected response and constructed response items while the remaining Departmental Examinations are solely selected response items. The constructed response items from both assessment programs are centrally scored by teachers. Classroom teachers participating in the AFL assessments are "encouraged to score student work, share feedback with the students and create student and classroom profiles prior to returning booklets for central scoring" (Saskatchewan Ministry of Education, 2007b).

The AFL program's goal is to provide data to teachers and education leaders "to provoke debate and inform decision-making in order to improve student learning" (Saskatchewan Ministry of Education, 2007c). These assessments are compulsory and provide diagnostic information for planning purposes. The AFL program currently administers student assessments in reading, mathematics, and writing. Plans are underway to include science in future assessments. Students in Grades 4, 7, and 10 complete the reading assessment in April and students in Grades 5, 8, and 11 complete the mathematics assessments in June and the writing assessments in April. All of the Grades 4, 5, 7, 8 and 10 students complete the assessments, whereas only those students enrolled in the appropriate mathematics or language arts Grade 11 courses are required to complete the assessment.

The Grade 12 Departmental Examinations are administered in core academic subject areas (biology, chemistry physics math, and language arts) to Grade 12 students instructed by

teachers who have not been accredited by the Ministry of Education. For these students, the examinations contribute 40% to their final course marks.

Schools receive detailed reports summarizing the AFL results for the school, district and province. The information includes overall results along with detailed information at the strand and format level. These reports are available to the public if requested. Annual reports from Saskatchewan Learning also include a summary of the assessment results in order to describe the state of the education in the province.

**Yukon.** Three provincial assessment programs are administered in the Yukon: the Yukon Achievement Test (YAT), the British Columbia Provincial Examination Program, and the Language Proficiency Index (LPI, Yukon Department of Education, 2007). The YAT and LPI are developed, administered, and marked by Yukon teachers. The assessments are based on curricular standards and include selected and constructed response items for the Grades 6 and 9 assessments and only constructed response items for the Grade 3 assessment.

The key purposes of the YAT program are to assess student learning to a) determine if students are learning what they are expected to learn, b) report to Yukoners how well students have achieved territorial standards at given points in their schooling and c) assist schools and the territory in monitoring and improving student learning (Yukon Department of Education, 2007). The YAT program is administered to student in Grades 3, 6, and 9 in mathematics and language arts. Students' scores on the Grade 9 assessment accounts for 25% of the students' final course mark.

The British Columbia Provincial Examination Program is used for students in Grades 10, 11, and 12. For these examinations, the results from the assessments account for 20% at Grade 10 and 11 and 40% of students' final course marks for Grade 12. The LPI program certifies Grade 12 student in language arts (Yukon Department of Education, 2007). This assessment is composed of performance items and is used by post-secondary institutions for admissions consideration.

Students and the school receive individual results from the YAT and provincial examination program. Schools receive a school level summary of the YAT results, providing aggregate results of the school in comparison to the overall results in the territory and in Alberta. School reports containing students' results in each particular school are also provided. Additional summary territorial reports are produced containing results across the assessment programs, further breakdowns of rural/urban, and First Nation/non-First Nation results, and cohorts of students who wrote current and previous assessments. (Yukon Department of Education, 2007)

### **The Current Assessment Programs in Canada**

Due to the provincial/territorial control of public education in Canada, there is no common curriculum across the country; however, many of the curricula have similar learner expectations. Further, provinces and regions have worked together to better align their curricula, for example the Eastern and Western provinces. Likewise, large-scale assessment programs are conducted by assessment and evaluation units within the provincial/territorial governments, or in the case of Ontario, a government organization. These provincial/territorial assessments are designed to reflect the foundational knowledge and skills contained within the respective curricula.

