An Effective and Critical History of Canada’s National Standardized Testing Program:

A Thesis
Submitted to the Faculty of Graduate Studies and Research
In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Educational Psychology

University of Regina

by
Margaret Anne Dagenais
Regina, Saskatchewan
March 31, 2011

© 2011 M.A. Dagenais
Ms. Margaret Anne Dagenais, candidate for the degree of Doctor of Philosophy in Education, has presented a thesis titled, *An Effective and Critical History of Canada's National Standardized Testing Program*, in an oral examination held on March 31, 2011. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner:  *Dr. Louis Volante, Brock University*

Supervisor:  Dr. Heather Ryan, Faculty of Education, Educational Psychology

Committee Member:  Dr. Douglas Brown, Faculty of Education, Curriculum & Instruction

Committee Member:  Dr. Paul Hart, Faculty of Education, Curriculum & Instruction

Committee Member:  Dr. Mhairi Maeers, Faculty of Education, Curriculum & Instruction

Committee Member:  Dr. Morina Rennie, Faculty of Business Administration

Chair of Defense:  Dr. Neil Ashton, Department of Physics

*Video Conference*
Abstract

This dissertation explores Canada’s national large-scale testing program, the School Achievement Indicators Program (SAIP) and its successor, the Pan-Canadian Assessment Program (PCAP) using a methodology built on Foucault’s theoretical concepts. The product of the research is a critical and effective history of the production of SAIP/PCAP within the Canadian social, political, and economic context. The study considers the work of Canadian scholars and commentators working in the field of large-scale testing and the work of those working in aspects of education that are proximal to the field of interest. The research recognizes the key role of the investigator, the critical importance of investigator’s proximity to the field, her struggles within the field; and it also recognizes the need of the investigator to step away from the obligation to argue a particular perspective. The study also explores standardized testing regimes operating in England and the United States as well as international testing programs to gather the effects of these correlative spaces on the Canadian experience. The outcomes of the research include the production of an understanding of how large-scale standardized testing was produced in Canada, its resultant effects of harmonization on curriculum, and the identification of possible research sites for further inquiry.
Acknowledgements

It is my pleasure to thank those who made my thesis possible. I would like to acknowledge the support and encouragement that I received from my family, friends, and colleagues. I would also like to recognize the moral support and guidance provided by my mentor and supervisor, Dr. Heather Ryan. My committee members - Drs. Doug Brown, Paul Hart, Vi Maeers, and Morina Rennie - offered invaluable feedback and support throughout the writing process.

Margaret Dagenais
Dedication

This dissertation is dedicated to the memory of my Father, Henry Abraham Friesen. My Father was my role model. His life was dedicated to research and inquiry, serving his community, and creating a loving and caring home at a time of crisis and adversity.
Table of Contents

Abstract ........................................................................................................................................i

Acknowledgements ..................................................................................................................ii

Dedication ...................................................................................................................................iii

CHAPTER ONE
Background and Introduction to an Effective History of Canada’s National Standardized
Testing Program ..........................................................................................................................1

Overview ....................................................................................................................................1

The Research Space ....................................................................................................................10

Why This Study and Why Now? ...............................................................................................19

Purpose ......................................................................................................................................22

Research Questions ....................................................................................................................23

Methodology ..............................................................................................................................24

Method .......................................................................................................................................28

Assumptions ...............................................................................................................................30

Limitations ..................................................................................................................................31

CHAPTER TWO
The Voice of the Self: An Ontological Perspective ....................................................................34

Las Meninas: A Contemplation ..................................................................................................36

The Gap .......................................................................................................................................37

A Synthesis of Possibilities ........................................................................................................42

The Self as a Field of Knowledge ...............................................................................................42

Memory .......................................................................................................................................44

Method .......................................................................................................................................47

Meditation ...................................................................................................................................50

Synthesis .....................................................................................................................................52
Event 5: No Child Left Behind Act 2001: Standardized Testing
Remains in Force ................................................................. 132

Summary ................................................................................. 136

The English Experience with Standardized Testing: .......................... 138
Government, Governance and Choice
  Governmental Engagement with Standardized Testing .................. 140
  Governance ........................................................................... 143
  Choice: Placing the Delivery of Educational Services in
  the Marketplace ...................................................................... 148

PISA, TIMSS, and PIRLS: Global Order, Regulating Statements and the Effects
  of Culture ................................................................................ 150
  TIMSS ..................................................................................... 153
  PIRLS .................................................................................... 158
  PISA ...................................................................................... 163
  Gender ................................................................................... 168
  National IQ Scores .................................................................. 172
  Summary ................................................................................. 175

CHAPTER 5
Speaking the Canadian Experience into Existence ................................ 178
  The Interwoven Threads .......................................................... 179
  Conversations and Interpretations .............................................. 187
    Testing: Science and Measurement; Economics and Social
    Restructuring ......................................................................... 187
    Accountability ........................................................................ 197
  Technologies of Control ........................................................... 205
    Curriculum-as-a-Document ................................................... 206
    Curriculum-as-Taught ............................................................ 212
  The Production of Useful Citizens and Docile Bodies .................... 214
Summary .................................................................................................................................................. 218

CHAPTER 6
A Genealogy of Canada’s National Regime of Large-Scale Standardized Testing................................................................. 220

Forces and Lines of Tension .................................................................................................................. 226

The Beginnings: 1964 – 1970 ............................................................................................................... 229

Education: The Forces and Lines of Tension ...................................................................................... 229

The Middle Years: 1971 – 1980 .......................................................................................................... 233

The First Steps toward a New Agenda: 1981 – 1990 ....................................................................... 236


The First Testing Cycle ...................................................................................................................... 255

The SAIP Mathematics assessment – 1993 ....................................................................................... 258

What Did the Results Tell Us? ......................................................................................................... 259

The SAIP Reading and Writing Assessment – 1994 ....................................................................... 265

The Results ........................................................................................................................................ 267

The SAIP Science Assessment – 1996 .............................................................................................. 271

Summary ......................................................................................................................................... 281

Contextual Canadian Conversations ............................................................................................... 283

Cycle 2 ............................................................................................................................................. 290

The SAIP Mathematics Assessment – 1997 .................................................................................... 290

SAIP Mathematics – 1997 - Student Questionnaire ..................................................................... 299

Summary ......................................................................................................................................... 300

The SAIP Reading and Writing Assessment – 1998 ..................................................................... 301

The Test Results ............................................................................................................................... 304

Summary ......................................................................................................................................... 310

The SAIP Science Assessment – 1999 ............................................................................................. 310
The Contextual Questionnaire ................................................................. 318
The Evolution of SAIP Over the Decade ................................................... 322
Conversations as the Decade Ends .......................................................... 324
The SAIP Mathematics Assessment – 2001 ............................................ 334
    SAIP mathematics – 2001 Academic Assessment ............................ 336
    The Canadian context – the 2001 Contextual Questionnaire ............. 340
The SAIP Writing Assessment – 2002 ..................................................... 346
    A Secondary Study: Habits of Mind .................................................. 352
    Summary ......................................................................................... 354
The SAIP Science Assessment – 2004 ..................................................... 355
    The SAIP Science – 2004 Academic Test Results ............................ 357
    The SAIP Science – 2004 Contextual Analysis .................................. 361
The Cycle Ends ....................................................................................... 366
Contextual Conversations in the First Decade of the New Millennium ... 367
    Philosophical and ideological conversations .................................... 368
    Meta Research on SAIP ................................................................. 374
Pan-Canadian Assessment Program: PCAP-13 2007 ................................ 380
    PCAP – 13 2007: The Academic Assessment .................................... 381
    PCAP - 13 2007: The Learning Context Analysis ............................... 389
    Supplemental Analysis: Performance of Minority/Majority Language Groups on PCAP – 13 2007 ............................................................. 394
Closing Out the Decade: Contextual Conversations – 2004 – 2010 .......... 397
    Conversations about Assessment, Accountability and Harmonization ............................... 398
    Secondary Studies and Meta Analysis .............................................. 404
The End of the Decade and the End of the Journey .................................................. 408

CHAPTER SEVEN
Understandings and Questions ............................................................................. 411

The Two Journeys ................................................................................................. 412

Understandings ..................................................................................................... 415

The Production of My Understandings Based on My Research Questions ...... 416

Questions .............................................................................................................. 436

Closing Comments ............................................................................................... 439

References ............................................................................................................ 441
CHAPTER ONE

BACKGROUND AND INTRODUCTION TO AN EFFECTIVE HISTORY OF CANADA’S NATIONAL STANDARDIZED TESTING PROGRAM

A dissertation serves as a piece of evidence about the process in which a student learns to think in a particular domain about some of the critical questions that percolate through the discourses that operate in that domain. The production of a dissertation creates opportunities for reflection of the writer/scholar engaged in its production. My dissertation is produced on two levels: My formal scholarly work addresses my research questions about Canada’s national standardized testing program, and my journal-type entries written in italics chronicle some of my thoughts as I reflect on the works of other commentators, my thoughts about how I have come to think in a particular way about the domain, and reflections on my lived experience in the field of standardized testing.

Overview

Education has served and supported different paradigms upon which societies have been organized and managed. Historically, education served the social paradigm, and through the uneven distribution of educational opportunities, individuals were able to maintain or advance to a particular status in society. With the advent of mass universal education in Western countries, education became the servant of a political paradigm whose goal was the democratization of educational opportunities. Currently, education serves the economic paradigm by providing educational services that respond to nations’ demands for skilled, creative and imaginative workforces that are able to compete in a global marketplace. In the globalized economy of today, intelligence is the raw material just as iron ore was the raw material of the steel industry of the industrial age. Just as steel plants put in place processes and tools to manage quality control of its product
and to provide customer assurance of value for money spent, the globalized knowledge-based economy has instituted similar quality control and product assurance requirements for the educational plants of the twenty-first century. Standardized testing, the analysis of test results, and the various reporting mechanisms instituted by these regimes of standardized testing now constitute the quality control mechanisms and customer assurance for the production of Human Capital.

The examination is not a new device for measuring learning, and its history is long; however, its purpose and its role in the educational process have changed as societies have moved through different organizational paradigms. While most developed nations now operate under a worldview shaped by an economic paradigm; worldviews can change and with that change, the role of human capital and the need for standardized testing will also change. The purpose of this dissertation is not to explore the worthiness of operating under an economic paradigm or whether standardized testing is good or bad; but rather it is to make explicit the tactical shifts in the lines of force in a particular domain of standardized testing in Canada and to identify the gaps and fissures in the taken-for-granted surfaces of the administration, interpretation and evolution of standardized testing in Canada.

The examination, as a domain of study, came to Foucault’s attention in his writing in the 1980s. According to Foucault (1984b), the introduction of “the examination” into the European educational process in the sixteenth century marked the transition of instructional activities from an art form that supported the status and birth right of an individual into a quasi science that fulfilled the functions of both ritual and science. The examination permitted the collection and tabulation of data about the transference of
knowledge, the differentiation of one student’s performance from another’s, and the facilitation of the production of statements related to cause and effect. As Foucault reflected on this transformation, he posited this notion, “Does their [examinations’] very technology, this tiny operational schema that has become so widespread … this familiar method of examination, implement, within a single mechanism, power relations that make it possible to extract and constitute knowledge?” (Foucault, 1984b, p. 197)

According to Foucault (1984b), the examination combines the disciplinary techniques of an observing hierarchy - parents, teachers, administrators, provincial and federal agencies of education, and the general public – and a normalizing judgement that determines the privileges or punishments to be awarded based on the performance of an individual or a group – admission to post secondary educational institutions, the receipt of scholarships, or marginalization to menial jobs, and/or public recognition of a school or jurisdiction as “excellent” (Foucault, 1984b). These disciplinary elements of the examination exert relational power by prescribing choices, making individuals visible, classifying and marking performances, but also by producing new capacities through the production of new realities, domains of objects and rituals of truth (Foucault, 1984b).

Although Foucault (2007a) has produced critical and effective histories of madness and the asylum, punishment and the penitentiary, language and the role of grammar, and the development of biological systems of classification of living organisms and the development of the life sciences, neither Foucault nor any other Foucauldian scholars have produced a general history of the examination and the educational system. For Foucault (1984b), the examination is a technique that harbours a “whole domain of knowledge and a whole of power” (p. 197). Although individuals have written specific
histories relating to the examination of particular students such as the blind or particular
tests such as the Scholastic Aptitude Test (SAT), no one has produced a general history
of the examination - its rituals, methods, the interaction of questions and answers, the
characters (administrators, teachers, and students) and the roles that they play. This
dissertation contributes another piece to the larger landscape of the general history of the
examination. Its perspective is limited by geography and by the type of examination. Its
focus is directed to the forces that produced the events rather than the individuals who may
have played key roles.

Subsequent to Foucault’s identification of a gap in our understanding of the
capacity of social systems/culture to engage in relations of power and their reciprocal
regimes of knowledge, and production of truth statements, Lemann (1999) produced a
total history of the Scholastic Assessment Test 1 (formerly the Scholastic Aptitude Test).
Lemann’s history focused on Henry Chauncey’s role in promoting standardized
assessment and in particular the Scholastic Assessment Test 1. According to Dean
(1994), a total history, such as Lemann’s history of the SAT, focuses on events, times,
places, and heroes, and villains. Lemann wrote this history around the actions of a central
heroic figure, Henry Chauncey, and his battles to instill the notion in the American public
that college/university admission decisions based on standardized test scores were
superior to decisions made on the merits of a preparatory school graduation diploma.
Lemann described the transformational process of the American postsecondary
educational system from one based on social merit/wealth to one based on test
scores/achievement. Essentially, the introduction of standardized testing into the
American post secondary educational system redistributed privilege. Brilliant students,
such as former President Bill Clinton, were able to qualify for admission and scholarships based on their test scores, an opportunity unavailable prior to the introduction of the SAT (Lemann, 1999).

In Canada, the nationally sanctioned standardized testing protocol for 13-year and 16-year olds, School Achievement Indicators Program (SAIP)/Pan-Canadian Assessment Program (PCAP), has a relatively short history; however, scholars such as Earl (2002), Fournier (2000), Goddard (2000), Moll (2004b), and Robertson (2007) have produced critiques of the testing instruments, the methods, the quality of the data, and the utility of the subsequent reports. But scholars have not, as yet, taken up Foucault’s (1984b) challenge to produce a general history of the examination or even a general history of a particular examination like SAIP/PCAP.

Dean (1994) uses the term general history to describe a history that is by definition, both critical and effective. A general history differs from a total history because it identifies the forces that brought a particular phenomenon into existence, offers a critique of its potential to produce regimes of knowledge, and describes the possibilities for generating rituals of truth that can operate within a particular local environment. And, for Foucault (2007b), resident within this type of critique, there is something “akin to virtue” as it strives to produce the truth about some thing (p. 43).

The researcher can approach this activity of the production of a general history through the use of the genealogical methodology (Foucault, 1984a). The genealogist operates in the world of documents, sifting through ignoble and locally produced documents to discover the point at which some thing is written into existence and the discourses that produced its existence. Foucault (1984a) describes genealogy as “gray,
meticulous, and patiently documentary” (p. 76). This process operates on “a file of entangled and confused parchments, on documents that have been scratched over and recopied many times” (Foucault, 1984a, p. 76). Genealogy seeks out production of understanding of present rather than the production of generalized statements about cause and effect (Foucault, 1984a). The production of a general history of a domain yields different knowledge from a subject-focused history; a general history yields knowledge about the forces that produce particular effects rather than knowledge about an individual or individuals who were viewed as instrumental in key events.

Foucault (1984a) justifies the need to produce such critical and effective histories of broad thematic domains, as these histories provide an opportunity to understand how the present has come to be. In illustration, Prime Minister Stephen Harper placed Foucault’s sentiment in a Canadian context when he quoted Izzy Asper in his reflections on the need for Canadian citizens to understand that the past produces the present: “To understand why a country is worth having, you have to know where it came from,” (Winnipeg Free Press, December 20, 2008, p. D12). Just as we need to know how our country came into existence and to understand how it functions today, we also need to know how our educational system came into existence. We need to know how its curricula were shaped and formed; how its assessment tools were fashioned and administered; and how the Canadian educational experience was shaped, differentiated and homogenized in response to the forces that influence Canadian culture (Robertson, 2007).

Standardized assessment tools like SAIP/PCAP are not neutral. They are political tools that influence educational decision making in the education ministry, the school
division, and even classroom (Barlow & Robertson, 1994). Standardized assessments create opportunities for a normalizing gaze to be directed at students, teachers, school administrators, and even parents (Barlow & Robertson, 1994). McAdie (2004) describes the transformational effects of standardized testing on Canadian classrooms: Competition has replaced cooperation as the key value in the Canadian learning environment.

Although Frank McCourt is an American teacher/writer, his comments about working in an environment shaped by standardized testing illustrate its disciplinary forces at work in the classroom, the resultant production of knowledge, and the development of the teacher’s capacity to recognize that not all learning can be captured by the standardized questions developed by other teachers. However, the results of standardized assessments can open doors for some students, but they can also limit opportunities for many others (McCourt, 2005). McCourt also illustrates the dilemma that many teachers face when the products of their instruction (students) are assessed by an external agency. According to McCourt, many of his students struggled to pass the New York Regents’ examination; yet, they contributed meaningfully to class discussions, improved their ability write, and mastered many elements of the curriculum.

“Every June during eight years at McKee, the English department met in a classroom to read, evaluate, and grade the New York State English Regents’ examination. Barely half the students at McKee passed the examination. The other half had to be helped. We tried to inflate the failure grades from high fifties to passing, the mandated sixty-five (McCourt, 2005, p. 107).

Unlike the New York Regents’ examination, the Canadian national standardized testing protocol, SAIP/PCAP, does not determine which students will pass or which ones can attend postsecondary educational institutions. However, SAIP/PCAP is a nationally mandated standardized assessment protocol whose self-identified purpose is to measure
student learning, facilitate the reporting of results to jurisdictions, set standards for student performance, interpret trends in student performance, and correlate the assessment results with environmental variables (Council of Ministers of Education, 2007). PCAP and its predecessor, SAIP, were instituted by Council of Ministers of Education, Canada, (CMEC) because “Canadians have long been interested in how well their educational systems have been meeting the needs of students and society” (Council of Ministers of Education, Canada, 2008d).

In reference to Foucault, an examination is a “form of discipline that consists of a normalizing process instituted by an observing hierarchy combined with that of a normalizing judgement” (Foucault, 1984b, p. 197). The examination serves as a form of surveillance that has both the potential to make individuals, schools, or jurisdictions visible through the process of differentiation and to construct them as objects through the process of subjection (Rabinow, 1984). SAIP/PCAP exercises its normalizing gaze and produces visibility through its reports that describe student performance by jurisdiction: The reports highlight the jurisdictions whose students performed better than the norm and those who performed below the norm (Council of Ministers of Education Canada, 2003b). The media reinforce the visibility through the process of selection and intensification of the performance differences and by drawing connections to hypothetical causes for the differences (Stack, 2006). Through the processes of norming, statistical comparison, and ranking, stakeholders are now able to analyze and compare the performance of jurisdictions, genders, language groups, and test questions (Robertson, 2000). The CMEC and the respective provincial ministries are able to use the data to pose possible and carefully-crafted hypotheses that explain the differences in performance.
(Robertson, 2000). CMEC encourages the jurisdictions to identify and consider opportunities for collaboration on curriculum issues, but it does not direct such collaboration (Fournier, 2000).

From Rabinow’s perspective, it is through the process of examination, students, teachers, principals, schools, and jurisdictions are transformed into objects. A standardized assessment process like SAIP/PCAP facilitates the quantification of the characteristics of students, teachers, the teaching and learning process, and schools through the delivery a standardized measurement process. The ritualized processes of the examination create an environment where the relations of power and knowledge are made visible; the examination exerts power through the forces of discipline by restricting freedom and limiting possibility while simultaneously producing capacity through the production of knowledge (Foucault, 1984b). Foucault summarizes the power of the normalizing process of the examination in these words:

> The power of the Norm appears through the disciplines. The Normal is established as a principle of coercion in teaching with the introduction of a standardized education and the ... The power of normalization imposes homogeneity, but it individualizes by making it possible to measure gaps, to determine levels, to fix specialities and to render difference (Foucault, 1984b, p. 197).

Foucault (1984) points to an inherent danger in the examination as it facilitates through its properties the promotion of the collection of data, production of summarizing statements, and publishing of norms. The collection of data and its analysis can lead researchers and policy makers to the mistaken belief that correlation is synonymous with causation; thus, Foucault warns his readers that mistaking “causes for effects” can often lead to detrimental outcomes for those who are affected by the decisions based on the outcomes of data collection and analysis (Foucault, 1984a, p. 80). As Crocker (2002a)
clearly identifies, the positive correlation between variables does not imply that the variables are causally related and that much damage can be done to both children and to the educational system when policy writers misinterpret correlation for causation.

The Research Space

Standardized testing is a complicated and emotional field that exists at the nexus of the theoretical fields of psychology and pedagogy, and the practices and products of standardized testing collide with the theoretical and practical fields of the economy and statistical modelling. Foucault (1980) posits that as societies accumulate a surplus of wealth and a surplus of people, the major task for any government is to construct a means to circulate wealth and to supply the production apparatus of society with a source of qualified labour to ensure the continuous circulation of wealth. Education, as a productive activity, is positioned as the servant of the economy; its task is to prepare the next generation with the skills and knowledge that serve the national agenda and the economy (Foucault, 1980). Thus, standardized testing programs such as SAIP/PCAP provide mechanisms for the policy makers to extend their gaze into classrooms without their need to be present and to provide opportunities for policy makers to exert disciplinary forces on curriculum developers, school administrators, educational practitioners, and students. The results of SAIP/PCAP have been used to support agendas of school accountability, educational standards creation and maintenance, accessibility, corporatization of the educational activities, and homogenization of school curricula (Barlow & Robertson, 1994). Simultaneously, the activities and results of standardized testing through SAIP/PCAP create and support competition between jurisdictions, schools, and students; and they encourage and support initiatives that lead to greater homogeneity in the
Canadian school curriculum (Robertson, 2007; McAdie, 2004). However, the Canadian educational system was founded on principles of local leadership, stewardship, and consultation; and this notion of local stewardship was preserved the British North America Act 1867, the act that established Canada as a nation (Barlow & Robertson, 1994).

Economic theorists like those at the Hoover Institute have postulated the tenets of the Theory of Human Capital Development which state that the economic productivity of a nation is linked positively with standardized test scores (Grineski, 2005). In A Learning Alberta, the Alberta government puts forward its 20-year plan for the development of a skilled and knowledgeable workforce for the province of Alberta; and in this plan, progress toward this goal will be measured, in part, by scores on standardized tests (Alberta Advanced Education, 2006). The Fraser Institute, a Canadian conservative economic think tank, has set for itself the task of monitoring, reporting, and commenting on the relationship between the standardized test scores earned by children in British Columbia schools and the performance of the Canadian and the British Columbia economies (Cowley, Easton & Thomas, 2010).

Educational policy makers in Canada face challenges, unlike those faced by their counterparts in other developed nations and member countries of the Organization for Economic Cooperation and Development (OECD), because the responsibility for education in Canada has been allocated constitutionally to the provinces (Council of Ministers of Education Canada, 2003b). Without a national ministry of education, CMEC struggles to organize and report a national perspective for education to the international community; however, Canadian students do participate in standardized testing and survey
initiatives supported by member nations of the OECD (Robertson, 2007). The perspective of the OECD member countries supports the view that education is a critical component of any national economic strategy, and it deserves the focus and attention of the national government (Robertson, 2007).

Since the Canadian educational system is a product of the provisions of Section 93 of Constitution Acts 1867 to 1982, the following statement governs the delivery of education in Canada: “In and for each Province the Legislature may exclusively make Laws in relation to Education, subject and according to the following provisions (Department of Justice Canada, n.d.). Because education is a provincial responsibility, curricula taught in Canadian schools have developed and evolved to meet the local economic needs of each region and responded to the cultural diversity of each jurisdiction (Morton, 1994). This approach functioned well when Canada was a collection of local economies with divergent cultural values and beliefs (Vidich & Bensman, 1960). At the time of confederation, the protection of religious interests in education superseded the importance of economic interests; furthermore, education represented one of the smallest expenditures in most provincial budgets and was relegated to a relatively low priority (Robertson, 2007). Currently, the provision of educational services in Canadian provinces and territories consumes a much larger portion of their respective budgets. To illustrate this point, the province of Saskatchewan allocated more than $1.2 billion of its 2009-10-provincial budget to the provision of educational services which was matched by municipal governments (Gantefoer, 2009). The provincial allocation for education was second in size only to the allocation for health services.
Because education is determined by agencies of provincial and territorial governments, Canadian students’ educational experiences, as viewed from a pan-Canadian perspective, may have many elements in common, but there are also significant regional differences attributable to provincial legislation, local values and beliefs, and religious influences (Goddard, 2000). Because the amending formula for the *Constitution Acts 1867 to 1982* is complicated and support from all jurisdictions is difficult to garner, the political solution to the development of a national approach to education has been to sidestep the development of national policies on education and to execute a national school assessment program, SAIP/PCAP.

The first steps toward an informal national policy on education began as a response to the need to be competitive in the “space race.” In 1957, the former Soviet Union launched the first rocket into space which stimulated concern in the United States and Canada about space exploration and science education (Robertson, 2007). The Canadian federal government responded in two ways: investing in one of the most progressive space programs of that era, Alouette I, a communications satellite launched in 1962, and enhancing policies for science education for Canadian children (Goddard, 2000). In 1967, the ministers of education for the provinces and territories organized themselves into a small secretariat, the Council of Ministers of Education, Canada, (CMEC) to preserve their provincial and territorial jurisdiction over education (Goddard, 2000) and to provide a “quasi-national home for educational discussions that were national in scope” (Robertson, 2007, p. 94). This secretariat, funded jointly by the provinces and territories and by Human Resources and Skills Development Canada (HRSDC), served as an access point for HRSDC to enter into education policy and
administration; and it provided a vehicle through which HRSDC could influence Canadian educational policy (Goddard, 2000; Council of Ministers of Education, Canada, 2007). The presence of CMEC remained uncontroversial; its capacity to provide leadership was limited by its unofficial status in the constitutional composition of Canada.

In the 1970s, parents began to criticize their public schools as they feared that the Canadian educational system was not preparing children for an increasingly uncertain future marked with the economic challenges of stagflation, the combination of uncontrolled inflationary prices and declining economic activity (Barlow & Robertson, 1994; Morgan & Robinson, 1976). During the same timeframe, the OECD described Canada’s educational system was an economic risk to future generations of Canadians, and it recommended that Canada move toward a harmonized educational system (Robertson, 2007). The OECD also identified that CMEC could not provide transparency or accountability for the provision of educational services: key features of a modern educational system (Robertson, 2007).

In 1981, the federal government assumed a more active role in education; however, the federal government did resist the urge to embark on an agenda of constitutional reform that would results in one or more of the following outcomes: the establishment of a national policy on education, the formation of a national ministry of education, or the establishment of an official research agenda on education (Robertson, 2007). At the close of the decade the federal government chose to exert its influence through the CMEC by directing funds to CMEC to establish a national testing program (Robertson, 2007). The concern about the Canadian educational system was echoed in
the American report, *A Nation at Risk*, which details the link between the decline in the performance of American school children on standardized tests and the decline in the American economy (National Commission on Excellence on Education, 1983). The report describes knowledge, learning, information and skilled intelligences as the raw materials of international commerce and the knowledge economy; thus, the national government has an interest in establishing standards for education as it is the primary process by which these raw materials are developed. Because Canada’s economic and cultural markets are linked to those in the United States, Canadian citizens and policy makers were also looking for the source of Canada’s economic woes during the recession of 1983 (Barlow & Robertson, 1994). Thus, education became a target for reform and a source of blame for economic hardship (Barlow & Robertson, 1994).

In response to the rising international concerns about education and its importance as an economic indicator of the potential of a country to perform in the global economy, the OECD and the United Nations Educational, Scientific and Cultural Organization (UNESCO) initiated the development, delivery, and analysis of several international measures of educational performance (Council of Ministers of Education, Canada, 1996b). The activities of UNESCO and OECD in the 1980s signalled the beginning of the urgent need to collect data, to describe student learning as a number, and to correlate school performance with economic output (Earl, 2002).

As documented by Goddard (2000), the philosophy of the CMEC migrated in the 1980s from an organization of ministers of education bounded by a common interest in preserving the integrity of their local systems of education that responded to the needs of their communities and provinces into a cohesive group of education ministers. CMEC
recognized the value of collective action in fulfilling their common goal, the provision of school accountability data through the administration of a standardized assessment program (Goddard, 2000). Thus, in 1989, the CMEC arrived at consensus on the elements of a national assessment program and defined its goal: “To measure the effectiveness of Canadian public education systems by measuring student achievement in school subjects” (Council of Ministers of Education Canada, 1999, p.1). CMEC initiated the development of a comprehensive measurement protocol as a means to achieve its “wish to provide the highest degree of effectiveness and quality to their [educational] systems” (Council of Ministers of Education Canada, 1999a, p.1).

This decision to proceed with a national testing program was extraordinary for many reasons: The Constitution Act 1867 to 1982 specifies education as a provincial responsibility. The establishment of a national assessment program could have been perceived as the first step toward nationalizing education and the erosion of provincial responsibilities laid out in the constitution for education; and it could have lent some credibility to the belief that Canada is moving away from values embedded in the notions of a cultural mosaic produced through diverse and regionally-determined educational systems toward a more nationalistic and homogenous curriculum and school system (Morton, 1994). However, acting in a spirit of consensus, CMEC identified the subject areas for its national testing program - reading, writing, and, mathematics - and established a test schedule for the next decade (Council of Ministers of Education Canada, 1999a).

*Could collaboration over the establishment and administration of a national assessment protocol also serve the goal of preserving provincial autonomy concerning decisions related to education? (Dagenais, December 27, 2008)*
However, the decision to move forward with a national testing program was not based only on issues affecting teaching and learning, but on economic issues put forward by the Business Council on National Issues (BCNI) and by parents and citizens who lacked confidence in the local school principal to make sound academic decisions for their children and for Canadian students as a whole (Barlow & Robertson, 1994). In addition, the public had developed an appetite for charts, graphs, and statistics generated by large scale studies on school performance. Corporate Canada - companies such as Air Canada, AT&T, Bechtel, Bombardier, Canadian Pacific, Cargill, Dupont, General Motors (Business Council of National Issues (1998)) - relied on the appetite of ordinary Canadians for statistics in its use of the media and its reliance on the power of statistics to launch messages targeted at everyday Canadians about the health of their educational system and the need for measures of accountability (Robertson, 1999). These charts, graphs, and statistics portrayed a picture of the Canadian educational system in relation to the school systems of other countries; however, the studies rarely addressed the significant differences between the Canadian educational system and that of any other country (Robertson, 1999).

In December 1991, CMEC on behalf of the Ministers of Education for the provinces and territories announced the signing of a memorandum of agreement that outlined a national assessment strategy for 13- and 16-year olds (Council of Ministers of Education Canada, 1998). Thus, the SAIP was born; this assessment protocol would render results based on school jurisdictions, but it would not support the comparison between the performances of individual students (Council of Ministers of Education Canada, 1998). Originally, SAIP was envisioned as a criterion-referenced tool that would
support the comparison of school performance of particular jurisdictions with a set of standards or benchmarks (Robertson, 1999). However, meaningful criteria were difficult to develop, the analysis was difficult to conduct, and the appetite of public consumers for a normative analysis was high (Robertson, 2004a).

In 1993, CMEC administered the first SAIP mathematics assessment to 13- and 16-year olds, followed by reading in 1994 and science in 1995 (Council of Ministers of Education Canada, 1998). CMEC continued to administer the assessments in rotation until 2003. The assessments were revised based on the functioning of the tools, priorities of CMEC, and international influences on measurement of school performance and student learning (Council of Minister of Education, 2003b). In 2003 Human Resources and Skills Development Canada (HRSDC) announced a major shift in its priorities and that its programs and research initiatives would incorporate test scores from the Program for International Student Assessment (PISA) rather than SAIP (Canadian Education, n.d.). CMEC stated that a revised assessment tool, PCAP, would “reflect changes made in jurisdictional curriculum within the past decade” and would reflect the increased jurisdictional emphasis now placed on international assessments” (Council of Ministers of Education Canada, 2008d). However, CMEC failed to comment on the significance of the name change; “school” and “achievement” have been replaced by “Pan-Canadian” and “Assessment.”

*Does the disappearance of school and achievement indicate a change in direction? Is the school no longer the object under the gaze of the CMEC? Is CMEC no longer looking for achievement? Has CMEC’s gaze shifted to measurement and has the power of statistics overtaken our ability to understand and describe the notion of achievement? (Dagenais, December 27, 2008)*

The name change represents one, and perhaps the most visible shift, in the statements and the terms used to describe Canada’s national testing protocol; it is but one example of a
Foucauldian shift that forms the substrate of a general history that is both effective and critical.

Why This Study and Why Now?

