

LARGE-SCALE ASSESSMENT OUTCOMES IN BRITISH COLUMBIA, 1876-1999

Helen Raptis & Thomas Fleming

For over 20 years, educators and administrators across North America have heatedly debated the value of large-scale student assessment. Throughout the history of schooling in British Columbia, large-scale student assessment outcomes have traditionally served to inform broader societal goals. Realistically, "assessment of" group learning (as opposed to classroom-based "assessment for" individual learning) will continue as the government's key focus. We also raise several unanswered questions such as the scope of assessment practices that include only numeracy and literacy or the achievement of elementary versus secondary students between 1976 and 1999, questions to move the decades-old debate beyond its current stalemate.

Key words: history, education, testing, achievement

Depuis plus de 20 ans, pédagogues et administrateurs de tous les coins de l'Amérique du Nord débattent avec passion de la valeur des épreuves communes. Tout au long de l'histoire de l'éducation en Colombie-Britannique, les résultats des épreuves communes ont servi à réaliser des objectifs sociaux plus vastes. En fait, l'évaluation des apprentissages d'un groupe (par opposition à l'évaluation de l'apprentissage de chacun au sein d'une classe) va demeurer une priorité pour le gouvernement. Les auteurs soulèvent en outre plusieurs questions laissées jusqu'ici sans réponse, tels l'éventail des pratiques d'évaluation portant seulement sur le calcul et la littératie ou le rendement scolaire des élèves du primaire comparé à celui des élèves du secondaire entre 1976 et 1999, autant de questions qui peuvent faire avancer le débat vieux de plusieurs décennies au-delà de l'impasse actuelle.

Mots clés : histoire, éducation, épreuves, réussite scolaire.

During the past two decades, education systems around the world have experienced unprecedented increases in reform initiatives (Calderhead, 2001; Holt, 2001; Massell, 1998), largely dependent upon large-scale student assessment programs serving as the “vehicle of choice” for promoting accountability within public schools (Earl & Torrance, 2000, p. 114). Some educators and researchers favour the use of large-scale assessments, arguing that these tests can “enhance learning within and across systems of education” (Plomp & Loxley, 1994, p. 176) by enabling policy-makers at all levels to assess their systems and to initiate reforms and policies based on the practices of higher achieving jurisdictions (cf. Beaton, Martin & Mullis, 1997; Schmidt & Burstein, 1993; Schmidt & McKnight, 1998; Stigler & Perry, 1998).

Others have opposed large-scale testing practices, asserting that they have negatively impacted teachers and children in schools. Some researchers argue that increased large-scale testing has narrowed the curriculum by overemphasizing performance in mathematics and reading, while marginalizing all other curricular subjects (Nichols & Berliner, 2005; Pedulla et al., 2003, Volante, 2004). Still others contend that large-scale assessments may encourage people to contemplate change, but insist that rarely do test results increase opportunities to implement reform (Firestone, Mayrowetz, & Fairman, 1998). In 2002, for example, Amrein and Berliner (2002) found no evidence that high-stakes testing led to improved test scores. More recently, Nichols, Glass and Berliner (2006) studied 25 American state-wide accountability systems and discovered no relationship between early accountability pressure and later non-ethnic cohort achievement in math and reading at the fourth and eighth grade levels. In the words of Clarke, Harris, and Reynolds (2004), “despite the dramatic increase in education reform efforts in most countries, their impact upon overall levels of student achievement” has not been “as successful as anticipated” (pp. 2-3).

Curiously, contemporary debates over the value of large-scale student assessment continue to rage as if assessors were unaware that most provincial governments across Canada have administered large-scale student assessments since the late nineteenth century. In British Columbia, large-scale assessment of students’ learning began in 1876 and has continued almost uninterrupted until today. In light of this long-

term commitment to testing, this research study addressed the following questions: a) When and why did the B.C. government become interested in measuring student achievement? b) How has the B.C. government measured student achievement from 1876 to today? c) What purposes have large-scale government assessments served? d) What have we learned about the assessment of student achievement after over a century of large-scale government assessments? That is, are there any discernable long-term trends in student achievement?

METHOD

Although the scholarly literature chronicling the history of educational achievement in British Columbia is sparse, sufficient archival files exist (in the form of assessments, result data, and related documents) from which to reconstruct our testing past. Data for this study were obtained from records in the British Columbia Archives and the B.C. Ministry of Education. Of particular value in this regard are the student achievement data published as part of the *Annual Reports of the Public Schools*. We examined and analyzed these reports for the period 1876 to 1999. In addition, the B.C. Ministry of Education provided access to 16 boxes of files containing tests, results, and reports from 1925 to 1972. These data reside in the B.C. Ministry of Education storage warehouse and have not yet been transferred to the B.C. Archives. Supplementing these reports were tests and results data for the Provincial Learning Assessment Program (PLAP) from 1976 to the late 1990s. These data enabled us to track student performance during the last quarter of the twentieth century. Finally, we interviewed several education ministry officials to obtain information that could not be derived from the documents.

Historical research differs from experimental approaches in that historians are involved in discovering data as opposed to generating it. When data sources are incomplete or unavailable, scholars can construct only a partial picture from existing sources. For various reasons, British Columbia's historical record concerning test outcomes is far from complete. Prior to the establishment of the provincial archives, materials were not always carefully preserved. Likewise, educational materials were not always perceived as valuable and were discarded for reasons of space, collection priorities and, sometimes, indifference. Therefore, the

portrait presented here is, at best, a representative and composite portrait from a somewhat fragmented archive.

Because we sought principally to identify broad achievement trends among the general student population, the following analysis does not describe the impact of testing on teachers and children, a topic discussed in the secondary literature. Nor does it examine the outcomes of specific sub-populations, such as boys compared to girls, urban and rural students, minority, aboriginal, or grade-12 exit or scholarship exams. Analyses of achievement results for these specific sub-populations promise interesting findings; we anticipate undertaking such research to provide some useful comparative data in the future.

STUDENT ASSESSMENT IN BRITISH COLUMBIA, 1876 – 1999

The province of British Columbia began administering large-scale student assessments in 1876. Since that time, there have been four testing “eras” which we have termed the “Departmentals” Era (1876-1924); the Standardized/ I.Q. Testing Era (1925-1972); the Provincial Learning Assessment Program (PLAP) Era, (1976-1999); and the Foundation Skills Assessment (FSA) Era (1999-today). The following discussion describes and summarizes the historical record of these testing eras.