Table 1 provides a summary of the grades each province and territory assesses its students, while Table 2 indicates the subjects assessed by grade level for each provincial/territorial program. For example, Alberta uses large-scale testing in Grades 3, 6, 9, and 12 (see Table 1). Language/Literacy is tested at Grades 3, 6, and 9, mathematics in Grades 3, 6, and 9, science and social science in Grades 6 and 9 and specific academic subjects are examined in Grade 12 (see Table 2). Despite the provincial/territorial control of education, there are strong similarities across large-scale assessment programs. For example, two structural trends can be observed. First, provincial assessments typically occur every three years starting in the primary division and continuing until early secondary school. The most common starting point for such an assessment program is either Grade 3 or 4, although Saskatchewan does begin its program in Grade 2 and New Brunswick does have an assessment for kindergarten students in the Francophone system. This appears due primarily to the variance in academic divisions across the provinces/territories rather than being based on the cognitive developmental stages of students.

These provincial assessments provide ongoing feedback regarding student achievement in critical subject areas, most commonly literacy and mathematics. Literacy is typically defined by reading and writing skills. The range of topics for the mathematics assessments is somewhat more varied, falling under the names of numeracy or mathematical literacy and focusing on number sense, problem solving, computation, or communication. While these two assessment areas are common, the structure and focus of the assessments vary. Regardless, they provide information on program and system effectiveness, although student level results are commonly produced and provided to students and their parents.

Table 1: *Current Provincial/territorial assessments by Grade*

Province/territory	Grade										
	k-2	3	4	5	6	7	8	9	10	11	12
Alberta		X			X			X			X
British Columbia			X			X			X	X	X
Manitoba		X	X			X	X				X
New Brunswick-A	X		X			X	X	X			
New Brunswick-F	X	X		X			X	X	X	X	
Newfoundland & Labrador		X			X			X		X	X
Northwest Territories		X			X			X			X
Nova Scotia		X			X			X			X
Nunavut		X									X
Ontario		X			X			X	X		
Prince Edward Island		X			X			X			
Quebec					X				X	X	
Saskatchewan			X	X		X	X		X	X	X
Yukon		X			X			X	X		X

Table 2: *Provincial/territorial assessments by subject and grade*

Province/territory	Subjects per grade				
	Language/ Literacy	Math/ Numeracy	Sciences	Social Science	Academic 11 /12
Alberta	3, 6, 9	3, 6, 9	6, 9	6, 9	X
British Columbia	4, 7, 10,12	4, 7, 10	10	11	X
Manitoba	3, 8, 12	3, 4, 7, 12			
New Brunswick-A	2, 4, 7, 9	5, 8			
New Brunswick-F	K, 2, 5, 8, 10, 11	3,5, 8, 11	5, 8, 10	11	
Newfoundland & Labrador	3, 6, 9	3, 9			X
Northwest Territories	3, 6, 9	3, 6, 9			X
Nova Scotia	3, 6, 9	3			X
Nunavut		3			X
Ontario	3, 6, 10	3, 6, 9			
Prince Edward Island	3, 6, 9	3, 6, 9			
Quebec	6	6			X
Saskatchewan	4, 7, 10 (reading) 5, 8, 11 (writing)	5, 8, 11			X
Yukon	3, 6, 9, 10	3, 6, 9, 10	10	11	X

Second, there is a trend to combine early student testing (i.e., between Grade 2 and Grade 6) without any direct associations with students' grades, followed by early secondary examinations having relatively minor impact on students' grades, and ending with exit certification examinations in Grade 11 or Grade 12, with relatively large direct impact on students' grades. This model appears to provide early ongoing monitoring of students' strengths and weaknesses so that appropriate programming, support, and remediation can be provided, focused mainly at the system level. In the case of the Grades 10, 11 and 12 examinations, the assessments are not commonly explicitly used as program or system indicators but they do have a direct impact on academic course grades and graduation.

Similarities also exist in test development, format, administration, marking, and reporting. Teachers are typically involved in item development, administration, and marking. Nine provinces and two territories enlist teachers to help develop items along with the Ministry of Education or in the case of Ontario, EQAO. Assessments typically have both selected response and constructed response items with few assessments containing performance-based items. The assessments in Quebec are a notable exception as they only contain constructed response items. The assessment items are based on curricular expectations stated in provincial/territorial curriculum documents, although literacy examinations are often considered to be cross-curricular. Further, assessments not contributing to students' grades and having a three-year cycle are based on the curriculum of the current and the preceding grades not having large-scale assessments. In contrast, assessments used to determine students' course marks are based on the curriculum of the corresponding course. A notable exception is the OSSLT in Ontario which is considered a cross-curricular test of literacy.