The national standardized testing protocol, SAIP, has adjusted its scope several times since its inception in 1989 and its first administration in May 1993 (Council of Ministers of Education, Canada, 1998) to the current version, PCAP, and its first administered in May 2007 (Council of Ministers of Education Canada, 2008d). These transformations have included the dropping of the practical science skills assessment; the inclusion of contextual survey questions for students, teachers and school administrators; the setting of standard levels of expected performance for each age group; revisions to the academic focus of the assessment protocol, and a shift in the target age group from 16-year olds and 13-year olds to only 13-year olds (Council of Ministers of Education, Canada, 2007). Some commentators have criticized the merits of the shifts and the level and type of consultation that informed these shifts; however, the shifts, themselves - as plays of power, products of knowledge, and rituals of truth - have not been contested from a systematic, analytical, and historical perspective (Moll, 2004b; Robertson, 1999). While the process that led to significant transformation of SAIP into PCAP began in 2003, the effects were not felt until 2007 with the launch of PCAP and the initiation of research activities related to best practices in teaching and learning by the Canadian Council on Learning (Canadian Council on Learning, n.d.; Moll, 2004b).

The research activities initiated by the Canadian Council on Learning, an advisory body to both the CMEC and the Canadian government through its department, HRSDC are informed by the theories and practices of The 21st Century Learning Initiative. In
Canada, *The 21st Century Learning Initiative* is led by a group of “academics, researchers, policy makers, and practitioners” who are interested in engaging Canadians in conversations about new learning systems, brain research, and the functioning of human societies and whose goal is “to make sense of research on learning and learning processes that were fragmented in many different disciplines, and embedded in many different universities, research institutions and businesses around the world” (Canadian Council on Learning, n.d). Kaser and Halbert (2008), working under the auspicious of Canadian Council on Learning, chose to use “evidence from attitudinal surveys from the Program for International Student Assessment (PISA) [to]...paint a disappointing picture of students’ current levels of satisfaction with their learning experience” rather than evidence produced through the administration of SAIP, a Canadian assessment tool (p. 56). The PCAP assessment instrument was aligned to PISA and designed to provide data useful in comparative analysis with that generated by 15-year olds on PISA (Council of Ministers of Education, Canada, 2008a).

The previous paragraph illustrates some abbreviated examples of the complexities of the discourses that occupy the taken-for-granted space of the Canadian national assessment protocol. Robertson (1999) attributes the complexity of education and its assessment processes and strategies to the fact that education is not a neutral activity; its intention is the production of useful citizens for our national and the global economy. However, what constitutes a useful citizen is often in contention as is the means for the production of useful citizens. Thus, education is a political activity with an international agenda fuelled at times with issues generated by transnational corporations; its curriculum represents the “official knowledge” that captures the goals and aspirations for
Canadian youth; and the outcomes of education determine whether the youth of Canada will be able to take up key roles on the world stage as “active citizens, flexible workers or as needy consumers” (Robertson, 1999, p.1).

Foucault (1980) would argue that the task of the historian is to make “all these discourses visible in their strategic connections” (p. 38). This means the bringing together of the discourses about the delivery of education and the administration of educational assessment with the discourses that arise within the educational, political and economic systems and with the decisions and regulations which govern that the field of educational assessment. Foucault (1980) describes the regulations and decisions as the “ruses” of the field; these ruses are not enacted or played by any particular person, but they assure permanence and the functioning of the examination (p. 38).

In the construction of a history of any field of knowledge, the writer must decide if the story is about heroes, those who have made a significant contribution to the field, have status within the field or may have been visible at particular turning points or shifts in the way the field is apprehended; or if the story is about the forces, the lines of tension, the relations of power that produced the knowledge that we use today as part of our everyday taken-for-granted toolbox (Foucault, 1980). This study focuses on the field of standardized testing, and in particular, SAIP, and its current iteration, PCAP. It identifies as its goal as the production of a critical and effective history of the field that documents the effects of power, the lines of force, and the reciprocal production of knowledge and the rituals of truth (Foucault, 1984a).
The Purpose

As a student, I wrote standardized assessments, I worried about getting the right grade, getting into the right class or the right faculty at university or the evidence that I was suited for the career of my choice. As a teacher, I prepared my students to write standardized assessments – departmental exams, GED exams, and standardized entrance exams. I worried about their performance. Did I teach them the right stuff? Were they adequately prepared? Would they be able to take the next step? As a test administrator and assessment coordinator, I researched the development of standardized assessment protocols, developed cut scores for alternative admission criteria for postsecondary educational programs and provided advice to students and instructors. As I worked on this project, I was constantly reminded of the concern that instructors had about what the cut scores meant. Who would be excluded? Who would be included? Why couldn’t we (instructors) just create our own exam as we were trained to do it? How would the cultural bias affect students’ scores? Is the test sensitive to our students’ needs? I was always amazed that many students entering postsecondary educational programs thanked me for reviewing their scores, commented on how they appreciated the opportunity to participate in the process, and said that they felt more confident that they could be successful in their program. For me, the purpose of this dissertation is to develop an understanding of forces that impact on the emotional space that surrounds standardized testing by unpacking one test, SAIP/PCAP. I want to understand the nature of the power that produces the visceral reaction in teachers and sometimes in students. (Dagenais, December 29, 2008)

According to Foucault (1984b), the examination is a “repeated ritual of power” that permits the “teacher to transform his pupils into a field of knowledge,” and it serves as a guarantee that knowledge has moved from the teacher to the pupil (p. 198).

Foucault’s deliberations about the examination considered its formation as a gathering of “extracted knowledge from the pupil destined and reserved for the teacher” rather than the collection of data through the norming process of standardized assessment for statisticians, policymakers and the media (p.198). However, a standardized assessment like SAIP/PCAP transforms the role of the teacher from the creator, administrator, and marker of the examination into a participant in the examination process. The performance of the teacher, as well as his or her students, becomes visible through examination process; and some external authority, CMEC, performs the role of the teacher in
Foucault’s analysis of examination process (Grineski, 2005). Thus, the purpose of this study is to produce a critical and effective history of the production of the discourses and discursive structures surrounding and contained within Canada’s national regime of standardized testing, SAIP/PCAP. Through the production of this critical and effective history, the purpose is to offer an understanding of how we, as Canadians, have come to be and know in this space of standardized testing and to make visible the intersections of lines of force between and among the fields of education, politics and economics.

Research Questions

As I read - and re-read - the literature on the social implication of standardized testing regimes and as I listened to colleagues speak about the issues concerning the implementation of a standardized testing protocol as an accountability strategy for the business of teaching and learning, I became interested in the underlying assumptions and beliefs that led the CMEC to their decision to institute SAIP and to subsequently modify, adjust, revise and finally transform SAIP into PCAP. As I consider the documents discussing the relative merits of standardized testing, the implication for policy development, and the potential impact on curriculum decisions, I am beginning to question how the data generated by the administration of SAIP/PCAP informs regional and provincial curriculum decisions and how the political spaces of provincial education identities are constructed.

My research will consider the following questions:
• How do culturally, economically and politically derived regimes of truth telling shape the beliefs and assumption about what counts for knowledge in the intersecting domains of education, assessment, and accountability?

• How does the knowledge production and truth telling surrounding the development, administration, and data gathering related to SAIP/PCAP inform, influence, and, shape national and provincial educational priorities; and conversely, how do culturally- and contextually-produced factors as well as regional, provincial, national, and international priorities inform and shape SAIP/PCAP?

• How do the SAIP/PCAP results serve the normalizing agenda of accountability?

Methodology

According to Phillips and Hardy (2002), Foucault’s approach to discourse analysis is appropriate when mega discourses such as race, gender, class and governmentality are analyzed. While Foucault (1984b) describes the examination, itself, as a “slender technology,” he also comments that the examination has become so “widespread” (p. 197). It has the capacity to “extract and constitute knowledge” and to constitute this knowledge at a level that makes possible the transformation of that knowledge into political investment…” (Foucault, 1984b, p. 198). Governmentality is part of a chain that is constructed by these elements: “power relations, governmentality, government of self and of others and the relationship of self to self” (Foucault, 2007b, p. 252). These elements are connected to “questions of politics and questions of ethics” (Foucault, 2007b, p.252). Canada’s national standardized testing protocol migrates
through the links on Foucault’s chain as a totalizing thematic space; and it enables power
relations and normalizes relationships between the individual, groups of individuals, and
government. This protocol also sets parameters for the government of the self (students
and teachers) by self.

In Graham and Neu’s (2004) Foucauldian analysis of the Alberta grade 12
departmental examinations, standardized testing is described as a form of
governmentality which defines the thematic space encompassing the power relations that
exist between individuals and groups of individuals and government policy and
regulations. However, Foucault (2007b) defines governmentality as a “strategic field of
power relations in the broadest and not merely in the political sense of the term ... a
strategic field of power relations in their mobility, transferability and reversibility” that
passes through the subject in its relationship of the self to the self (p. 252). For Foucault
(2007a), the recognition of the problem of the subject in relation to the construction of
knowledge about the human sciences is critical. Foucault (2007a) postulates the problem
of the subject in any knowledge project in this question:

Can one speak of science and its history (and therefore its condition of existence,
its changes, the errors it has perpetuated, the sudden advances that have sent it off
on a new course) without reference to the scientist himself – and I am speaking
not merely of the concrete individual represented by his proper name but of his
work and the particular form of thought? (p. xiii).

Foucault (2007a) postulates that subjects, who produce scientific discourses, are
determined by the possibilities and conditions within that domain or field of knowledge.
Thus, Foucault (2007a) recommends that the production of a general history
characterized as both effective and critical should examine the rules that come into play,
the conditions that the scientist must fulfill to make his/her discourse coherent and
perceived as true. The subject is decentred, and the rules of discourse become the focus of analysis.

Foucault (2007a) enters his methodological approach through discourse analysis (Best & Kellner, 1991). His application of the tenets of discourse theory are used to analyze the institutional bases for the production of discourse in sites such as the asylum and penitentiary, to identify the viewpoints and positions from which people speak, and to identify the power relations which these discourses can allow (Kellner & Best, 1991). Foucault (1980) developed two methodologies, archaeology and genealogy to facilitate his approach to discourse analysis. He described these methodologies as useful in the analysis and the production of understanding about the discourses that are produced by the unstructured and unplanned nature of human social life (Scheurich & McKenzie, 2005). Foucault (1980) uses the term, archaeology, to describe a specific analytical process that searches for the rules that govern a discourse. These rules reside in the sediment and layers of documents that a culture produces over time (Best & Kellner, 1991). Foucault (1980) uses the term genealogy to describe methodological tactics that justify the redirection of the researcher’s gaze toward the context of the subjectifying practices. These tactics are used to draw out the political consequences of subjectification (Best & Kellner, 1991). Through the application of genealogical methodologies, “the individual is interpreted not only as a discursive construct but also as an effect of political technologies through which its identity, desires, body, and soul are constituted” (Kellner & Best, 1991, p. 47).

Through the application of the archaeological method of discourse analysis, the moment of naissance or emergence of any thing can be located; this is the point in time at
which the thing appears in records, documents, and discussions (Scheurich & McKenzie, 2005). At this time the thing begins to have attributes and characteristics (Scheurich & McKenzie, 2005). An analysis of the moment of emergence for the Council of Ministers of Education, Canada, would reveal the productive and resistive forces that brought this institution into existence and its relative importance to the construction of Canadian educational, economic, and political spaces. These constitutive forces, assumptions, beliefs, and values set the agenda of accountability and bring its mechanism of visibility, standardized testing, into being (Barlow & Robertson, 1994). Institutional knowledge, the process of an institution’s coming into existence, and the rules governing the application of an institution’s locally-produced knowledge – these constitute the discourses of savoir while the knowledge embedded in specific disciplines such as education constitute the discourses of connaissance. The interplay between the regimes of knowledge constituted in savoir (regulations governing the production of knowledge) and connaissance (the body of learning) construct the regimes of knowledge implicated in the mega discourse of governmentality (Scheurich & McKenzie, 2005). The task of the genealogist is to pursue the origin things and their characteristics while recognizing the impossibility of achieving this goal (Foucault, 1980). The archaeological method involves attempts to re-constitute and re-enmesh a discourse within the historical period in which it was produced (Foucault, 1980). The archaeological method is also application of historical methods of discourse analysis to determine the origins of a thing, its emergence, and the events surrounding its being thought into existence (Foucault, 1980).
Method

Foucault’s methods have been criticized as ill-defined, ambiguous, and lacking a systematic approach (Best & Kellner, 1991). Foucault invites scholars and researchers who are interested in using his methods to adapt his ideas, to stretch them, and to pick and choose the elements that are useful (Foucault, 1980). However, Scheurich & McKenzie (2005) warn researchers that picking and choosing elements such as the Foucauldian thematic of discipline while abandoning other elements of his historical methodology is inappropriate.

The question for me, as the researcher, becomes just how should I incorporate Foucault’s ideas? How do I determine if my method is consistent, has integrity, and can yield an understanding of the field of standardized testing? It is important for any researcher to recognize that Foucault’s methodology evolved from his quest to understand these basic epistemological problems: How can the knowing subject also be the subject of the knowing and be produced as an object? How is discourse connected to rituals of truth? How do the statements we produce about a thing represent how we can know that thing? (Foucault, 2007a) Although Foucault (2007a) refuses to connect himself to a particular methodological perspective, he does recognize that discourse is no longer connected to an essential truth. Discourse can produce rituals of truth, and the knowing subject is implicated in the production of truth (Foucault, 2007a). Foucault adds more ambiguity to this issue of research methods by stating that

[D]iscourse is so complex a reality that we not only can, but should approach it at different levels with different methods. No single theory or method of interpretation can by itself grasp the plurality of discourses, institutions and modes of power that constitute modern society” (Best & Kellner, 1991, p. 39).
While these statements still leave the researcher to work with a only a few Foucauldian tactics or toolkit, Foucault does put forward this critical requirement for researchers if their work is to have “political meaning, utility and effectiveness”; the researcher must have “some kind of involvement in the struggles taking place in the area in question” (Foucault, 1980, p. 64). As I reflect on my experience as a student, teacher, and educational researcher, and administrator, I realize that I have struggled in this space of standardized testing. It is a difficult space, filled with emotion, quasi-scientific practices, secrecy, issues of confidentiality, and ritualized methods.

According to Foucault (1980), understanding the production of taken-for-granted, everyday operation of the relations of power that govern the processes and procedures related to standardized testing and the regimes of knowledge produced by these relations of power is not accomplished by re-examining the sovereign power that legislated these effects of power. But rather, this understanding begins by disrupting the effects of power and the discourses they produce. According to Foucault (1980), it is these discourses that produce the subjectivities; and for this study, it is the production of the subjectivities of teachers, students, parents, and bureaucrats that are of particular interest.

An apparatus, such as standardized testing, is not a natural element of the educational process; however, standardized testing is a tool constructed through human activity to ensnare the educational process, trap it, and hold it up for examination (Foucault, 1980). It is the documents produced as a result of the processes and products of standardized testing that become the artefacts of relations of power, the products of regimes of knowledge, and evidence of culturally-derived rituals of truth (Foucault, 1980). Archaeological and genealogical activities disrupt the continuities of taken-for-
granted to reveal the disturbances and the redirection of statements covered over and rewritten with time; archaeology and genealogy consist of the activities for revealing these documents, making visible their statements of knowledge, uncovering the relations of power, and disturbing the rituals of truth (Foucault, 1980). Just as Foucault revealed that prisons were constructed as a means for the protection of wealth; schools and the educational process are viewed as vehicles for the production of wealth according to the theory of Human Capital Development (Grineski, 2005, Council of Ministers of Education, Canada, 2003a).

Assumptions

Foucault applies his methodology to the analysis of transformations in the construction of society over periods of centuries (Foucault, 1980). The period, as a methodological concept, is critical to production of a Foucauldian critical and effective history. The Council of Minister of Education, Canada, have a relatively short history (approximately 40 years), and SAIP/PCAP assessment protocol has an even shorter history (approximately 20 years). Based on a preliminary genealogical task, there appears to be sufficient shift in the statements used to describe CMEC and SAIP/PCAP and their procedures, processes and methodologies, key indicators of changes in truth telling.

The production of a critical and effective history relies on the availability of documentation: official and unofficial, local, and ignoble (Foucault, 1984a). These documents include CMEC formal reports, and scholarly documents that critique and analyze SAIP/PCAP, CMEC, and CMEC’s procedures, methodologies, and research instruments. However, few classroom teachers have written articles about their experiences with standardized testing, and even fewer students have written about their
experiences. The student experience is more frequently written about; students’
experiences are reported and commented on, but rarely shared from the first person
perspective (Fedore, 2005).

The challenge facing any Foucauldian scholar is the recognition of what might be
considered to be a rupture, a discontinuity, or a shift in direction and having the ability to
interpret its epistemological significance (Best & Kellner, 1991). These ruptures are
never clear or absolute breaks or shifts in direction; they always include elements from
the former epistemological regime in combination with new elements (Best & Kellner,
1991). The task is to develop the genealogical skills to sort and shift through the
totalizing narratives that dominate our everyday social reality to discover the suppressed
but autonomous, discourses, the knowledges, and the voices that have permeated the field
of standardized testing in Canada (Best & Kellner, 1991).

Limitations

I believe that what is essentially local character of criticism indicates in reality is
an autonomous, non-centralised kind of theoretical production, one that is to say
whose validity is not dependent on the approval of the established regimes of
thought. (Foucault, 1980, p.81).

From my perspective, Foucault’s (1980) theorizing leads away from the
construction of theories and laws that serve to explain the phenomena of the social world,
but his theorizing does focus on what is happening in a particular situation at a particular
time and identifies what truths can be told at a particular time and in a particular place.
For Foucault (1980), truth telling; that is, understanding who can tell the truth, under
what circumstances can the truth be told, and how the truth evolves over time is more
useful than the determination of congruence of a particular phenomena with a set of theories about the social world.

Thus, Foucault’s methodology reflects the genealogy of his own history of thought – the evolution of his own truth telling. The conceptualization of critical and effective histories of present – the documentation of truth telling activities over time - represents the production of his own histories of systems of thought (Dean, 1994). Like the genealogy of a system of thought, Foucault’s own methodology is at times seemingly progressive; each successive concept appears as an evolution of the previous, but at other times, disruptions, re-constructions, re-configurations appear as his methodology swerves, veers and moves in un-expected trajectories (Dean, 1994). The swerves and disruptions may appear as a casting-off, a rejection of what came before; however, from my perspective, Foucault pushes the boundaries that crisscross the limits of what truth will permit and re-conceptualizes new possibilities and problematizations. Dean describes the move from archaeology to genealogy as a methodological improvement; however, from my perspective genealogy is a logical extension of his methodological approach to the production of an understanding of the present. Genealogy opens up of new possibilities for the production of meaning (Scheurich & McKenzie. 2005).

Foucault’s critical and effective histories encourage educational researchers to look beyond the obvious and consider the possibilities (Scheurich & McKenzie, 2005). The thematic topic of standardized testing has spawned a resistive literature that documents the negative effects produced by relations of power that include limiting the scope of the curriculum, the teacher’s ability to instruct the curriculum in an imaginative and creative manner, and the educational experience for the student (Robertson, 2000;
Robertson, 1999). However, Foucault’s methodology provides encouragement to ask “What is being produced?” Are there positive (productive) effects of power (not necessarily good effects) and products of knowledge that escape traditional analysis? Does a perspective shift reveal new possibilities?
CHAPTER TWO

THE VOICE OF THE SELF: AN ONTOLOGICAL PERSPECTIVE

As I consider this space of standardized testing and my engagement in this knowledge project, I realize that how I have come to know and be in this space shapes and colours what I can produce as knowledge. Foucault (2007a) states that knowledge is shaped and formed by the effects of culture and experience. So just what effects do my culture and my experience exert on this project, and what gaps in understanding do they create for me as the researcher approaching the topic and for my audience?

As I consider this question, I am drawn to a conversation with a colleague. My colleague began the conversation by stating, “I have never known anyone who actually studied standardized testing; most people just have an opinion. So what is your opinion; just what do you think about standardized testing? Don’t you think that standardized testing would establish some minimum level of performance?”

I thought about the question and realized that the answer was not a simply “I am for standardized testing or against standardized testing.” My perspective is somewhat complicated and based on my experience and the cultural milieu in which I attended school.

My first real school activities began in June 1956 with a standardized test – for me, the first two years of formal education spent in kindergarten did not count as real school. It was an occasion – the Sunday best came out of the closet, my hair was set in ringlets, and my mother took me to school with my best friend. We sat in rows anticipating the distribution of the test booklets; our fingers playing with the newly sharpened pencils; we were nervous but excited.

My school administrator used the results to distribute the new batch of grade one students into four classrooms. I attended a streamed school; it was important to be in the “D” class – important for me and important for my parents. I secured a spot in the “D” class, but I did not shine early. I was bored with Dick and Jane,
and just didn’t get some of nuances of the school game. Towards the end of grade two, the school administrator arranged a meeting with my parents to convey the message “Her test scores are great; the performance is just not there. Mostly As and Bs; there should be some As, but mostly Hs (H stood for honours).”

Some months later, while I was on a vacation with my parents, I had the opportunity to talk to my father about school, and how I just didn’t understand this studying stuff. I am not sure of his exact words, but his comments related to my potential – the test results, the need to find meaning in what I was learning and the need to identify the importance to me of what I was learning. Somehow, that conversation between me, an eight-year-old, and my father about the results of a standardized test and my ability to learn made sense to me. I aced the social studies test for which I was studying; I figured how to learn and how to play the school game. (Dagenais, February 8, 2009).

For me, the relevance of standardized testing is locked into the meaningfulness of the conversations it stimulates. If the test taker learns critical information about the self through the process of standardized testing and is able to use that information in some positive transformational way, then the process of standardized testing, like any other data collection process, may have some value. This is not a simple or direct answer to my colleague’s question. It seems as though my response is characterized by embedded conditions of what the test might be; where it might be administered; what information a student, teacher, and/or administrator might gain from it; and how the data might be used (Aoki, 2003).

Through the narrative approach, I have engaged in an exploration of some of the elements of culture and experience that have shaped how I have come to think in this space of standardized testing. However, I recognize large gaps that emerge between the words, the present/the absent of my representation – the narrative/ the conversation - and what my reader might apprehend (Aoki, 2003).

Both Foucault (2007a) and Aoki (2003) contemplate the notion of the existence of the gap in understanding between the intention of the representation produced by the
writer or the artist and the meaning making of the audience. Foucault (2007a) illustrates this notion through his contemplation of the painting *Las Meninas* by Diego Valàzquez; and Aoki (2003), through his contemplation of the poem, *Stranger Music* by Leonard Cohen. Both theorists consider the opportunity that the gap creates for contestation of meaning making.

*Las Meninas*: A Contemplation

---

*La Meninas* by Diego Valàzquez 1656-57 Oil on canvas 318 by 276 cm (Valàzquez-Meninas.jpg retrieved on February 1, 2009 from http://en.wikipedia.org/wiki/Las_Meninas)

Foucault approaches the notion of the subject as constituted by a form of reflexivity from two perspectives: First is the representation of knowledge by the subject, the meaning making of this representation by an audience, and the emergent gap (Foucault, 2007a). Second is the self, as an ontological field, to be submitted to a possibility of action resulting in some form of truth telling (Foucault, 2007a). The
questions that Foucault (2007b) poses in regard to his historical study of the subject reinforce the notion that the subject is not a “field of subjective data to be interpreted” (p. 160), nor is it a field of data from which some buried truth can be uncovered from some obscure corners of the soul relating to the individualization of the subject. However, Foucault (2007b) states that the truth exists as an attractive force; a force that has the power to evoke what is good for everyone, and a force that exists in relations with the subject. Foucault (2005a) argues that it is through this imperative of *care of self* that the *will to know yourself* emerges as a form of reflexivity that attends to “what we think and what takes place in our thought” (p. 11). The specific nature of the relationship between one’s thinking about some *thing* and what constitutes one’s thought about some *thing* constitutes a form of reflexivity specific to this or that form of care of self. Reflexivity, like truth telling, is mutable and can be modified; thus, how one comes to know oneself in relation to the activities of truth telling can be modified. Therefore, the knowing subject’s disposition to the ability to engage in truth telling activities can vary with time and with situation (Foucault, 2005a). These modifications to relationship between the knowing subject and the truth are gaps or discontinuities (Foucault, 2005a). Locating the gaps is critical Foucauldian work because it is the gaps in the smooth patina of progress that signal changes in the truth telling activities.

The Gap

Foucault (2007a) approaches the problem of the representation of knowledge by the subject through his contemplation of the painting *Las Meninas* by Diego Valázquez. The artist, Diego Valázquez, produced the canvas, *Las Meninas*; and he appears in self-
portraiture within the painting as does a partially obscured canvas, a potential surface of representation (Atlee, 2003). In Foucault’s analogy, the artist serves as analogue for the knowing subject, the producer of knowledge; while at the same time, the image of artist serves as a known subject, an object, some thing to be known. Foucault (2007a) considers the representation of the images of painter, the canvas and its position (partially obscured), the invisible models, the models’ reflection in the mirror, and the observing audience as elements in an analogy for questioning the representation of knowledge by a knowing subject (the painter, the subject of the painting, and the audience). Foucault (2007a) concludes that the image of painter is constructed such that its gaze is locked into a relationship of reciprocal visibility with its audience. While on the surface this relationship appears to be simply that of “a mere confrontation, eyes catching one another’s eyes”; however, it is really “a whole network of uncertainties, exchanges and feints” that open into a space for the construction of meaning, interpretation and understanding (Foucault, 2007a, p. 5).

Foucault (2007a) points out that the gaze is never stable or neutral; the image of the painter appears to be staring toward the place of the invisible models and the place potentially occupied by the audience. While at the same time, the canvas, that the image of the painter is creating, remains unavailable to the audience. The audience focuses its gaze upon the painter who simultaneously grabs the audience and forces them to become part of the canvas (the field of knowledge). However, amid all the elements of painting - the faces, eyes, and gestures – a void or gap emerges between the image of artist and the image of the canvas. The image of the artist is available to the audience but not his
canvas; the painting, *Las Meninas*, is available to the audience but not the artist, Diego Valázquez.

Foucault (2007a) uses the analogy of the relationship between the artist and his creation (the concealed canvas) to illustrate the gap between the knowing subject and the subject as an object of knowledge production. In reference to the obscured canvas and the audience, Foucault (2007a) describes the gap as the “the manifest essence, the profound invisibility of what one sees [the model and its representation] is inseparable from the invisibility of the person seeing [the audience]” (p. 17). Foucault summarizes the challenge of representation by stating that “It is not possible for the pure felicity of the image ever to present in a full light both the master who is representing and the sovereign who is being represented” (p.17).

Foucault (2007a) used the imagery of the *Las Meninas*, a painting considered by most art historians to be the best painting in the world - but also the most puzzling - to illustrate the gap that exists between artist and his intention and the audience and its ability to understand the intention of painter. As Atlee (2003) points out, Valàzquez’ representation of the rigidly hierarchical structure of Spanish society leaves the audience with the question “What is going on here?” Painters were considered to be mere artisans and not worthy of inclusion in images of the royal court. Clark states (n.d.) that Valàzquez’ goal was to “tell the whole truth about a complete visual impression.” Atlee’s (2003) simple question draws attention to an emergent gap, and Clark’s statement draws attention to the goal of the painting, the production of truth. For Foucault (2007a), both the gap between the representation of the truth and the activities of truth telling were ontological concerns embedded in any knowledge production project.
This gap, to which Foucault (2007a) draws attention, relates to the researcher and the product of research, the thesis, the dissertation, the paper. The products of research are burdened with a gap, a place of invisibility, a murky hue caste by the researcher. The product of research is only a representation of what is. The dissertation is created by the author in whose eyes it is a resemblance of activities, discourses, and statements gleamed from others - arranged, ordered, and made into a fictive narrative of truth telling tainted by the author’s perspective and lived experience. Just as the artist uses frames, light, and the arrangement of objects to produce some representation of the truth, the researcher frames the research problem, selects a methodological approach, decides what should be included, and identifies what should be emphasized, and what should be omitted.

Berendt (2005) offers his notion of truth telling through his characters in his nonfiction work, *The City of Falling Angels*, when one of the minor characters states that “Telling the truth is the most anti-conformist act I know” (p.334). Berendt implies that the path to truth is not easy and can lead to difficult discoveries about the self. As I embark on this project of truth telling, I am aware of my own situated-ness in this research space as a test taker, test administrator and as a designer of regulations and policies relating to the uses and consequences of standardized tests. I am also aware that in researching this space I will engage not only in truth telling about the space occupied by the standardized testing but also about my-*self*.

As I engage in this work - this truth telling project - I am becoming increasingly aware of my active role in the construction of the story – this genealogy of standardized testing. I select the quotations; I grab the examples; I draw the connections between writers and commentators; I weave the scholarly and the everyday writing into my story of what has and is happening in this space of standardized testing; and I determine the boundaries – what gets said and what doesn’t. Is this truly a journey in truth telling? Whose truth am I telling? Is my writing so transparent that my reader will recognize the effects of my voice? Can
I represent the voices of others to truly illustrate the shifts in statements that characterize how we can know a thing?
(Dagenais, February 1, 2009).

Stranger Music: Meaning Making in the Space Between

Ring the bells that still can ring,
Forget your perfect offering.
There is a crack, a crack in everything,
That’s how the light gets in.
(Cohen, 1993, p. 373)

Aoki (2003) approaches the notion of the gap between the producer of knowledge and author of the representation and the audience or spectator from the perspective of opportunity. Aoki views the gap as an opportunity for illumination, an opportunity to re-understand what has already been produced. Using the notion of curriculum-as-a-plan and curriculum-as-lived, Aoki draws attention to the opportunities that the gaps between the representation and its enactment in the classroom present as fields for interrogation and knowledge production. Similarly, the gaps in understanding between the intention of Diego Valázquez in his production of Las Meninas and the apprehension of his intention by the audience of his work have resulted in copious works of literary criticisms and commentaries by art historians and philosophers. It is the gaps that created the potential for thought and knowledge production. For me, as researcher and author, the promise of positive outcomes emanating from the gaps between my representation of truth telling and the apprehension of this dissertation by the reader renders me some comfort. The questions, the concerns provoked by the gaps may stimulate further inquiry and more questions.
A Synthesis of Possibilities

Both Foucault (2007a) through his contemplation of Las Meninas and Aoki (2003) through his reflection on Stranger Music explored elements – the conditions and methods for the production of truth and knowledge – critical to Foucault’s analysis of the subject through their representations of knowledge. According to Foucault, an analysis of the subject involves understanding the conditions under which it is possible for the subject to produce true knowledge and the method by which we constitute ourselves as subjects of knowledge and truth (Davidson, 2005, p. xxiii). Foucault (2007a) provides a clue to the conditions and the method under which subjects are able to produce knowledge and truth:

History is probably the most erudite, the aware, the most conscious and possibly the most cluttered area of our memory, but it is the area from which all beings emerge into their precarious, glittering existence. It is the mode of being of all that is given us in experience; History has become unfavourable in our thought (p. 237).

The Self as a Field of Knowledge

Foucault’s interest in the self as a field of knowledge stems from his primary question concerning “the relations between the subject and truth” (Davidson, 2005, p. xx). The notion of subject gives rise to two sets of questions: questions about the subject – who is she/he? and questions about what does the subject know about her/himself? (Foucault, 2005a). An individual, according to Foucault (2005a), can often state what he/she thinks but is often unable to state how he/she has come to think this way or that way. For me this situation became apparent when I was asked to implement a standardized testing protocol for special admissions at a post-secondary institution. My colleagues aligned themselves on one side of debate or the other, but rarely did they offer more than an emotional response supported by an anecdotal situation as justification. When I would
question their thinking, they would respond with defensive statements and/or sweeping
generalizations.

In 2001, I was invited to present an overview of the new special admissions policy for this postsecondary educational institution to the student services professionals. They were sceptical about the merits of the approach: this new strategy would possibly qualify individuals for admission who lacked a grade 12 diploma. Their comments came in the form of questions: What would happen to our developmental programs? How could we be sure that these potential students could be successful? Who would possibly believe that the K-12 system would support this initiative? What would happen to our credibility as a post-secondary educational institution? We are already concerned that we don’t have the same public respect that other institutions in our province enjoy, so isn’t this just more evidence that points to our lack of rigour? If the public thought that we were second rate before, now they would know for sure, wouldn’t they?

The student services professionals were troubled by the change in the special admissions policy, and the role of standardized testing in the new policy may have been part of the problem that they were having with the change. They knew what they thought, but they did not know how they came to think this way. I was troubled because my logic and reason did not and could not operate in this emotion-filled atmosphere. Later, I read Lemann (1999) who described the emotional response to the thematic of the standardized test to be similar to that of religious believers: for those who believe; no explanation is necessary, and those who do not believe; no explanation is possible. (Dagenais, February 16, 2009)

While my challenges with the student services professionals can be reduced to that of membership in a particular group: They belonged to one, and I belonged to another. The questions concerning the truth and the relationship of the truth to the subject are erased through this tactic (Foucault 2005b). According to Foucault (2005a), it is the subject that speaks the truth into existence, and the nature of this truth is dependent upon the experience and the culture of the subject. Truth telling is dependent upon the care that the subject takes of him/herself through consideration of one’s reason, one’s capacity to tell the truth, and one’s soul (defined by Foucault (2005a) as the mind).