The “ Departmentals” Era, 1876-1924

With the 1876 establishment of Victoria High School, the only institution of higher learning west of Winnipeg, the province needed to implement a means of determining who was, and was not, worthy of admission into this citadel of knowledge (Annual Report of the Public Schools of British Columbia, hereafter, ARPS, 1874, pp. 21-22). Following the lead of other Canadian jurisdictions, particularly Ontario, B.C.’s government implemented its first large-scale examinations, the “departmentals,” also known as the junior matriculation exams. These high school entrance exams – held in March and July – were comprised of problem-solving questions and/or open-ended essay questions in four subjects: arithmetic, spelling, English grammar, and geography. Based on these exams, 31 boys and 23 girls, mostly from Victoria, were admitted to the inaugural high school class. (See Appendix, Table 1) For high school admission, the province ruled that the aggregate of a pupil’s marks “must amount to at

least 60% of the total mark assigned for all the subjects of examination [of which] at least 30% must be obtained in each subject" (ARPS, 1889, p. vii). Because of the importance placed on reading and writing as a prerequisite for all further study, candidates who failed to gain 50 per cent of the questions in the grammar paper were refused admission, a situation that enraged social reformers who viewed this as an unnecessary stumbling block for those students who were not necessarily linguistically inclined.

By the early 1900s, local community leaders, politicians, and educators were beginning to despair over how few students were able to pass the high school entrance exams. In 1876, 160 students wrote the exams and only 43 per cent passed. By 1910, although the number of students who wrote the exams had increased to 1,928, still only 55 per cent passed (or 1,051 students) (ARPS, 1910). Pressured by the employment demands of an expanding economy in 1911, the Education Office prepared different exams for candidates in urban and rural schools to increase pass rates. In 1918, regulations were relaxed to allow pupils in "cities of the first and second class" to be promoted to high school on the recommendations of teachers and principals.¹ All others were required to sit the exams until 1922 when the top 60 per cent of pupils in schools of seven or more divisions (grades) could be promoted by their schools on their principals' assessment (Johnson, 1964, pp. 68-69).

For most youngsters prior to the Great Depression, these high school entrance exams were, as one observer put it, "the only graduating exam they would write, the only certificate of achievement in school that they would get" (Cochrane, 1981, p. 57). One Ontario schoolmistress, upset with government exams and the memory work they encouraged, spoke for many other teachers across the country when she complained:

I taught arithmetic, I taught grammar, I taught spelling, and history, and geography and science, and if I taught Johnny and Mary and Peter, it was purely coincidental. On or about the 24th of May I divided my class into the sheep and the goats. I was sure that the sheep would pass and so they took their lessons from 9:00 to 4:00, but the goats came at 8:00 and we dr-i-l-l-ed spelling and grammar and arithmetic for an hour, and from 4:00 till 5:00 we drilled again. Of course, we called it review, but it was nothing short of cramming and resembled

the stuffing of a Christmas turkey. If I could pour enough facts into their heads and they could hold those facts long enough to regurgitate them on the examination I had achieved my purpose. (cited in Cochrane, 1981, p. 58)

The Education Office generally disapproved of allowing unprepared students to write exams, prompting teachers to “hold back” students until they were reasonably sure they could pass. Of 908 Vancouver pupils “in the senior grade” who were eligible to write the 1914 high school entrance exam, only 320 (or 35 %) did and, of these, 263 (or 82%) passed, a sterling result compared to the province as a whole (ARPS, 1914, p. A42). Although government officials sometimes protested the practice of “holding back” students, as well as the “mental indigestion” and “lack of power to do independent thinking” that resulted from teachers dictating notes and providing students with summaries of textbooks, the importance that government and everyone else placed on the departmentals virtually ensured that teachers would “teach to the test” and only advance their best candidates for examination (ARPS, 1914, p. A42). This essentially “Darwinian” view of schooling – in which only the best were selected for promotion to high school and eventual admission to the professions – ruled the day until the mid-1920s, when both British Columbia’s social context and educational thought shifted profoundly.

What conclusions have we drawn about student achievement during the “Departmentals” Era, from 1876-1924? First, usually fewer than half of the students who wrote the high school entrance exams in the early decades of such testing actually succeeded in passing them. The lowest pass rate was recorded at a mere 41 per cent in 1890-1891 and again in 1894-1895. Over time, however, greater numbers of students wrote the exams, and although increasing numbers of students were promoted into high school over time, this did not mean that students were more successful in passing the exams. Indeed, the highest pass rate (71%) occurred in 1877-1878. In 1922-23, when 93 per cent of the students who wrote the entrance exams were promoted to high school, 36 per cent were promoted on recommendation of their school staffs and only 56 per cent actually passed the government-set exams. More revealing still is the finding that the “average” pass rate over the era was roughly 55 per cent (ARPS, 1922-23). In short, high school departmental exams were

considerably more illustrative of what students “didn’t know” rather than of what they did know.

These findings prompt an interesting question. Of the students who entered high school in British Columbia during this era, what percentage eventually obtained a high school graduation diploma? A study of this nature would no doubt reveal some interesting findings about student success rates over a considerable length of time.

The Standardized/ I.Q. Testing Era, 1925-1972

In the mid-1920s, as a modernizing provincial economy and new educational ideas promoted greater access and higher levels of learning for all the province’s children, the departmental exams fell out of favour. By the early twentieth century, standardized intelligence (I.Q.) testing was being heralded throughout the Western world as a more objective instrument for assigning student grades – despite early opposition (Cubberley, 1934). According to American historian Gerard Giordano (2005), educators initially embraced the new, more “scientific” approach as a means of elevating “the lackluster status of their occupation” (p. xvi). For administrators, it provided a means of ensuring that teachers followed the curriculum at all grade levels (Giordano, 2005, p. 26).

Most instrumental in bringing British Columbia’s open-ended, subject-focused “departmental” exams to a close was the 1925 publication of Harold Putman and George Weir’s (1925) mammoth *Survey of the School System*, a report to government that advocated revolutionizing the structures used to direct and assess provincial schools. Dismissing the high school entrance exams as an anachronism, or a “Moloch,” to whom students were needlessly sacrificed, Putman and Weir recommended that the province adopt a regular program of standardized testing to obtain a more accurate assessment of students’ scholastic abilities. In their view, the “departmentals” were an “outgrowth of an educational system essentially Prussian, rather than British, in spirit” (Putman & Weir, 1925, p. 259). The commissioners argued that if the “traditional written examinations were an accurate test of intelligence or educational achievement, a strong defense for retaining them as an integral part of the provincial school system could be offered” (p. 260). As evidence this was not so, they pointed to wild

fluctuations in annual failure rates and concluded that there was “no valid reason why 75% of the pupils who enter grade one should not pass directly from grade six without the necessity of undergoing an examination ordeal” (cited in Johnson, 1964, p. 106). The Commissioners also noted the high costs of the examination system, stating that in 1924, “the total cost of the examination of teachers and high school entrance classes was \$25,373,” with only \$10,899.33 being recouped from the candidates (Putman & Weir, 1925, p. 261).

During the mid-1920s, the social moment was right for discussion about educational change. The enormous urbanization and industrialization that marked the province’s growth in the early decades of the twentieth century made new demands on schools to prepare ever-increasing numbers of students for trade and commerce. As cities and towns swelled with new buildings, streets, and immigrants, few of the myriad jobs associated with this new economy seemingly required the expertise in Latin, British history, or other classical subjects that a formal high school education bespoke. British Columbia had entered a new age characterized by energy, ambition, and industry (Reksten, 2001). This new era called for a new view of schooling and a new curriculum that would more practically prepare as many students of the general population as possible with knowledge and skills that were relevant to the realities of learners’ lives (Putman & Weir, 1925, p. 82).