Assessments are administered under standardized conditions to classes or cohorts of students, most commonly in the students' school. Administrations typically occur in April or May for assessments having no direct consequences to students and at the end of the course for assessments and exit examinations that contribute to course grades. These exit examinations have multiple administration dates throughout the year, depending on the different school schedules in operation (full year, semester, quarter). Two exceptions are the elementary literacy assessments in Nova Scotia (administered in September and October) and the OSSLT, a graduation requirement that is administered in March. Teachers and local school administrators proctor the assessments.

Scoring methods are somewhat variable across jurisdictions although teachers are always involved in marking the constructed response items. Assessments having a direct impact on students' grades or graduation requirements are marked in more central locations.

Assessments not directly impacting grades are either marked in a similar manner or by classroom teachers using marking guidelines provided by the province’s Ministry/Department of Education. Additional training is provided for scorers in most provinces/territories.

Assessment reports are most commonly produced at the student, school, school board and provincial level, and less commonly at the classroom level. When centrally scored, reports tend to be released in the summer for exit examinations and in the fall of the following school year in for other assessments. Freedom of Information laws impact the reporting of provincial assessment results. Hence reports are distributed in a manner protecting the confidentiality of individual students and teachers. Thus while the general public and policy makers have access to school, district, and provincial results, only the students, parents, and in some cases, the teachers or school administrators have access to individual student results. When public reports are published, caveats are provided noting the limitations and inherent errors of measurement.

### **Explicitly Stated Purposes**

The identification of the explicit purposes of the large-scale assessments across the provinces and territories is limited somewhat by the differences in accessibility of the relevant information, ambiguity of the purposes, and contradictions occurring across documents. However, what was gleaned revealed that these large-scale assessments serve several purposes. Further, the purposes continue to expand. Table 3 classifies the explicitly stated purposes that could be ascertained for the provincial/territorial assessments under the guiding framework of accountability, gatekeeping, instructional diagnosis, and monitoring student achievement over time. As shown in Table 3, many of the assessment programs have purposes within more than one of the categories. For example, the purposes of the Alberta Testing Program (ATP) can be classified under accountability, instructional diagnosis, and monitoring student achievement over time, whereas, the Alberta Diploma Examination Program (ADEP) serves the purposes of accountability and gatekeeping.

Table 3: *Provincial/territorial assessment programs by explicitly-stated purposes*

Province/territory	Accountability	Gatekeeping	Instructional diagnosis	Monitoring student achievement
Alberta	ATP, ADEP	ADEP	ATP	ATP
British Columbia	FSA, GPPE	GPPE		FSA
Manitoba	Senior 4 PEs	Senior 4 PEs	Grades 3/4, MYA	Grades 3/4

New Brunswick (Anglophone)	Grades 2, 3, 8, and 9 PEs	Grade 9 proficiency examination		Grades 2, 5, 7, 8, and 9 PEs
New Brunswick (Francophone)	Grade 2,3 PEs	Grade 11 French/ Mathematics PEs	Évaluation de la petite enfance	Grades 5, 7, and 8 PAs
Newfoundland & Labrador	CRT, PEs	PEs		CRT
Northwest Territories	ATP, ADEP	ADEP	ATP	ATP
Nova Scotia	PLAN, NSE	NSE	PLAN	PLAN
Nunavut	ADEP	ADEP		Grade 3 Assessment
Ontario	Grades 3, 6, and 9 assessments, OSSLT	Grade 9 Numeracy, OSSLT		Grade 3, 6, and 9 assessments
Prince Edward Island	Grades 3, 6, and 9 assessments	Grade 9		Grades 3, 6, and 9 assessments
Quebec	LES	CEs	LES	
Saskatchewan	AFL, DEs	DEs		DEs
Yukon	YAT, B. C. GPPE,	YAT (Grade 9), B. C. GPPE, LPI		YAT