Foucault’s work on the subject and the truth takes the form of a genealogy of thought about the subject and its relations with the truth - a genealogy that he believes
offers meaningfulness to contemporary society (Foucault, 2005a). Care of self comprises three general themes: a way of behaving in the world and of considering and engaging in relations with other people; a type of attention one places on oneself, a way of looking at one’s thoughts; and the “actions by which one takes responsibility for oneself” (Foucault, 2005a, p. 11).

In 2005 I enrolled in the Ph.D. program; I was a tentative student, still wondering if I could really “think the way I felt that I needed to think to become a competent scholar.” When my friends and colleagues would ask me why I would choose to do something like this when I was so close to retirement, I respond by saying, “This is my pre-retirement project; I just need to find out if I can still really think.” I believe that I not-knowingly embarked in a Foucauldian project of care of self by taking care of my ability to reason and by focusing on my ability to engage in truth telling. I simply used the structure and form of a Ph.D. program to guide and discipline my work on my self. For me this action was a form of my taking responsibility for my self. (Dagenais, February 15, 2009)

Foucault (2005b) argues that the work of the self refocuses these questions of relationship of the subject to truth by applying one’s mind to oneself through reflexive activity. The reflexive activity, the care of self, is always mediated “through the relationship to someone else who is the master” (Foucault, 2005c, p. 58). Thus, my seeking and selecting a mentored program of study not only provides the structure but also the guiding relationship for reflexive work on my self. Foucault (2005e) summarizes the problem of the subject and his/her relationship to the truth in these questions:

What is the price I have to pay for access to the truth? This price is situated in the subject ... in the form of ... work I must do on myself, fashioning of myself ... [modifying of myself] to be able to have access to the truth (p. 189).

Memory

It's like a dream you try to remember
But it's gone
Then you try to scream
But it only comes out as a yawn
When you try to see the world
Beyond your front door
Take your time, is the way I rhyme gonna make you smile
When you realize that a guy my size might take a while
Just to try to figure out what all this is for
(Barenaked Ladies, 2000)

The lyrics of the Barenaked Ladies (2000) hit single, “Pinch Me,” tackles the importance of memory, its illusive nature, and the possibility of individuals’ using memory to understand how they make meaning of the world. Foucault (2005c) transforms the everyday concepts of memory and its companion, the dream, into mechanisms for work on the relationship between truth and self. According to Foucault (2005c), the subject does not have automatic access to the truth, but through activities like those of memory, the subject can withdraw from the external world, detach oneself from illusionary sensations, review the activities of the day, and purge one’s self of one’s faults of reason.

The memory is like a reflecting surface on which the individual’s mind can recognize itself as form of thought and knowledge (Foucault, 2005d). Individuals are directed to turn their gaze inward, to remember what they have learned (Foucault, 2005d). As I reflect on my early experiences with standardized testing, an incident in grade 9 is particularly vivid in my memory.

_It was a spring day; the sun was shining, and I was sitting a grade 9 health and guidance class. Our teacher had just passed out the worksheets for the Strong Interest Inventory (Psychometrics, n.d.). My job was to color the circles that corresponded to the element of the pair that I liked most or was most like me. I really didn’t believe that an afternoon of coloring circles on worksheet could yield a viable career choice - there has to be more to it than that! What about passion and personal dreams and commitment._

_The results of the test would identify the occupation or career which best suited me. I decided that I wanted the results to produce “lawyer” as the best possible_
career choice. I hypothesized what type of person a lawyer might be and what types of activities a lawyer might like to do – I had watched a couple of television programs about lawyers and one of my parents’ friends was a lawyer. I coloured in the circles that matched my assumptions – I did not want teachers or nurses on my results list. The results were returned to us a few days later – my results listed “lawyer” as well as “clergyperson” as my top choices. I believed that I had figured out the test.

However, several years later, I prepared, argued, and received a favourable decision in the Court of Queen’s Bench. A real lawyer observed my presentation before the court, asked to speak with me afterward, commented on the quality of the presentation and recommended that I consider attending law school. Perhaps my attempts to manipulate the test outcome many years ago were not really as successful as I had previously believed. (Dagenais, February 16, 2009)

From a Foucauldian perspective, the act of turning my gaze on my self through acts of memory, I am able to recall what I have learned or what I have seen (Foucault, 2005e). These acts of looking and memory are connected; the gaze is directed toward the object, and the object is displayed in its entirety on the reflecting surface of mind (Foucault, 2005h). These thoughts and knowledge collide to form some sense of my own being. In this particular instance, the memory illustrates the early formation of my scepticism about the power of standardizing testing to reveal what is really in someone’s mind. As a ninth grade student, I did not have the language to speak eloquently about the limitations of a small sample of behaviour, concepts of validity, and the importance of reliability. However, I did sense that answers on a simple worksheet were insufficient to predict my future career. Yet I am somewhat troubled by the eerie possibility that the test was actually predictive of a suitable career choice given my courtroom success – were my responses more truthful than I had believed them to be, or was it all just happenstance?
**Method**

Foucault’s (2005g) conception of method describes the work that an individual must do to understand this relationship of the self with the self as analogous to the training that good athlete must do to be competitive. A good athlete must invest in the training, but the training must be strategic and focused. The athlete must know the key actions and movements, master them, and perform them flawlessly in competition (Foucault, 2005g). According to Foucault (2005g), the same principle of utility operates in the domain of knowledge; knowledge which deals with the human existence should be piled on the useful pile while that dealing with *things* should be placed on the useless pile. The argument that Foucault (2005g) adopts is one of utility: the objects of nature are available and present; the individual can simply observe nature and gather whatever information he/she needs to know. The knowledge about human existence is obscured from view; the individual must search within himself, engage in philosophical methods, and train like an athlete if his goal is to discover the riches of the human experience (Foucault, 2005g).

The first exercise Foucault (2005i) identifies is one of decomposition both by time and by element. Foucault uses the medium of music and urges the individual to listen to a composition note by note, to separate the melody from the harmony and then isolate each note so as to master the control that the piece of music has on the individual. Once the individual has gained control over the emotional pleasure generated by the composition, the composition will lose its attractive hold over the individual: “the law that there is no reality for the subject except what is given in the present instant – each note or movement will appear in its reality” (Foucault, 2005i, p. 302).
For me, it is the lure of the interwoven complexity of ideas, words and sentence construction of the poststructural writers such as St. Pierre (2000) in her treatment of subjectivity and the subject. How can she think this way? How does she think this way? And how can I possibly understand the complexity of her message? In response to my dilemma, I developed strategies to unlock the meaning, to find gaps to interrogate: locate the subject and verb, get rid of the qualifiers and all the background noise, find the crystal of meaning, and then bit by bit add the qualifiers - add the noise. As I became more proficient at using my strategy, my understanding of the seemingly complex text improved, and I seemed to need my strategy less. The emotionality of my response to the texts diminished because I no longer took them in all at once, and my awe transformed into respect.

Returning to the work of St. Pierre’s (2000) on the notion of subjectivity and the subject, I focus my attention on this statement:

Subjectivity is produced socially, through language in relations … [and it] illustrates poststructuralism’s double move in the construction of subjectivity: a subject that exhibits agency as it constructs itself by taking up available discourses and cultural practices and a subject that, at the same time, is subjected into subjectivity and by those same discourses and practices.(p. 302)

St. Pierre (2000) has taken up Foucault’s primary question about the subject: How can the knowing subject also be the known subject? (Foucault (2007a) and St. Pierre have situated this question in the domain of language relations, agency and power.

When I first approached St. Pierre’s work, I was stuck by the words, the ideas, so many of them all together in one sentence; however, by breaking them apart, examining each word, then groups of words, the meaning emerges and the awe departs. I am left with respect for her writing but the paralyzing fear of the text has departed. (Dagenais, February 20, 2009)
Foucault (2005k) offers a second exercise: reading for the purpose of constructing “an equipment of true propositions for yourself” (p. 358). Foucault argues that to truly know oneself, one must do more than read to understand another’s writing. One must read to construct one’s own set of true propositions that operate as a prescription for behaviour. Reading then becomes a stimulus for meditation, for reflective thought.

As I read the Vinson’s (1998) discussion of curriculum standards juxtaposition to Foucault’s work on disciplinary practices, I had to stop, and reflect on the role of curriculum as a coercive tool in the production of docile bodies and useful citizens. Vinson states that “educators must question the inherent disciplinary nature of national curriculum standards. They must ask: How do national standards coerce certain behaviours? How do they establish conformity?”(p. 23-24). At first glance, curriculum standards should be a good thing; they should encourage all schools to provide a minimum level of education to all students; however, as schools adjust their curricula toward meeting standards, the curriculum is often severed from the lived experience of the students. Thus curriculum standards can be a product of power relations and indicative of who gets to decide rather than the product of local collaborative decision making conducted by the communities - parents and local educational authorities that have a vested interest in what is important for the next generation of students. Determining a strategy to approach the business of educating the next generation of students so that they can assume their roles as socially-engaged citizens becomes more complicated.

Foucault (2005k) states that writing is a natural progression from reading. After collecting several readings, formulating ideas, talking about those ideas with others, the
individual should engage in personal writing. The act of writing encourages the individual to think, reflect, and formulate their thoughts (Foucault, 2005k). Through my writing, I worked with my ideas, shape them, played with them and assured them a place in the external world.

As I reflect on my own writing, I am drawn to Van Manen’s article, “Writing Phenomenology,” and his question: When does writing begin? Does writing begin when I touch the keys on the computer; is it when the ideas begin to roll around in my head; or is it when that first sentence emerges while I am putting the supper dishes away. Just when does writing begin; what marks my departure from reading the works of others into the world of making ideas plain? (Dagenais, February, 22, 2009)

Meditation

Foucault (2005j) appropriates an unconventional definition for the term meditation: “the appropriation of thought”... and development of a conviction about its truthfulness so that this thought can become a principle of action (p. 357). The goal of meditation is the acquisition of the true discourses that individuals require to establish a perfect relationship with themselves; to become subjects that tell the truth and to be subjects of a true discourse (Foucault, 2005i). Foucault describes the process for subjection of true discourse as follows: one must engage in listening, reading, writing, speaking, and note making so that the things one hears and reads become one’s own. Therefore, one must learn to listen - to really listen - to determine if the truth is being spoken, as the truth of the message is determined by one’s physical response. However, Foucault (2005j) urges individuals to take a different tactic for reading; it is important to read - but read few books and few authors - read them well, learn the principles of the works, and write notes.
As I study and learn, I read, I write, I reflect, and I read more texts – the same author or authors that informed the principle ideas on which I work. I take notes – sometimes verbatim – sometimes I paraphrase what the author is saying. I bring the words and ideas into my own work; I practice using the ideas; I try out the truth telling statements to determine their appropriateness. I select; I edit; and I contemplate the importance of words, the phrases and the sentences. The following words and ideas made a lasting impression on me; they are words that trouble my identity and my physical presence when I read them aloud.

*Everyone likes to give as well as receive. No one wishes only to receive all the time. We have taken much from your culture ... I wish you had taken something from our culture ... for there were some beautiful and good things in it* (Chief Dan George, 1989, p. 41).

As a child, I learned that it was better to give than receive and that we should focus our attention on what we can do for others, yet Chief Dan George draws attention to the need for reciprocity in giving and receiving. It is important for a culture, a people, an individual to develop the capacity to both give and receive and to be able to recognize the gifts as they are given. Taking care of oneself before taking care of others is a primary tenet of care of the *self*. As an illustration, Foucault (2005j) describes the situation of a distressed father who fails to care for himself, and subsequently, jeopardizes the health of his sick child.

*As I reflect on the role of standardized testing in the assessment of learning for all Canadian children, I often wonder how or if these types of processes can or should fit into a vision of a society that supports the engagement of all Canadians in both society and the economy. Are the differences in learning styles and approaches real, and are we as a dominant society just trying to give ourselves out of this very difficult dilemma of different learning outcomes for different segments of the Canadian population?*
We provide gifts of programs, modified outcomes, and services without asking the recipients what they perceive the real problem to be. We wonder why these initiatives fail. Perhaps we have failed to be open to the receipt of gifts from other cultures. Bouvier and Karlenzig (2006) remind us that in our attempts to measure the educational performance of Saskatchewan students, we limit our vision of what constitutes learning to those discrete elements of the approved curriculum; there is no place for wisdom and knowledge relationships born out of an ancient relationship with the land. (Dagenais, February 16, 2009).

Synthesis

The gap in understanding between the knowing subject – the author, the writer, the artist – and the audience – the reader, the spectator, the observer – can be perceived as a barrier to meaning making or as an aporia, a place filled with complexity, doubt and possibility. By engaging in philosophical exercises, the knowing subject can come to know not only what one thinks but also how one comes to think in this way.

For me, the process of reading and confronting texts, extracting quotations, and listening to the voices as I read and re-read the commentators’ words, emphasizes how my culture and experience influence what text I select, how I integrate the words and phrases, and how I make sense of the world. As I reflect on this knowledge project, I return to these words:

*Through the process of writing, I created the text, selected the voices and wove them together to make a particular picture of the discourses produced by standardized testing. The act of creating of a picture is a process of selecting, privileging, and including while simultaneously discarding, eliminating, and silencing. This picture of Canadian voices is imperfect partial and incomplete* (Dagenais, February 23, 2008).
CHAPTER THREE
FOUCAULT’S METHODOLOGY: POSSIBILITIES AND OPPORTUNITIES FOR BUILDING UNDERSTANDING

If one is interested in doing historical work that has political meaning, utility and effectiveness, then this is possible only if one has some kind of involvement in the struggles taking place in the area in question. (Foucault, 1980, p. 64)

Foucault’s (1980) directive to researchers is to “face the conflicts that traverse the domain [of your field] and confront [the conflicts] ... that can construct instruments which will enable you to fight on [your] terrain” (p. 65). In addition, he remarks that if “one or two of these ‘gadgets’ of approach or method that I have tried to employ… can be of service to you, then I shall be delighted” (p. 65). Thus, Foucault is urging researchers to face the conflicts and the struggles of which they have knowledge (lived experience) and identify the barriers, conflicts, and obfuscations embedded in the instruments designed to study these areas so that the surface patina of official knowledge can be disturbed to allow the local, the ignoble, and the contextual knowledge to burst through the surface. Recognizing the importance of selecting a field of study, in which I have struggled and faced conflicts, I have chosen to face the conflicts that have traversed my path as an educator in the field of standardized testing.

Foucault would label standardized testing as an apparatus of power because it operates within the spaces of education, economics, and politics; and the results generated by programs of standardized testing influence socioeconomic, socio-political, and educational agendas - locally, provincially, nationally, and internationally (Foucault, 1980). The implications for knowledge production and the relations of power produced by the activity and the results of standardized testing move beyond the limited space of the taken-for-granted testing process that regulates classroom activity, the teacher, and
the students who are compelled by national, provincial, or local policy to engage in the testing process. According to Foucault (1980), understanding the production of taken-for-granted, everyday operation of the relations of power that govern the processes and procedures related to standardized testing and the regimes of knowledge produced by these relations of power is not accomplished by re-examining the sovereign power (bodies of authority) that legislated these effects of power; but rather it begins by disrupting the effects of power and the discourses they produce. It is these discourses that produce the subjectivities of teachers, students, parents, and bureaucrats (Foucault, 1980).

Many commentators have approached the thematic space of standardized testing using the lens of resistance to authoritative, directional acts of power (Moll, 2004a; Robertson, 2000; Robertson, 1999). Foucault (1980) would argue that this struggle against the juridical apparatus of power or legislative power is critical; but using only this lens limits the opportunity for knowledge production; and it denies the presence of local capillary power that can disrupt the taken-for-granted patina of everyday accepted practices, procedures, and relations of power. By researching the ignoble, the local, and the unofficial forms of knowledge production, the constitutive discourses that brought the apparatus of standardized testing into being can be questioned and made visible. The strategic connections of these discourses to other fields of knowledge such as the economy, international competitive advantage, and politics can be revealed (Foucault, 1980). By changing the research lens from a negative, resistive focus to a positive, productive focus, the outcomes of research change from the need to identify solutions to the problem of inequity to the opportunity to describe the local production of knowledge,
a natural outcome of the relational forces of power operating between discursively produced subjects (Foucault, 1980).

The systems that we, as individuals, have come to accept as natural are often not natural or neutral but were conceived to address a particular, often-forgotten problem (Foucault, 1980). By questioning the effects of power – the classroom testing of Canadian students at ages 13 and 16 for reading, writing, mathematics, and science proficiency - and following the ascendancy of power from its capillary effects in the classroom to its constitutive source, the productive discourses and their agential power can be described. This research process is positive, productive, and disruptive; but it does not produce solutions, nor does it produce argumentation that testing Canadian students is either good or bad. Instead, Foucault (1985) argues, that the goals of research should be the production of possibilities represented in the form of an effective history that Dean (1994) describes as both critical and neutral. The production of an effective history of the present reveals the opportunities for the production of knowledge as well as the disciplinary effects of the relations of power (Dean, 1994).

Rather than describing his methodology, Foucault (1980) illustrates it through his critical and effective histories such as that of the prison system. This critical and effective history illustrates how Foucauldian methodology reveals the opportunities for knowledge production shift and change as what counts for the truth, who can tell the truth, and the power embedded in truth telling statements shift and change (Foucault, 1980). Foucault (1995) also illustrates through his example of the prison system of the 19th century that the discourses produced within the prison system are as important as the discourses produced about the prison system (Foucault, 1995). By making the discourses produced
within the prison visible, its modes of regulations, patterns of decision making, its means of functioning, and its ruses become available for questioning or problematization.

Foucault (1997) defines problematization as “the development of a domain of acts, practices, and thoughts that seem ... to pose problems for ‘politics’” (p.3). These problems, asserts Foucault (1997), are those about which politics should have some involvement, but can “never answer them completely” (p.3). The key shift in emphasis is that Foucauldian problematization asks “politics what it had to say about the problems with which it was confronted” rather than analyzing some thing “from the point of view of politics” (Foucault, 1997, p. 4).

Thus, just as prisons became sites of the apprenticeship for the production of criminals, the question can be posed about the regulatory effects of standardized testing and its effects on the nature of the production of students/citizens. McAdie (2004) states that the taken-for-granted school virtues of learning and co-operation have been replaced with new virtues of competition and individualism, while Scott (2007) counters that the true virtue of the school and its processes lies in its ability to produce useful employees and that entrepreneurs and creative thinkers evolve in opposition to the school and its processes. Both McAdie (2004) and Scott (2007) engage in forms of truth telling about what is happening in the field of standardized testing and the production of useful Canadian citizens. However, Scott (2007) engages in truth telling from a journalist’s perspective, and McAdie (2004) engages in truth telling from the perspective of a former elementary school teacher. Both live and work within different relations of power and may use different statements to determine what represents the truth.
Although advocates for standardized testing, such as CMEC, may rationalize that the results of standardized testing provide quality assurance to Canadian citizens that schools are consuming resources wisely and producing high quality products – educated students - demanded by society (Council of Ministers of Education, Canada, 2003a), schools regulated by standardized testing risk becoming factories for the production of test-takers (McAdie, 2004). Resistive forces have pounded at the sovereign power of the regulatory policies that implement and maintain regimes of standardized testing by arguing that the curriculum becomes restrictive and narrow: Students become anxious and overly competitive, teachers become stressed and are fearful about the implications for employment and tenure, and industry is concerned about students’ lack of creativity and inventiveness (Scott, 2007). However, the patina of relations of power remains uncontested, the productive elements of knowledge unrecognized, and the virtues of such an apparatus of power remain a largely uncontested and taken-for-granted “good” of social activity (Dean, 1994).

Most resistive studies follow the descending path of power that constitutes an effect (Foucault, 1980) exemplified by students’ scores or placements following the implementation of a regime of standardized testing. Foucault argues that production of an effective or critical history requires a different study, a study that includes both official or legitimate knowledge and local knowledge (Foucault, 1980). For example, the official knowledge produced in Foucault’s effective history of the French penal system is contained within the official documents constituting the policies and practices of prisons; local and ignoble knowledge is produced through the practices of the prisoners, the prisons’ employees, their local systems of environmental management, and their methods.
of transforming the convicted into the prisoner (Foucault, 1997). In case of Canada’s national regime of standardized testing, knowledge is not only contained in the official reports published by the Council of Ministers of Education, Canada, but also in journal articles, newspaper articles, press releases and memoranda about standardized testing and its operation within the Canadian school system and its effects on society.

Schools are subjected to systems of surveillance documented in official policies that include standardized testing and policies of accountability based on national and regional agendas (Robertson, 2000). Measuring the effectiveness and efficiency of schools in delivering the curriculum – a nation’s official knowledge – and the ability of schools to produce graduates who have the knowledge and skills required to support the economic agenda of a region and nation is a concern for industry, business, parents, and social-economic leaders (Moll, 2004b). Standardized testing reduces the complex task of measuring complicated multivariate social processes of education into a relatively time-effective task; however, the effects of these systems of measurement and prescription are felt by teachers, parents, students, and society (Moll, 2004b; Robertson, 2000). Because standardized testing, is not a natural element of the educational process, but a tool constructed through human activity to capture the educational process and to hold it up for examination, its processes and products become the artefacts of relations of power and the products of regimes of knowledge.

Through the production of effective histories, the taken-for-granted nature of regimes of standardized testing can be exposed for contestation and rendered visible by questioning the need to produce students, teachers, and schools as either disabled, deficient, below standard, or as enabled, endowed, and above standard (Dean, 1994;
Foucault, 1980; Moll, 2004b; Robertson, 2000). Foucault (1980) summarizes this position by stating: “The exercise of power creates and causes to emerge new objects of knowledge and accumulates new bodies of information” (p. 51). Thus, it is conceivable that standardized testing is operating as an articulation of sovereign power; and through its relations of power, it creates new objects of knowledge. These objects of knowledge create new mechanisms of power to control, limit or validate their expression.

Foucault’s Methodology and Method

Foucault argues that history is not a story of steady progress toward some form of increasing clarity, but it really comprises the statement and re-statement of systems of knowledge production in which individuals are produced as subjects through systems of domination (Dean, 1994). According to Foucault (1980), the shifts in our notions of subjectification are the result of the application of different rules regulating what counts as knowledge. Although Foucault was interested in discourses as rule-governed systems, his personal heuristic did not congeal into a formal method (Dean, 1994); thus, scholars using Foucault’s methodology, concepts, and methods are often forced to adapt, improvise, and consider the nature of possibilities that may arise from their problematizations. From Foucault’s perspective, it is the implementation of re-thought methodological statements that permit the re-shaping, re-positioning, and re-transforming of pre-existing statements that disrupts the taken-for-granted-ness and reveals opportunities for knowledge production (Dean, 1994).
Three Thinkers

In the twentieth century, the emergence of the three thinkers - Isaiah Berlin, Michel Foucault, and Richard Rorty - whose work straddles the difficult space between the history of ideas and philosophy began the task of making plain the challenges of the Project of Modernity with its notion of the steady progress of humankind and its dependence on rationality of Enlightenment as a modality for determining truth. The thinking and writing of Hegel, Kant, Nietzsche and Derrida influenced their thought projects; however, Berlin, Rorty and Foucault were reluctant to attribute recognition to other scholars or each other (Malachowski, 2002; Ignatief, 2000).

Berlin, a Russian-Jewish émigré to Great Britain in 1921, first articulated his challenges with the tenets of Enlightenment and its conflicts with Liberalism in his lectures on freedom (from 1935 – 1970) (Ignatief, 2000). Berlin opens the door on the project of intellectual history or the history of thought, as he traces the history of ideas about what constitutes freedom through the Greek, Roman, and European societies (Ignatief, 2000). Freedom, Berlin determines, operates in a space bounded by the human need to think and to choose and society’s corresponding need to control man’s inhumanity toward man (Ignatief, 2000). Ideas about what constitutes reasonable limits on a human’s basic need to have the freedom to think and to choose are shaped by the context of society in which individuals live (Ignatief, 2000). Ignatief’s quote from Berlin’s unpublished manuscript, *Freedom and its Betrayal* (from the Berlin Archive, Wolfson College) captures the essence of Berlin’s philosophical conundrum:

This is one of the most powerful and dangerous arguments in the entire history of human thought. Let us trace its steps again. Objective good can be discovered only by the use of reason; to impose it on others is only to activate the dormant reason within them; to liberate people is to do just for them which, were they
rational, they would do for themselves, no matter what they in fact say they want; therefore, some forms of the most violent coercion are tantamount to the most absolute freedom (Berlin in Ignatief, 2000, p. 202).

While Berlin recognizes the importance of reason and rational thought, he also recognizes that good-ness can be produced through the irrationality of religion, caring, and human connectivity and that uncontested rationality can bring horrific harm to humankind (Ignatief, 2000). Berlin also reminds us that the topography of things/ideas should not be forgotten: Things, like freedom, can be experienced at the surface as just freedom; and it does not have to become something else – something deeper like justice or ethics or anything else that individuals may think of (Ignatief, 2000).

Foucault’s first major work, The Order of Things, was first published in 1970 after Berlin presented his most memorable lectures and essays. In The Order of Things, Foucault (2007a) embarked on his first knowledge project: to determine “if the history of non-formal knowledge had itself a system” (p. x). By seeking an answer to this question, Foucault began comparative historical studies of fields of knowledge such as living beings, language laws, economic facts, political systems, and ethics in relation to the contemporary philosophical discourse (Foucault, 2007a). Foucault’s (2007a) initial questioning led him to a contemplative space that did not reside in the traditional space occupied by philosophy nor by traditional history – a space much like that occupied by the thought projects of Berlin. Foucault (2007a), like Berlin, struggled with the notions of modernity and the rationality of Enlightenment. Rather than rejecting or accepting the tenets of Enlightenment, Foucault refused to be labelled, categorized, or confined by artificial boundaries. While Berlin’s thought projects waffled between the tenets of Enlightenment and Liberalism, Foucault (1980) released himself from the tension-filled
space of belonging-ness to a particular philosophical perspective and focused on discourse, the statements that regulate what counts for the truth in a society or culture.

Foucault (1984b) also picks up Berlin’s concern about freedom, but he re-names and re-frames the construct using the terms discipline, coercion, and power. For Foucault (1980), discipline is an outcome of relations of power that results in the production of knowledge, the statements of a discourse that regulate truth telling, and the productive force of coercion that can produce some social action. While Berlin constructs the concept of negative liberty, the justifiable limits placed on individual freedom for the greater good of society, Foucault (1984a) reframes discipline as a positive productive act of power, necessary to produce useful citizens for the perpetuation of a society.

Both Foucault and Rorty were influenced by the work of Nietzsche, and the similarities in their focus on culture, language, truth, truth telling and change are evident; however, their foundational conceptualization of essential reality differs (Malachowski, 2002). While Rorty (2000) ascribes to a form of pragmatism that opposes essentialism and the philosophical notion that reality and things, themselves, have some intrinsic nature, Foucault (2007a) affirms the possibility of the existence of some external essential reality; and things, themselves, may have some unknowable essential nature. Rorty however, does emphasize the importance of discourse, which he renames conversation, as the means by which speaking subjects speak themselves into existence (Malachowski, 2002).

Rorty, like Foucault, values the importance of the metaphor over argumentation as a philosophical tool and attributes the changes in a culture’s metaphors or beliefs to historical accidents. However, Foucault (1985) ascribes shifts in truth telling to changes
in the statements used to determine whether information is true or false. Rorty’s position avoids the need to label statements as true or false (Malachowski, 2002). Using the following example to illustrate how Rorty’s position and Foucault’s position differ, a shift in a cultural metaphor might be the following: “A student’s results on a standardized test predict his/her performance in university” shifts to “A student’s performance in high school (grade point average) predicts his/her performance in university.” According to Rorty, the truthfulness of the belief statements does not change, but rather the first belief statement has less support than the second (Malachowski, 2002). Foucault (1984) would describe this change as an event, as evidence of a rupture, a discontinuity in the statements used to determine whether statements are true or false. Berlin would state truth telling is influenced by the period in time that produced the truth statements (Ignatief, 2000). Rorty would state that the later statement serves greater utility than the first (Malachowski, 2002).

Berlin, Foucault, and Rorty tackled the historicity of culturally-produced truth telling through the production of some form of intellectual history, critical and effective history, or pragmatic history of the change of culturally-produced metaphors. They also struggled with notions of the Project of Modernity and its focus on the steady progress of society, and they contested pre-eminence of rationality and reason of the period of Enlightenment (Dean, 1994; Ignatief, 2000; Malachowski, 2002). The presence of three eminent scholars working in this marginalized space between history and philosophy legitimizes the work of other scholars who are working in this somewhat uncharted space and are attempting to construct method out of metaphors and tactics (Foucault, 1980) and who recognize that having a history is a fundamental mode of being.
Determining the ontological and epistemological grounding of any theory is critical to its use in knowledge projects. The researcher must know what questions can be asked and what methodological limits are placed on what can be known. However, throughout Foucault’s writing, he has consistently resisted and evaded ontological labels, categories and structures and has offered simple resistive statements to the notion that his musings are philosophical in nature and that his thinking is the result of manifestations and manipulations of structuralism (Foucault, 2006). Scholars, writers and commentators such as Heyes (2007), St. Pierre (2000), Dean (1994), Hall (1997) and Dreyfus and Rabinow (1984) have all considered elements of Foucault’s methodology, his genealogies and his commentaries and have slated his work into various ontological positions or categories. However, Foucault (2007a) does cast some light on his ontological positioning in the framing of his initial knowledge project.

Foucault (2007a) introduces his work in *The Order of Things* by stating that his studies are comparative in nature as opposed to descriptive. His histories resist the urge to focus on major events that describe the progress of humanity or analyze the great clashes of humanity. Initially, Foucault resists the question of ontological positioning by stating that his historical work is not based on a “particular type of knowledge or body of ideas” (Foucault, 2007a, p. x). However, as he articulates his knowledge problem concerning the suddenness and thoroughness of change, he reveals his contemplations concerning the possibility of the existence of a reality or of an essential order, the scientific or philosophical explanation for the existence of order and the universal laws it obeys (Foucault, 2007a). However, it is a third domain, culture - situated between the notion of
an unknowable essential order and the philosophical and scientific laws - that accounts for order that constitutes the domain of Foucault’s problematization (Foucault, 2007a). This domain, according to Foucault, hosts actions characterized as confusing and obfuscating, and the results of these actions are the production of deviations from the primary codes of the essential empirical orders.

Foucault (2007a) asserts that there exists some essential order whose rules and laws are obfuscated by the actions of culture, and it is the work of scientific inquiry and philosophical ponderings to uncover these universal laws and rules that govern the essential order of things. Culture, argues Foucault, operates on the activities of the social world and is made visible and comprehensible through the actions of language. The unintentional role of culture is to mask, alter, and confuse the apprehension of this essential order (Foucault, 2007a). The effects of culture are not perceived in the same way by everyone; thus, culture is not all-encompassing, “all-pervasive, and everywhere the same” (Foucault, 1999a, p. 370). However, if one is to look, to scrutinize, and to examine, “the gaps, the crevices, the gaffs and the sour notes of discontinuity will play for the searching audience” (Foucault, 1999a, p. 370). Culture, through its uneven hand, disrupts truth telling, creates gaps or ruptures, and lays the groundwork for the fabrication of new truth telling activities, which are produced through the construction of new lines of tension (Foucault, 1980).

Although Foucault firmly denies structuralism as his philosophical position, Mills (2003) categorizes Foucault’s early archaeological work as simply structural analyses of autonomous discourses; and Dreyfus and Rabinow (1984) concur that Foucault’s archaeological work is a product of structuralist theorizing. Although Foucault does resist
building mental models and structures typical of structuralist theorizing, he does provide metaphors and analogies that can be interpreted as structural representations of mental modes or models of knowing the world (Mills, 2003). Foucault (2006) counters the structuralist label by pointing to the significance that he places on the event as the evidence of discontinuity and rupture in the statements that serve as the foundation for truth telling. This emphasis on discontinuity and rupture opposes the structuralist emphasis on the continuity of signifiers and the signified (Foucault, 2006).

Hall (1997), in his distillation of Foucault’s work, categorizes the archaeologies and genealogies as constructionism, a philosophical position that asserts that essential reality does exist but remains inaccessible to the knowing subject. In Foucault’s writings, the existence of ultimate reality is acknowledged but remains unavailable to human understanding (Hall, 1997). Although Hall captures Foucault’s conceptualization of ultimate reality and essential order, Hall fails to emphasize the importance that Foucault places on this third domain, culture and its influence on the formulation of the statements that regulate and justify the production of knowledge. It is the historical shifts and influences of culture on the statements and rules that regulate how a culture can come to determine what constitutes true knowledge that interests Foucault.