If school goals were to meet the needs of changing times, then it was necessary to change the instruments used to assess student performance. As one of Canada’s first surveys of a provincial education system, the Putman and Weir Commission proposed that students be promoted by subject on the recommendation of principals and their staffs – a practice already well underway in some schools since 1921 – and that the Department of Education’s junior matriculation exams be replaced by a high school accreditation system. Featuring prominently in this new and progressive assessment scheme was the application of intelligence (I.Q.) tests, which were gaining support throughout Western education systems. Putman and Weir suggested that the provincial school inspectorate could still be charged with assessing the overall effectiveness and efficiency of the system through school visits, as it had in the past, but such visits could be augmented by the use of I.Q. and

other standardized tests (Putman & Weir, 1925, pp. 239-45). As part of the Commission, University of Toronto's Dr. Peter Sandiford was hired to assess B.C. students' mental capacities which were reported by ethnicity, a reflection of nativist sentiments prevailing at the time. Satisfaction with Sandiford's work led the Commissioners to recommend I.Q. tests as the means by which educators could compare students' intellectual potential with their subject-specific achievement and rates of progress, the ultimate goal being to bridge the gap between students' predicted capability and actual, subject-specific achievement. Mirroring educational developments in other jurisdictions, provincial policy-makers hoped that the new tests would help to narrow the capacity-achievement gap and would eliminate the "age-grade retardation" that resulted from "holding students back" until they were considered capable of passing the matriculation exams. As historian Brian Simon (1974) notes, British policymakers at the time were also eager to implement assessments "which could allow for disadvantages attendant on poor home background, poor schools or teaching, and this ruled out testing mere attainment in favour of seeking to gauge 'capacity and promise'" (p. 235).²

Records from the I.Q. testing era suggest that, in addition to I.Q. testing, British Columbia's education department introduced standardized, multiple-choice style tests in the 1920s and used them extensively for multiple purposes until 1972. Tests were developed in many areas of the curriculum, including mathematics, geography, handwriting, languages, physical education, science, spelling and commerce. First, government officials compared outcomes on various standardized subject-area tests with students' mental and chronological age as measured by standardized I.Q. tests. Gaps between mental ability and students' subject area progress could thereby be identified and responsibility could be attributed to various factors, including poor teaching, poor placement, or other organizational factors that encouraged the worrisome and costly problem of student "age-grade retardation." Government officials then alerted teachers to students who were age-grade "retarded" and encouraged school staffs to provide such learners with appropriate remediation. In addition, government also communicated provincial and district averages so that individual school

staffs would know where they stood compared to other schools. During this era, student outcome data were also used to compare British Columbia's student achievement to that of American learners for whom the standardized exams had originally been normed. Finally, such data enabled officials to compare rural averages with urban results and, from time to time, to compare the relative achievement of boys and girls.

Between 1925 and 1972, 34 different types of I.Q. test forms were administered to students across grades one to eleven in British Columbia (see Appendix, Tables 2 to 11). In 1948 alone, public schools enrolled 137,000 students, of which 40 per cent (or 56,000 students) wrote I.Q. tests, 55 per cent (77,000) wrote subject-specific exams and only 5 per cent (7,000) were spared the testing ordeal (Figures calculated from Conway, 1949, p. 64).

What conclusions can be drawn about student achievement testing in British Columbia from 1925 to 1972? First, the multiple-choice format of I.Q. type testing instruments used during this era differed markedly from the format of the "departmentals," which generally consisted of open-ended, essay-type and problem-solving questions. Second, during the half century in which these tests were popular, youngsters in British Columbia tended to outperform their American counterparts in most examination subjects except language arts and languages. Third, some assessment instruments obviously occupied a privileged place in the hierarchy of tests. For example, during this era far more I.Q. type tests were administered than subject-specific tests. This practice was likely due to the desire of government officials to identify students whose subject area scores were not in line with their "potential" capabilities, as determined by I.Q. tests. One of the main aims of policymakers who advocated the use of I.Q. testing was to signal age-grade retardation to teachers and inspectors who were then expected to provide remediation so that as many learners as possible could succeed and contribute to the province's economic development.

Appendix Tables 4, 6, 7, 9, and 11 also illustrate that many curricular subjects (including history, science, languages, physical education and commerce) were clearly under-assessed during this era, and that some subject areas were completely un-surveyed (such as art, music, industrial arts). These findings prompt an intriguing and important question: Why

were certain types of tests and certain subject areas perennially considered to be more important for government to test and measure? Likewise, who determined such priorities in assessment and upon what criteria?

The Provincial Learning Assessment Program (PLAP) Era, 1975-1999.

By 1972, the I.Q. assessments so highly favoured during the previous era had fallen from grace, largely as a result of changing social and political views.³ Although criticism had dogged I.Q. testing since its inception, 1960s civil rights protests over the importance of social equality moved the topic of educational testing from the confines of “academic debate” to the “center of public controversy” (Giordano, 2005, p. 146). With the 1972 election of the New Democratic Party in British Columbia, education minister, Eileen Dailly, phased out the use of I.Q. assessments and entrusted local authorities with responsibility for curriculum development and assessment initiatives (ARPS, 1972, pp. D33-35). However, after a three-year testing hiatus, marked by considerable public and professional opposition, Dailly proposed in 1975 a new testing initiative that differed appreciably from what had constituted testing over the previous century. Instead of focusing on student achievement, the new province-wide assessment program initially took aim at two larger targets – informing the public about global achievement and assisting curriculum committees, as well as teachers, in improving courses in reading, writing, literature and oral communication. Curriculum design and professional development for teachers, along with providing ideas for research and resource allocation, had emerged as the government’s new assessment focus (Dept. Assessing Language Arts, May 1975, p. 1).

Gone was the old-fashioned emphasis on large-scale measures of student achievement or the diagnostic assessment of individual students. A new era of “program evaluation” in education and, indeed, in other ministries of government was dawning. For a century, provincial education officials had focused on measuring individual students’ inadequacies. Now educational government was taking a page from corporate practices of the day by assessing the adequacies of its own programs and support services rather than that of students in the system.

Following a change in government that saw the restoration of the Social Credit Party, the new education minister, Dr. Patrick McGeer announced the ambitious “Provincial Learning Assessment Program” (PLAP) in October 1975 that was also generally “programmatically” and “public” in emphasis. The new, multi-purpose program that had initially been conceived under Dailly’s tenure, would, *inter alia*, evaluate student progress over time, account to the public for the strengths and weaknesses of the K-12 system, provide individual districts and schools with performance data, support curriculum change and educational research, and provide management information for resource allocation (British Columbia Ministry of Education Records, Box 307-04-145/89-1839-25).