All of the provinces/territories explicitly state that at least one of their assessments is for system accountability purposes. The accountability framework is largely accomplished through public reporting of assessment results to the different educational stakeholders, especially at the school, school board, and provincial levels. Gatekeeping was found in 12 of the jurisdictions, and is most commonly associated with exit examinations contributing between 20

to 50% to students' final grades. At the extreme, Ontario and New Brunswick (Anglophone) have a graduation requirement attached solely to successful performance on a literacy test.

While all of the provinces/territories used language that suggested the results of their assessments could be used for instructional diagnoses, only Alberta, Manitoba, New Brunswick (Francophone), the Northwest Territories, Nova Scotia, and Saskatchewan have explicit statements clearly identifying the use of assessment results for instructional and diagnostic purposes, although the statements were often vague. Manitoba, Prince Edward Island, Quebec, and Saskatchewan mention that their assessments can be used to help teachers examine the effectiveness of their classroom practices, although this appears to be related to the increasing popular notion of "Assessment for Learning." Given that the administration of the assessments generally occurs later in the school year, these instructional and diagnostic uses appear to be directed primarily to support future teacher practices rather than supporting the individual students who have completed the assessment. Nonetheless, the majority of the jurisdictions do use language also suggesting similar purposes.

Monitoring student achievement over time is explicitly-stated as a purpose of 10 assessment programs across Canada. The most common approach is to compare performance a cross-sectional design. Provinces with individual student numbers are also conducting longitudinal cohort analyses (e.g., same students in Grades 3, 6, and 9).

While accountability is the most common purpose for these large-scale assessments, three of the four categories are commonly found in each jurisdiction. The only exception is instructional diagnosis, which is only explicitly mentioned in six of the provinces/territories. However, based on the scheduling and marking of the examinations, the achievement of this mandate would often be difficult to accomplish.

Monitoring student achievement is common at the elementary level, while gatekeeping is common at the secondary level. However, given the current emphasis on accountability, schools and districts are expected to use their assessment results from many of the programs to support data-based decision making and develop their own accountability frameworks. For example, elementary schools in Ontario with consistently low performance on the provincial assessments are being provided extra funding or resources to help support improvement efforts designed to improve overall student performance.

None of the educational jurisdictions in Canada attach any negative consequences for teachers, schools, or districts based on assessment results; however, poor performance on several high-school examinations could have negative consequences for students. And there are reported cases of teachers not teaching, through request or transfer, the students in grades

or classes for which there are assessments. Some jurisdictions (Alberta, Ontario, Prince Edward Island) encourage or mandate the use of assessments primarily designated for monitoring purposes to be included in students' grades.

It is often not clear which are the primary purposes and which are the secondary purposes or if the purposes are all considered equally important. The most common examples are those assessments in which the results are not only used as part of the students' grades but also serve an accountability function. There also appears to be purposes that are attached to assessment programs that were employed well after the assessment was first implemented. Based on our review, the assessments themselves have not changed in substantive ways to address these increasing purposes. Examples of expanding purposes can be found in the provinces that now list "assessment for learning" as a purpose for at least one of their large-scale assessments.

The various audiences to whom results are reported have become the dominant mechanism to meet the explicit purposes of these assessment programs. Nonetheless, reporting of assessment results to various educational stakeholders requires forethought and caution if results are to be interpreted in appropriate, valid, and meaningful ways. Further, given the ongoing changes to purposes, there is a real need to carefully examine and the psychometric assessment properties of the current assessments and the interpretations of the resulting scores to ensure particular inferences are valid and not open to misinterpretation. When assessments constructed for one purpose are used to draw inferences about a newly added purpose, the validity of the inferences drawn may be in jeopardy and as such, the resulting educational decisions and reforms may be misdirected.