In contrast, Dean (1994) places early Foucault’s work within a positivist paradigm that is “haunted with remnants of the internal meaning of truth and value it sought to overcome” (p. 132). Subsequently, Dean (1994) re-visits, re-thinks and re-categorizes Foucault’s later work on genealogy as a contestation of the themes of modernity and Humanism. Dean’s analysis focuses on several key aspects: Foucault’s refusal to ascribe to the duality of the Enlightenment debate (Foucault’s questions the necessity to be either
for or against Enlightenment) and his contestation of the limits imposed by the
philosophy of Enlightenment. Dean describes Foucault’s movement away from
positivism and the philosophy of Enlightenment as experimental crisscrossing of limits
and transgressions, which works somewhat in accord “with the attitude of modernity”
(p.54). For Foucault, transgression is not liberation or global emancipation but the raising
of possibilities within a work of criticism (Dean, 1994).

St. Pierre (2000) uses Foucault’s work on language, power, and truth as a
touchstone in her feminist poststructural knowledge projects. Although St. Pierre does
not specifically categorize Foucault as a poststructural philosopher, she places herself in
that category and draws attention to similarities in her thought projects to that of
Foucault. For St. Pierre, human subjects speak themselves into existence through
language; essentializing truth statements are impossible to produce. Heyes (2007),
another feminist poststructural writer, incorporates Foucault’s work on power and
discipline into her work; however, she draws attention to the limits of Foucault’s
perspective. Heyes recognizes that application of Foucault’s methodology can disrupt the
continuity of a cultural system to reveal relations of power imbedded in regimes of the
knowledge production of an institution situated within a culture. However, the
methodology, according to Heyes, does not support the production of solutions to
problems of inequality. Although the application of Foucault’s methodology, does not
lead directly to solution making, it does reveal sites for further inquiry which would
otherwise remain obfuscated by the artefacts of culture.

The field of standardized testing is comprised of cultural elements that manifest
themselves as regulations: regulations about the administration of tests, the preparation of
the test subjects, the gathering of the testing materials, the publication of data, and the use of the data. Foucault (1984b) categorizes these processes and procedures as constituting “a branch of knowledge” and creating “a hold for a branch of power” (p. 203). These regulations, policies, and procedures are artefacts of culture, produced by the society in which they operate as statements that govern the production of knowledge and access to the truth (Foucault, 1984b).

Individuals who are “described, judged, measured, and compared with others” through the operation of these statements perceive; and they their effects as the acts of training, classification, normalization and exclusion (Foucault, 1984b, 203). The examination in its general form raised the individual to the level of description by “turning ordinary lives into writing ... a procedure of objectification and subjection” (Foucault, 1984b, p. 203). The special examination, the standardized test, extended the opportunity for normalization, categorization, and corrective treatment beyond a single classroom or school to regional and national jurisdictions. The procedures and processes of the standardized testing established statements that governed the production of data and formalized the reporting process and structure to facilitate comparison and determination of corrective treatment such as curricular modifications. According to Foucault (2007a), the design, administration, and analyses of a standardized test such as SAIP/PCAP are indicative of the activities of a cultural domain on the formulation of statements concerning truth telling, knowledge production, and disciplinary forms of power. The categories produced by effects of standardized testing are cultural artefacts rather than essential categories of ultimate social reality. Foucault (1984b) summarizes
the effects of the examination, its role in knowledge production, and its utility as an apparatus of power as follows:

[T]he centre of the procedures that constitute the individual as effect and object of power, as effect an object of knowledge ... by combining hierarchical surveillance and normalizing judgment ... [The examination] assures the great disciplinary functions of distribution and classification. (p. 204).

The application of Foucault’s methodology to the field of standardized testing situates this project as a cultural study of the forces that lead to the inception of some thing like a national standardized test, regulate its production, and document its effects on the cultural and economic institutions within the culture. Testing is a regulated and regulatory activity; it produces the types of statements that limit what can or cannot be known; thus, this type of domain provides a rich field of statements whose modifications alter what can or cannot be known. It is the changes, shifts, or discontinuations in statements that signal the effects of cultural changes, priorities, values, and beliefs.

**Discourse**

Discourse was of interest to Foucault as data because discourse - words, phrases, and statements – is a form of similitude that bears a relationship to knowledge but is no longer the essential truth (Foucault, 2007a). These similitudes portrayed through discourse take up the forms of analogies, juxtapositions, emulations, and sympathies (attractions) that illustrate the relationships of things to other things (Foucault, 2007a). Their signatures constitute the mark or sign that indicates that a thing can produce a particular effect – how else would we recognize that “ground walnut mixed with wine can cure a headache?” (Foucault, 2007a, p.29). Foucault (1999b) also specifies that the
[O]bject of all critical discourse must be: the relation ... of the speaking subject to that being – singular, difficult, complex, and profoundly ambiguous – we call language. ... [A] true literary work disrupts and questions the being of language over a period of time and illustrates through criticism what language has become and how the relation between the speaking subject and the being of language has changed. (p. 234)

In the discourse of the ancients, the words and signs were images of what they expressed – the essential truth of things (Foucault, 2007a). As the activities of culture affected relationships between words and signs and the images of things, themselves; words and signs became similitudes or representations of the things, but these similitudes or representations were not an expression of the essential truth of things (Foucault, 2007a). This mutable relationship between the truth of things, the manner by which humans speak these representations of things into existence, and the changes in these representations formed the foundation of Foucault’s questions. Foucault’s interest is related to the discontinuities or shifts produced when a culture changes the way it speaks things into existence through the production of statements that enable truth telling activities (Foucault, 2007a). Since the essential truth of things is no longer available to any particular culture, it is the activities related to the determination of what counts for the truth that is of interest to researchers engaged in the production of effective and critical histories using Foucault’s methods. In the field of standardized testing, representation of truths are produced from the generalization of data generated from students’ test writing. However, these representations of truth shift as the cultural context of Canadian society and education as field of inquiry shifts its priorities; and as economic and social agendas of the world shift. Conducting a critical history of the field of standardized testing as it is practiced at a national level in Canada has the potential to lay bare the shifting representations of truths and truth telling activities.
According to Foucault, each society has a regime of truth and its general politics of truth: “the types of discourses which it accepts and makes function as true” (Rabinow, 1984, p. 72). Foucault ties discourse to the production of truth telling statements and to the power that these statements have to invoke knowledge and to produce effects within the society or culture. It is the actions of these statements in determining which discourses function as true or false. As I consider the field of standardized testing occupied by SAIP/PCAP, I contemplate how the statements concerning what can be offered as truth telling have changed or shifted. As an illustrative example, the renaming of the testing instrument from SAIP to PCAP signals a change, a shift, a rethinking, and resets priorities to a pan-Canadian perspective with a normative assessment focus and away from a school perspective with criterion-referenced focus. What are the implications for what we can or cannot know about schools and their achievements if our lens has shifted and what forces produced the shift?

Foucault (2007a) reminds us that language as an element of discourse stands in representation of thought, and thought stands in representation for the thing itself. The representative nature of discourse links language vertically to the foundation of what is designated or horizontally to other things that exist in a similar mode of thought (Foucault, 2007a). As thought shifts its focus, its external representation through language also shifts its focus; and subsequently, what can be said to be true or false also shifts or changes. The problem manifested in discourse is the presentation of these shifts. Just how and why did the CMEC change the name of its national school testing protocol? What does this shift represent? What social forces caused the change? Such changes or shifts formed the primary concern for Foucault’s work as evidenced through the
production of his archaeologies and genealogies. These shifts on the production, administration, and interpretation of the SAIP/PCAP form the focus of my work. Change or shifts in the way things are spoken into existence can be sudden and thorough as evidenced by the title change for Canada’s national assessment instrument (Foucault, 2007a).

What does the name change imply? Does it mean that schools are no longer the orchestrator and determiner of learning but merely the delivery agent? Does it mean that the Council of Ministers of Education Canada by default has assumed the authoritative power for determining what constitutes official knowledge in Canada? (Dagenais, September 14, 2009)

Changes or shifts in the way things are spoken into existence are key indicators that a society has changed the way in which they think things into existence. This phenomenon is evident in both the social and natural world (Foucault, 2007a). For Foucault (2007a), these changes in the way things are spoken into existence are important and should not be reduced through acts of generalization and reduction into some form of continuity. These shifts and changes create opportunities to understand what is producing the worded world in which we live. In the scientific and philosophical thought, theories, principles and laws explain why things behave the way that they do. They provide guidance to a society to determine which statements are consistent with truth telling and which are not (Foucault, 2007a). These theories, principles, and laws are not immutable; and they transform with time (Foucault, 2007a). Foucault draws illustrations from grammar and the taxonomy of living things to illustrate how changes in knowledge and technology change how a society comes to think a particular thing into existence (Foucault, 2007a).
Although all words, phrases and statements have the potential to serve as a discourse for truth telling, it is the words, phrases, and statements that shift, change, or transform as a result of shifts or changes in the way a society thinks things into existence. For my study related to the Canadian national standardized testing protocol, I am particularly interested in changes and shifts in the way the CMEC, commentators, critics, and popular writers choose and change the way they use words, phrases, and statements to engage in truth telling activities about standardized testing.

*Discourse Analysis*

The result of Foucauldian discourse analyses is the production of effective and critical histories that introduce and promote discontinuities (Rabinow, 1984). The degree of efficacy of these histories is measured by their ability to destabilize the self, remove the assurance of continuity, and divide emotions (Rabinow, 1984). Foucault places his lens on the specific, the local, and the nearest; and he challenges the goal of traditional histories to examine that which is general, remote, and distant (Rabinow, 1984). Foucault urges researchers to study those fields in which they are immersed from the perspective of the forces that are produced within the field. Following Foucauldian advice, the production of a critical and effective history of the domain of standardized testing populated by SAIP/PCAP includes more than a chronology of events populated by major milestones and significant heroes that forged the domain. The production of a critical and effective history of the domain populated by SAIP/PCAP requires the description or interrogation of forces that produced the testing protocol, the statements that structure the truth telling, the production of knowledge, and establishment of regimes of power. Such
studies are bounded by time and consider the events in the context that produced them; the critical backward looking *gaze* that considers the past in terms of the present is under sanction (Rabinow, 1984).

Foucault’s conception of discourse analysis constitutes a different response, a critical and historical response, which attempts to interrupt and problematize the taken-for-granted. His interest is focused on *things said*,

their rules of regularities in what is said at a given time and place, and that these rules govern not just the kinds of things talked about but also the roles and positions of those talking about them. (Chomsky & Foucault, 2006, p. x)

Foucault also argues that

historical regularities in utterance cannot be explained by innate structures in the minds or brains of language learners…[but] they instead condition and constrain the actual use or exercise of our minds across a series of practices at once material and practical (Chomsky & Foucault, 2006, p.x).

Thus, speaking subjects speak themselves into existence through established rules of discourse, and regimes of standardized testing produce rules of regularities concerning what can be said by whom at a specific time and place.

To illustrate his conception of discursive formation Foucault (1980) argues that “historical work that is meaningful, useful, and effective” evolves from the personal engagement of the researcher with the struggles in that area; thus, he chose to focus on psychiatry (p.64). For me, the creation of Foucauldian problems in the area of standardized testing is personal and has evolved from my personal engagement with “the struggles, the lines of force, tensions and points of collision” in this area (Foucault, 1980). Politically important work does not, as Foucault (1980) warns, arise from studying a prestigious, correct, or clean area.
Foucauldian discourse analysis identifies as its object, the history of change in a particular domain or field of knowledge perceived as a science (Foucault, 2007a). Foucault (1980) exemplified his method as he traced the history of knowledge production in fields of psychiatry, the penal system, and human sexuality. Knowledge production in these fields was perceived as a science, but it lacked some of the rigour of the physical sciences. According to Foucault (1980), the examination is a rich field as it is both an organizer of knowledge and regulator of knowledge production, and it bares many traits similar to the physical sciences. The examination invokes the need to compare, to establish norms, to sort, and to validate what serves as official knowledge.

The event is the signal of change and is a critical element to the Foucauldian method; however, causality of change remains as an embarrassing methodological problem because “there are no definite methodological principles on which to base such an analysis” (Foucault, 2007a, p. xiii). These changes marked by the event may alter a science as a whole; yet these changes in the way we come to think some thing into existence do occur suddenly and completely (Foucault, 2007a). This methodological difficulty is exacerbated in the empirical sciences by the plethora, presence, and role of “instruments, techniques, institutions, events, ideologies, and interests” whose enterprise and purpose is the discovery of causality of things (Foucault, 2007a, p. xiii). Rather than attempting to offer solutions, Foucault purposefully chooses a methodological goal that sets the “problem of causes to one side.” (p. xiv). The goal of any Foucauldian critical and effective history is the description of transformation, a necessary developmental step in the journey towards the construction of a possible theory of scientific change and epistemological change (Foucault, 2007a). Thus, this study is about the transformations...
associated with the development, design, implementation, and delivery of SAIP/PCAP. It focuses on the description of the transformations so that we, as Canadian cultural consumers and producers, can understand how our present educational culture has been produced by the past.

The notion of the study of transformations is developed so that the boundaries include some elements of traditional history but exclude others. Foucault proposes that it is possible to develop a history of a science “without reference to the scientist himself ... his work ... [or] his form of thought” so then it must be possible to develop a history of SAIP/PCAP without reference to particular test developers or particular members of the CMEC (Foucault, 2007a, p. iv). While the elements of traditional history – biographies, theories, concepts, and themes – are important, there may be other systems of regularities or rules that operate at the margins and outside the domain of traditional historians: Rules that emphasize the specific, local, and temporal elements of a scientific discourse such as regimes of standardized testing occupy (Foucault, 2007a).

Foucault describes his first methodological approach as archaeology because it seeks to determine the system of rules that appear across systems of knowledge (Foucault, 2007a). Foucault employed directed his archaeological methodological to produce understandings in the domains of grammatical formations, the production of wealth, and the development of the taxonomy of living things. I directed my archaeological focus to the domain of standardized testing as it relates to the space occupied by SAIP/PCAP and to writers such as Berger, Epp, and Møller (2006), Bouvier and Karlenzig (2006) and Wasserman (2001) who speak about standardized testing using the language of possibilities consistent with the scientific writers and theorists in the
domain of quantum mechanics rather than the language of statistical predictability common in the texts of Lemann (1999) and Herrnstein and Murray, (1994). A Foucauldian archaeological question would ask what culturally-situated forces, lines of tensions, and temporality caused writers in the field of standardized testing to abandon their metaphors and analogies related to statistics, predictability and certainty for the ambiguity and uncertainty of quantum mechanics.

Foucault’s second approach, to the study of transformations, genealogy, appears to be layered over or a replacement of archaeology. However, Scheurich and McKenzie (2005) interpret genealogy as an additional layer in Foucauldian thought as opposed to a replacement; it is more like a re-thinking (Dean, 1994). Foucault (1980) appropriated the term genealogy from biology and uses the term as a metaphor for his approach to the “unmasking of truth claims of all knowledge” by determining the emergence and transformation of truth claims (Dean, 1994, p.29). The genealogical methodology forces the researcher “to instrumentalize the past in terms of the needs of present” (Dean, 1994, p. 29). The product of inquiry using a genealogical methodology is a history of the present that contains analysis of “those objects given as a necessary part of our given reality” while simultaneously revealing “the positive and productive forms of power underlying every movement of institutional or discursive limitations of statements” (Dean, 1994, p. 33).

In terms of my study of SAIP/PCAP as a regime of standardized testing, a genealogical methodology has the potential to reveal some understanding of how this national strategy was instituted, how it has transformed itself, how it has been transformed by cultural forces; and how it has transformed the activities of schools,
teachers, students, educational policy and national economic expectations based on the students’ performance on SAIP/PCAP.

An effective and critical history produced using archaeological and genealogical methodologies is designed to consider the past as it implicates the present, disrupts trans-historical conceptualizations, and considers the singularity of events rather than effacing them to create the comfort of continuities (Dean, 1994.) “An effective [and critical history] refuses to use history as a vehicle to assure us of our own identity and necessity of present” (Dean, 1994, p.18); thus, as I consider the field occupied by SAIP/PCAP, it is possible that some of our collective assumptions about the Canadian educational system may be challenged as well as some of our assumptions about what it is to be Canadian

The problem is: how do things happen? And what happens now is not necessarily better or more advanced, or better understood, than what happened in the past. (Foucault, 1980, p. 50)

Archaeology

Archaeology, as a research methodology, is not the study of antiquities; but it is the study of “mechanisms of power which have invested in human bodies, acts and forms of behaviour” (Foucault, 1980, p.61). Foucault uses the metaphor of archaeology to refer the researcher to the archive to begin the search for the beginning of some thing (Foucault, 1980). The object of re-discovery, according to Foucault (1980), is the disciplining and normalizing of the social body, the segregation of the normal from the abnormal, and the unveiling of the historically produced rules of regularities (Foucault, 1980).
The archaeology, as a methodology, is loosely guided by a complex set of interrelated concepts that include the following conceptual elements: knowledge (rules that determine truthfulness of statements), bodies of learning, power, positivity, enunciation of regularities of correlative spaces, level, limit, periodization, event, discontinuity, statements, archive, science, discursive formation, and gaze (Scheurich & McKenzie, 2005). Foucault begins to formulate his conceptual elements and their interrelationships in *The Order of Things*. These elements are expanded, enriched, and teased at the edges as his theorizing moves beyond historical perspectives and methodology into three clearly articulated domains of theorizing: rationality, discourse, and the production of truth; the practices of government, power and domination; and subjectivity, the self and ethical practice (Dean, 1994).

Foucault (2007a) introduces the notion of the enunciation of correlative spaces in *The Order of Things* when he discusses the possibility of placing a number of elements in juxtaposition – the knowledge of living things, the knowledge of laws of the language, and the knowledge of economic facts beside contemporary philosophical discourse. The correlative space emerges as a critical thematic organizer in the production of a critical and effective history of SAIP/PAC because the results of standardized testing are used as key economic indicators of the economic potential of a society and as key cultural indicators (Alberta Advanced Learning, 2006; Grineski, 2005). Common analogies emerge when different domains of knowledge that are bounded tightly by time but loosely by space are analyzed in a side-by-side fashion (Foucault, 1980). Following Foucault’s illustrative exemplar, archaeology of geography, the thought space for the analysis of SAIP/PCAP is designed to be loosely bounded by indeterminate and vague
demarcations, but the periodization (the demarcation of specific periods of time) is designed to be specific.

Periodization is a critical element in Foucauldian archaeological analysis as “Each periodization is the demarcation in history of a certain level of events, and conversely each level of events demands its own specific periodization” (Foucault, 1980, p. 67). If the period is shorter, the granularity of the events is finer; however, if the period is longer the granularity of events is courser. Foucault chose to conduct his analysis of particular domains such as psychiatry, the care of self, geography, and the penal system with a periodization of a century, which, subsequently, dictated the level of his analysis as certain events become visible with this type of lens while others are reduced through generalization to part of the taken-for-granted continuity of human existence.

In transferring Foucault’s concept of periodization and event to the domain of SAIP/PCAP, I recognize that both CMEC and SAIP/PCAP have a relatively short history; thus, establishing a periodization of a decade will facilitate a level of analysis consistent with the testing cycle. From my preliminary investigation of this field, I discovered that the granularity of the decade functioned well as an organizing tool. Periodization is the research tool that enables the event to make itself be known; the event is that bump in the continuity of human existence that signals transformation or discontinuity in how we, as society, living within certain cultural boundaries have come to know and be in the world (Foucault, 1980).

These Foucauldian events appear to be anomalies in the continuity of human existence; however, Foucault describes them as discontinuities and possible critical nodal events signalling changes in the way things are thought into existence (Foucault, 1980).
These discontinuities, defined as “sudden take-offs, hastenings of evolution, [and] transformations,” signal a failure of the continuity of the existing mode of truth telling (Foucault, 1980, p. 112).

Discontinuities captured Foucault’s attention as he formulated the questions that directed most of his philosophical writing: “[W]hat governs the statements and the way they govern each other to constitute a set of propositions which are scientifically accepted as true” (Foucault, 1980, p.122). From this initial question, Foucault examined the way Western societies organize knowledge emanating from the philosophical and scientific statements that guide truth telling and the effects of culture on the production of truth telling statements (Foucault, 2007a).

The construction of SAIP the assessment and the goals established by the CMEC illustrate Foucault’s theory. The initial conception of SAIP was based on the knowledge statement or theory that the outputs of an educational system can serve as a measure of the effectiveness of the educational system (Council of Ministers of Education, 1993). Measurement of students’ learning (on a standardized academic achievement test) and graduates rates contributed to the body of learning emanating from that the theory that outputs predict a quality educational system. When educators and policymakers began to question the validity or truthfulness of the statements produced by this theory, the theory was expanded to included inputs, process, and outputs - all contained within a learning context (Council of Ministers of Education, Canada, 2005b). Graduation rates and students’ performance on a standardized achievement tests were no longer sufficient bodies of learning to support the truth telling activities prescribed by the new model of student learning. The shift in the CMEC’s theorizing (Foucauldian statements of
knowledge) was culturally-produced re-thinking about what constitutes effectiveness of an educational system. Statements describing the shifts of cultural values and beliefs are found in the archive of the domain of interest (SAIP/PCAP large-scale standardized procedures and testing results) and also in the domains that occupy the correlative spaces juxtaposed to the regime of standardized testing.

The event is the bellwether, a nodal event that escapes classification, that signals a shift in the construction of statements used by a society to determine the truthfulness of statements; however, the event often “escapes out of rational grasp” (Foucault, 1980, p. 113). In a Foucauldian analysis, the event is the symptom of change; a particular society is no longer using the same statements to distinguish true statements from false statements (Foucault, 1980).

The events are revealed through the analysis of the archive of a discipline such as standardized testing but also of its correlative spaces – economics, politics, and education (Foucault, 1980). The archive consists of primary documents, traditional histories, journals, and public communications – anything that reveals how a society determines the truth about a particular domain. For example, the Council of Ministers of Education, Canada (2003a) draws attention to the notion that Canadians want to know how well their public schools are performing and that SAIP provides a measure of school performance. If this statement is operationalized as a true statement, there will be other statements or examples of events that echo the same sentiment; however, if this statement is not generally accepted by society as a true statement, there will be examples of other events that do not conform to this sentiment.
These discontinuous events are not the result of random acts, but they are evidence of the productive force of power that operates as a result of the production of knowledge (Foucault, 1980). Thus, the discourses of philosophy and science that produces knowledge are tied inextricably to the production of change through regimes of productive and positive power (Foucault, 1980). Foucault defines power as a force or mechanism whose “exercise ... creates and causes to emerge new objects of knowledge and accumulates new information for bodies of learning [that] operates on the social body” (Foucault, 1980, p. 51). This power is coercive and operates below the level of recognized authoritarian power (Foucault, 1980). As an example, Foucault draws attention to the power of the gaze operating as a form of surveillance in schools and prisons – one individual can observe and control many students/prisoners. Graham and Neu (2004) identify standardized assessment as a form of surveillance or gaze. The government can observe and compare the work of thousands of teachers and students through the instrument of the standardized assessment, a tactic and strategy of power that ultimately controls the choices and decisions of citizens.

The concepts of knowledge and bodies of learning are critical to the understanding and use of an archaeological methodology. The composition of knowledge includes philosophical ideas, institutional and commercial practices and legal regulation (police) while bodies of learning include formal bodies such as physics, chemistry, psychiatry or education (Scheurich & McKenzie, 2005). Knowledge creates the possibility for the production of bodies of learning such as those related to the development of education, a discipline with its own body of learning; thus, bodies of learning cannot be studied independently of the knowledge that created the possibility for
their existence (Scheurich & McKenzie, 2005). Thus, it is impossible to study the body of learning contained in the thematic topic like standardized testing without considering the knowledge – the policies, the regulations, the institutional procedures, and the regulations – that created the possibility for standardized testing as a body of knowledge to exist (Scheurich & McKenzie, 2005).

Dean (1994) describes Foucault’s archaeologies as an exploration of “how it is possible to think in a certain way and how far a specific language can be used” (p. 2) rather than the imposition of a structural analysis onto how it is possible to think in a certain way. Archaeology is not a method for putting the unsaid into discourse, nor is it a matter of identifying the relationship between continuous and discontinuous elements (Dean, 1994). Archaeology concerns itself with the organizing, dividing, ordering, and arranging documents. The archaeologist, as a researcher, engages in some or all of these tasks: the establishment of series, the discovery of elements, the identification of the relevancy between and among elements, the definition of unities, and description of relationships (Dean, 1994). The principle product is the production of a critical and effective history (Dean, 1994).

According to Dean (1994), Foucault’s contribution to historical sociology can be “best understood as a delineation of a form of history which is both critical and effective” (p.14). Although Foucault was preoccupied with the production of knowledge through the regulating effects of discourse, he chose not to apply the same rules of governance to his own heuristics; therefore, formal, fixed methodological rules have not congealed and are not available to subsequent researchers (Dean, 1994). Instead Foucault offers his technologies, techniques, and concepts to those who might adapt them to their own
knowledge projects by considering his examples and employing his ethic of self-writing (Foucault, 2006).

The task of the archaeologist/historian focuses on the creation of problems around the gathering, ordering and selection of historical resources to produce what society recognizes as its history; the result is a historically-based shift in perspective (Dean, 1994). History’s problems are constructed as themes and the archaeologist/historian questions the purposes to which history is put to use (Dean, 1994). The documents or data are not the raw materials of events past but the processed, conserved, and organized products of historical production; and the activities of archaeology consist of organizing, distributing, and establishing relevance of the historical document (Dean, 1994). Foucault (1972) argues that archaeology is not simply another iteration of hermeneutic inquiry, but the goal of an archaeological project is to seek a “pure description of discursive events” (p.27), and “to treat discursive facts as an irreducible reality. This irreducible reality is constituted as fields or systems of relations that form the conditions of discursive production (Dean, 1994). Archaeology as a methodology does not seek out the origins or the hidden depths of a discourse, but it does attempt to describe the system of formation and transformation of statements (Dean, 1994).

Foucault (1980) redirects the historical gaze away from the production of total histories focused on monuments, heroic events and reductionist generalizations. Unlike Lemann’s (1999) smooth, progressive history of standardized testing in the United States, which situates Henry Chauncey in the lead role as its promoter, a critical and effective history problematizes the statements provided within texts and searches for disruptions and transformations. A Foucauldian critical and effective history would displace the
heroic figure, Henry Chauncey, with an analysis of the shift in cultural views and beliefs that supported the replacement of birth right, social standing and wealth with performance on a standardized test as a key factor in the admission decision making process for postsecondary institutions.

In summary, Foucault (2007a) contests the notion that events and processes of history can be reduced to general statements about outcomes, and he describes archaeology as a methodology that preserves the factual aspect and singularity of events and processes that define the level of analysis. The archaeology of a discourse such as “a regime of standardized testing” creates a problem related to the relationship “between what we do, what we are obliged to do, what we are allowed to do, what we are forbidden to do” (Foucault, 2006, p.125-126). Because Foucault has resisted the temptation to allow his methodology to congeal into a discourse governed by rules and was determined to leave his methodology open to disruptive possibilities, researchers are directed to consult his metaphors, his examples, and the writing of other scholars while at the same time remaining open to possibilities (Dean, 1994; Foucault, 2006; Heyes, 2007).

**Genealogy**

The archaeology would be the appropriate methodology of this analysis of local discursivities, and genealogy would be the tactics whereby, on the basis of descriptions of these local discursivities, the local subjected knowledges were thus released would be brought into play. (Foucault, 1980, p. 85)

Although genealogy is often discussed in the absence of archaeology, they are complementary, additive, and somewhat intertwined. Genealogy should not be conceived as a new and improved version of archaeology but rather as a continuation and an extension (Scheurich & McKenzie, 2005). Archaeology pertains to the locating, the
isolating, and the determining of the expression of some event; whereas genealogy includes the mechanisms or tactics for tracking the expressions of local discursivities (Foucault, 1980). The motivation for Foucault’s shift from archaeology toward history, and his subsequent reflection on the usefulness of history, its purposes, and historical analysis remains somewhat uncertain (Dean, 1994). For Foucault’s own theorizing, the move toward genealogy represents a shift or transformation in the systems of rules that govern his own knowledge production.

From Dean’s (1994) perspective, the significance provided by archaeology is vested in its “rethinking of the place of study of discourse and rationality within historical sociology” (p.17), although archaeology fails to describe the “relationship between discursive and non-discursive practices” (p.17). The fundamental flaw with archaeology is the “way it perceives the rootedness of historical study in present problems, and the way it neglects to explicate the “how of the historical description of positivity of discourse is to be mobilized in terms of its current purposes and issues” (Dean, 2004, p.17). The positivism of archaeology provided a technique for determining the correct level of analysis of its objects; however, it failed to articulate its own purposes and its own relationship to the present (Dean, 1994). Thus, from the perspective of Foucault’s critics, his move to genealogy resolved the problem that was created by the rarefied form of positivism embedded in archaeology (Dean, 1994).

Foucault (1980) employs the metaphor of genealogy, the study of change, to describe his analysis of local discursivities and their respective relations of power. Through genealogy, biologists can follow both the persistence of traits (continuities) and the mutations (discontinuities). Thus, the metaphor of genealogy is an effective
representation of the study of the production of knowledge and its relations of power and regimes of truth (Foucault, 1980). For Foucault (1980), genealogy provides a subtle metaphor that aptly describes his knowledge projects. The task of the genealogist is the pursuit of the origin, the beginning, the instantiation of some phenomena (Scheurich & McKenzie, 2005). Foucault (1984) describes this pursuit as “an attempt to capture the exact essence of things, their purest possibilities, and their … original identity”; however, the essential ultimate expression of things is masked by the artefacts of culture (Foucault, 1984, p.371). Foucault (1984a) argues that what is behind things is not some essential secret but rather a piecemeal construction that represents an assemblage of alien components - not the result of some logical progression toward some goal.

To bring his genealogies into play, Foucault employs conventional historical analysis; however, he remains attentive to the pitfalls and foibles of these historical methodologies (Dean, 1994). His genealogies attack the presentism of Humanism by avoiding the temptation to analyze the past in terms of present (Dean, 1994; Foucault, 1980). In my critical and effective history, the contextualizing documents are organized by publication date, and reflective summaries written in the present about the past are not part of the study. Genealogies are produced as histories of the present because they contest the “necessary objects of our present reality” (Dean, 1994, p.33). In essence, national standardized testing regime was a response created in the past to address problems that existed in the past; however, its effects persist today.

Dean (1994) interprets the term ‘genealogy’ as “the union of two forms of subjected knowledge: the erudite functionalism of global theory and history; and the popular knowledges and local memories” (p.33). The opportunity for knowledge
production revealed by genealogy is located in the historical knowledge of human
struggles that inform the present (Dean, 1994, p. 33). Genealogy is born out of some
modern day struggle such as those struggles that teachers and students face when they
discover that the curriculum is shaped by the forces of standardized testing (Dean, 1994).
Although genealogies are born out of struggles, their outcomes are not emancipating or
liberating; nor can it be said that liberation or emancipation gives rise to the production of
a genealogy (Dean, 1994). According to Dean, genealogy is the approach for analyzing
“multiple, open-ended, heterogeneous trajectories of discourses, practices, and events,
and of establishing their patterned relationships, without recourse to regimes of truth that
claim pseudo-naturalistic laws or global necessities” (p. 35).

While an archaeology reveals the site of attack, genealogy “undertakes the
analysis of those objects given as necessary components of our reality” (Dean, 1980,
p.33) to reveal the fabrication of their origins, to show the imprint of their history on the
social body and to describe the system of subjection (Scheurich & McKenzie, 2005). The
production of moments of discontinuity indicate sites of transformation: At certain
moments some fields of knowledge appear to be taking-off, actively evolving,
confounding the expectation of smooth and continuous progress toward a goal (Chomsky
& Foucault, 1985). The features of disruption – the speed and extent - are the signs that
the rules that govern the formation of knowledge are shifting (Chomsky & Foucault,
1985). The genealogical question is not how did we progress, but how did things happen?

The Foucauldian Triangle: Truth, Power, and Self
Foucault’s writing produced the systematic displacement of the ideas, thought,
and science that comprise the history of truth (Dean, 1994). Foucault’s (1980) work
populated three domains and located three elements within each domain. The first
domain, truth, is described as the analysis of the regularities and formation of discourse
and is comprised of reason, truth and knowledge (Foucault, 1985). The second domain,
power, is described as the embedded-ness of discourses in institutional practice and the
power relations produced by those discourses, and is comprised of the power of
domination and government (Foucault, 1980). The third domain, self, is described as the
relationship of discourse to the self and forms of ethical conduct, and is comprised of
ethics, self, and freedom (Foucault, 1985). Flynn (1988) describes these triangular
relationships as the Foucauldian triangle.