From 1976 to 1999 as part of the PLAP, the provincial government tested various areas of the curriculum on what was initially intended to be an alternating four-year cycle. The 1976 language arts assessment alone produced over 100 recommendations relating to curriculum development and how reading and writing should be taught. “The fundamental purpose of the program,” the Ministry advised, “is to facilitate educational decision-making in areas such as curriculum development, fiscal management, teacher education and research” (Dept. Assessing Language Arts, May 1975, p. 1). Subjects examined during this era included: language arts, science, mathematics, social studies, physical education, and French (see Appendix, Tables 12 to 17).

With new assessment purposes came new formats and new approaches. Although the multiple-choice format was retained for most assessments, some test questions – in modern languages and language arts to name two – became more open-ended and essay-type. In addition, unlike exams of the previous era, students’ PLAP outcomes were not assessed relative to students’ overall intelligence scores, nor were they compared to norms from other learners’ scores from other districts or other countries. At the exam preparation stage, government officials enlisted the services of various individuals from the field (generally teachers and professors) to constitute “Interpretation Panels.” Based on the provincial curriculum, these panels determined what knowledge and skill levels could be expected of learners at various grade levels. The panels could then evaluate students’ average test scores (or

mean percentage correct) relative to standards they had set.

From our reading of the PLAP era data, we have decided that the most illustrative way to present these data is within their subject areas, comparing the results (both mean per cent correct and interpretation panel judgments) across grade levels. Achievement outcomes from the Provincial Learning Assessment Program, (Appendix, Tables 12 to 17), reveal several interesting trends. First, over the years from 1976 to 1999, British Columbia's elementary learners proved generally more successful at meeting the interpretation panels' expectations than their secondary counterparts, except in the areas of social studies and physical education (see Appendix tables 12 to 17, judgment column). Why this appears to be the case is a question that clearly merits further research.

Another broad trend revealed by the data is that some subject areas have been assessed far more frequently than others. For example, during the PLAP era more language arts assessments were administered than assessments in any other subject area, prompting the question as to why this was the case. Likewise, as witnessed during the I.Q. testing era, some subject areas were under-represented (such as social studies, see Table 15) and several (such as fine arts or industrial arts) were not assessed at all. Additional research might reveal why some subject areas held more prominence during the era of the Provincial Learning Assessment Program.

The Foundation Skills Assessment (FSA), 1999-present

During the latter half of the 1990s, parental pressure, as well as requests from various district administrators, caused government to reconsider the adequacy of the PLAP (J. Mussio, personal communication, June 13, 2006).⁴ In 1999, the provincial government added to its regimen a new testing program, known as the Foundation Skills Assessment (FSA). Once per year, students in grades four, seven, and ten write a standardized, province-wide examination in the areas of literacy (reading comprehension and writing) and numeracy. According to the B.C. Ministry of Education website, this annual ritual provides a snapshot of students' academic skills (see Foundation Skills Assessment website, <http://www.bced.gov.bc.ca/assessment/fsa>). Families receive their individual child's standings reported with the following

descriptors: Meeting Expectations; Exceeding Expectations; Not Yet Meeting Expectations. School, district, and province-wide scores are reported on the Ministry website. And although the FSAs were never meant to supplant the PLAP, provincial authorities have not administered any PLAP exams since 1999, reflecting their belief that the FSA is a more reliable indicator of an individual student's ability because it measures fewer skills in more depth, with more items. With the introduction of the FSA, breadth of curricular coverage was traded for more information about individual cases. (Barry Anderson, personal communication, June 12, 2006).⁵

Although the FSAs are too recent to determine broad trends in overall student achievement, their format is notable when considered in historical perspective. Beginning in 1876, the testing programs have generally focused on assessing students' subject-specific performance. Today's FSA tests, however, focus on two areas: literacy and numeracy, which are considered to be "foundational" to all other learning. The reason for this narrowed focus is certainly worthy of future inquiry.

CONCLUSION

The results of this research are significant for several reasons. First, this historical review of large-scale provincial testing adds a new dimension to a contemporary debate that has been largely ahistorical. Although the stand-off between the critics (cf. Nichols, Glass, & Berliner 2006; Volante, 2004) and advocates (cf. Schmidt & McKnight, 1998; Stigler & Perry, 1998) of large-scale testing drags on, this historical research study sheds new light on the purposes and outcomes of government-initiated, large-scale testing over time. Indeed, over the past 130 years, large-scale government assessments have generally acted as accountability mechanisms serving broad societal goals. In this regard, government has been mainly interested in the *assessment of* group learning and not in the *assessment for* individual learning – the latter has been generally left to the classroom teacher. Furthermore, this research clearly demonstrates that large-scale assessment is a well-entrenched and traditional government practice in the province of British Columbia and, as such, forms part of what U.S. historians David Tyack and Larry Cuban (1995) refer to as "the grammar of schooling," (p. 85) that is to say, part of the

core organizational structures, rules, and practices that endure over time. Large-scale testing is anything but new. It pre-dates by a century ideological shifts within government that have become associated with testing and accountability in recent popular and scholarly literatures. The historical record should serve to inform recent critics of large-scale testing that this enduring “grammar” has long been embodied in the foundations of public schooling and, therefore, prevents anti-testing reforms.

Despite the continuity of large-scale assessment, however, the means of assessing student achievement have been anything but static. Instruments, purposes, and reporting formats have changed significantly over this time and four discernible testing epochs can be identified between 1876 and 1999: the “departmentals” era from 1876 to 1924; the standardized/ I.Q. testing era from 1925 to 1972; the Provincial Learning Assessment Program era from 1976 to 1999; and the Foundation Skills Assessments from 1999 to today.

These changes have reflected broader societal shifts over time. During the first era, the favoured approach to testing was on open-ended, essay-type or on problem-solving exam questions, requiring learners to use subject matter knowledge to make connections, develop generalizations, and support solutions. Although initially, candidates who failed to gain 50 per cent of the questions in the grammar paper were refused high school admission, this practice was dropped by the 1890s with the broadening of the list of examinable subjects. Furthermore, few students were permitted to write the exams and of those who wrote, not many succeeded. These approaches were appropriate for their time, as turn-of-the-twentieth-century Canadian policy-makers were preoccupied with nation-building when most citizens supported themselves through farming. Government used assessment instruments as a means of selecting the gifted few learners who would continue their studies and enter the traditional professions of teaching, law, and medicine.

During the second era, from 1925 to 1972, the most popular assessment instruments were multiple-choice, IQ-type tests that facilitated the marking of exams for vastly increased numbers of students who were now being assessed due to newly imposed

mandatory school attendance laws. Using standardized tests, government officials predicted students' intelligence at various points of their education and, by comparing I.Q. scores to students' subject-specific test results, they determined situations in which students were "age-grade" retarded. Armed with this information, teachers then provided remediation activities for learners who were not "living up to their potential." I.Q. and subject-specific scores also helped efficiency-conscious administrators to direct pupils toward the career choice that best fit their capabilities, thus facilitating students' transitions from school to work.⁶ This test-type reflected societal concerns of the time about Canada's capacity to compete in an increasingly industrial world (Dunn, 1979, p. 236). Overall, students in British Columbia tended to do well relative to American norms, with the exception being in the language arts and languages.