### **Conclusion**

The evolving culture of large-scale assessment programs in public education throughout Canada results in a diverse terrain of programs designed to meet several purposes and needs. Educational jurisdictions have created structures and protocols to attempt to meet these purposes. The results are a general set of assessments that are relatively similar in the subject areas and grade levels tested, and the structures and timing of the assessments. However, notable exceptions exist, and more fundamental differences are found in the scoring, reporting, and use of the assessment results.

Several psychometric, practical, and philosophical issues arise from such diverse uses of large-scale assessments. While we focused on the explicitly stated purposes of the assessment programs, it is likely there are also implicit purposes evolving for the various programs. Hence we also see a need to examine these assessment programs in relation to

developing implicit purposes and the alignment of the assessment methods with their intended uses.

The *Principles for Fair Student Assessment Practices for Education in Canada* identifies the need for test developers and users to clearly indicate the intended purposes of assessments (Joint Advisory Committee, 1993). Explicitly-stated purposes for large-scale assessments are essential if valid inferences from assessment results are to be deduced. Further, in attempting to maximize the use of these assessments, their purposes appear to be increasing over time. The most recent example is the use of particular assessments to support “assessment for learning.” Given the increasing reliance on large-scale assessments in Canada, it is critical for provinces and territories to fully articulate the various purposes of their large-scale assessment programs and ensure their format and use can meet these purposes.

Large-scale assessments and provincial/territorial assessment programs are an active part of Canadian education. Inferences drawn from these assessments help shape and guide instruction, curriculum, and policy, and inform student-based decisions. As such, a number of educational, social, and political issues arise from their use. In response, provincial/territorial assessment programs must continue to respond to societal demands while ensuring assessment integrity if they are to positively inform educational decisions. The ongoing debates regarding the purposes and value of such assessment programs serve to highlight the need for ongoing examination of the large-scale educational assessment programs in Canada and the society in which these assessments operate. Further, the findings of the present study reveal current practices and changing trends in large-scale assessments and inform current debates regarding the roles and functions of large-scale educational assessment not only in Canada but also in other countries that are expanding their use of large-scale assessments to inform and monitor education.

Our results also indicate a need for more research into the framework used to classify the purposes of large-scale assessment programs. For example, the purposes of accountability and monitoring student achievement over time seem to be almost synonymous in most jurisdictions. Further, the increasingly occurring purpose of “Assessment for Learning” may not fit under any of the current categories.

#### Notes:

This study has been supported by a standard research grant from the Social Sciences and Humanities Research Council (SSHRC).

The authors would like to thank the anonymous reviewers for their helpful advice and comments. Thank you also to representatives from each of the provincial ministries of education for clarifying the information about their respective assessment programs, specifically, Tim Caleval, Ken Clark, Andre Corbet, Anne Doucet, Bob Gardiner, Robert Laurie, Joanne McGrath, Brenda Neufeld, John Rymer, Dwight Tranquilla, and Cindi Wood.

## References

- Adams, J. E. & Kirst, M. W. (1999). New demands and concepts for educational accountability: Striving for results in an era of excellence. In J. Murphy and K.S. Louis (Eds.), *Handbook of research on education administration* (2nd ed., pp. 463-489). San Francisco: Jossey-Bass.
- Alberta Education: Ministry of Education. (2007). Retrieved October 1, 2007, from <http://www.education.gov.ab.ca/>
- British Columbia Ministry of Education. (2007). Retrieved October 9, 2007 from <http://www.bced.gov.bc.ca/exams/>
- British Columbia Ministry of Education. (2007). Retrieved November 15, 2007 from <http://www.bced.gov.bc.ca/assessment/fsa/>
- British Columbia Ministry of Education. (1993). *Annual report: 1992-1993 school year* (122nd ed.). Victoria, BC: Queen's Printer.
- British Columbia Ministry of Education and Ministry Responsible for Multiculturalism and Human Rights. (1991). *Science in British Columbia: The 1991 British Columbia science assessment provincial report*. Victoria, BC: Queen's Printer.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- Crundwell, 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 1-21. Retrieved October 11, 2007 from <http://www.umanitoba.ca/publications/cjeap/pdf%20files/crundwell.pdf>
- Earl, L. M. (1999). Assessment and accountability in education: Improvement or surveillance. *Education Canada*, 39(3), 4-6, 47.
- Education Quality and Accountability Office. (2007). Retrieved October 5, 2007, from <http://www.eqao.com>
- Graham, C., & Neu, D. (2004). Standardized testing and the construction of governable persons. *Journal of Curriculum Studies*, 36, 295-319.
- Hodgkinson, D. (1995). Accountability in education in British Columbia. *Canadian Journal of Education*, 20, 18-26.
- Johnson, F. H. (1964). *A history of public education in British Columbia*. Vancouver, BC: University of British Columbia.