Truth

What is the hazardous path the Truth has followed? What is the history of the will
to truth, what are its effects? How is all this interwoven with relations of power?
(Foucault, 1980, p.66)

According to Foucault, the production of the truth and the path to the will to truth
in Western societies is constrained by the ritualized processes of reason that prescribed
the requirements for authentication and attribution of knowledge (Foucault, 1980). The
path to the will to truth reveals itself in historical work as obstacles to its appearance are
disturbed and removed (Foucault, 1980). In modernity, the truth exists as the result of
exclusion; projects of science set hypotheses and reject those statements, facts and
observations that do not conform to what is expected (Foucault, 2006). The truth is
attributable, recognizable, documented, progressive, and teleological (Foucault, 2006).
Shifts in the path to the will to truth in human science are produced as systems of
statements are shifted and reorganized to reflect the opportunity for new trajectories for
knowledge projects (Foucault, 2006). Foucault (1980) describes the production of the
truth in Western societies as follows:
Truth is centred on the form of scientific discourses and the institutions which produce it; subject to constant political and economic incitement …; it is the object, under diverse forms, of immense diffusion and consumption (circulating through apparatuses of education…); it produced and transmitted under control … it is the issue of political debate and confrontation. (p. 132)

Thus, knowledge is not stable or definite; it is discontinuous, local, and rooted in the relations of discourse; and it is differential and incapable of unanimity (Foucault, 1980). Knowledge is produced through the questioning of the “universal historical subject and the continuous philosophy of history” (Dean, 1994, p.92) and “through the rediscovery of struggles and the memory of their conflicts” (Foucault, 1980, p.83). The goal of a Foucauldian critique is the contestation of the status of certain forms of knowledge and the privilege given to their modes of production (Dean, 1994).

Foucault contested the type of rationality employed in the projects of modernity: The human subject was both the investigator and the object of the investigation (Foucault, 1980). In a Foucauldian approach, the primacy of the human subject is de-centred and removed from the production of effective and critical histories because the purpose is not in the production of a new human science through application of a different system of rationality but in the contestation of the given-ess of the contents and the assumptions underlying the existing human sciences (Foucault, 1980).

The rational trajectory for knowledge production constructed by sciences such as biology and psychiatry does not follow the neat path of rational emergence but follows a path of irrationality characterized by messiness and complexity (Scheurich & McKenzie, 2005). Foucault’s contestation of the nature of reason moves beyond the bodies of learning to the very nature of the master narrative of the project of modernity as he enumerates its inherent flaws: its validity based on its own supremacy, its dogmatisms
contained in its structural autonomy, and its refusal to recognize its lack of refinement and increasing rationality (Scheurich & McKenzie, 2005). Examples of the limits of reason appear as ruptures in bodies of learning; these breaks and discontinuities in the regular production of knowledge further illustrate the limits of reason. The emergence of specialized fields of knowledge such as standardized ability testing created new bodies of learning; however, teachers working in classrooms have already developed informal tools (tests) to assess student performances and have learned to use the information generated by these tools to refine their teaching practices. The emergence of standardized testing as a body of learning was the result of political policy and social action external to the classroom’s processes.

Power: Sovereignty, Discipline, and Governmentality

Knowledge and power are integrated with one another, and there is no dreaming of a time when knowledge will cease to depend on power; this is just a way of reviving humanism in a utopian guise. (Foucault, 1980, p. 52)

Mechanisms and strategies of power are rarely the subject of research; history studies the individuals, the events, and the economic structures of a society or culture. The rules of right - authoritarian power - formally delimit power; but relational power produces effects of truth (Foucault, 1980). Knowledge produces relations of power; and relations of power produce regimes of truth (Foucault, 1980). Power does not exist apart from the discourses that produce it (Foucault, 1980). Thus, Foucault (1980) states that the formation of discourses is not defined “in terms of a particular object nor a style, nor play of permanent concepts, but ... in the form of a regular dispersion of statements” (p.63). Therefore, power resides in the formation of statements, and their ability to produce
action. Graham Neu (2004) illustrate the effects of such statements in the following example: After Alberta high school students’ parents read statements published in newspapers describing the results of standardized school leaving exams (summarized and posted by school), they had the opportunity to exercise their freedom to choose the school for their child (Graham and Neu, 2004). These regularly-produced statements produced action.

Power is constantly producing its effects on the forms and production of knowledge and the creation of information, and knowledge wages its effects on power by producing new relations of power (Foucault, 1980). Conducting a multifaceted analysis of knowledge is the key to unlocking how knowledge functions as a form of power (Foucault, 1980). Knowledge must be analyzed at its site of production (region), on its body of learning (domain), and through acts of displacement or disruption (examining the consequences of disturbing the taken-for-granted assumptions) (Foucault, 1980). The following serves as an example of Foucault’s analytic methodology. After the administration of SAIP/PCAP, the CMEC (domain) produces reports (bodies of learning) describing the performance of the Canadian students. These reports rank the summarized performance of students for each jurisdiction and rank the cumulative results of students in each jurisdiction. In most CMEC reports, the performance of the students in Quebec and Alberta is better than the performance of the students in many other provinces (Saskatchewan, Prince Edward Island and Labrador and Newfoundland) (CMEC, 2008). On the surface, it would appear that the curriculum or teaching approaches used in the provinces with lower test results were inconsistent with the testing protocol; however, by disrupting the site of knowledge production (examining the consequences of disturbing
the taken-for-granted), questions can be asked about sampling strategies for each province/territory and the motivations of CMEC relating to their decision to publish the results of a criterion-referenced testing tool in a norm-referenced manner (CMEC, 2008d). What response/action did CMEC hope to motivate from the provincial ministries of education?

Power was conceived as an original right of individuals; however, individuals have relinquished this power through historical contractual agreements in which individuals surrendered some of their power in return for state provision of safety, security and other social benefits such as education (Foucault, 1980). Sovereign power administered by the state flows downward through the hierarchical structure; its path is obvious and traceable (Foucault, 1980). However, disciplinary, relational power is circulatory and occupies the space below that which is visible; its presence is masked by the effects of discourse (Foucault, 1980).

Disciplinary power exists as a circulatory force of something that functions as a chain; it cannot be appropriated or localized; it exists as coercion (Foucault, 1980). Disciplinary power invades a space and exerts its coercive forces through the acts of normalization (Foucault, 1980). Considering the example of standardized testing regime, predetermined test scores serve as bars for achievement. If students or schools or provinces are unable to achieve the normative bar, they may be coerced into action through sanctions or psychologically imposed imperatives (Grineski, 2005). The coercion can be relatively subtle like the publishing of test results and distributing them to the appropriate ministries of education (Goddard, 2000). Or it can be more overt like the establishment of merit pay for teachers and substantive rewards for students that perform
above the expectation of the local school board or ministry of education (and of course, punitive consequences for teachers whose students perform below the standard) (Grineski, 2005). Robertson (1999) illustrates how the force of disciplinary power shapes the actions of teachers: They amend their curriculum and reorganize their teaching practice so that their students have an enhanced opportunity to achieve the normative performance standard – earn rewards and avoid sanctions. Disciplinary power produces knowledge such as that produced by teachers, learning experts, and students who describe strategies for succeeding at the testing game (Hansen, 2007).

And between the two, the games of power and the states of domination, you have technologies of government…To denote the two, Foucault used the term governmentality. It ... implies the relationship of self to itself, and … [covers] the ranges of practices that constitute, define, organize and instrumentalize the strategies which individuals in their freedom can use in dealing with each. (Foucault, 2006, p. xvii).

**Governmentality**

The concept of governmentality represents a refocusing of Foucault’s interest away from the local, the particular, and the singular to the macro discourses and discursive structures of government and conditions that facilitate the emergence of the mentality of government. In this paradigm, government is defined broadly to include any “rational activity undertaken by a multiplicity of authorities and agencies, employing a variety of techniques and knowledge, that seeks to shape conduct” (Dean, 199, p. 11). Therefore, based on this definition, CMEC would be described as a form of government. Mentality refers to the ways individuals’ think about the conduct of government and the strategies “that individuals in their freedom can use in dealing with each other” (Foucault, 2006, p. 300).
As governments evolved from structures of direct rule into their modern forms that rule at a distance, governing bodies have increasingly relied on implicit contractual arrangements with their citizens that support the individual’s right to act freely but also to act responsibly in every aspect of one’s life (Hay, 2003). These governments assume “that social subjects are not and should be subject to direct forms of state control,” but these subjects should be governed through “programs that shape, guide, channel – and upon responsible, self-disciplining social subjects” (Hay, 2003, p. 166). Foucault (1998) describes these actions of government as conduct. In the modern state, self-sufficiency and a type personal freedom can exist only if they moderated by self-responsibility and self-discipline (Hay, 2003). This moderation of human action in response to the indirect actions of government delivered at a distance is described as conduct (Foucault, 1998). Thus, agencies such as CMEC, operating as Foucauldian governments are responsible for developing programs and initiatives (governmental conduct) that have the potential to shape and guide the conduct of individuals and social groups toward to some socially, politically or economically important goal. Individuals have the responsibility to exercise their freedom to make decisions that support the maintenance of social order and the government’s economic, political and social goals (Hay, 2003).

From a research perspective, employing the concept of governmentality and the analytics of government would focus on the “mechanisms of government” (Bennett, 2003, p. 47); so a question might be how does CMEC get things done, rather than how does CMEC legitimate its use of power. The role of culture rises in importance in this type of research, as culture affects an organization’s ability to get things done. Drawing on the work of Graham and Neu (2004), the concept of governmentality emerges in their
discussion of the conduct of the Alberta government and the corresponding actions of parents residing in rural jurisdictions. Initially, the Government of Alberta excluded children residing in rural school districts from the opportunity to participate in provincial standardized exit exams because their populations were small and the administration of exams would have been difficult (Graham & Neu, 2004). The parents of the affected children protested to the government and demanded that they be treated equally and that their children should also write these exams. The parents demanded the opportunity to be placed under the surveillance of government; the actions of these parents illustrate how knowledge (information about school exit exams) has the potential to produce power (the demand to participate) that results in action (a form of conduct) and the subsequent production of knowledge (information about performance of students educated in rural communities) (Graham & Neu, 2004).

Although societies have changed and the styles of government have changed, governments continue to use the arts and technologies of pedagogy, politics, and economics to produce docile bodies and useful citizens (Foucault, 2007b; Foucault, 1997). According to Foucault (2007), the instrument or tool of critique poses the analytical question: “How does one not want to be governed?” (p.46). It is interesting that Foucault asks what one does not want rather than asking what one does want. Consistent with this thinking about what one does not want, most of the commentators on standardized testing write from the perspective of what one does not want. As an illustration, Robertson (2000) rejects the statements of the CMEC concerning the meaning of the data produced by the national testing protocol; she states her unwillingness to be governed by policies that from her perspective, limit the professional
duties and responsibilities of the teachers she represents. Unlike the parents in rural Alberta, Robertson is unwilling to demonstrate obedience because she does not accept the truth that CMEC is promoting.

Foucault links governmentality with regimes of truth by stating that an implication of rejecting the need for government or setting boundaries on the extent of government is a rejection of the truths promoted by government. CMEC activities related to the regulation of conduct continue through acts of relational power, informal agreements, and shared cultural beliefs concerning the importance, at this time in our history, to have a body that speaks on educational issues of national interest. However, the Canadian people did not petition or vote on the issue of a national standardized testing protocol, nor did parliament pass a law to implement a national regime of standardized testing. (Council of Ministers, of Education, Canada, 2003a).

Foucault’s concept of governmentality is enriched, supported, and served by the use and exploitation of the concept of population. The aggregation of individuals into communities have created problems for governments; as the government must be able to provide and predict the need for services, deliver education, protect wealth, and create employment for these communities; however, the conduct of a population can be shaped by the introduction of a program (Foucault, 2006). Since groups of individuals need quantities of services and can provide quantities of labour, governments have developed statistical procedures to quantify, enumerate, and reduce individuals into populations for the purpose of monitoring and shaping their conduct (Foucault, 2006).

Standardized testing is based on the notion of population; its products are generalized results presented in a reductive fashion. These data collection projects
quantify groups of individuals; and through statistical analysis, individual group members are reduced to a population with specific characteristics distributed in a particular pattern that permits comparison with other populations (Graham & Neu, 2004). As an illustration, the test scores from SAIP/PCAP are aggregated by gender, age, and province to facilitate ranking and the production qualitative statements about the relative performance of each population (Council of Ministers of Education, Canada, 2003). Individuals disappear; they are no longer part of the discussion. Their contribution to the data pool is enmeshed, collated, summarized, and reported as part of a population’s characteristics that may not be representative of any one individual.

The concept of population is just one example of a social technology that governments use to conduct or “lead, direct or to guide” in some moral sense the ordering and managing of risk in fields of social enterprise like the economy or education (Dean, 2010, p. 2). Governments-at-a-distance get things done through the exercise of conduct on the conduct of others. Their practices attempt to “shape, mobilize, and work through the choices, needs, wants, and lifestyles of individuals and groups” (Dean, 2010, p. 3). Individuals use knowledge such as that produced by through the activities of standardized testing to regulate and discipline their decision making and the exercise of their right to choose. Individuals, like governments, want to maximize order and minimize risk in the conduct of their lives. Therefore, parents who read the league tables of standardized test scores released by the media are concerned about the placement of their child’s school on that roster, and they use the league table information to manage their child’s future risk and to ensure order in their daily lives. Although the thinking “involved in the practices of government is explicit and embedded in language and technical instruments,” this way
of the thinking about conduct on conduct is part of the taken-for-granted mentalities of government (Dean, 2010, p. 10).

Methodological Perspective: Ambiguities and Interpretations

Foucault’s methodology, methods, and concepts have been re-interpreted, re-written, and re-conceptualized by scholars and commentators working in various fields of human inquiry. Foucault, himself, would have endorsed this adaptation, re-working, and re-writing of his techniques and technologies in the production of effective histories of the present (Foucault, 1980). The greatest tribute to the thought of a writer is “to use it, to deform it, to make it groan and protest …and then say if I am being faithful or unfaithful…that is of absolutely no interest” (Foucault, 1980, p.54). Thus, writers and scholars should appropriate Foucault’s methods and concepts, stretch them, make them moan and protest as they struggle with their knowledge projects. With this general permission, to take Foucault’s methods, concepts, and ideas, Scheurich and McKenzie (2005) issue this gentle admonition to be careful: Selecting often-used concepts like disciplinary power, but neglecting Foucault’s foundational ideas can lead to disastrous outcomes. For me, this admonition led me to consider Foucault’s original questions pertaining to the nature of truth, the production and nature of truth telling discourses, and the relationship of culture to truth telling; and led me away from the singularity of a notion such as discipline.

Because Foucault remained resistant to the notion of articulating a specific methodology with a set of strategies that articulate the processes and methods for executing research based on his writing, researchers have appropriated his concepts or
methods, have rewritten his methodology in comparative studies, and have blended his methodologies with theories and tactics of other scholars. Foucault (1972) relies on the power of a metaphor embedded in illustrative concepts and examples to convey his ontological position, and he leaves the reader questioning his intentionality and his need to remain open to possibilities. Kolenick (2006) states that “one of the drawbacks of conducting a discourse analysis using Foucault’s archaeology and genealogy is that there is no ready-made recipe or procedure on how to analyze text” (p. 63). Dean (1994) illustrates this concern by stating Foucault fails to offer a “rationalistic plan to put into practice his analysis” (p.2).

Researchers – Heyes (2007), Scheurich (1997), and Dean (1994) - have purposively illustrated the elasticity of Foucauldian concepts in a variety of domains and their subtlety to withstand the twisting and squeezing of re-writing and re-thinking through continued acts of self-writing. Thus, the conundrum for researchers is the lack of boundaries, limits, guidelines, and recipe books that can govern what can and cannot be said. The determination of a methodological lens becomes a methodological problem: The lens determines what lies within and between the metaphors, what historical processes are possible, and what knowledge can be produced.

**Method**

In framing my research method, I have relied on Foucauldian concepts and definitions to set the parameters for my study of Canada’s national standardized testing program (Foucault, 2007a; Foucault, 1980). I have used his genealogies of the French penal system (Foucault, 1984b) and ethics (Foucault, 2006); Dean’s genealogy of poverty
(Dean, 1994); Heyes’ work on self-transformations (Heyes, 2007); and Graham and Neu’s work on the system of provincial school leaving assessments in Alberta (Graham and Neu, 2004) as my illustrative examples. I have heeded the cautionary tones of Scheurich and McKenzie (2005) to avoid selecting only one or two of the conventional and convenient elements of Foucault’s work while neglecting other critical elements cogent to the conceptualization of Foucauldian research and the production of a critical and effective history of SAIP/PCAP.

Foucauldian Concepts

Throughout Foucault’s writing, concepts and examples illustrate his historical method; these concepts are often integrated within his work and are not carefully articulated as a specific aspects or elements of his research method. The Foucauldian method begins with the fundamental concept of history as a somewhat neglected or taken-for-granted discipline, as he states that history is “the mother of all sciences of man, and … as old as human memory,” and for that reason “we have … passed it over in silence” (Foucault, 2007a, p. 400). Foucault’s thematic of the taken-for-grantedness permeates not only his conceptualization of history as a passed-over discipline, but also his directive to researchers to trouble or problematize the taken-for-granted assumptions about a thing and how a thing operates in society (Foucault, 1980). Taken-for-grantedness can take the form of a simple appeal to all Canadians as the following example illustrates: The Council of Ministers of Education, Canada, urges us to embrace this regime of truth: “All Canadians want to know how well their public schools are performing and that standardized test results are the manner by which most Canadians want to receive that feedback” (Council of Ministers of Education, Canada, 2007, p. 1).
As I contemplate this assertion by the Council of Ministers of Education, Canada, I wonder when I was asked or how I gave my consent to this process. I wonder if all Canadians believe that our schools should be measured by standardized testing or has the need for accountability become a concept to which we have become conditioned. Does the Council of Ministers of Education, Canada, believe that accountability has a value in and of itself? Has accountability and measurement become one of the taken-for-granted assumptions of the Canadian educational system? (Dagenais, July 6, 2009)

Throughout this study, it is my task to trouble those assumptions though critical analysis.

History involves “ordering the time of human beings upon the world’s development,” the sequencing and ordering events into some chronology (Foucault, 2007a p. 400). The documents and bits and pieces of detritus that form the archive of human experience serve as the evidence of this ordering (Foucault, 2007a). The archive, according to Foucault (1980), is not a place but rather documents – texts, journal articles, newspaper articles, and memoranda - that contain the discursive formation indicative of the field. The discursive formations of field represent the statements that regulate what can be known of a discipline and how it can be known: the discourse defines the code of a “natural rule, a norm” (Foucault, 1980, p. 106). My task is to uncover these natural rules or codes, ask how they were produced; what event or discontinuity or eruption marked a change in the easy flow of history that led to their production and to their change or discontinuation (Foucault, 1980).

Foucauldian research method privileges temporality over space or geography. Foucault (1980) argues that it is the event, the disruption of discursive formation that is the focus of research rather than the place in which it occurs. There is a need to ponder this type of question: Is this a change that is localized and specific or is it global and more general in its appearance? As I consider the Canadian national standardized testing
program, I am constantly reminded that standardized testing in the United States has a long history, and the American experience informs and influences the Canadian experience. The Canadian national standardized testing program is intertwined with the international testing protocol Programme for International Student Assessment (PISA). Human Resources Development and Skills Canada (HRDSC) funds the national testing program (SAIP/PCAP) and PISA, facilitates the administration of PISA, and uses the results of PISA in its Longitudinal Youth in Transition Survey (Looker & Thiessen, 2008). Although SAIP/PCAP is a Canadian invention, its production and its outcomes are influenced by factors outside the limits of Canadian geography. The academic performance testing of school-age children is an international phenomenon that influences and is influenced by economic and political policy making within Canada and beyond its borders (Looker & Thiessen, 2008).

The second characteristic of temporality is periodization, the segmentation of the flow of history into manageable chunks to facilitate a particular level of analysis and to create opportunities for the discontinuities to reveal themselves through the archaeological analysis (Foucault, 1980). In my study of SAIP/PCAP, the period length is a decade because the discussions about the formation of the CMEC are recent – beginning in the 1960s, and those about a tool like SAIP/PCAP began to populate the literature in the 1990s. The length of period affects the level of visibility of disruption – if the period is longer, the disruptive event must be larger to gain visibility against the cultural landscape of society; if the period is shorter, smaller disruptive events may reveal themselves (Foucault, 1980). Because the event – a disruption of discontinuity - is the indicative marker of a change in codification of a rule or norm, the period length must
facilitate the appropriate level of analysis and the revelation of these critical events that demonstrate changes in discursive formation (Foucault, 1980). My task, as the researcher, is to determine the period length, the granularity or level of the analysis, search out the event markers that denote a change in the discursive formation of statements, and describe the forces that produced the changes.

Both Foucault (1980) and Dean (1994) employ the concept of historicity, the belief that social structures, such as the CMEC; events such as the administration of SAIP/PCAP; and texts such as the tests, the reports, scholarly journal articles, and CMEC reports are understood best within the context of their historical development (Foucault, 2007a). Historians should resist the urge to understand the past in terms of present. To facilitate understanding the past in terms of past, Foucault (1980) urges researchers to explore the correlative spaces that surround a discipline of interest – for standardized testing, these spaces include the fields of education, politics, the economy and globalization. Drawing on the Canadian experience with standardized testing to illustrate Foucault’s point, I must consider not only the Council of Ministers of Education, Canada’s decision to implement a national program of standardized testing in 1991, but also, what was happening to the Canadian economy at that time, the Canadian public’s attitudes towards schools and teachers, and their perceptions about the overall health of Canadian education system, and Canadian attitudes toward a global economy and trade integration with the United States (Barlow & Robertson, 1994).

Dean (1994) describes Foucault’s histories as critical and effective. The notion of critical comes from of Foucault’s discussion of criticism in The order of Things (2007a) as a comparative term in contrast to commentary: Criticism is a type of analysis that
operates at the surface, and its goal is to establish the understanding of how things are produced and what forces cause their production while the goal of commentary is to dig below the surface to reveal the reasons for a particular event or phenomenon.

Subsequently, Foucault (1980) refines his notion of criticism into archaeology, a form of “analysis of local discursivities” at the site of their production” (p. 85) which means that discursivities are fixed in a particular period – time - and operate in and between a number of other fields or domains. My research task is to examine the periods for changes in production of these discursivities related to regimes of standardized testing in Canada and to resist the urge to compare these period-specific norms with the present.

_Tactics and Examples_

Foucault (1980) refers to genealogy as “tactics” by which local discursivities are revealed; however, he does not present a practical toolbox filled with gadgets and tools that a researcher can use in a tactical assault on a research problem. He does, however, introduce these broad organizational concepts: discipline, a coercive force of normalization; governmentality, relationship of the individual with the technologies of government; and care of self, the relationship of the individual with the self (Dreyfus & Rabinow, 1983). These broad organizational concepts or tactics provide a lens to analyze locally-produced discursivities. Graham and Neu (2004) use the concepts of discipline and governmentality as the theoretical organizational framework through which they reveal the production of knowledge, relations of power and regimes of truth in the Alberta school leaving testing program. Heyes (2007) uses the lenses of discipline and care of self to reveal the knowledge production and truth telling by participants in a
Weight Watchers program. These examples inform my analysis of the activities of the CMEC and their administration of SAIP/PCAP, but these examples and organizing lenses do not provide a ready-made rubric for analysis, but rather possible models that can offer some useful bits and pieces of thoughtful analysis.

In my study of standardized testing, I will appropriate a model delineated by Dean (1994) as Foucault’s triangle containing the concepts of truth, power and knowledge and their productive relationships. From the perspective of this study, the truth and acts of truth telling are culturally produced and have the capability to shift in response to the production of knowledge and the effects of power. The field of standardized testing can be perceived as a rational field dedicated to seeking of causality and causal relationships between teaching, learning, curriculum, social factors, and students’ ability to perform on the testing tool. This study occupies the space between the correlative spaces of economics, cultural sociology, and politics, and the rational space of standardized testing.

Limitations

All methodological perspectives place limitations on what can be studied, how it can be studied, and the possibilities for knowledge production and truth telling (Foucault, 2007a). The results of this Foucauldian research will produce an understanding of how SAIP and PCAP were produced as effects of power, how SAIP/PCAP may function as a regime of power, how SAIP/PCAP may serve as agents of knowledge production, and how SAIP/PCAP may operate as a discursive formation that supports possibilities for truth telling about the functioning of educational system in Canada (Foucault, 1980). This research is not only about revealing possibilities for understanding how standardized
testing in Canada operates, but also about revealing the possible sites for other types of research activity (Foucault, 1980).

While Foucault labels his methodology as critical, it does not address the need to shift or change power relations by offering solutions to the complex social problems; however, it offers the production of understanding as a necessary precursor to any solution seeking action (Foucault, 2006). Heyes (2007) states that Foucault’s concepts of discipline and care of self are useful analytic tools; however, the failure to include solution making as part of the methodology is a short fall. For me, the openness of Foucault’s methodology to imagine new possibilities of meaning making and the production of understanding is strength. Scheurich and McKenzie (2005) label Foucault’s critiques of the social forms in which we live are unrelenting and totalizing, and they comment further that these critiques fail to provide an exit/possibility (hope). Foucault does provide a number of concepts and tactics that can serve as methodological lenses; however, he does not provide the toolkit or recipe book (Kolenick, 2006). Because Foucault was a prolific writer and his concepts evolved with his methodological perspective, these concepts and tactics are embedded in a number of texts; thus, the researcher must patiently piece together the concepts and examples to derive a methodological approach.

As I consider writing a critical and effective history of Canada’s standardized testing program, I am drawn back to Foucault’s urging to researchers to become engaged in domains in which they have struggled (Foucault, 1980). To have struggled in this field of standardized testing means that I, as the researcher, have been affected by the power relations, have engaged in truth telling, and have had the opportunity to witness the
functioning of discursive formation in the determination of authentic culturally-produced knowledge (Foucault, 1980). That means, as I reflect on the field of standardized testing, I am in that field of knowledge; I am both a producer of knowledge and an object to be known (Foucault, 2007a). The archaeology and genealogy that I produce is just one of many possible genealogies and archaeologies situated in this field. Although Foucault states that he is producing *an archaeology* and *a genealogy* as he refers to his work on geography, the indefinite articles may easily be missed.

I am uneasy as I read and reflect on Dean’s (1994) statement that describes Foucauldian’s histories as neutral, but how can these archaeologies and genealogies be neutral? Foucault (1980) sets out as a condition: The researcher must be embedded in the struggles of research domain – how can the researcher be both neutral and embedded in the struggle? Foucault, (2007a) himself, began his knowledge projects by asking how subjectivity is produced by the knowing subject who is also an object to be known. Does this mean that my genealogies and archaeology of standardized testing is really a product of my own subjectivity? As I read and re-read books and journals, I relate the commentaries and journal articles about standardized testing to my own experience. I select the texts based on my situated-ness in the domain. Is this a methodological flaw, or this situated-ness a critical element of Foucault’s methodology? Scheurich and McKenzie (2005) remind us that Foucault’s emphasis on decentring the subject is the critical element in Foucault’s methodology? Is this notion of decentred-ness the focal point of my anxiety with the methodology? (Dagenais, July 7, 2009)
SAIP/PCAP is just one example of a national testing protocol operating within a larger space of standardized testing regimes and the economic and political agendas of nations. Some of the regimes of standardized testing like *Programme for International Student Assessment* (PISA) are global in nature and seek to normalize the academic performance of students from 66 countries (Organization for Economic Co-operation and Development, n.d.). In Canada PISA is coordinated with the administration of PCAP, the Canadian standardized assessment test for 13-year olds. Other regimes of standardized testing are local in nature like the *Texas Assessment of Knowledge and Skills* (TAKS) and are focused on state school accountability and the provision of school performance information for parents, but these local regimes influence or are influenced by the testing activities of other jurisdictions. Texas, the second largest state in the union with its growing political power, is often considered to be a bellwether for activities in other jurisdictions in the United States and other countries because “what happens in Houston affects the nation” (McNeil, 2000, p. xxi).

Similarly Ontario’s large population and economic activity sets the trend for Canadian education; however, the messages distributed by Ontario’s departments within the Ministry of Education differ in tone from those of Texas and from each other. Jackson (2007) describes the Education Quality and Accountability Office (EQAO) as a “catalyst for student achievement in Ontario by measuring student’s foundational skills of reading, writing and mathematics [through the administration of large-scale assessments] according to the expectations established in Ontario’s curriculum” (p. 1). However, the
Ontario Ministry of Education website offers these almost apologetic comments about its limited ability to provide information to parents about the performance of Ontario schools: “We don’t have standardized tests, we don’t have high school leaving exams, we have little by which to measure school progress (or the lack thereof), and little useful data for parents” (Keller, April 9, 2009). These are examples of some of the global activity in the field of standardized testing and its correlative spaces; they are illustrative of Foucault’s (1980) description of the need to remove territorial boundaries from the production of a critical and effective history; and they demonstrate the need to attend to what is happening in the correlative spaces surrounding the domain of the study.

As I reflect on McNeil’s (2000) comment “accountability by standardized testing is a panacea for school reform” (p. xxi), I wonder why one department of the Ontario Ministry of Education promotes a form of standardized testing (without naming it as such) with data collection, stakeholder reporting, curriculum reform, and school reform attached to it, while the Ontario Ministry of Education website offers statements which seem almost apologetic because it states that it does not have a regime of standardized testing and school accountability. How is standardized testing perceived in relation to school reform in Ontario as compared with Texas? Why does the Ontario government feel the need to be somewhat apologetic to its citizens about the lack of standardized school leaving exams? Is this a deficit that the Ontario government feels that it should fill, or is this a perceived gap that it wants to fill? Is the government trying to gather a sense of the political appetite for Texas-style knowledge and skills assessment tests with a similar capacity to publish school performance results? Is this essentially an attempt to sort the schools and provide an opportunity for the market forces and the media to do the rest of the job? Is the Ontario government concerned about using the word standardized in conjunction with the word testing? What does this say about the cultural impact on the production of similar phenomena in different territories? Why is reporting school performance using standardized test scores appropriate in Texas while in Ontario, it appears to be both appropriate and inappropriate? What types of truth telling are happening and how does the “who” part of the truth telling activity affect what truth can be told? (Dagenais, September 30, 2009)

In summary standardized testing transcends the boundaries of communities, provinces, and countries to occupy a space in the fields labelled as global or international;
and standardized testing transcends the field of education and influences the fields of politics, cultural sociology, and economics. In many ways standardized testing is a global phenomenon as it seeks either intentionally or unintentionally to serve as a “multidimensional set of social processes that create, multiply, stretch, and intensify worldwide social interdependencies and exchanges while at the same time fostering in people a growing awareness of deepening connections between the local and the distant” (Streger, 2003, p. 13). Steger’s definition assists us to understand that fields like standardized testing transcend geopolitical boundaries, and it provides confirmation that geography or territory does not limit the production of knowledge in a particular field or domain. According to MacGillivray (2006), the phenomenon of globalization is marked by events - such a spread of standardized testing to countries all over the world - that intentionally embrace the globe. These events, notes MacGillivray, are not slow and gradual in their appearance; but they are short, jerky, disruptive, pulses that create “planet-shrinking moments when people from all walks of life seize ... [a] new opportunity” (p. 17). MacGillivray’s definition of globalization echoes sentiments embedded in Foucault’s (1980) description of the event as some thing that is by nature disruptive.

Standardized testing has permeated domains of education, economics and political policy construction as the results from tests like PISA and SAIP/PCAP influence how we, as Canadians, have been encouraged by the Council of Ministers of Education to think about our own education system in relation to our political and economic advantage and to consider how the educational system contributes to human capital development (Council of Ministers of Education, Canada, 2008). In particular, the integration of
regimes of standardized testing and the use of standardized tests results as an indicator of progress have influenced the construction of Alberta’s 20-year plan for the development of an educated and technologically savvy society that is not only an economic competitor but also significant contributor to global research and development (Alberta Advanced Education, 2006).

These illustrations provide some glimpses at the global nature of standardized testing; the influences that one regime of standardized testing may have on another; and the interdependence of the fields of education, economics, political policy construction, and cultural sociology within the domain of standardized testing. Exploring the discourses related to standardized testing from a global perspective exceeds the boundaries for this study; however, in an attempt to provide some sense of the correlative spaces surrounding SAIP/PCAP and the nature of the conversations in those spaces, a snapshot of the of American, English, and international regimes of standardized testing will follow.

American Voices: Who Speaks about Standardized Testing and What Do these Voices Say?

The American perspective on standardized testing occupies one of the correlative spaces surrounding the implementation and delivery of SAIP/PCAP. The development and implementation of regimes of standardized testing has been part of the American educational landscape and economic policy construction since the inception of universal education in the United States (Haladyna, Hass & Allison, 1998). As the United States is not only a geographic neighbour to Canada but also economic partner, political ally, and
a contributor to North American cultural industries, the influence of the American
educational sector is both incidental and direct (Barlow & Robertson, 1994).