Test types during the third era of assessment (PLAP) tended to be multiple-choice, but over the years, more open-ended formats were also introduced. These tests reflected government's interest in determining how well students performed as a whole to inform revisions to provincial programs and curricula. Results from the tests administered during this era were never revealed directly to students or teachers. Still reeling from the protests that arose through the human and civil rights movements of the 1960s and 1970s, governments proceeded cautiously and reported only aggregated data that would assist government to revise its own practices. Interestingly, during the PLAP era, elementary students tended to perform closer to expert panels' standards than did secondary students, except in the areas of social studies and physical education. This finding is certainly worthy of future inquiry, particularly in light of recent World Bank allegations that both secondary and tertiary education have been neglected over the past few decades due to an emphasis on primary education (World Bank, 2002, 2005).

Because the province continues to administer Foundation Skills Assessments (FSAs) today, it is premature to assess any long-term trends arising from the FSAs. However, the current testing era attributes prominence to numeracy and literacy as "foundational" to all other learning. Why numeracy and literacy have been attributed such prominence is certainly worthy of future scholarly inquiry. It is also

curious to note that the PLAP has not been administered since the introduction of the FSAs in 1999, despite the fact that the FSA was not meant to supplant the PLAP but to augment it.

From these analyses, it is also apparent that certain kinds of knowledge have dominated the testing agenda from time to time. From 1925 to 1972, the majority of government-administered assessments were I.Q. tests, with mathematics exams following in prominence. Since 1976, on the other hand, language arts and literacy assessments have dominated the provincial testing agenda rendering these skills as a kind of proxy measure for all other learning – perhaps not unlike the I.Q. scores of yesteryear. Today's FSAs measure literacy and numeracy implying that these are foundational to learning in other un-tested subject areas. The assumption that the FSAs are valid "proxy" measures of all learning is an assumption that has neither been challenged nor debated but is certainly worthy of future scholarly inquiry.⁷

Although these broad trends are important in themselves, they prompt further questions about the performance of constituent groups that comprise the general provincial population. What assessment trends might be identified or confirmed by examining specific sub-populations such as minorities, girls compared with boys, urban learners compared to those in rural settings, or, for that matter, grade 12 learners' results on exit and scholarship exams? How can we explain historically why provincial officials have preferred to assess certain subject areas over others? All these questions are worthy of consideration in any agenda for further research.

Predicting the future is fraught with danger, particularly in human endeavors such as public policy-making. Nevertheless, we would be remiss in concluding a study on what has been learned from 130 years of large-scale testing without a brief discussion of possible future directions. This research has indicated that, historically, testing developments in British Columbia have been similar to initiatives elsewhere in the western world (cf. Simon, 1974, in Britain and Giordano, 2005, in the U.S.A.). With more than 35 American states now having introduced more authentic, performance-based measures, it would be tempting to predict that Canadian provinces will soon follow suit and relegate paper and pencil measures to the dustbin of passé pedagogy.

Indeed, various researchers have begun to advocate for reform of large-scale assessments in order to support classroom learning (cf. Chudowsky & Pellegino, 2003; Ungerleider, 2007; Volante, 2005). Nevertheless, the history of testing in British Columbia suggests a different future.

Over the past 130 years, testing purposes have varied considerably, reflecting different government concerns, from nation building to maintaining international competition in industry, to informing curricular reform. What has remained relatively constant over time has been the use of paper and pencil instruments focusing largely on basic academic skills. A second area of constancy has been the divide between government-initiated testing and teacher-designed, classroom-based testing. Although this research has not determined the reasons prompting such constancy, paper and pencil assessments of basic skills have most likely prevailed due to their low costs and ease of administration when compared with more authentic, performance-based approaches. Indeed, we were able to discern only six occasions when more performance-based approaches were used – and then abandoned – over the entire 130-year period under examination. Based on past trends, a more likely future scenario is a continued focus by government on paper and pencil approaches to assess foundational academic skills. Classroom-based, teacher-generated tests, on the other hand, will likely continue to diversify, moving toward more authentic, performance-based approaches.

ACKNOWLEDGEMENTS

The British Columbia Ministry of Education supported this research. In particular, we would like to thank Barry Anderson, Donna Coward, Gerald Morton, and Nancy Walt. We are enormously grateful to Dr. James Gaskell, formerly of the B.C. Ministry of Education who provided us with his complete set of PLAP instruments and achievement reports. We presented an earlier version of this paper at the 2006 Canadian Society for Studies in Education conference in Toronto. The authors also thank Charles Ungerleider, Ruth Childs, and four anonymous CJE reviewers for their helpful feedback.

NOTES

¹ Cities designated as first class had populations greater than 10,000 inhabitants. Second class cities had populations greater than 5,000 inhabitants but fewer than 10,000.

² For another perspective on the goals of I.Q. testing, see C. Ungerleider (1987).

³ I.Q. tests are still used in British Columbia to determine eligibility for services for students with special needs.

⁴ Dr. Mussio served as the director of the Ministry of Education Assessment Branch at the time when the FSA was introduced.

⁵ Dr. Anderson is the most senior Ministry of Education employee and has served as lead director of the Assessment Branch and the Information Branch.

⁶ For discussions of the impact of the mental hygiene movement on education, see Cohen (1983) and Thomson (2006).

⁷ Literacy and numeracy have not always had such prominence. During the departmentals era, for example, government assessed subject areas individually, reflecting an assumption that each discipline had different demands that could be mastered by the willing student. This view was somewhat akin to Gardner's current notion of multiple intelligences. During the I.Q. era, reformers downplayed this view and promoted intelligence as something unitary, innate, and immutable to be represented by one omnibus (I.Q.) score. Although this notion fell out of favour by the 1970s, we believe that it has resurfaced in the form of omnibus literacy and numeracy scores assessed by the FSAs.

REFERENCES

Primary Sources:

British Columbia Archives, GR 0452, Box 1, File 7, Letter 366 and Box 2, File 3, Letter 71.

Government of British Columbia, *Annual Reports of the Public Schools, 1876-1999*, Victoria, BC: Government of British Columbia.

Government of British Columbia, B.C. Ministry of Education, Records: Boxes: 307-04-145/ 89-1839-25; 307-04-143/ 89-1839-23; 307-04-146/89-1839-26; 304-04-145/89-1839-25; 307-04-143/89-1839-26; 307-04-140/ 89-1839-20; 307-04-142/89-1839-22; 307-04-143/ 89-1839-25; 307-04-1461/ 89-1839-26; 307-04-1451/ 89-1839-25. BC

Government of British Columbia, Provincial Learning Assessment Examinations and Technical Reports, 1976-1998. (From the Collection of Dr. Jim Gaskell).