- Joint Advisory Committee. (1993). *Principles for Fair Student Assessment Practices for Education in Canada*. Edmonton, AL: Author.
- Kurial, R. (2005). Excellence in education: A challenge for Prince Edward Island: Final report of the task force on Student achievement.  
[http://www.gov.pe.ca/photos/original/task\\_force\\_edu.pdf](http://www.gov.pe.ca/photos/original/task_force_edu.pdf)
- Levin, B. (1998). An epidemic of education policy: (What) can we learn from each other? *Comparative Education*, 34(4), 131-141.
- Madaus, G. F., & Kellaghan, T. (1992). Curriculum evaluation and assessment. In P. W. Jackson Ed.), *Handbook of Research on Curriculum* (pp. 119-154). New York: Maxwell Macmillan International.
- (Manitoba Department of Education, Citizenship and Youth, 2007). (2007). Retrieved October 9, 2007, from [http://www.edu.gov.mb.ca/k12/assess/assess\\_program.html](http://www.edu.gov.mb.ca/k12/assess/assess_program.html)
- Mazzeo, C. (2001). Frameworks of state: Assessment policy in historical perspective. *Teachers College Record*, 103, 367-397.
- McCutcheon, J. M. (1941). *Public education in Ontario*. Toronto, ON: T. H. Best Printing.
- McEwen, N. (1995). Accountability in education in Canada. *Canadian Journal of Education*, 20, 1-17.
- Millar, J. (1893). *The educational system of the province of Ontario, Canada*. Toronto, ON: Warwick & Sons.
- Ministère de l'Éducation, du Loisir et du Sport. (2007a). Retrieved October 9, 2007, from [http://www.meq.gouv.qc.ca/GR-PUB/m\\_englis.htm](http://www.meq.gouv.qc.ca/GR-PUB/m_englis.htm)
- Ministère de l'Éducation, du Loisir et du Sport. (2007b). *Information document: Compulsory examination, English Language Arts*.  
[http://www.mels.gouv.qc.ca/DGFJ/de/pdf/2007/ela3\\_07.pdf](http://www.mels.gouv.qc.ca/DGFJ/de/pdf/2007/ela3_07.pdf)
- Nagy, P. (2000). The three roles of assessment: Gatekeeping, accountability, and instructional diagnosis. *Canadian Journal of Education*, 25, 262-279.
- National Assessment Agency (2008). Retrieved April 8, 2008 from: <http://www.naa.org.uk/>
- New Brunswick Department of Education. (2006). *Provincial examination results: Francophone school districts*. Retrieved October 11, 2007, from <http://www.gnb.ca/0000/publications/evalf/Rapportpublic2006AN.pdf>
- New Brunswick Department of Education. (2007a). Retrieved September 29, 2007, from <http://www.gnb.ca/0000/anglophone-e.asp#e>
- New Brunswick Department of Education. (2007b). English Language Proficiency Assessment Information Bulletin. Retrieved October 9, 2007, from