The critical and effective history of the American experience with standardized
testing can be characterized by shifts or changes produced by five Foucauldian events or
disruptions (Haladyna et al, 1998). These disruptive events have produced knowledge as
the truth telling statements that determine what counts for truth, and these statements
have shifted to reveal new understandings of the world of assessment, education,
economics, and political policy making (Foucault, 1980). Although these events seem to
be ruptures to the way in which American universal education was thought into existence,
Haladyna et al (1998) and Sacks (1999) argue that the shifts or events have done little to
disrupt the taken-for-granted space educational assessment. From a Foucauldian
perspective, standardized testing has worked as a program of governmental conduct that
has shaped and directed the conduct of citizens towards a mode of self regulation and
self-discipline that supports the orderly achievement of American economic and political
goals. The events represent acts re-writing or revising the program of governmental
conduct.

Event 1: The Production of Public Education

Public education began in the United States as a grand policy experiment in the
1800s; the goal was to provide educational opportunities to all citizens while maintaining
cost efficiency (Haladyna et al, 1998). This mass education initiative shifted a few public
dollars towards the education of the masses while the American meritocracy continued to
participate in privately-funded preparatory schools and colleges outside the scrutinizing
gaze of government policymakers (Lemann, 1999; Sacks, 1999).
The emergence of standardized achievement testing in the United States has been linked with the democratization of education during the period from 1840 – 1875 (Haladyna et al, 1998). As the American policymakers turned their focus from educating the elite to educating the masses, the results of standardized achievement testing provided the evidence of student learning and school performance (Haladyna et al, 1998). The government used the introduction of achievement tests as a means to provide public assurance that all children, regardless of status, could be assured of similar educational outcomes and that schools and education were serving as an economic and social equalizing force (Sacks, 1999).

However, the intention and purpose of the administration of standardized tests migrated away from the measurement of academic achievement of students to the comparison of “schools and children without regard for non-school influences” and to the provision of data to support arguments for school reform (Haladyna et al, 1998, p. 263). Although writing about a much later period, McNeil (2000) reiterates that standardized testing has and continues to be a panacea for school reform and school reformers; the standardized tests results provide the data for school-to-school comparisons upon which the cry for school reform and teacher criticisms are often based. This demand for data is fed, in part, by an American fascination with data, mental measurement and IQ, and the American cultural obsession with the quantification and representation of an individual’s intelligence by a single number (Sacks, 1999).

Test scores formed a vital and integral element of the decision making process for school and educational services budget construction and resource allocation at the local, state and national levels (Sacks, 1999). The administration of achievement tests formed
an integral component of governmental surveillance of the educational system (Vinson & Ross, 2001). However, to advance their goal of efficiency, some American schools used standardized test scores to limit opportunities for some students – particularly those living in poverty or those who lacked the language skills necessary to succeed in a regular American public school classroom (Haladyna et al, 1998). Typically these students required additional resources, specialized instructional expertise, and additional time in the educational environment to experience some form of measurable success on a standardized test.

This policy focus on providing efficient education worked well in the 20th century; the United States led high school completion rates for developed nations. However, in the 21st century the United States high school completion rate dropped to 21st position out the 27 reporting nations (College Board Advocacy, 2008). Many commentators, researchers, teachers, and parents point to an over reliance on the results of standardized testing and their use in educational policy construction, curriculum design, local school management, budgeting, and resource allocation as one of the causes for declining high school completion rates. While other nations are investing in the development of strong curriculum and services to meet the needs of a diverse student population, the American public school system is still invested in ranking and comparing schools, teachers, and students through the application of standardized testing tools (College Board Advocacy, 2008).

At the beginning of the 20th century, the focus of American standardized assessment regimes shifted to include both achievement testing and ability testing (Haladyna et al, 1998). The SAT (Scholastic Achievement Test), an achievement test,
was introduced to Harvard University by Henry Chauncey in 1901 as decision making tool to facilitate the awarding of scholarships to students from disadvantaged socioeconomic backgrounds (Lemann, 1999). In 1916, Binet’s scale, “a diagnostic tool for assessing the developmental progress of children,” was published in the United States where it was enthusiastically greeted at the University of Chicago (Sacks, 1999, p. 25). Under the guidance of Spearman and Terman, the Binet scale was refined; the results classified; and American norms established using test data produced by children in American public school classrooms (Sacks, 1999).

According to Sacks (1999), Terman extended the current understanding of what intelligence could be or might be through his definitions of potential ability measured by intelligence and aptitude tests and performance or achievement measured by course grades and other performance measures. According to Sacks (1999), Terman’s work on Binet’s simple scale “would become to the measurement of human minds what Einstein’s famous expression about energy, matter, and the speed of light would become to human warfare” (p. 26). Although Terman discovered a high correlation between the results from the Binet Scales and social status, he offered that American public schools were the great equalizer and eliminated any negative impacts of social status on test results; thus, the results accurately reflected children’s potential (Sacks, 1999). These revised Binet Scales provided an opportunity to sort, categorize, and prescript educational offerings based on students’ ability to benefit; thus, the provision of mass education could increase its efficiency as educational opportunities could be provided to those who would benefit most, and alternatives to public education could be offered to those who seemed to be unable to benefit (Sacks, 1999).
Event 2: The Introduction of the Multiple Choice Test Item

Following the introduction of the Binet Scales to the American educational system, the next shift or rupture in the assessment of children’s learning and assessment of school performance was the introduction of multiple choice test items and the abandonment of essay-type questions (Haladyna et al, 1998). The introduction of multiple choice test items supported American education policy makers’ priority of providing efficient mass education. The introduction of multiple choice items reduced the cost per student and increased assessment capacity while facilitating school-to-school, teacher-to-teacher, and student-to-student comparisons (Haladyna et al, 1998). In 1923, the Stanford Achievement Test was introduced to assess reading and mathematics skills in elementary school children; and following its introduction, the University of Iowa launched the Iowa Test of Basic Skills in 1935. Both of these achievements tests have been revised and updated to take advantage of the digital technology (Coffman & Matthews, 1973).

According to Haladyna et al (1998), these tests captured the imaginations of efficiency experts who sought to reduce school funding, define inputs, and measure outputs through administration of externally-produced standardized tests which were unconnected with the classroom curriculum. Despite multiple revisions, their critics still argue that these tests do not measure the students’ ability to perform real classroom work today, yet the data generated by these tests is still considered by many educators to be a reasonable representation of student, teacher, and school performance (McNeil, 2000).

From the 1940s through the mid ‘60s, American students’ test scores on the Iowa Test of Basic Skills, Stanford Achievement Test, and the SAT rose; American citizens
felt comfortable that the school system was working; students were learning more each year than they had the year before (Harnischfeger & Wiley 1976). However, the static or declining test scores of the mid 1960s sounded a warning – American citizens wondered if something wrong with the educational system, the tests or the students (Harnischfeger & Wiley, 1976). Harnischfeger and Wiley’s research pointed to changes in students’ in- and out-of-school activities and behaviour, as well as, changes in the test items – perhaps these factors may have contributed to the declining scores. According to Vidich and Bensman (1960), the role of the school began to change as youth were no longer focused solely on jobs and careers in their local communities, and policymakers became more focused on the national economy and the preparation of national labour force in the late 1950s. These factors, in combination, seem to work as a transformative force – youth were no longer performing on standardized tests in the way they used to perform.

In reaction to the declining test scores, Summers and Wolfe (1977) postulated that any child’s academic performance could be determined based on a quantification and subsequent analysis of the following factors: genetic capacity and the child’s social economic status, teacher quality, non-teacher school factors, and peer group characteristics, and a constant factor $f$. If the factor responsible for the decline could be identified through statistical analysis, it could be remediated – and the system, the child, or teacher could be fixed (Summers & Wolfe, 1977). Data from standardized test scores and the accompanying statistical analyses could provide the information that would lead to the solution of the declining test scores produced by American children.

The American public was encouraged to believe that improving and increasing the nation’s capacity to assess children’s learning through standardized testing accompanied
by statistical analysis would identify and isolate the issue/problem responsible for the decline in test performance (Sacks, 1999). The federal government under the leadership of Lyndon Jonson enacted Title 1 of the Elementary and Secondary Education Act (1965), a well-meaning federal measure to assist schools in poor neighbourhoods to improve their ability to provide appropriate education for impoverished children; however, all funding under this act was tied and continues to be tied to production of standardized test scores on the Iowa Test of Basic Skills (Sacks, 1999). The federal government’s implementation plan for the distribution of Title 1 funds provided for rigorous scrutiny of test scores (Sacks, 1999).

**Event 3: The Introduction of Criterion-Referenced Testing**

Tyler’s work in the 1930s sparked an interest in criterion-referenced assessment approaches which was reinforced with Glaser’s work in the late’60s. Glaser (1966) proposed that sound instructional design for any domain (reading, mathematics, writing, etc.) should be based on an analysis that would yield the subject-matter competencies or learning outcomes. Glaser specified that an educational technologist should conduct an analysis that would be characterized as scientific using the language from the natural sciences to frame the analysis. Key elements of Glaser’s approach included specifying behavioural competencies and conditions in terms of learning outcomes and the interlinking of assessment and instruction. Key to Glaser’s model of instruction and assessment was the linkage between the two as instructional should drive assessment and assessment should drive instruction. Both the instruction and the assessment were produced by the teacher or instructional leader and responsive to the needs of the learner.
Offering internally-produced criterion-referenced assessments was a dramatic departure from the norm-referenced externally-produced standardized testing regime required for Title 1 school funding.

The focus of assessment specialists’ interests shifted from seeking definitions for criterion-referenced testing and advocating its adoption in the 1960s to the analysis of the statistical problems produced by norm-referenced standardized testing in the 1970s (Meskaukas, 1976). Campbell (1988) sounded an academic warning urging social scientists to be cautious in the ease with which they embraced generalizations about human performance and behaviour in fields like education based on the statistical analysis of data produced through administration of standardized tests. According to Campbell (1988), these generalizations, often proffered by policymakers and standardized testing advocates, have limited relationship to the school curriculum and reduce or obfuscate the effects of pluralism, multiple paradigms, and temporality of the performance. The setting in which the students learn and live and essentially the individual disappear (Campbell, 1988). Campbell stated that “Subjectivity is the norm ......not objectivity” (p. xii) for production of a social science; thus, producing millions of children as objects of a standardized testing data gathering process will, at best, only approximate “what we hope to achieve” (p. 27). According to Campbell, multiple measures of the same phenomenon – like students’ responses on teacher-generated assessment tools based on the specifications of the curriculum - are more likely to bring us closer to understanding what we want to know than a single measurement of some narrowly defined variable. Over the decade, the focus had shifted from advocating for an alternative to norm-referenced assessment to identifying the limitations of norm-
referenced standardized test results and the flawed nature of the statistical analysis of the data.

During the 1970s, Wise (1977) reflected on the merits of rationalizing school services and instructional delivery through the administration of regimes of high stakes standardized testing. According to Wise, the school system, the school, and the provision of educational services are delivered through a system that operates with its own internal and contextual logic. The addition of external systems of high stakes standardized testing to determine the efficacy of the curriculum, the professional competence of the teachers, and the learning gains of the students is an example of hyperrationalization. *Hyperrationalization* as related to standardized testing and the delivery of instruction manifests itself through the following:

- The engagement of policymakers in the prescription of outcomes such as the provisions and requirements of Title 1 funding (mandatory student testing using the Iowa Test of Basic Skills) (Sacks, 1999),

- The development of inappropriate solutions such as funding the development of minimum-competency testing as opposed to investing in strong curriculum and teacher preparation (Shepard, 1991),

- The design of first order solutions (using the words in the perceived problem to design the solution) such as the introduction of accountability programs based on externally-derived goals and measurement tools rather than consulting with teachers to produce locally-produced accountability measures that yield contextually relevant data.
The engagement in wishful thinking may lead policymakers to direct solutions to problems that are not realistic for a school to solve given its resource allocation and its neighbourhood wealth such as prescribing test score outcomes and punitive consequences for schools in impoverished neighbourhoods (Grineski, 2005).

Popham (1978) comments on the American educational policy dilemma of the late ‘70s in this statement:

There are also scores of decisions currently being made regarding whether to retain, revise or scrap particular educational programs. In many instances these decisions are reached, in part, on the basis of [standardized] test results. If the wrong programs are scrapped, then countless pupils will be robbed of an effective instructional program (Popham, 1978, p.6).

Given the concerns raised by researchers, policy analysts, and statisticians about the merits of using high stakes standardized testing as a measurement tool to evaluate student learning and gather data on the functionality of the American public school system, why did criterion-referenced, teacher-constructed assessment experience such a short life? The argument against criterion-referenced testing is similar to the argument levied against standardized testing: “its outcomes seemed easy to teach and easy to test” (Haladyna et al, 1998, p. 264). However, criterion-referenced tests do not provide the norming data that support the student-to-student, teacher-to-teacher, and school-to-school comparisons (Haladyna et al, 1998). Simply stated, if most of the students achieved mastery, how will teachers, school administrators, and policy makers determine the following: Who is best? Who is first? Who is last? And who is worst? However, the questions that precede this series of questions may be more important: How is the best determined? What is the nature of the tool that determines who is best? Who constructed
the tool and under what context was that tool produced? What truths are available to be told, and who is able to tell these truths?

Popham (1987), an advocate for criterion-referenced assessment, sums up the dilemma in this statement: “If properly constructed and implemented, measurement-driven instruction currently constitutes the most cost-effective way of improving the quality of public education ... in United States” (p.679). The other methods that are equally effective but more costly include the introduction of current, high-quality instructional materials, and highly skilled, well-paid teachers (Popham, 1987). However, in the United States, the overwhelming urge to compete and to rank the competitors has relegated all of Popham’s solutions to the back burner. By the end of the 1970s, standardized testing with norming capability regained its momentum and favour with the American people and America’s policymakers (Halydyna et al, 1998).

Event 4: A Nation at Risk – Identifying Who is Responsible

In the late 1970s and early 1980s the American people were struggling with the negative impacts of the most severe recession since the end of World War II; interest rates were high, profits, and wages were falling; and programs and benefits crashed (Sacks, 1999). The government blamed the educational system for failing to produce the types of graduates who could compete in the current economy (Sacks, 1999). According to Sacks (1999), the Reagan government argued that the opportunity for American citizens to achieve the American Dream was threatened by the economic engines of South Korea, Japan, and Germany; and the policy document, A Nation at Risk, outlined the solutions to the problem. As the National Commission on Excellence in Education
was preparing its findings on the American educational system, I was completing my first
teaching contract in an American international school.

*From 1981-83, I was an elementary school teacher in an American International
School. Twice per year, we had test week – this was a lot of time devoted to
testing given. Prior to testing the children on the Iowas, the principal accessed all
the English-as-a-Second Language students’ reading ability. The students were
grouped by reading ability rather than age or grade. The students prepared for
the tests on trial questions; then sat for hours coloring circles on answer sheets
while the teachers watched the minutes tick by. We scored the results, passed
them out to the students – no envelopes. The American students traded their
results as if they were playing a card game. The principal bundled the up the
results, placed in the diplomatic pouch and sent them of to the National Office of
Education. The American principal and teachers were agitated and muttered
concerns about how K’s low reading score or S’s low math score may impact
their employment opportunities back in the States. At the time, I did not have a
reference point for their concerns and did not perceive the necessity to devote so
much teaching time to testing and worrying about testing. Although students’
parents paid more than $2700 per child for their children to attend the school; it
was not considered to be a private school; the school, its teachers and the
academic performance of its students did not escape the gaze of the State
Department. (Dagenais, October 1, 2009)*

In 1983 the National Commission on the Excellence in Education completed its
penning of an *Open Letter to the American People* entitled *A Nation at Risk: The
Imperative for Educational Reform*. The impetus for the production of this policy
document came from the Secretary of Education’s belief that the American people were
concerned about issues at all levels of the American public school system (National
Commission on the Excellence in Education, 1983). The commission was charged with
conducting an assessment of the quality of teaching and learning in all American schools
and with comparing the performance of American schools with that of other advanced
nations (National Commission on the Excellence in Education, 1983). The commission
also assessed the relationship between high school grades and college admission criteria,
established the criteria for excellent programs, and identified programs that met the
criteria (National Commission on the Excellence in Education, 1983). Critical to their
analysis was the documentation of social changes over the past 25 years that may have had an impact on students’ school performance and future contribution to American society (National Commission on the Excellence in Education, 1983). From a Foucauldian perspective, the actions of the National Commission represented the government’s need to design programs that would enhance the opportunities for individuals and populations to engage in activities of self-regulation and self-discipline so that the America’s economic woes could be redressed through citizens’ participation in relevant educational opportunities.

The preamble of this document presents the challenge of the declining quality of American public education as an act of war on America’s pre-eminence in the fields of commerce, science, industry, and technological innovation (National Commission on the Excellence in Education). America’s former system of education worked well when the American economic and political policy operated in isolation from the rest of the world; however, in 1983, there were well-educated and motivated competitors in the world markets – both products and ideas (National Commission on the Excellence in Education, 1983). The policy document cites the perceived decline in the education system of the United States as a breach of promise: Not all individuals would have a fair chance and access to the tools for developing their minds and their spirits in service of themselves and America (National Commission on the Excellence in Education, 1983). Key to this concern is the recognition that the former American economy based on the exploitation of natural resources required its people to have a different education and skill set. A few educated citizens were sufficient to ensure economic growth; however, by 1983, there were new competitors in automobile manufacturing, electronics, and textile
manufacturing sectors who recognized the importance and economic advantage of a highly-skilled and literate workforce (National Commission on the Excellence in Education, 1983).

National Commission of Excellence in Education (1983) identified several instances of educational system failure based on the results on standardized test scores: International test scores results placed the performance of American students below that of many other industrialized nations, the decline in the average standardized test score of American students since the launch of Sputnik, the inability of gifted students to meet their test score potential, decline in SAT scores, and decline in scientific literacy based on the National Science Test. While the commission did consider other quantitative measures of student performance, standardized test scores comprised a significant component of the data. The document writers did not question whether these test scores have something meaningful to contribute, or what they represented. The test scores were beyond contestation. The focus was clearly on recovering America’s competitive edge, recovering the trend of increased wealth generation by each successive generation and to affirm standardized tests scores as a mechanism to report that American students were indeed on track (National Commission on the Excellence in Education, 1983).

The policy paper extolled Americans to demand excellence of their educational system; to hold it to the highest standard; to recognize its critical importance in the renewal of human resources; and to lobby for federal, state and local funding, expertise, and ingenuity (National Commission on the Excellence in Education, 1983). The National Commission on Excellence in Education (1983) put forward one recommendation pertaining to the integration of a national, state-administered
standardized testing program to provide quality assurance on the teaching and learning process at all major transition points in a student’s career. Standardized tests results would now fulfill the purposes of certifying student achievement and credentialing; identifying a student’s need for remedial intervention; and determining the opportunity for accelerated progress (National Commission on the Excellence in Education, 1983).

The release of *A Nation at Risk* generated both sympathetic commentary and sharp criticism. Sacks (1999) described the policy as a scapegoating document that blamed the American educational system for the ills of a floundering economy, and which reassigned local control and authority for the school system to state officials. In 1983, the school accountability movement was small; however, the publication of *A Nation at Risk* transformed that small movement into a national cause (Sacks, 1999). Although programs of rewards and sanctions based on standardized test scores existed in some state jurisdictions as early as the 1960s, they not prevalent nor widely accepted (Payne, 2000). However, the release of the policy document clearly tied governmental responsibility for educational policy and standard setting with the outcome of failing schools (Smith & Fey, 2000). Although the American government had a number of policy choices – fiscal equalization across schools, professional development, lower class size and mandatory preschool – it chose mandatory standardized achievement testing as the sole mechanism through which schools could demonstrate their accountability (Smith & Fey, 2000).

Another key transformation triggered by the publication of *A Nation at Risk* was the tying of instructional practice to test scores; this decision was based on the
assumption that quality teaching produced satisfactory test scores and that unsatisfactory test scores could be ameliorated by good teaching (Payne, 2000).

The politicization of the education process became more evident as policy makers instituted minimum competency tests in 36 states, used their ability to reward and sanction schools to exact more rigorous performance, and established processes to certify student performance (Payne, 2000). The power to reward and sanction students had migrated to schools, school budgets, and in some instances teachers. The power of rewards and sanctions is based on the belief that instructional outcomes will improve in relation to the attractiveness of a reward or the consequences of a sanction (Payne, 2000). In the quest for performance, some principals have introduced remedial multiple choice activities to enhance the test scores, some teachers devote up to four weeks to test preparation activities, and some students are rewarded if their test scores meet or surpass the state requirements (Payne, 2000).

State testing programs were introduced with haste. Michigan was the first state to launch a standardized testing program; the tests were launched before assessment specialists had been able to determine validity, reliability, purpose, and other messy details (Sacks, 1999). Although the state education agency had promised to control the release of the results, they were eventually made available to the media (Sacks, 1999). Teachers complained that the publication of test results was harmful to students and only served to confirm the economic disparity among the school districts in Michigan (Sacks, 1999). Sacks (1999) described the relationship between standardized testing and accountability as follows:

Testing and accountability were now rendered virtually undebatable, and became the essence of school reform... They [school districts] now had to prove to
taxpayers and business organizations – with precision offered only by percentile scores – that schools deserved the funding they sought. (p. 79).

In 1989 American children devoted up to 20 million teaching days to testing activities and as much as 10-20 times that amount of time in test preparation activities (Payne, 2000). While this reallocation of resources and human effort to the testing project seems large, in comparison with other nations, American students spend much less time writing standardized assessments (Payne, 2000). As I consider the American drive for excellence through the production of high test scores, my thoughts are drawn to the following conversation.

As I write about the American dedication to testing and sorting, I am drawn to a conversation that I had with a private secondary school owner/administrator in Malawi about his programs and teaching regime. Malawi is a country with few high schools and even fewer postsecondary educational institutions so students compete for the few places on standardized exams. Increasing the students’ opportunity to enter postsecondary education drives the programs at his school. The students rise at 3:30 am (to sound of the headmaster’s bell) to have a cold shower and study for two hours. They have breakfast, attend school for from 7:30 am to 4:30 pm, do their chores, eat supper, attend mandatory study session and fall into bed. It seems to me that American policymakers would like students, teachers and school administrators to embrace this ethos of competition for a scarce commodity, higher education; however, higher education in the United States is not scarce and not limited. (Dagenais, October 21, 2009)

Donlevy (2000) tackled the dilemma produced by the overdependence on standardized test scores as the sole measure of school accountability and student achievement. While most citizens embrace and support high academic standards and first-rate student achievement, Donlevy (2000) took a different tactic. His examination focused on questions about the purposes of education and the identification of what has been lost in this dedication to the pursuit of high test scores and superior educational outcomes. In Donlevy’s (2000) and McNeil’s (2000) counter-discourse, they lament the loss of creativity and spontaneity in the classroom. From their perspective creativity and
spontaneity have fallen victims to control and routine, the hallmarks of a good standardized test score production curriculum. The curriculum has become narrow, test skill-item content squeeze out the interesting and local content; and the individual learning opportunities are often erased as too time-consuming (McNeil, 2000). In this paradigm of standardized testing, individual potential is lost in the frenzy to prepare for the test (McNeil, 2000).

At the end of the century, a tension was emerging in schools between the forces that recognized the importance of the school’s role in preparing youth to take their positions in the economic engine of the next generation and those that recognized the importance of the school in preparing students to recognize their potential, to transmit cultural traditions, to participate as citizens, and to make ethical decision (Donlevy, 2000). McNeil (2000) exemplified this tension in her questioning of the use of standardized test results as the sole accountability measure for schools and the educational process. From her perspective, teaching and learning are complex, multidimensional processes that are dependent on a myriad of difficult to measure variables whose interrelationships are complex and difficult to understand.

Hilliard (2000) added another dimension to this tension-filled space in his justification of the need for accountability, but he questioned the relationship of standardized testing to school performance and student learning. Hilliard (2000) stated “Children, parents, communities, and the public at large, including policy makers, have every right to raise the difficult question of whether schools achieve the appropriate goals” (p. 295). High-stakes standardized testing may be connected to accountability, but according to Hilliard, it may not. The quality of the curriculum, the competence of the
teachers, variety and appropriateness of the instructional methods, and the recognition and responsiveness of the program to the needs of the students – these are key contributors to a quality educational environment (Hilliard, 2000). High-stakes testing may be a useful tool if it has meaningful validity and appropriate standards to accurately reflect the academic growth of the students (Hilliard, 2000).

In conclusion Hilliard (2000) draws attention to two critical gaps in the American national, state-administered testing program. Although the testing program is national in scope, curriculum is locally determined; and inequalities in educational services between jurisdictions are tolerated and maintained (Hilliard, 2000). In addition, school districts select learning resource material based on local budgets and teachers’ preference so the question can be posed: What do the differences in the test scores really mean, and how can they effect change? After considering the interplay between the standardized advocates, the policymakers, and critics of A Nation at Risk, Sacks (1999) asserts that “Learning itself would become a political act in this brave new world” (p. 77).

Event 5: No Child Left Behind Act 2001: Standardized Testing Remains in Force

President Bush positioned educational reform, and specifically the eradication of low educational expectations for minority students (students with disabilities and students with limited English language proficiency), as a major plank in his 2000 presidential campaign bid (Dillon, 2009). After election, the Bush administration sponsored its only bipartisan bill, the No Child Left Behind Act (2001) (NCLB) to introduce a test-based accountability system designed to improve school performance, to enhance the role of the federal government in a primarily state-run function, and to re-authorize the Elementary
and Secondary Education Act (1965) including Title 1 funding (Dillon, 2009). The Act
sets out four pillars of the school functionality: to enhance results-based accountability
measures, to increase flexibility and choice in a state’s or school district’s decisions to
use federal funds, to support rigorous scientific research on program effectiveness, and to
provide parents with more opportunities to choose their children’s schools based on
report cards for parents on the functioning and efficacy of their neighbourhood schools
and consequences for schools that fail to reach preset performance targets (U.S

To comply with the NCLB, states were required to set standards for each grade
and develop standardized assessment instruments to measure students’ and schools’
progress towards accomplishing those standards and to provide choice for parents
through the establishment of charter schools (U.S Department of Education, 2004). In
2001, the federal government provided support for the state education agencies to
establish 1780 charter schools and a $1-billion Reading First program – a program proven
to assist early readers develop grade level appropriate reading skills (Mynster, 2001). In
2003 the U.S. federal government provided additional funding ($17 million) to assist the
state education agencies improve the quality of existing achievement tests and another
$370 million to support the delivery of the achievement tests (Malico, 2003). Through the
enactment of NCLB, the American federal government moved into the zone of
hyperrationalization: The federal government was not only allocating inputs but also
prescribing outcomes, the tools to measure the prescribed outcomes, and the
consequences of failing to meet the prescribed outcomes (Wise, 1977).
The American federal government has maintained its vision that education is a vehicle for training a skilled workforce for the American economy. According to the U.S. Department of Education (2004), the results from standardized tests will provide assurance to policymakers and the American public that schools are doing their job. 

NCLB categorized the large number of minority students, who have been marginalized by the education system, as a liability to be ameliorated (Dillon, 2009). The approach to solving the problem was directive, prescriptive, and dependent on the production of standardized test scores (Dillon, 2009).

NCLB outlined goals for the state education programs that according to Linn, Baker and Betebenner (2002), created several challenges for state education agencies. These challenges arose from the lack of consistent content standards, standardized achievement tests reflecting similar rigor, and stringent performance standards (Linn et al, 2002). These factors act as barriers to the capacity of individual states to achieve the predetermined 2014 performance targets (Linn et al, 2002). By 2014 all students taking the state achievement tests must demonstrate 100 percent proficiency for their assigned subgroup which would be based on factors such as English language proficiency, racial-ethnicity grouping, and socioeconomic grouping (Linn et al, 2002).

In the United States the number of immigrants has been increasing steadily since 2000; there are more than 34 million foreign-born Americans living in the United States and one in every five children attending elementary school was not born in the United States (Capps, Fix, Murray, Ost, Passel & Hernandez, 2005). Limited English proficiency (LEP) is a significant barrier to academic achievement for immigrant children (Capps et al, 2005). Advocates for immigrant children and their families look to the NCLB with
hope; as it mandates specialized testing, additional instruction, and institutional choice for children facing language barriers (Capps et al., 2005). However, Abedi (2004) speaking from the school performance perspective expresses concern that the inaccurate placement of students into LEP subcategories may threaten the validity of the yearly progress indicators. Furthermore, schools with large numbers of students in the lowest LEP categories are often placed under the most pressure to demonstrate improvement. Thus, the intention of the government’s program to enhance the proficiency of marginalized students has been hampered by its need for performance data.

The promise of NCLB is to provide “equal access to an education that allows them [all children] to enjoy the freedoms and exercise the responsibilities of citizenship in our [American] democracy” (Wood, 2004, p. vii). Several commentators put forth arguments that NCLB will not be able to keep that promise because the American government has grossly underfunded the initiative (Kim & Sunderman, 2005; Fusarelli, 2004; Wood, 2004). Wood (2004) and Kim and Sunderman (2005) add that the pressures placed on ethnically-diverse and impoverished schools to meet the predetermined proficiency levels by 2014 will further diminish students’ learning experience because these students will experience a narrow, test-driven curriculum. The disaggregation of test scores may, in fact, serve to emphasize the deficits of some schools that may subsequently lead to imposition of sanctions by government authorities (Wood, 2004; Kim & Sunderman, 2005).

According to Fusarelli (2005), this simplistic definition of the complex problem, the perceived under achievement of the American school system and its students, was not likely to respond to the implementation of a rigorous testing program that included a
series of rewards and sanctions. Wood (2004) stated that many Americans did not see the need for another test-and-punish act to draw attention to the challenges facing American public schools, as average Americans attempted to respond to issues of growing poverty and increasing diversity among the most vulnerable of American citizens.

Abandoning NCLB was not an option; however, Wood stated that Americans have had the opportunity to reinvent NCLB from a school punishment program to school improvement. While accountability is still a cornerstone of NCLB, accountability is no longer provided to the local taxpayers, students, parents, and the communities who have invested interest in their community schools (Wood, 2004). Instead state agencies and federal organizations have become the recipient of standardized test scores and the brokers of accountability meted through terse statements: “meets the proficiency standard or does not meet the proficiency standard” (Wood, 2004, p.6).

Summary

Using the Foucauldian tool, the event, to analyze the changes in the ways American citizens and policymakers thought the complex domain of standardized testing and education into existence reveals the operation of tensions and lines of forces within that domain. The American education system is novel because the federal government exerts control through a national education act (a juridical use of power) that includes a number of funding options (relational exercise of power) for state agencies of education that are charged with determining curriculum, setting standards, and reporting school progress toward those standards (Haladyna, et al, 1998). The federal government has attached a number of conditions to its funding that have included initially measures of
school performance (1840s), followed by accountability measures for school performance (A Nation at Risk), and finally accountability coupled with consequences and school choice (No Child Left Behind Act, 2001). These events represent a re-thinking about the relationship among the American public school systems, the economy of the nation, and the vesting of responsibility for the production of those outcomes. Although the force lines seem to be shifting their trajectories somewhat, Haladyna, et al (1998) and McNeil (2000) assert that the challenges and issues with the integration of the standardized testing with the provision of mass universal education have largely remained consistent. The short flirtation with criterion-referenced assessment represents a rupture, an abrupt change, in how testing and curriculum were thought into existence. Thinking about teaching and learning shifted away from the need to compare students, teachers, and schools toward the appropriateness of the curriculum and how well students were learning the curriculum (Haladyna et al, 1998).

The emergence of increased computing capacity coupled with the government’s perception of an economic crisis - so severe that the American way of life and America’s economic pre-eminence was threatened - shifted the thinking about education, its role in society, and its measurement toward data analysis based on standardized test results. The recognition that the number of poorly educated minority children and youth represented an economic liability – both lost productive potential and a drain on the social welfare system - led to the promises in the No Child Left Behind Act (2001). Few politicians and educators disagree with the premise that every child in America should have an opportunity to benefit from the American universal mass education system; however, many educators argue that the strategies for data collection; the consequences for
students, teachers and schools; and support for the establishment of private charter
schools may, in fact, impede America’s desire to achieve the goals in *No Child Left
historian, iterates the taken-for-granted importance of the relationship between
standardized testing and the economy:

> Although the relationship between students’ outcomes [on standardized tests] and
economic growth may vary to some degree ... there is evidence to suggest that
teacher efforts in translating strong academic outcomes among youth, in turn,
translate into a higher standard of living in the future” (p. 366).

From Jeynes’ perspective, the smooth curve of progress has been disrupted by the
recent decline in economic growth; however, this is only a momentary stutter in the role
of standardized testing and its relationship to the production of effective and efficient
schools designed to produce a competent, creative, and imaginative workforce for the
future.