Secondary Sources:

- Amrein, A. L., & Berliner, D. C. (2002). High-Stakes testing, uncertainty and student learning, *Education Policy Analysis Archives*, 10(18). Retrieved February 19, 2007, from <http://epaa.asu.edu/epaa/v10n18>
- Beaton, A. E., Martin, M. O., & Mullis, I. V. S. (1997). Providing data for educational policy in an international context: The Third International Mathematics and Science Study (TIMSS). *European Journal of Psychological Assessment* 13, 49-58.
- Calderhead, J. (2001). International experiences of teaching reform. In V. Richardson (Ed.), *Handbook of research on teaching* (pp 777 – 800). American Educational Research Association: Washington, DC.
- Chudowsky, N. & Pellegino, J.W. (2003). Large-scale assessment that supports learning: What will it take? *Theory into Practice*, 42(1), 75-83.
- Clarke, P., Harris, A., & Reynolds, D. (2004, April). Challenging the challenged: Developing an improvement programme for schools facing extremely challenging circumstances. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, California.
- Cochrane, J. (1981). *The one-room school in Canada*, Toronto: Fitzhenry & Whiteside.
- Cohen, S. (1983). The mental hygiene movement, The development of personality and the school: The medicalization of American education. *History of Education Quarterly*, 123-149
- Conway, C.B. (1949). Research and testing in British Columbia *Canadian Education*, IV(3): 59-79.
- Cubberley, E. (1934). *Public education in the United States: A study and interpretation of American educational history* (rev. ed.). Boston: Houghton Mifflin.
- "Dept. Assessing Language Arts," (1975, May) *Education Today*, p.1.
- Dunn, T. (1979). Teaching the meaning of work: Vocational education in British Columbia, 1900-1929. In D. C. Jones, N. M. Sheehan, & R. Stamp (Eds.), *Shaping the schools of the Canadian West* (pp. 236-256). Calgary, AB: Detselig.
- Earl, L., & Torrance, N. (2000). Embedding accountability and improvement into large-scale assessment: What difference does it make? *Peabody Journal of Education*, 75, 114-141.

- Firestone, W., Mayrowetz, D., Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2), 95-113.
- Giordano, G. (2005). *How testing came to dominate American schools: The history of educational assessment*. New York: Peter Lang.
- Holt, M. (2001). Performance pay for teachers: The standards movement's last stand? *Phi Delta Kappan*, 83, 312- 317.
- Johnson, F. H. (1964). *A history of public education in British Columbia*. Vancouver, BC: University of British Columbia Press.
- Massell, D. (1998, July). State strategies for building local capacity: Addressing the needs of standards-based reform. *CPRE Policy Briefs*, University of Pennsylvania.
- Nichols, S. L., & Berliner, D. C. (2005). *The inevitable corruption of indicators and educators through high-stakes testing*. Tempe AZ: Education Policy Studies Laboratory.
- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006, January 4,). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14(1). Retrieved February 19, 2007, from <http://epaa.asu.edu/epaa/v14n1/>
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003, March). Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers. Boston, MA: National Board on Educational Testing and Public Policy. Retrieved February 19, 2007, from <http://www.bc.edu/research/nbetpp/statements/nbr2.pdf>
- Plomp, T. and Loxley, W. (1994). The Contribution of International Comparative Assessment to Curriculum Reform. In OECD (Ed.) *The Curriculum Redefined: Schooling for the Twenty-First Century*, Paris: OECD, 173-186.
- Putman, J. H., & Weir, G. M. (1925). *Survey of the school system*. Victoria, B.C.: Charles F. Banfield.
- Reksten, T. (2001). *An illustrated history of British Columbia*. Vancouver: Douglas & McIntyre.
- Schmidt, W. H., & Burstein, L. (1993). Concomitants of growth in mathematics achievement during the populationA school year. In L. Burstein (Ed.), *The*

- IEA study of Mathematics III: Student growth and classroom Processes* (pp. 309 – 327). Oxford, UK: Pergamon Press.
- Schmidt, W. H. and McKnight, C. C. (1998, December). What Can We Really Learn from TIMSS? *Science* 282,1830-31.
- Simon, B. (1974). *The politics of educational reform, 1920-1940*. London, UK: Lawrence & Wishart.
- Stigler, J. W., & Perry, M. (1998). Cross cultural studies of mathematics teaching and learning: Recent findings and new directions. In D. A. Grouws (Ed.), *Perspectives on research on effective mathematics teaching* (pp. 194 – 223). Reston, VA: National Council of Mathematics Teachers.
- Thomson, G. (2006). 'Through no fault of their own': Josephine Dauphinee and the "subnormal" pupils of the Vancouver school system, 1911-1941. *Historical Studies in Education*, (Spring), 51-73.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- Ungerleider, C. (1987). Testing: Fine tuning the politics of inequality. In L. L. Stewin & S. J. H. McCann (Eds.), *Contemporary educational issues: The Canadian mosaic* (pp. 126-136). Toronto: Copp Clark Pitman.
- Ungerleider, C. (2007). Reflections on the use of large-scale student assessment for improving student success, *Canadian Journal of Education*, 29(3):873-883.
- Volante, L. (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, 35. Retrieved February 19, 2006, from <http://www.umanitoba.ca/publications/cjeap/articles/volante.html>
- Volante, L. (2005). Accountability, student assessment, and the need for a comprehensive approach, *International Electronic Journal for Leadership in Learning*, 9(6). Retrieved February 21, 2007 from <http://www.ucalgary.ca/~iejll/volume9/volante.htm>
- World Bank (2002) *Constructing knowledge societies: New challenges for tertiary education*. Washington, DC: World Bank.
- World Bank (2005. October 9,) *A world of opportunity*, 376 (8843), Washington, DC: World Bank.

APPENDIX

“Departmentals” Results: Data for this table were derived from the *Annual Reports of the Public Schools of British Columbia*, for the years 1876 to 1924.

Table 1. Student Achievement during the “Departmentals” Era, 1876-1924¹

Date	Number of students who wrote H.S. entrance exam	Number & percentage who passed H.S. exam	Number & percentage recommended for high school	Grand Total & percentage who went on to high school
1875-1876	160	68 (43%)		
1876-1877	101	50 (50%)		
1877-1878	129	106 (71%)		
1878-1879	34	20 (59%)		
1879-1882	Not recorded in annual reports	Not recorded		
1882-1883	Not recorded	40		
1883-1884	Not recorded	56		
1884-1885	Not recorded	71		
1885-1886	Not recorded	80		
1886-1887	Not recorded	78		
1887-1888	Not recorded	99		
1888-1889	225	122 (54%)		
1889-1890	282	143 (54%)		
1890-1891	279	114 (41%)		
1891-1892	378	203 (54%)		
1892-1893	402	228 (57%)		
1893-1894	523	257 (49%)		
1894-1895	596	245 (41%)		
1895-1896	485	264 (54%)		
1896-1897	516	287 (56%)		
1897-1898	620	290 (47%)		
1898-1899	686	400 (58%)		
1899-1900	Not recorded	238		
1900-1901	Not recorded	610		
1901-1902	873	503 (58%)		
1902-1903	1101	496 (45%)		
1903-1904	1056	574 (54%)		
1904-1905	1068	604 (57%)		