- <http://www.gnb.ca/0000/publications/eval/ELPA%20Information%20Bulletin%20September%202006.pdf>
- Newfoundland and Labrador Department of Education. (2007). Retrieved October 9, 2007, from <http://www.ed.gov.nl.ca/edu/>
- No Child Left Behind Act. 2002: Pub. L. No. 107–10.
- Northwest Territories Department of Education. (2007). Retrieved October 9, 2007, from <http://www.ece.gov.nt.ca/>
- Nova Scotia Department of Education, Evaluation Services Division. (2007). Retrieved October 5, 2007, from <http://plans.ednet.ns.ca/>
- Nunavut Department of Education. (2007). Retrieved October 9, 2007, from <http://www.gov.nu.ca/education/eng/index.htm>
- Organization for Economic Co-operation and Development. (2004). *Chile: Reviews of National Policies for Education*. Centre for Co-operation with non-members. Paris.
- Ontario Department of Education. (1968). *Living and learning: The report of the provincial committee on aims and objectives of education in the schools of Ontario*. Toronto, ON: Ontario Department of Education.
- Prince Edward Island Department of Education. (2007). Retrieved October 9, 2007, from <http://www.gov.pe.ca/educ/index.php3?number=1017793&lang=E>
- Putman, J. H., & Weir, G. M. (1925). *Survey of the school system*. Victoria, BC: Charles F. Banfield.
- Qualifications and Curriculum Authority. (2008). Assessment and reporting arrangements. Retrieved April 8, 2008 from: [http://www.qca.org.uk/eara/documents/KS2\\_v07aw-2.pdf](http://www.qca.org.uk/eara/documents/KS2_v07aw-2.pdf)
- Raphael, D. (1993). Accountability and educational philosophy: Paradigms and conflict in Ontario education. *Canadian Journal of Education*, 18, 29-45.
- Rogers, W. T., & Klinger, D. A. (2006, May). *Have the provincial achievement tests programs in Alberta and Ontario promised too much?* Paper presented at the Annual Conference of the Canadian Society of the Study of Education in Canada, Toronto, Ontario,
- Royal Commission on Learning. (1994). *For the love of learning: A report of the Royal commission on Learning* (Vols. 1 & 2). Toronto, ON: Queen's Printer for Ontario.
- Ryerson, E. (1868). *A special report on the systems and state of popular education on the continent of Europe, in the British Isles, and the United States of America, with practical suggestions for the improvement of public instruction in the province of Ontario*. Toronto, ON: Leader Steam Press Est.

- Santrock, J. W., Woloshyn, V. E., Gallagher, T. L., Di Petta, T., & Marini, Z. A. (2004). *Educational Psychology* (1st Canadian ed.). Toronto, ON: McGraw-Hill Ryerson.
- Saskatchewan Ministry of Education. (2007a). Retrieved October 05, 2007, from <http://www.learning.gov.sk.ca/departement-overview/>
- Saskatchewan Ministry of Education. (2007b). *Assessment for Learning Information Package*. Retrieved October 9, 2007, from [http://www.learning.gov.sk.ca/adx.aspx/adxGetMedia.aspx?DocID=801,608,615,200,135,107,81,1,Documents&MediaID=1061&Filename=info\\_pkg\\_2007.pdf](http://www.learning.gov.sk.ca/adx.aspx/adxGetMedia.aspx?DocID=801,608,615,200,135,107,81,1,Documents&MediaID=1061&Filename=info_pkg_2007.pdf)
- Saskatchewan Ministry of Education. (2007c). Retrieved October 05, 2007, from <http://www.sasked.gov.sk.ca/branches/aar/afl/aflreading.shtml>
- 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.
- Volante, L. (2007). Educational quality and accountability in Ontario: Past, present, and future. *Canadian Journal of Educational Administration and Policy*, 58 1-21. Retrieved March 11, 2007 from <http://www.umanitoba.ca/publications/cjeap/pdf%20files/volante.pdf>
- Yukon Department of Education. (2007). Retrieved October 9, 2007, from [www.education.gov.yk.ca/pdf/2006-2007\\_yukon\\_education\\_annual\\_report.pdf](http://www.education.gov.yk.ca/pdf/2006-2007_yukon_education_annual_report.pdf)