The English Experience with Standardized Testing: Government, Governance, and
Choice

The examination of any phenomena requires a lens, a lens that shapes, and to
some degree, directs the collection of information towards some understanding of the
*thing* under consideration. In their writing about the English educational system and its
adoption of standardized testing, both Hudson (2007) and Williams (2004) draw on
Foucault’s (1980) concepts and metaphors of disciplinary power, Panoptism, and the gaze
as appropriate methodological tools for unpacking the taken-for-granted spaces of the
English educational system and its relationship with standardized testing. Foucault
describes how governments exercise their sovereignty through the production of laws
such as the *Education Reform Act 1988* and policies such as the *Green Paper, Every Child Matters 2003* (DfES, 2003). Their enactment, however, is dependent on disciplinary constraints exercised through mechanisms of dominations such as the League Tables that display the test results for each school in England. Mechanisms like the League of Tables are legalized components in the technology of power, Panoptism (Hudson, 2007).

Foucault (1980) describes Panoptism as “the dissemination of micro-powers, a dispersed network of apparatuses without a single organizing system, centre or focus, a transverse coordination of disparate institutions and technologies” in which institutional surveillance is used to produce useful citizenry (p. 71). The *gaze* is an integral Foucauldian tactic in Panoptism, which “watches for disorder, anticipates the danger of crime, penalizing every deviation” (Foucault, 1980, p. 72). By providing the opportunity for school choice, the government has provided parents with the freedom choose the school for their child, but also the responsibility to select the school optimizes their child’s opportunity to become a successful and productive citizen (Hudson, 2007). The government has maintained a watchful gaze through the implementation of a regime of standardized testing, has relinquished its centralized authority over the delivery of education in England, and adopted a system of prescribed outputs and a complex network of micro systems to ensure the delivery of those outputs (Hudson, 2007; Williams, 2004).

Foucault’s (1980) definition of disciplinary power and Panoptism provides a theoretical conceptualization for understanding the reorganizing of the lines of governmental authority over education. For Panoptism to be effective, the individuals targeted for coercion must perceive that they are under surveillance – the ever watchful *gaze*. In the English educational system, the function of the *gaze* is performed by the
administration of standardized tests, the publication of the test results, and the parental right to choose the school for their children (Hudson, 2007).

**Governmental Engagement with Standardized Testing**

During the 1980s, England struggled with rising unemployment, a slumping economy, and the abandonment of Keynesian economics in favour of Friedman’s laissez-faire doctrines of minimal government and private enterprise solutions (Klein, 2007). With the cover and distraction offered by the Falklands War, the Thatcher government promoted the doctrine of the ownership society, encouraged private ownership of public housing, reduced social welfare programs, and introduced a program of “radical capitalist transformation” (Klein, 2007, p. 162). Against this backdrop of growing acceptance for free-market responses to social problems, a general concern about the productivity and competitiveness of English economy, and the recognition that the economy of the future would be that of ideas and the production of knowledge, the English government passed the School Reform Act 1988 (Klein, 2007). The tenants of this act established the National Curriculum and assessment protocol (for ages 7, 11, and 14), provided for the publication of standardized test scores aggregated by school, and provided for consumer choice in a free market environment (Hudson, 2007). Education was transformed; it was no longer considered to be a social good, but rather the collection of measurable activities to prepare the citizens of the future with the skills and knowledge that they would require to participate in the new information age economy (Hursh, 2005).

The government not only mandated that 70 percent of the instructional activities must pertain to the National Curriculum, but also the acceptable instructional
methodology if a school wished to maintain its grant-eligibility status (Hursh, 2005). The reconfiguration of curriculum refocused the lens from society and the needs of society to the individual and the needs of the individual (Hursh, 2005). The individual was given the right through legislation to select the optimal school for his or her need; therefore, the responsibility rests with the student and his/her family to ensure that the right decision is made so that the student has the opportunity to mature into a useful citizen-of-the-future (Williams, 2004).

In 1999 Prime Minister Tony Blair greeted the citizens of England with this New Year’s message: “Education is the best economic policy we have ... successful nations will see that education is the key economic and social imperative for us all” (Blair, 1999). Blair’s rhetoric and priorities are consistent with that of the governments of other Nordic European countries that emphasize that the “importance of education [lies] in stimulating growth and competition [with the production of] social cohesion and the generation of wealth” (Hudson, 2007, p. 168). Blair made this speech in which he both connected education to the economic prosperity, and he established education for economic productivity as a key element of his political agenda just prior to his introduction of the Education Act 2002 (Reid, 2004). The Education Act 2002 maintained the role of the National Curriculum, the mandatory assessment of student progress at key stages, the publication of test results, and consumer choice in school selection while adding strong language about child welfare and safety and the reduction of class size (Reid, 2005).

The publication of the Green Paper, Every Child Matters (DfES, 2003) produces children as resources for state investment to assure the economic productivity and prosperity of England (Williams, 2004). Children now occupy centre-stage; children’s
care inside and outside of school is now a social responsibility; and penalties are levied against those who contravene their social responsibility for the care of children (Williams, 2004). The Green Paper identifies appropriate, effective, and streamlined assessment practices tied to accountability and effectiveness of high-quality education as a means to ensure that the citizens-of-the-future receive the training and education that they require today.

In 2006 the English government revisited the provisions of the Education Act 2002 and the White Paper, Higher Standards, Better Schools for All when it passed the Education and Inspections Bill (Burgess, Briggs, McConnell & Slater, 2006). The Education and Inspections Bill retained the National Curriculum and assessment protocol of previous Acts, and it also introduced language to promote the development of private options for the provision of public education, school choice, and responsive programming for children with disabilities and special needs (Burgess et al, 2006). The philosophical positioning or truth telling embedded in this legislation is the notion that competition between and among schools improves the quality of education and that the free market and school choice are the best vehicles to ensure the development of a competitive market for educational services (Burgess et al, 2006). The publication of the test scores ensures that students and parents have the data to select the best school if the test scores really do reflect mastery of the critical learning outcomes for success as a citizens-of-the future.

The English national government has asserted its sovereignty forcefully through the passage of a series of education acts that affirm the national government’s right to transform education from a social good into a servant of the economy and future.
economic prosperity of the nation (Williams, 2004). The *Education Act 1988* established the key elements - the National curriculum, standardized external assessment, and public reporting and consumer choice (Hursh, 2005). Subsequent acts pertaining to education have clarified this philosophical positioning, refined the initial position, affirmed responsibility, and assigned duties. The tenets of these acts have not re-thought the transformation of education into a public service rationalized by market forces to support the wealth generation project of England. The drive to satisfy the demands of economy and the perceived future needs of the economic engine of England have left little space in the political minds of the politicians, economists, and businessmen to consider what has been erased when the local voices – the teachers, the students, the parents and the local education authorities - are silenced.

**Governance**

Governance is composed of the micro-systems of disciplinary power that complement the sovereignty of government legislation; these heterogeneous systems perform the regulatory functions such as self-monitoring and societal self-organization that make the behaviours and actions of schools, teachers students, and parents visible (Hudson, 2007). The governments in England and elsewhere face the challenges of needing to control the delivery and outcomes of education while at the same time needing to remain open to the possibility of new and creative options for the delivery and management of educational services (Hudson, 2007). The micro-systems of disciplinary power set up the monitoring systems that make the performance of schools visible and available for shaping.
As governments shift their priorities from the allocation of resources to the prescription of outcomes, governance systems have evolved through the need for local education authorities to deliver the core curriculum, for students to perform at or above the pre-set standard, to compete for students in educational services market, and most importantly to influence and shape their public image (Hudson, 2007). Although the National curriculum is set, local education authorities have choices to make: They can choose to deliver the required 70 percent of the National curriculum and qualify for grant funding, deliver more elements of the National curriculum in belief that students may do better on the standardized external assessments and the public image of the school may rise, and/or follow the prescribed teaching strategies for remedial activities and qualify for additional funding (Hudson, 2007).

Unlike other countries in northern Europe, the education system in England is characterized by a drive for homogeneity – all schools should be producing graduates who know and can do similar things at a similar level; consequently evaluation of learning is external, centralized, and standardized (Hudson, 2007). The visibility produced by charting the school-by-school results on the League Tables invites commentary and comparison without context and background about the students, their neighbourhoods, or the communities’ expectations (Hudson, 2007). According to James (2005), the situation in secondary schools “appear to be more political and public (or at media) interest in league table positions than in any of the many other possible (and arguably, more interesting and informative) articulations of the results in learning” (p. 87). The micro-systems created by the emphasis on League Table rankings and erased or
made invisible other - perhaps more meaningful or genuine - improvements in students’ learning due to changes in teaching methods (James, 2005).

While the neoliberal agenda of England has promoted market-driven responses to education, the government has responded to the rise in capitalistic initiatives through implementation of quality assurance systems that regulate both the production side of the education - curriculum, schools, and the delivery of instruction - and the quality of the educational products of that system through the a series of standardized tests (Hursh, 2005). The implementation of mandatory standardized testing and public reporting of the results “makes it possible for market forces to operate by providing a currency of information which fuel competition between schools” (Hursh, 2005, p. 6). Privatisation of the provision of educational services and the encouragement of competition among service providers is viewed by government as a strategy to ensure quality and to promote equality of service (Hudson, 2007). The increase in the number of schools with above standard ranking on the League Table, the increase in international exam scores in 2002, and the reduction of the number of children living in poor households – these are all factors that policy makers and the public welcome as signs that the transformation of the English school system from the philosophical notion of providers of a “social good” into preparatory institutions for citizens-of-the-future is working (Hudson, 2007; Reid, 2005; Williams, 2004).

Regulation is the key is the English transformation: self-regulation imposed by students, teachers, and institutions (Hudson, 2007). Hudson labels this regulation as a form of Foucauldian discipline. Institutions choose particular policies or a set particular priorities based on the consequences of League Table results such as: media attention,
public perception about the school, and implications for enrolment. Teachers select
teaching materials and make curriculum choices based on the potential impact of that
decision on the standardized test scores rather than on the potential usefulness in
designing a meaningful lesson that may impact students’ potential for learning,
motivation, or intellectual development. Students select schools based on their
reputations, League Table results, and their attractiveness to other students (Hudson,
2007; James, 2005).

All of this testing has generated data – data about school, teacher, and student
performances as compared to a benchmarked standard (Hudson, 2007). This data has
been used to produce school rankings, school performance benchmarks, gender-based
performance analyses, and research problems (Hudson, 2007). Test scores have been
used to set performance standards for each assessment level, and these performance
standards have been rising (Tymms, 2004). Although policymakers have publicized the
rise in the test score results, the question must be posed: Are children learning more,
learning more effectively; or have they learned to write the test; or is the evolution of the
test contributing to the rising tests scores? However, The Trends in International
Mathematics and Science Study (TIMSS) results and the International Association for
Evaluation of Educational Achievement (IEA) results do not confirm the increase in
achievement levels by British students (Tymms, 2004). Tymms attributed the discrepant
results – rise in student test scores on the English statutory tests, but constant results on
the TIMSS and IEA tests - to cut score setting procedures and annual changes test items
on the English statutory standardized tests. According to Tymms, the English statutory
standardized tests are reviewed annually, test items are reviewed and revised and the
benchmarking process is revisited to re-establish the required scores for each performance level. Tymms (2004) expressed concern that results of year-to-year comparisons based on the proportion of students scoring at each level may be influenced by the revision process as much as it is influenced by changes in student performance. Although English policy makers are quick to label the rise in test scores on the British statutory tests since 1995 as real evidence of school improvement, the question about the meaning of the test score increases remains (Tymms, 2004). Can the rising test scores be considered as evidence of students’ increasing mastery of the curriculum or is something else at play, and are the new school policies a causal force in the production of this phenomenon? In response to these questions, Teddie, Reynolds, and Sammons (2000) note that statistical analyses may obfuscate the real educational issues because “At a certain point, such complex statistical applications appear quite unrelated to substantive issues regarding school effects” (p. 113).

The prevalence of test score data and increased digital computing capacity have facilitated the disaggregation of data into specific population categories such as gender, ethnicity, and social class; thus, researchers are able to pose more complex questions about the variability of the test scores and their underlying factors such as gender-based performance differences. While the assumption that girls outperform boys on standardized tests tends to be true for white middle-classed students, the real issue is the variability in boys’ test scores as compared with those of girls (Tinklin, 2003). However, the media and the public have constructed the poor performance by boys on standardized assessments as a high profile issue, which has sustained in the public’s mind by the publication of English National testing program results (Strand, Deary, & Smith, 2006).
The underachievement by boys is described as the number one public burden. This disturbing problem has motivated policy makers to establish targets for improvement and form task teams to address the issue so that the economic and social burden associated with underperforming boys can be ameliorated (Strand, Deary, & Smith, 2006).

*Choice: Placing the Delivery of Educational Services in the Marketplace*

The introduction of the *School Reform Act 1988* set the groundwork for the privatization of educational services in England and the creation of the public appetite for market-driven educational services delivery (Hudson, 2007). By creating the option of choice within a highly-regulated educational services market, the English government believed that it could address the issues of improving both quality and equality in the provision of educational services (Hudson, 2007). By creating an environment in which students have the right to choose the schools that they attend and the teachers from which they learn, the English government has transferred responsibility for achieving the prescribed outcomes laid out by the national curriculum assessed by the standardized tests to the students and their parents (Hursh, 2005). The English government has minimized its risk related to the introduction of market-driven solutions through its policies and regulations concerning the administration of standardized testing and the reporting of results (Perry, 2006).

Choice in the English educational system is not equally distributed for either students or schools as barriers limit students’ ability to exercise their choice, and schools face limits in their ability to compete for students based their public image and their
ability to generate revenue (Hursh, 2005). Advantaged schools - those with excellent programs, resources, and teachers - are able to attract good students from schools in poor neighbourhoods (Hursh, 2005). The schools in the poor neighbourhoods struggle to attract students because they suffer the consequences of declining revenues and poor reputation based on their past students’ performances on the national standardized tests (Hursh, 2005). Students from middle-class families seem to benefit most from introduction of choice into the provision of educational services as their parents have the resources and environmental knowledge to maximize their children’s options (Hursh, 2005). The issue is further complicated by the activities of corporations that may choose to endow a school for the right to control the membership on the board of directors; thus, corporations are able to shape the admissions policy and academic policies indirectly (Hursh, 2005; Hudson, 2007).

In summary, the English educational transformations have emphasized the “principles of individual, economic and political freedom” while the government asserts that maintaining diversity, competition, and choice” are its primary goals (Hursh, 2005, p. 10). The importance of English educational system lies in its responsibility for “creating and maintaining the national identity... [and] for economic development” (p. 266); however, “a standardized, mass-produced education system has been increasingly portrayed, by both government and business, as inappropriate” for producing the type of educational product that is useful in a globalized national economy (Hudson, 2007, p. 269). The production of the modern English educational system is the result of several forces operating in a global and a national way: the growth of the knowledge economy, the need for pluralistic and individualistic responses from educational systems, and a
diminishing demand for standardized and uniform solutions while at the same time ensuring that self-regulatory forces produce appropriate and visible outcomes (Hudson, 2007; Williams, 2004).

PISA, TIMMS, and PIRLS: Global Order, Regulating Statements, and the Effects of Culture

International standardized testing is a domain like grammar, economics and the prison system, in which psychometricians, sociologists, and psychologists have “employed rules to define objects proper to their own study, to form their own concepts and build their theories” and through the meta analysis of the application these theories and concepts certain foundational regulating rules pertaining to teaching, curriculum, school effectiveness, and resource inputs for formal educational systems may become plain (Foucault, 2007a, p.xii). The codes of culture, in large part, govern how individuals and societies use language to produce discourses about educational practices, establish hierarchies of practice related to teaching and curriculum design, and determine truths about how schools function and students learn (Foucault (2007a). The impetus for most international standardized testing programs is the enhancement of our understanding of the foundational relationships that underpin the conceptualization of teaching, learning, curriculum design, and the organization of educational systems (IEA, n.d.). The results of international standardized testing projects provide the empirical data upon which researchers can seek out truths about the functionality of teaching and its practices in a cross-national environment, identify the statements or rules that govern the production of knowledge, and examine the effects of culture on the production of those knowledge
statements (Foucault, 2007a). The survey data generated by international standardized testing programs create an opportunity to contemplate issues such as the effect of gender on test scores, role of technology in the learning environment, the production and value of a national IQ scores, and the contribution of a nation’s educational system in the development of stable and prosperous societies in a globalized world (OECD, 2007; Rindermann, 2007).

The formative development of international standardized assessment tools, data collection, hypothesis generation, and meta-analysis of test results interpretations illustrate the search for ultimate explanations (reality), the identification of statements that govern the production of truth and what can count as truth (rules), and the role of culture in shaping the rules that determine what can be told as a truth. Two organizations, the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic Cooperation and Development (OECD) have provided the necessary leadership for the research, development, analysis, and truth telling about the performance of schools and students in several countries in the developed and developing worlds.

These activities began in 1958 at the UNESCO headquarters in Hamburg when a group of psychologists, sociologists and psychometricians met to discuss their perceptions about the nature of problems within schools, education systems, and student evaluations (IEA, n.d.). In 1967 this informal group evolved into a formal organization, the IEA, whose thesis is that “effective evaluation requires not only consideration of the inputs to education and the conditions under which schooling is conducted but also the examination of such educational outcomes as knowledge, attitude and participation”
The founders of IEA operated on this assumption: The world could serve as a "natural educational laboratory" for educational research (IEA, n.d.). Optimal educational outcomes could be attained through different methods, and comparative research would yield insights into the performances of individual educational systems that might otherwise be obscured in regionally or nationally conducted research (IEA, n.d.).

In the years 1959-62, the informal organization (precursor to IEA) conducted preliminary research to evaluate the utility in conducting cross-national academic testing and to identify the challenges of administering this type of academic survey in a cross-cultural environment (IEA, n.d.). The results of this research provided findings that had both practical and academic merit, as well as, critical and positive feedback about the feasibility of conducting cross-cultural studies about the roles of schools, teachers, students, and parents; and about the functionality of educational systems – their policies and their governance (IEA, n.d). The outcome of the preliminary research was the development and delivery of two internationally-delivered assessments administered on a multi-year cycle (IEA, n.d.). IEA launched the Third International Mathematics and Science Study (TIMSS, later renamed Trends in Mathematics and Science Study) in 1995 (and subsequently in 1999, 2003, and 2007) and the Progress in Reading Literacy Study (PIRLS) in 2001 (subsequently in 2006). Both TIMSS and PIRLS are tied to internationally-agreed upon curriculum outcomes for fourth and eighth graders. The fourth and eighth years of schooling were chosen as a key transition periods in students’ school careers because they are beginning to use mathematical and scientific concepts in other subject areas and in everyday situations or they are transitioning into more sophisticated applications of mathematical and scientific concepts (Gonzales et al, 2008).
Canadian students participated in 1995 and 1999 administration of TIMSS; Alberta, Ontario and Quebec students participated in 2007 TIMSS benchmarking administrations; and students in the Alberta, British Columbia, Nova Scotia, Ontario and Quebec educational systems participated in PIRLS 2006.

The second organization responsible for mass international student testing is the OECD. The mission and the goals of OECD have evolved from that of an aid distribution agency to an organization that uses its data collection and research capacity to help its member governments build economic capacity, generate prosperity, fight poverty, and ensure stability (OECD, n.d.). As economies change from resource-driven to knowledge-based, the demand for highly skilled and creative human capital increases (OECD, n.d). The OECD and its consultants have worked with experts from its member countries to produce the testing protocol, Programme for International Student Assessment (PISA), and to provide the support for its administration and analysis (OECD, 2007).

**TIMSS**

At the time of its launch in 1995, TIMSS was deemed the most “ambitious study of comparative education” (IEA, n.d.). Based on more than 20 years of comparative studies in the cross-national educational environment, TIMSS became a reality through the collaborative work of researchers and policy-makers to marry a research vision with the practical needs of policy development (IEA, n.d). TIMSS is administered on a four-year cycle to several samples of the student population in the participating countries and regions. Prior to participating in the TIMSS assessment program, the participating countries must conduct a preliminary probability sampling exercise (Gonzales et al,
The characteristics of the student sample must meet those identified by TIMSS: the population size represented by the sample must meet a prescribed threshold (small samples are excluded from TIMSS cross-country analysis); the number of exclusions (students or schools exempted from taking the TIMSS assessment) must not exceed the prescribed 5 percent; and the students must meet the prescribed 9.5 years, the age threshold (Gonzales et al., 2008).

The TIMSS test consists of 14 mathematics and science units combined into student test packets consisting of two units of science combined with two units of mathematics (Mullis & Martin, 2009). Because administering and completing 390-minute test is an unreasonable task for most schools, students complete a 90-minute test package and the scores for several samples are combined to form a jurisdictional score (Mullis & Martin, 2009).

The TIMSS survey developers realized that the teaching and learning environment can reach beyond the boundaries of the classroom (Mullis & Martin, 2009). Since students can be actively engaged in learning activities in their homes and neighbourhoods, the survey sample extended beyond students to teachers, principals, and national research coordinators who completed survey questionnaires about students’ homes, neighbourhoods, and school environments (Mullis & Martin, 2009).

Scores for participation in the mathematics section are presented on a 1000-point range with a mean of 500 and a standard deviation of 100 with five prescribed benchmarks and accompanying descriptors (Gonzales et al., 2008, p.5). The instrument design employs item-response theory to weight the items and scale the scores (Mullis &
The survey instrument is designed to facilitate country-to-country comparisons and cycle-to-cycle comparisons (Mullis & Martin, 2009).

According to Gonzales et al (2008), country-to-country comparison studies have revealed that the TIMMS’s mathematics score for American students has been rising since its first administration (1995), and the mathematics score has been rising more quickly for American students than for students in 13 other countries. Gonzales et al attributed the gains in mathematics scores to the positive impacts of NCLB. However, Ravitch (2008), a former Assistant Secretary of Education, labelled the gains as “small” and consistent with gains made by American students prior to the implementation of NCLB.

Bracey (2008) describes the international assessment business as unrelated “to anything important to a national economy” because Americans lead the world in innovation. However, Bracey (2008) contends that it is America’s obsession with standardized testing and the production of instruments like TIMSS that stifle American creativity and limit the imagination of its students (Bracey, 2008).

In 1995 the performance of Canadian students in grades 4, 8, and 12 on the TIMSS assessment ranked at the top of the G-8 countries; however, Canadian students have not participated as a country in subsequent administrations (Kilpatrick, 2009). In 2007, students living in Alberta, Ontario, and Quebec participated as benchmarking regions (Mullis & Martin, 2009).

TIMSS has been described as a mathematical derby in which the winner takes home the trophy; however, the membership of the participant pool has not remained constant. The perceived strength of participant field continues to shift and change –
Canada, Belgium, France, and Ireland participated in 1995 but dropped out of subsequent administrations; and they have been replaced by countries like Ghana, Botswana, and El Salvador (Kilpatrick, 2009). The real or perceived overall effects of the shifting membership of participant pool on production of scores and the ranking of a particular region or country has not been assessed. Kilpatrick (2009) notes that participating in TIMSS is expensive and time consuming; schools and students are generally not willing participants so their investment in their performance is low. The complicated sampling criteria disadvantage countries like Canada with large geographical land mass and small population (Kilpatrick, 2009). The imposition of a virtual international curriculum for school mathematics advantages those countries whose curriculum is closely matched with the expert-driven TIMSS assessment curriculum.

Given all the perceived challenges that accompany the administration of a cross-national assessment like TIMSS, students in developed countries tended to perform well on TIMSS 2007; however, the relationship between real performance level and the potential of a country’s or region’s economy and the efficacy of its educational remain unclear (Kilpatrick, 2009). On the TIMMS 2007, year-four students living in Hong Kong (607), Singapore (599) and Japan (568) performed better than students living in the United States (529) while students living in Quebec (519) and Alberta (504) performed somewhat below the level of American students (Government of Alberta, 2009). The scores for year-four students in the United States and Ontario increased as compared to the scores these students posted in 1995 while scores for Alberta students dropped between 1995 and 2007 (Government of Alberta, 2009). Although the scores for students living in Alberta changed, the importance of those changes and the implications for the
students’ future education and the economy remain unclear. At this point it is difficult to understand if falling TIMSS scores are a signal to the Alberta Government that the possibility of achieving the goals set out in its 20-year plan, *A learning Alberta* (2006), is slipping out of reach as this plan for economic prosperity identifies the importance if rising standardized test scores as a key indicator of success. Alternatively, the scores may be somewhat irrelevant to global competitiveness, provincial economic development, and the development of creativity and ingenuity in the citizens that live and prosper in a particular society.

Riley and Torrance (2003) state that many political leaders, like the fictional President Bartlett on *West Wing*, do take the TIMSS results seriously, but they also caution that few leaders move beyond the country-to-country ranking tables and that not all leaders ascribe the same importance to placement of their country’s students. From a Foucauldian perspective TIMSS produces knowledge that has a differential capacity to invoke relations of power. The acts of aggregating, grouping, categorizing, and reducing human performance to some number illustrate the effects of the Foucauldian concept of *population* (Foucault, 1980). Individual students, classes, and schools are placed under erasure; Alberta’s students are a “504,” their measured aggregate mathematics skills are just below those students living in Ontario (Government of Alberta, 2009).

**PIRLS**

PIRLS situates itself in the essential space of the fundamental-ness of reading to every form of “personal learning and intellectual growth” (Mullis, Kennedy, Martin & Sainsbury, 2004, p. v). IEA describes reading as crucial to maximizing human potential, maintaining global competitiveness, and improving the quality of life for all (Mullis et al,
Improving the reading and literacy capacity of a nation, argues IEA, is positively correlated with improvements to a nation’s prosperity and well being; thus, all nations should be interested in how well its citizens can not only read text but also use text in a variety of life’s pursuits (Mullis et al, 2004). PIRLS is designed to provide information to participating countries about how well its year-four and year-eight students can use reading and literacy skills to solve problems and make decisions in a simulated world created by test designers and reading experts (Mullis et al, 2004).

PIRLS is based on the following definition of reading literacy: “the ability to understand and use those written language forms required by society and/or valued by the individual” (Mullis et al, 2004, p.3). From a PIRLS theoretical perspective, readers are viewed as individuals who are “actively constructing meaning,” using effective reading strategies, engaging in reflection, and interacting with text (Mullis et al, 2004, p.3). Like TIMSS, the PIRLS assessment instrument includes a supplementary questionnaire designed to gather information about children’s reading habits and attitudes towards reading (Mullis et al, 2004).

PIRLS, like TIMSS, is administered on a regular five-year cycle to participating countries from both the developed and developing worlds (Naumann, 2005). Currently, the number of developing countries - such as Ghana and Senegal - remains small; however, the number is growing and is expected to grow as these countries begin to develop modern information-age economies (Naumann, 2005). Experts have designed PIRLS based on these assumptions: All children, regardless of cultural and social environment, move through similar developmental phases and transitions zones at similar age-specific times (informed by Piaget’s developmental theory) (Naumann, 2005). Based
on these assumptions, the experts and test designers at IEA postulate that it is possible to gather data on the reading performance of the world’s children and that educational policy makers can conduct cross-national comparisons of the efficacy of educational policies and systems of participating member countries (Naumann, 2005).

Aggregated test results are standardized on a 1000-point scale which is divided into levels of reading competence (Naumann, 2005). On the 2001 PIRLS test, Canadian students living in Ontario and Quebec performed at competency level 3 and demonstrated the ability to understand advanced texts and derive implied meaning (Naumann, 2005). In 2001, Canadian students from Ontario and Quebec placed sixth overall and just behind students from Sweden (score of 561 and within the Level 3 range). Because only Quebec and Ontario students wrote PIRLS, Canada did not meet the rigid sampling requirements for PIRLS country-by-country comparisons (Naumann, 2005). However, in 2006 students from Alberta, British Columbia, and Nova Scotia joined students from Ontario and Quebec in writing PIRLS (Mullis, Martin, Kennedy & Foy, 2007). The PIRLS score card placed students from Alberta (560) in third place – just behind the first place students from the Russian Federation (565) and second place students from Hong Kong (564) ((Mullis et al, 2007). Both Ontario (555) and Quebec (533) improved their scores over the previous administration but fell to seventh and twenty-third places respectively (Mullis et al, 2007). Alberta Minister of Education Ron Liepert described the performance of Alberta’s students in these words:

These results show that the achievements of Alberta students continue to be among the best in the world. Our students are developing skills that will serve them well throughout life. We can attribute this success to the excellent quality of teaching in Alberta, outstanding learning and teaching resources, high quality centralized curriculum and standardized assessment programs. (Alberta Education, November 28, 2007)
The difference between the results posted by both Alberta and British Columbia and the Russian Federation are not statistically significant; however, the difference between the scores posted by Ontario, Nova Scotia, and Quebec and the Russian Federation are statistically significant (Mullis et al, 2007). It is tempting to assume that Alberta and British Columbia have better schools and that we as Canadians should pay attention to what is happening in those jurisdictions; should copy their policy framework and curriculum models; and should ask ourselves what is not working in the schools in other provinces. However, cross-national assessments are based on a virtual international curriculum to which some jurisdictions are matched more closely than others, and the impact of regional differences in neighbourhoods, families, and school wealth are reduced to a single national Human Development Index number that blurs the effects of regional differences.

Overall, most countries have students who exhibit strong reading skills and those who exhibit weak reading skills; the variability within countries – about 250 points – is similar to the variability of mean scores between the highest and lowest performing countries (Mullis et al, 2007). For the PIRLS 2006 administration, the United Nations Human Development Index accompanied the league table results. The United Nations Human Development Index is a much broader concept than the reporting of fluctuations average national income (UNDP, 2009). Measures of women’s equality, income inequality and consumption, and poverty – these are just some of the variables that contribute to the creation of the following:

[A]n environment in which people can develop their full potential and lead productive, creative lives in accord with their needs and interests ... And it is thus
much more than economic growth, which is only a means – if a very important one – of enlarging people’s choices. (UNDP, 2009).

By introducing the UN Human Development Index with the reporting of PIRLS results, researchers have the opportunity to consider the combined impact of broad societal factors in relation to the overall performance of a country’s students and its educational system.

Generally, countries with higher Human Development indices post higher achievement scores; however, the Human Development Index for the Russian Federation (.797) is significantly lower than any other country with achievement scores above the PIRLS Scale average (500) (Mullis et al, 2007). The participating Canadian provinces have a much higher Human Development Index (.950) and a somewhat lower PIRLS scores (Mullis et al, 2007). It may be reasonable to hypothesize that the importance of the strength of Human Development Index is influenced by other culturally-produced factors.

Naumann (2005) contends that children in most developed countries adhere to the ages and stages concepts outlined in the Piagetian developmental model, the theoretical basis for the development of the PIRLS assessment. He uses as evidence the predominance of Level 3 scores (526 – 600); however, he also states that Piagetian theory does not match the developmental process that most children in developing countries experience as these children are not initiated into the process of attending school in the same manner as children in developed countries. Thus, interpreting and comparing the results for students in developing countries with those in developed countries remains problematic for these reasons: these types of empirical studies are relatively novel in developing countries and most students learn and speak a lingua franca that differs from the language of instruction and assessment (Naumann, 2005).
A few fledgling educational projects have begun in Burkina Faso and Mali that offer initial reading and literacy training in the lingua franca followed by formal education in the official language in year three; the test score results for these countries show marked improvement (Naumann, 2005). MacGillivray (2006) in discussions of globalization affirms the importance of African economies, their potential for growth, and their need for educated workforces. As African countries begin to implement changes to their education policies such as the introduction of outcomes-based education curriculum projects and mandatory primary education, their policy makers will become interested in participating in cross national surveys such as PIRLS.

As I think about Naumann’s concerns about developing countries, my thoughts are drawn to my recent experiences with Malawian post-secondary educational system, educators, secondary school teachers and administrators, and students and children. The opportunities for a secondary school education are limited – less than 4 percent of the graduates from primary school have the opportunity to attend secondary school, and admission is both competitive and costly. Students must compete for seats, and parents must be able to purchase a school uniform. In many instances private secondary school is the only option. Secondary schools in Malawi are fiercely competitive – students compete for just a few seats in post-secondary educational institutions and often they are slotted into programs in which they have no interest. Parents and students are willing to endure a rigorous study schedule because they understand that education is the only way that Malawi will achieve its president’s dream of transforming itself from a nation of importers and consumers into a nation of exporters and producers. These students who study and compete for limited and precious seats in post-secondary represent the elite and the brightest minds in Malawi, but also the future test writers and contributors to the PIRLS data base. I wonder if their developmental experience will continue to be substantively different as these developing countries transform themselves into more productive economies. Will education be reframed from an opportunity for some to a right for all as the country begins to recognize that a marginally educated workforce is a national liability and reduces the opportunity for prosperity for all?
In the 1990s the leaders of the OECD member countries came to the realization that the skills, knowledge, and creative abilities of their labour forces affect their nations’ ability to enjoy the fruits of dynamic and prosperous economies (OECD, 2007). In the future, the quality of a nation’s workforce will have an even more pronounced effect on the standard of living its citizens enjoy (OECD, 2007). This observation coupled with the real life examples of the economic recovery of Japan and South Korea: two nations - devastated by war and destruction and without the benefit of significant stores of natural resources – that rebounded to build vibrant and productive economies (OECD, 2007).