1905-1906	1217	766 (63%)		
1906-1907	1330	652 (49%)		
1907-1908	1383	965 (70%)		
1908-1909	1675	971 (58%)		
1909-1910	1928	1051 (55%)		
1910-1911	1782	1029 (58%)		
1911-1912	2189	1484 (68%)		
1912-1913	2440	1420 (58%)		
1913-1914	3124	2062 (66%)		
1914-1915	3366	2394 (71%)		
1915-1916	3440	2450 (71%)		
1916-1921	No reports			
1921-1922	4231	2168 (51%)	1417 (33%)	3585 (84%)
1922-1923	4939	2788 (56%)	1791 (36%)	4579 (93%)
1923-1924	Not recorded	2887	1895	4782
1924-1925	Not recorded	3568	2130	5698

Standardized/ I.Q. Testing Era Results: Data for Tables 2 to 11 were derived from B.C. Ministry of Education Records, Boxes: 307-04-145/ 89-1839-25; 307-04-143/ 89-1839-23; 307-04-146/89-1839-26; 304-04-145/89-1839-25; 307-04-143/89-1839-26; 307-04-140/ 89-1839-20; 307-04-142/89-1839-22; 307-04-143/ 89-1839-25; 307-04-1461/ 89-1839-26; 307-04-1451/ 89-1839-25. BC Archival records drawn upon in this paper are from GR 0452, Box 1, File 7, Letter 366 and Box 2, File 3, Letter 71.

We thank Barry Anderson, Donna Coward, Deanna Miller, Gerald Morton, and Nancy Walt for access.

Table 2: *I.Q. Outcomes During Standardized/I.Q. Testing Era, 1925-1972²*

TEST & DATE	GRADE(S)	BC NORM	US NORM
1925 - Pintner-Cunningham Primary Mental Test	1	31.4*	28.0
	2	41.9*	41.0
1925 - National I.Q. Test	3	53.2*	43.0
	4	76.8*	66.0
	5	93.9*	86.0
	6	105.3*	104.0
	7	122.7*	120.0
	8	141.1*	131.0
1925 - B.C.	9	75.0	85.0*

Intelligence Test			
	10	87.4	109.0*
	11	100.9	120.0*
1945 Progressive Achievement Test B	10	12.2*	10.6
1946 Otis	8	Mean Mental Age 14-2.1*	US not reported
1946 – Reading & Scholastic Aptitude	11	105.0* 9 mo. ahead of US	US not reported
1947 – Henmon Nelson	8	105.2*	US not reported
1948 – Metropolitan Achievement Test A	7	BC not reported BC 4.5 mo. ahead of US*	US not reported
1948 – Metropolitan Achievement Test T	7	BC above US*	US not reported
	12	99.6 BC below US	US not reported*
1948 – Henmon-Nelson B	8-12	BC not reported BC below US	US not reported*
1949 – Iowa Work Study Skills O	8	BC not reported BC below US	US not reported*
1958/9 Metropolitan Achievement Test T	8-12	BC not reported BC 2-14 months ahead of US*	US not reported
1958 Metropolitan Achievement Test A	7	BC not reported BC 1-15 months ahead of US*	US not reported
1962 Henmon-Nelson	8	110.8*	US not reported
1965 Iowa Workstudy Skills O	8	BC not reported BC above US*	US not reported

*Thirty-four general I. Q. assessments were administered. British Columbia students outperformed United States' norms on 22 out of 34 tests.

Table 3: Arithmetic Outcomes during Standardized/I.Q. Testing Era, 1925-1972

TEST & DATE	GRADE(S)	BC NORM	US NORM
1925 - BC Test of Fundamentals in Arithmetic	3	26.72*	Not reported
	4	39.69*	Not reported
	5	35.44*	Not reported
	6	55.12*	Not reported
	7	62.94*	Not reported
	8	70.76*	Not reported
1925- Woody-McCall Test of Arithmetic	3	15.5*	7.3
	4	17.4*	13.5
	5	21.4*	18.2
	6	25.2*	22.7
	7	26.9*	26.1
	8	30.3*	28.0
1925 – Holz Algebra Test	10	24.36*	23.6
1945 Progressive Achievement Battery	8	2.7% below US norms	Not reported*
1947 – Arithmetic Computation and Reasoning	6	Not reported	Not reported*
1964 – Stanford Arithmetic Test L	9	Reasoning: 1.8 years above US* Computation: 1.1 years above US*	Not reported

*Sixteen arithmetic assessments were administered and British Columbia students outperformed United States' norms on 14 out of 16 tests.

Table 4: Geography, History & Social Studies Outcomes During Standardized / I.Q. Testing Era, 1925-1972.

TEST & DATE	GRADE(S)	BC NORM	US NORM
1925 - BC Geography Test	7	Not compared to US	-
	8	Not compared to US	-
	9	Not compared to US	-
	10	Not compared to US	-
	11	Not compared to US	-
	12	Not compared to US	-
1925- BC Test in History	7	Not compared to US	-
	8	Not compared to US	-
	9	Not compared to US	-
	10	Not compared to US	-
	11	Not compared to US	-
	12	Not compared to US	-
1925- BC Test of Geography	6	Not compared to US	-
	7	Not compared to US	-
	8	Not compared to US	-
1925 – BC Test in History	8	Not compared to US	-
1953 – BC Test of Social Studies	10	Not reported BC outperforms US*	Not reported

*Nine assessments were administered in geography; 7 in history and one in the combined area of social studies. British Columbia students outperformed United States' norms on one of 17 tests.

Table 5: Handwriting Outcomes During Standardized/I.Q. Testing Era, 1925-1972.

TEST & DATE	GRADE(S)	BC NORM	US NORM
1925 – BC Handwriting Test	2	42.0*	38.0
	3	49.0*	42.0
	4	53.0*	46.0
	5	59.0*	50.0
	6	62.0*	54.0
	7	63.0*	58.0
	8	66.0*	62.0
	1949 – BC Handwriting Test	5	39.0
7		47.0	58.0*
1952 – BC Handwriting Test	7	48.0	58.0*
	8	48.0	62.0*

*Eleven handwriting assessments were administered. British Columbia students outperformed United States' norms on seven out of 11 tests.

Table 6: Languages Outcomes During Standardized/I.Q. Testing Era, 1925-1972

TEST & DATE	GRADE(S)	BC NORM	US NORM
1925 - Henmon Latin Test	10	85.4	104.00*
	11	128.3	132.50*
1925 – Henmon French Test	10	86.49	91.0*
	11	135.75	149.50*
1950 – BC Spanish Test	10	Not reported	Not reported*
	11	Not reported	Not reported*
1950 – BC German Test	10	Not reported	Not reported*
	11	Not reported	Not reported*

*Eight languages assessments were administered. British Columbia students failed to outperform United States' norms on any of the eight tests.

Table 7: Physical Education Outcomes During Standardized/I.Q. Testing Era, 1925-1972

TEST & DATE	GRADE(S)	BC NORM	US NORM
1955 – BC Test of Physical Education	10	Results not reported	

*One physical education assessment was administered and no comparison was made with United States' norms.

Table 8: Language Arts Outcomes During Standardized/I.Q. Testing Era, 1925-1972

TEST & DATE	GRADE(S)	BC NORM	US NORM
1925 – Thorndike-McCall Silent Reading Scale	6	49.07*	47.9
1925 - Ayres-Burgess Test in Silent Reading	3	42.64	50.0*
	4	43.77	50.0*
	5	50.29*	50.0
	7	53.3*	52.8
	8	58.2*	58.1
1945 – Progressive Achievement Battery	10	Not reported	Not reported*
1946- Reading Comprehension Test	11	Not reported* BC 9 months above US	Not reported
1947 – BC Intermediate language arts Test	6	Not reported	Not reported*
1950 – BC language arts Test	4	Not reported	Not reported*
1973 – Stanford language arts Test	6	Not reported	Not reported*

*Eleven language arts assessments were administered. British Columbia students outperformed United States' norms on five out of 11 tests.

Table 9: Science Outcomes During Standardized/I.Q. Testing Era, 1925-1972

TEST & DATE	GRADE(S)	BC NORM	US NORM
1925- Ruch-Popenoe Test in General Science	10	38.93*	34.70
	11	41.94*	Not reported.
1948 – BC Test of General Science	8	Grade 9.2*	Not reported.

*Three science assessments were administered. British Columbia students outperformed United States' norms on three out of three tests.

Table 10: Spelling Outcomes During Standardized/I.Q. Testing Era, 1925-1972

TEST & DATE	GRADE(S)	BC NORM	US NORM
1925 – BC Spelling Test	2	72*	73
	3	85*	73
	4	81*	73
	5	81*	73
	6	86*	73
	7	82*	73
	8	87*	73
	1928 – Metropolitan Achievement Test	7	Not reported* BC 4.5 months ahead of US norms

*Eight spelling assessments were administered. British Columbia students outperformed United States' norms on eight out of eight tests.

Table 11: Commerce Outcomes During Standardized/I.Q. Testing Era, 1925-1972

TEST & DATE	GRADE(S)	BC NORM	US NORM
1968 – BC Typing Test	11	22.8 wpm	Not reported
	12	29.8 wpm	Not reported

*Two assessments in commerce were administered. No comparisons were made with United States' norms.

Provincial Learning Assessment Program Results: Data for Tables 12 to 17 were derived from Dr. Jim Gaskell's collection of PLAP assessments and reports, 1976 to 1999. We thank Dr. Gaskell of the BC Ministry of Education for access.

Table 12: Literacy Outcomes During the PLAP Era, 1976-1999

TEST & DATE	GRADES	JUDGMENTS*
1976 – BC English language arts Assessment	8 & 12	Gr. 12 > Gr. 8
1977 – BC Reading Assessment	8 & 12	Gr. 12 > Gr. 4
1978 – BC Assessment of Written Expression	4,8 & 12	Gr. 4 > Gr. 8 > Gr. 12
1980 – BC Reading Assessment	4,8 & 12	Gr. 4 > Gr. 8 > Gr. 12
1984 – BC Reading Assessment	4,8 & 10	Gr. 4 > Gr.8 + Gr. 10
1988 – BC Reading & Written Expression	4,7 & 10	Gr. 4 > Gr. 7 + Gr. 10

Assessment		
1993/4 – Provincial Assessment of Communication Skills	Gr.K-3, 4-7, 8-10 & 11-12	Gr. K-7 > Gr. 8-10 + Gr. 11-12
1994 – Writing for Specific Audiences and Purposes	11	n/a
1998 – BC Assessment of Reading Comprehension & First Draft Writing	Gr. 4, 7 & 10	Gr. 10 > Gr. 4 & Gr. 7

*Symbol (>) indicates that grade level has more successfully met expectations than other grade levels. Nine literacy assessments were administered. Students in elementary grades were assessed more favourably than secondary students in five out of nine tests.

Table 13: Numeracy Outcomes During the PLAP Era, 1976-1999

TEST & DATE	GRADES	JUDGMENTS
1977 – BC Math Assessment	4, 8 & 12	Gr. 12 > Gr. 4
1981 – BC Math Assessment	4, 8 & 12	Gr. 4 > Gr. 8 > Gr. 12
1985 – BC Math Assessment	4, 7 & 10	Gr. 4 > Gr. 7 + Gr. 10
1990 – BC Math Assessment	4, 7 & 10	Gr. 7 > Gr. 10 > Gr. 4
1995 – BC Math Assessment	4, 7 10 & 10A	Gr. 4 > Gr. 7 > Gr. 10 > Gr. 10A

Five numeracy assessments were administered. Students in elementary grades were assessed more favourably than secondary students in four out of five tests.

Table 14: Science Outcomes During the PLAP Era, 1976-1999

TEST & DATE	GRADES	JUDGMENTS
1978 – BC Science Assessment	4, 8 & 12	Gr. 4 + Gr. 8 > Gr. 12
1982 – BC Science Assessment	4, 8 & 12	Gr. 4 > Gr. 8 + Gr. 12
1986 – BC Science Assessment	4, 7 & 10	Gr. 4 > Gr. 10 > Gr. 7
1991 – BC Science Assessment	Gr. 3-5, 6-8, 9-10	Gr. 10 > Gr. 4 > Gr. 7
1995 – BC Science Assessment	4, 7 & 10	Gr. 4 > Gr. 10 > Gr. 7

Five assessments in science were administered. Students in elementary grades were assessed more favourably in four out of five tests.

Table 15: Social Studies Outcomes During the PLAP Era, 1976-1999

TEST & DATE	GRADES	JUDGMENTS
1977 – BC Social Studies Assessment	4 & 12	Gr. 12 > Gr. 4
1989 – BC Social Studies Assessment	4, 7 & 10	Gr. 10 > Gr. 7 + Gr. 4
1996 – BC Social Studies Assessment	4, 7 & 10	Regular Program Students > Sub-Populations (ESL worst)

Three assessments in social studies were administered. Students in elementary grades were assessed more favourably in none of the three tests. In two out of the three tests, secondary students were assessed more favourably.

Table 16: Physical Education Outcomes During the PLAP Era, 1976-1999

TEST & DATE	GRADES	JUDGMENTS
1978 – BC Physical Education	3, 7 & 11	Gr. 11 > Gr. 3 + Gr. 7

One assessment in physical education was administered. Students in secondary grade were assessed more favourably than elementary students in this one test.

Table 17: French Immersion Outcomes During the PLAP Era, 1976-1999

TEST & DATE	GRADES	JUDGMENTS
1987 – BC French Immersion Assessment	4 & 7	Gr. 4 > Gr. 7

One assessment in French immersion was administered. Students in primary grade were assessed more favourably than intermediate students in this one test.