The accomplishments of these two countries seemed even more remarkable when their economic well being is compared to countries like Nigeria which has the benefit of vast oil reserves or Burkina Faso and its gold reserves (OECD, 2007). According to the UN Human Development Index (2004), Canada, with the benefit of both a skilled workforce and natural resources placed fourth; Japan, ninth; South Korea, 28th; Nigeria, 151st; and Burkina Faso, 175th (ABC NEWS ONLINE, July 15, 2004). Although, natural resources can provide a significant boost to a nation’s economy, natural resources alone cannot compensate for the negative impacts of an underprepared workforce in today’s globalized economy. However, an educated, skilled, and creative workforce can overcome the potential economic disadvantage imposed by limited natural resources. Thus, the OECD member countries, in consultation with experts and policy makers from the respective member OECD countries, posed these research questions: “Are students well prepared to meet the challenges of the future? Are they able to analyze, reason, and communicate their ideas effectively?” (OECD, 2007, p. 16) Emerging from the posing of the primary research question was the survey tool, PISA, a rigorous, complex, multi-faceted
international assessment program designed to gather data on students’ academic performance, school environments, instructional practices, school resources, school neighbourhoods, and student/family wealth (OECD, 2007).

PISA is designed to assess the readiness of 15-year-olds (15 years 3 months to 16 years 2 months) to make the transition into the workforce and/or into institutions of higher learning (OECD, 2007). Each academic assessment focuses on one of the three major subject areas: reading, math, or science and treats the other two subject areas as minor foci (OECD, 2007). The determination of the characteristics and construction of sample student population (400,000 students representing 20 million eligible 15 year-olds in 57 countries) is critical to the interpretation of PISA results. The students are controlled for birth year rather than the number of years spent in formal schooling (TIMSS and PIRLS control for years of formal schooling rather than student age) with a caveat that they must have completed at least six years of formal schooling (OECD, 2007). In many school systems, students who are 15 years old are completing their last year of mandatory formal education, a scheduled transition from mandatory student to student-by-choice and/or engagement in the workforce. Since the test population constitutes the employees and entrepreneurs of the near future, their characteristics – both academic and ethnographic - are of interest to researchers, educators, and policy makers (OECD, 2007).

Although the OECD does not state explicitly that PISA results are predicative of future economic development for nations or economies, the OECD does cite examples of studies that correlate PISA results with future outcomes for students (OECD, 2007). The Canadian Longitudinal Youth in Transition Study illustrates that 88 percent of the
students who achieved level 5 (highest level) on the PISA 2000 Reading assessment were enrolled in postsecondary educational institutions while only 28 percent of the students who achieved level 1 (lowest level) were enrolled in a postsecondary educational institution (Knighton & Bussiere, 2006). If the rate of participation in formal postsecondary educational programs is a factor in economic prosperity, then countries and economies have a vested interest in increasing the size of the student population that can achieve a level 5 in any subject and decreasing the size of the population that achieves Level 1 (PISA uses a five-level scale with level 1 designated as the lowest threshold and level 5 designated as the highest threshold). Countries have prioritized investment in education differently. Korea improved its ranking on PISA 2006 by investing in its top performers while Poland improved its ranking by investing its low achievers; however, insufficient time has passed to determine which investment strategy will optimize the long-term socioeconomic benefit for the country (OECD, 2007).

As I read about the approaches of Poland and Korea, I am drawn to the recollections of my experiences consulting with one of the Tribal Councils. The Aboriginal leaders shared their stories of growing prosperity on one of the reserves in the region. The reserve was able to finance and build several new houses on the reserve because they were beginning to reap the benefits of a reserve owned and managed welding company. Most of the employees were not members of the reserve; however, reserve members were benefitting from the profits. I wonder if this system of continuing to invest in the most skilled while the profits take care of the least skilled is a viable economic plan for the future. Are Korea and the reserve increasing their liability and incurring the possibility of a much larger redress in the future if they fail to invest in their most needy citizens? (Dagenais, November 19, 2009)

PISA has a nine-year cycle; and within each nine-year cycle, the PISA assessment is administered three times – each administration has a different focus (OECD, 2007). The test items fall into three main types: multiple-choice items (52 percent), constructed
response items (40 percent), and closed response items (8 percent). Item response theory is used to weight the 390-minute assessment; for PISA 2006 (science focus), more assessment time was devoted to science (210 minutes) than either mathematics (120 minutes) or reading (60 minutes); the time devoted to each component and relationship to test items from previous years was sufficient to permit year-to-year (OECD, 2007). The experts recognized that a 390-minute assessment would not respect either the students or the quality of the test data; thus, the assessment is packaged in 13 different combination-booklets to be randomly distributed in each test writing class. The results generated from PISA assessments are not individual but rather the collective result of a country’s students’ efforts (OECD, 2007).

Some economic regions such as Quebec, Ontario, and Alberta chose to participate in PISA 2006 assessment as separate jurisdictions to facilitate the comparison of their students’ performance with that of students in other countries and to the mean performance of Canadian students (OECD, 2007). Although Canada as a whole placed fourth, Alberta’s second place finish was just that much closer to the coveted first place finish (OECD, 2007). The Minister of Education for the province of Alberta announced that “Alberta students are among the world’s best” and that “Alberta’s students have jumped from fourth to second place in science” just behind Finland (Government of Alberta Education, 2007). The Minister of Education also recognized that student success is a “collaboration of the efforts of teachers, parents, students and all educational stakeholders” (Government of Alberta Education, 2007). The minister’s comments identify the importance of placing as close to the top of the league table results as possible and the importance of recognizing that educating youth is a collaborative project.
The ranking of academic performance often overshadows the data gathered about educational environment, instructional practices, the wealth, and support of parents (OED, 2007).

The PISA experts identified literacy as an innovative key concept in the construction of the PISA academic assessments; literacy, in the PISA context, is “concerned with the capacity of students to apply knowledge and skills in key subject areas and to analyse, reason and communicate effectively as they pose, solve and interpret problems in a variety of situations” (OECD, 2007, p. 16).

According to the OECD, literacy functions on a continuum; it is not a set of skills, knowledge, or attitudes that one has or doesn’t have; but rather it is the development of those skills, knowledge, and attitudes that support the preparation of the future workforce and lifelong learning. This definition of literacy forms the foundation for the development of the types of test items and their weighting (OECD, 2007). To illustrate this definition of literacy, PISA defines scientific literacy as the acquisition of the “knowledge and use of that knowledge to identify questions, to acquire new knowledge, explain scientific phenomena, and to draw evidence-based conclusions about science related issues” (OECD, 2007, p. 34).

PISA, like TIMSS and PIRLS, sets stringent sampling criteria and limits the proportion of schools and students that can be excused from writing the assessment (OECD, 2007). Most countries exclude less than 1 percent of the potential population from the PISA assessment; the threshold is set at a maximum of 5 percent of the population test population (OECD, 2007). Canada excludes 4.3 percent of its sample from the PISA assessment without explanation or justification (OECD, 2007).
PISA documentation elaborates on the importance of the stringent sampling strategy and addresses the need to ensure the comparability of the data gathered from the member and participant countries (OECD, 2007). Some of the exemptions are explained by the presence of small, remote, and isolated school communities that do not fit the PISA sample criteria; however, other exceptions consist of schools and students that have been identified as special needs (OECD, 2007). Through the standardization process, scores for jurisdictions, such as Canada, with an above average exclusion rate are adjusted to compensate those of countries and regions that adhere more rigidly to the sampling protocol (OECD, 2007).

*Gender*

Interest in gender and gender-based effects of schooling have been fuelled historically by the low levels of success for females in the subject areas of mathematics and physical science, and more recently by the perception of disengagement by males and by their consistently low and falling performance in reading and literary subjects (Watanabe & Ischinger, 2009). The gender-based differences in performance do matter and are important as research questions related to the design and administration of standardized tests and as political and economic questions. Males and females who are marginalized from the educational system represent a social and economic liability for nations and regions. Through the exploration of PISA results disaggregated for gender, researchers and policy makers have asked critical questions about the magnitude and scope of gender-based performance differences, identified opportunities to address these differences, and have furthered our understanding about how students learn (OECD, 2009).
PISA 2000 results illustrated that “large gender differences observed in both student performance and student attitudes” for reading and reading-related life competencies (Watanabe & Ischinger, 2009a. On PISA 2000, females outperformed males by an average of 32 points (about half the range of a competency level) and by 38 points on PISA 2006 (the difference is not statistically significant). Finland, the country with the highest achievement score in the PISA 2000 reading assessment, also posted the largest discrepancy between females and males (51 points) (Watanabe & Ischinger, 2009). The largest gaps in performance were on the reading skills of reflection and evaluation while the smallest discrepancy was in the reading skill of interpretation (Watanabe & Ischinger, 2009).

Attitudinal differences were also marked; more females than males identified themselves as diversified readers (Watanabe & Ischinger, 2009). Females also reported investing more time in reading than males (Watanabe & Ischinger, 2009). Males, on the other hand, were more likely to engage in reading for a specific purpose rather than for pleasure (Watanabe & Ischinger, 200a). Females tended to be more effective in their use of their reflective and evaluative skills along with their persistence to work out solutions to problems while males tended to use their reflective skills and purposeful reading skills to process information more effectively (Watanabe & Ischinger, 2009). The question is philosophical: Are the differences in skills that males and females bring to society enriching human experience or should all males and females bring the same compendium of skills and attitudes to society and the workplace?

The tradition gender-based differences for performance on mathematics assessments were reversed on the PISA 2003 mathematics test: The males on average
outperformed the females by 11 points for all countries and regions with the exception of Iceland which posted a 15-point gap in favour of females (Watanabe & Ischinger, 2009). Of the four content areas evaluated by PISA 2003 mathematics, the discrepancies (statistically significant) were most pronounced in the area of spatial and geometrical relationships (Watanabe & Ischinger, 2009). The gender-based discrepancy trend maintained itself in the PISA 2006 administration (mathematics is a minor focus but is sufficiently substantive to facilitate comparisons); however, there was increase in the proportion of males in both extremes of the proficiency levels - below Level 1 and above Level 3 (Watanabe & Ischinger, 2009). While the performance differences were somewhat modest, the attitudinal differences between males and females were substantive; females expressed much lower levels of interest and enjoyment in mathematics and perceived themselves as having low self-efficacy in mathematics (Watanabe & Ischinger, 2009). Just as negative male attitudes toward reading and reading-related activities seemed to be a more significant factor in the development of a reading culture than the actual discrepancy in reading performance, female attitudes towards mathematics seemed to be more detrimental to the development of an inclusive and technologically savvy culture and society than the real discrepancy in the performances between females and males.

On PISA 2006 (science), gender-based differences in performances were marginal and not statistically significant; however, the variation in scores for males is much larger than for females (Watanabe & Ischinger, 2009). Females tended to outperform males on the competency related to identifying scientific issues, while males tended to outperform females on the competency related to explaining phenomena scientifically (Watanabe &
The differences in male-female performance on PISA 2006 may, in fact, be related to the differential performance and attitudinal differences between males and females on the reading and mathematics test components. Males tended to prefer task-specific reading and mathematics concepts which may have more relevance to explaining phenomena scientifically, as many scientific phenomena are explained with the use of mathematics. Females tended to be more diversified readers and more interested in reading for pleasure and interest; therefore, scientific issues such as global warming may be of more interest to females as these types of topics connect the social and scientific worlds.

The underperformance of male students marginalized by the negative effects of poverty, racial difference, and ethnicity and the underperformance of women and girls in science and mathematics have been a concern for decades. According to Good, Aronoson and Inzlicht (2003), theses concerns are “not likely to disappear soon” (p. 3). Good et al (2003), psychologists at Columbia University, have put forward the theory of stereotype threat to explain the underperformance of girls and women in science and mathematics and the underperformance of marginalized males in reading and writing. Stereotype threat is defined as the “negative performance outcomes (lower standardized test scores and less engagement with academics) because they [test takers] are burdened by the prospect of confirming cultural stereotypes impugning their intellectual and academic abilities” (p. 4). The concept of stereotype threat includes self–generated pejorative interpretations of one’s performance based on perceived stereotypes and the need to reconcile one’s performance in accordance with the statements contained within a particular stereotype myth (Good et al, 2003). The negative impacts of these pejorative
interpretations on standardized test results can be overcome through positive messaging from a trusted source prior to the student’s participation in a testing event and through the provision of accommodations to the testing environment (Good et al, 2003).

Key credible professionals such as teachers can ameliorate the negative impact of _stereotype threat_ by sending positive messages about the malleability of intelligence and by creating a testing environment that reduces the visibility of students who appear to be culturally different from the majority (Good et al, 2003). Good et al state that anxiety produced by the testing-taking situation is higher in marginalized students who are made visible. However, Cadinu, Maass, Rosabianca and Kiesner (2005) have measured no significant increase in physical symptoms indicative of anxiety. They describe threat stereotyping as “pervasive” and that “relatively little is known about the processes underlying performance deficits under stereotype threat” (Cadinu et al, 2005, p. 572.) In their research, Cadinu et al (2005) identify the importance of thought intrusion and its relationship to the production of wrong answers on standardized tests.

_National IQ Scores_

Notions about a national coefficient of cognitive ability, a national IQ, have surfaced periodically throughout modern history. The basic questions concerning the effects of race, ethnicity, and culture on the development of cognitive capacity have surfaced and then disappeared as these types of questions are thorny, complicated, and challenge our very conceptions about what it means to be human. Yet these issues of cognitive capacity affect our choices, our job satisfaction, and life opportunities. Rindermann (2007) building on the work of Herrnstein and Murray (1994) and Lynn and
Vanhanen (2006) incorporated the results of PISA, PIRLS, and TIMSS with national IQ tests to determine if, in fact, there are significant societal differences in cognitive abilities.

Rindermann (2007) put forward these basic assumptions that achievement tests (PISA, PIRLS and TIMSS) and IQ tests “measure cognitive ability at the macro-societal level” and that intelligence and knowledge are measured differently within assessments and “these differences are often greater than between assessments” (p. 668). However, Rindermann based his findings on a meta-analysis of existing data, gathered to support other research questions. Meta-analytical studies create opportunities for researchers like Rindermann to use existing data to answer other research questions; however, the quality of this type of research is often hampered by the structure and design of original studies, sampling ambiguities (adherence to age and years of schooling requirements), statistical methods, and lack of clarity concerning what constitutes a nation (Rindermann, 2007).

Rindermann (2007) determined that the following factors affect the production of intelligence as measured by PISA, PIRLS, TIMSS, and national intelligence tests: family background and the number of years and type of formal education. According to Rindermann, “Intelligence and knowledge are both a result of successful education” (p. 689); however, Hart and Risely (1995) argue that parental influence exerted in the areas of education and language development can overshadow the influence of socioeconomic factors. PISA, PIRLS, and TIMMS assess what is described as school–near knowledge; that is, knowledge that students would acquire while they are in a formal educational setting (Rindermann, 2007). From Rindermann’s perspective this knowledge is mutable,
and developing societies can alter their performance as they introduce universal education policies that transform education from a privilege for some into a right for all.

Rindermann (2007) noted that wide-scale cognitive ability research is often disrespected and deemed questionable because the terms - genes, intelligence, and ability - are brought together in the same scientific study. Because cross-national assessments are funded by political institutions, there is often suspicion about the goals and agendas of such research (Rinderman, 2007). For example, in whose interest is it to make the statement that the average IQ for Canadian students is 102 as compared to 105 for Japanese students (Rindermann, 2007). These types of studies often stimulate the production of more questions than they provide solutions or answers to socioeconomic problems.

As I contemplate Rindermann’s work, I find myself generating questions for which the answers are difficult. Who benefits and who is harmed by these statements? Do these types of statements and analyses assist us as educators to produce better programs or understand our students in some unique way? Does the calculation of national IQs assist governments to plan the distribution of resources and their investment in the delivery of educational services? Is the national IQ yet another possible predictor of economic well being? (Dagenais, March 15, 2010)

Demetriou (2007) posed significant questions related to PISA, PIRLS, and TIMMS and their contribution to the calculation of national cognitive ability scores: “Is the g-factor [the general intelligence factor] malleable and can it be educated?” (p. 720). If the g-factor is malleable and educable and PISA, PIRLS and TIMSS have sufficient rigour to record with accuracy real fluctuations in cognitive ability, then it may be possible to rank societies according to cognitive ability (Demetriou, 2007). However, any such enterprise must be undertaken with extreme caution and care; and furthermore, programs and methods designed to raise the cognitive abilities of individuals, groups, and
nations should ameliorate the causes of “poverty, misery, and war within and between nations” (p. 722).

Summary

Countries, regions, and individual school boards have committed resources, time, and ingenuity to the development of standardized assessments to provide accountability for educational spending, to gather data on the nature of the competitive advantage, to gather comparative data on curriculum efficacy and instructional effectiveness, to provide vigilance over the educational process, and to gather the stuff that Ministers of Education need to brag to their electorate that their government is doing a good job. As national and regional economies shift their focus to become global and information-based, the dependence on human ingenuity, cognitive ability, and creative genius will increase. Tools like national regimes of testing, PISA, PIRLS, and TIMSS will become the measuring sticks of human capital development; however, with the introduction of any metric, the identity of individuals that produce the measurement disappear into the aggregate number and rank placement on a roster. Consequently, governmental programs and initiatives are designed to respond to the needs of the population and to shape the conduct of the population.

Large questions remain about the influence of national and international testing and the goals of meta-analysis of the data generated by large national and cross-national assessment programs. As Rindermann (2007) illustrated, it is possible to aggregate the scores generated by different instruments, create country IQ profiles, and show significant difference between countries. Rindermann’s work signalled the emergence of a new form of research that makes use of test scores generated by instruments designed
and delivered by another agency for a different purpose. While limiting his discussion to the national perspective, Ungerleider (2006) urges Canadian national assessment developers to consider the research goals and potentialities now possible through meta-analysis. According Ungerleider, researchers engaged in meta-studies should have input into the design and structures of these large national and cross-national instruments.

Demetriou (2007) and Pratt (2004) raise the issues of standardized testing to a higher level by combining the webs of peace and stability with those of knowledge production and relationship building. Meta-studies conducted with care and caution using data generated by the plethora of standardized testing protocols can not only bring difference into view, but it can also support strategies to ameliorate difference (Demetriou, 2007). Pratt (2004) adds that such studies designed to address human problems should employ multiple methodologies and be collaborative in nature. However, neither Pratt nor Demetriou address the compulsion of human nature to compare and to rank the performance of others with their own performance. Not everyone can be first; however, all the education ministers, regardless of the jurisdiction, would rather place first on the league table of test results than second or third or last.

As I contemplate the literature from the American, English and world perspective, I am drawn to the notion that the variation within schools and within regions and nations is often equal to or greater than cross-national variation. In our exuberance to celebrate our relatively high place on the roster, the variability of the scores that comprise the mean national or regional score escapes our view. Just as there are students who perform very well on these tests, we in Canada also have students who struggle. What is the nature of their school experience? How is this difference produced in Canada, and in our local school? How should we look beyond the aggregate number to the complexity of the composition of the number? (Dagenais, December 23, 2009)
CHAPTER 5

SPEAKING THE CANADIAN EXPERIENCE INTO EXISTENCE

Correlative spaces, the discursive structures, and the discourses that speak them into existence are key elements of Foucauldian research (Foucault, 2007a). The forces
and lines of tension produced by the American, English, and international spaces of standardized testing produce effects of knowledge and regimes of truth that operate in varying degrees on the Canadian space of standardized testing. However, due to geographical proximity, shared language, and somewhat common cultural experiences, the effects of the American voices and their scholarship, legislation, and policy development have influenced and have been incorporated to some degree into the Canadian experience (Lipset, 1990). According to Lipset (1990), a comparative historian, American voices, those voices that speak from the experience of occupying the same continent, can offer insight into the Canadian experience; as these voices can mirror both common experiences and the difference of two societies conceptualizing education under different societal principles (Lipset, 1990). The more populous American society was built on populist and meritocratic principles and has a relatively lengthy history and extensive scholarship in the field of standardized testing (Lemann, 1999). The less populous Canadian society was built on principles of peace, order, and good government and has a much shorter history and less extensive scholarship in that field (Lipset, 1990). From Lipset’s perspective both Canadians and Americans can come to a better understanding of themselves by considering what each has to say about a field of knowledge production.

Throughout this section American voices such as Donald Campbell (1988), American social scientist; Asa Hilliard (2000), a world renowned educator, historian, and psychologist; and Gustafson (2001), a Canadian professor writing from an American university, offer their perspectives in spaces not currently occupied by Canadian voices. Lemann (1999) provides an example of a history of the American standardized testing
experience, and McNeil (2000) writes about the harsh effects of the Texas standardized testing program on students who live and learn in poverty or struggle with disabilities. These are just some examples of the storied experiences that inform, influence, and shape the Canadian experience.

The Interwoven Threads

Canadians have long been interested in how well their education systems are meeting the needs of students and society. To provide information on this issue the provinces and territories, through CMEC, developed SAIP and PCAP to assess the performance of students in mathematics, reading, and science. (Council of Ministers of Education Canada, 2007)

The CMEC (2007) has identified standardized test results as the most appropriate strategy for providing assurance to Canadians that their educational system is meeting its performance goals; that is, equipping students with the knowledge and skills required to compete in a global economy of the 21st century. The assessment strategy, part of Five-Point Plan to develop a literate Canadian society, is built on an “analysis of pan-Canadian and international assessment programs of literacy skills (Council of Ministers of Education, Canada, 2009, p. 14). CMEC describes the importance of the task of improving literacy skills for all Canadians as follows:

[It is an] investment that will benefit every member of society, whether it be from the perspective of economic competitiveness, societal progress, or personal empowerment and fulfillment. Improving the literacy skills of Canadians means tapping into our collective strengths, creativity, and knowledge to create a truly Canadian vision for a better future. (Council of Ministers of Education, Canada, 2009, p. 21)

The CMEC views the participation by Canadian students in both national and international regimes of standardized testing as a positive sign that Canada is in tune with the global information economy and that Canadian students will be competitive both in national and international labour markets (2007). Participation in regimes of national and
international testing is both an accountability strategy and a communication/marketing strategy designed to demonstrate the strength of Canada’s overall performance in the education and training sector (Council of Ministers of Education Canada, 2007). CMEC takes pride in its effort to coordinate its own regime of standardized testing, PCAP, with other regimes of standardized testing such as PISA and TIMSS.

The public comparison and ranking of schools, provinces, and nations through the media publication of aggregate test scores has the potential to promote an environment of academic competition (Moll, 2004). The data has the potential to serve the multi-million dollar testing and study skills development industry as the press releases frequently paint a picture of a failing public education system (Moll, 2004; Stack & Kelly, 2006). According to Klein (2007), the public’s perception that schools are unable to rise to the task of educating today’s youth for tomorrow’s labour market, supports the development of public acceptance of a ready-made market-driven responses such as charter schools (publicly-funded schools managed by corporations). The introduction of charter schools, parental choice, and corporate investment in Canadian schools – these are all themes connected to the discourses of standardized testing and national and provincial competitive advantage (Klien, 2007; Stack & Kelly, 2006; Council of Ministers of Education, Canada, 2003).

CMEC (2003) has justified its interest in assessing the performance of Canadian schools and students by referring to the theory of Human Capital Development, which attributes increases in the Gross Domestic Product over a 20-year period to significant increases in standardized test scores of school-aged children. The Theory of Human Capital Development also permeated the tenants of A Learning Alberta, the 20-year plan
forwarded by Alberta Advanced Education that outlines Alberta’s strategy for becoming a world class competitor in the domains of science and technology and its reliance on the outcomes of standardized testing to measure the overall literacy gains made by its citizens (Alberta Advanced Education, 2006). Provinces such as Alberta are beginning to move voluntarily into positions of synchrony with the federal government initiatives. Because education is a provincial responsibility, the federal government must manoeuvre its entry into the field of education as an outcome of its interest and jurisdiction over the development of economic policy and labour force (Council of Ministers of Education, 2007).

CMEC (2007) assumes that all Canadians value measures of accountability for their publicly-funded educational system and that standardized test scores provide the best measures of accountability. Processes of accountability are assumed by the CMEC and by many Canadians to be a social “good” and that the government and its agencies should provide feedback concerning the use of their resources and the achievement of their predetermined goals; however, the mechanisms for providing accountability and the human and resource costs of providing accountability are in contestation (Moll, 2004b). In addition, CMEC (2007) fails to communicate the true costs of administering these tests although it does publish its overall budget allocation of $7.5 million for the administration of PCAP, which is relatively modest in comparison to the $49 million expenditure by the Ontario government for its provincial testing program (Council of Ministers of Education, Canada, 2007; Stewart, 2004). These large expenditures for regimes of standardized testing are often made at the same time that instructional budgets are cut and parents are asked to participate in fundraising activities to support the local
school (Stewart, 2004). These assessment costs do not include the costs attributable to lost instructional days - days devoted to test site preparation, student preparation, test administration, and managing anxiety (students, teachers and parents). Curricular costs are more difficult to calculate and more variable in nature. These include losses in breadth and scope due to curricular narrowing, loss of diversity due to the centralizing effect of standardization, loss of professional capacity (teachers leaving teaching because they no longer direct their classrooms), and the possible loss of public education to private enterprise (the proliferation of and acceptance of private tutorial alternatives such as Sylvan Learning Centres and charter schools).

According to Canadian sociologist, Scott Davies (2004), the pressures related to the “competition for educational credentials is rising across North America” and the importance that students’ and parents’ perceived need to achieve high scores on standardized tests also rising. These factors influence parents’ school choice decisions (p. 233). Although the participation of Canadian students in private schools has risen by more than 20 per cent in the past decade and by more than 40 percent in Ontario, the participation of Canadian students in private schools remains relatively low – about 5 percent of elementary and secondary student population (Davies, 2004, p. 234). Davies also notes that the Canadian public consistently identified non-affiliated academically intensive private schools with high-quality educational programming (61 percent of Canadians and 66 percent of Ontarians on the Angus Reid polls of 1999 and 2000 respectively). Because tuition costs for private schools exceed the capacity of most Canadian parents’ ability to pay, private tutoring companies provide an affordable
strategy for many parents who want to assist their children improve their school performance and their scores on standardized tests (Davies, 2004).

The assessment of the effectiveness of Canadian public schools, the quality of teaching, and the proficiency of student learning is a complex thematic topic in which the Foucauldian concepts of governmentality, regimes of knowledge, relations of power, and truth-telling statements are operationalized. CMEC operates as a Foucauldian form of government-at-a-distance with the responsibility to create programs (conduct) to shape the self-regulation of citizens, parents, students, and teachers. These programs are directed toward the goals of ensuring economic growth and prosperity, minimizing social risk due to global market fluctuations, and ensuring social order. According to CMEC (2007), the production of an educated citizenry is the means to achieve these goals.

Individuals and groups of individuals can choose to regulate their behaviour in a compliant manner which produces one form of knowledge (supportive discourses that identify with the truth statements produced by CMEC) or in a non-compliant manner which produces a different form of knowledge (the resistive discourses that put forward other types of truth telling statements). Ultimately, things get done through the relational capacity of power to produce action which Foucault (1980) describes as the productive capacity of power. Throughout the discourses of standardized testing, things are getting done, discourses are produced, and individuals and groups are being produced as subjects of discourses, but they are also being produced as objects of these discourses.

This productive capacity of power produces the discourses of assessment through the activities of standardized testing. These activities are regulated by systems of statements that define what can be said, by whom, and under what conditions. Teachers,
parents, students, policy makers, philosophers, and business leaders are constituted within the discourse of standardized testing that produce the interwoven threads of the understanding. Their public writing, the journal articles, newspaper articles, and speeches provide evidence of their subjection within the regime of standardized testing (Foucault, 2007). As one reflects on this domain of standardized testing, Foucault’s (2007) rhetorical question about the cultural production of knowledge is a useful touchstone: “What false idea has knowledge gotten of itself and what excessive use has it exposed itself to, to what domination is it therefore linked?” (p. 59) His response to the question reminds us that truth telling statements are culturally-produced and the products of political processes: “The political problem, to sum up, is not error, illusion, alienated consciousness, or ideology; it is truth itself” (p. 17).

Thus, these individuals constituted as subjects within discourses of standardized testing produce their own history, “in terms, of how it [history] would be traversed by the question of the relationships between structures of rationality which articulate true discourses and the mechanisms of subjugation which are linked to it” (Foucault, 2007, p. 56). Stewart (2004), speaking as the president of the Prince Edward Island Teachers’ Federation, illustrates how teachers and students are produced within regimes of standardized testing. According Stewart, participation in standardized assessment protocols such as SAIP and PISA wreak havoc on teachers and students.

Stewart’s resistive discourse is an oppositional play of power that advocates the construction of authentic, performance-based assessments, and professional development for teachers to hone their assessment and evaluation skills. Stewart’s comments are valid if the movement toward standardized testing were indeed a response to the level of
classroom teachers’ assessment skills and issues of pedagogical competence. However, if the move to standardized assessment is part of some larger economic strategy, improving the capability of classroom teachers to assess students’ learning will probably have limited effect on the overall decision to maintain regimes of standardized testing.

Gustafson (2001) employs a resistive form of discourse as he speaks from his position as a professor in an American university; he cautions Canadians about the terrors that accompany regimes of standardized testing as they promote general school decay, loss of instructional resources, loss of capable teachers, and a narrowed curriculum. He also warns Canadians about the subsequent and sometimes hidden consequences of standardized testing that can include the implementation of pay-for-performance plans for teachers and school administrators, resource allocation schemes based on systems of rewards and punishment, and the introduction of corporate solutions to instructional issues (Gustafson, 2001). Gustafson’s goal is to use the productive power of the truth-telling statements embedded in his discourse to motivate Canadians to resist the implementation of programs of standardized testing.

Wideen, O’Shea, Pye, and Ivany (1997) take up a resistive discourse from inside the Canadian classroom. According to Wideen et al, the implementation of regimes of standardized testing limit teachers’ opportunities to recognize classroom diversity, restrict their opportunities to use a variety of instructional strategies, and force them to teach to the test. Standardized testing from their perspective causes teachers to self-regulate their instructional practice to comply with the goals and objectives of the test.

The discourses of parents are either resistive or compliant and supportive based on their experiences with their local schools and their perceptions about the social “good”
of standardized testing (Canadian Press Newswire, 2000; Hesch, 2000). Parents want their children to reap the benefits of a solid education that will lead them seamlessly to employment or further educational opportunities; however, parents may differ in their opinions about what constitutes evidence that their children are learning what really matters. Some parents support the view that “teachers and schools should use standardized instruments” to serve the interests of their students more effectively (Zwaagstra, Clifton & Long, 2010, p. 37). Other parents believe that standardized assessments create a competitive classroom with a narrow curriculum focused on rote learning (Hesch, 2000).

Canadian philosophers, Klein (2007) and Ollman (2000), categorize standardized testing as an outcome of corporatism (an extreme form of capitalism) and not a legitimate element of the educational process. Ollman offers a resistive discourse that encourages students to manipulate the tests so they can achieve the outcomes that are beneficial for them. Klein offers a different form of resistive discourse that constitutes cautionary statements about the co-optation of the public school system by corporations.

These writers illustrate that they have produced different discourses which serve a threads that crisscross the domain of standardized testing. The governmental programs made possible by relational acts of power designed to shape the conduct of citizens have produced different discourses – some compliant, some resistive and some hesitant. This production of discourses leads us back to Foucault’s (1980) question about the nature and form of relational acts of power produced by governmental agencies:

What conception of politics itself is supposed by the existence of an active critical intelligence that can be contained in no given form of government; no established form of knowledge, no constituted types of political rationality? (p. 16).
Conversations and Interpretations

There is an enormous amount of information that has nothing to do with anything. There is a sort of civic landfill and you ought not get into the business of civic landfill. If your aim is to change conditions, then there is a certain amount of information needed, but not more. After that, one needs to address the questions, “Why does nothing happen? Why do some proposals that seem fairly reasonable, workable and sensible never get beyond the lip service stage?” That requires a very different sort of knowledge. That is the knowledge of the structure of power. (Franklin, 1996, p. 8)

Testing: Science and Measurement; Economics and Social Restructuring

Conversations about standardized testing in Canada have been shaped and influenced by the events and decisions that took place more than hundred years ago in France and the United States. In 1905 Binet developed the first intelligence test to identify French children with learning problems (Lemann, 1999). Binet’s test crossed the Atlantic with the assistance of testing promoters, Terman and Thorndike, who were immediately enthralled with the proposition that innate human traits could be measured through the use of a test (Lemann, 1999). From the perceptions of Terman and Thorndike, Binet’s test held the promise of transforming the quasi-science of psychology into a real hard science with the rigor of statistically analyzable data (Lemann, 1999).

Lemann (1999) uses the metaphor of the glass slipper to describe the promise held within Binet’s standardized assessment scales. Just as the glass slipper held the promise of a different future for Cinderella; standardized testing held out the promise of a different future for humanity as psychologists would now be able to see the invisible working of people’s minds and would be able to predict the future for individuals who took the tests. In the United States during the 1930s, four distinct mental testing camps
argued about the vision and purposes of standardized testing: the progressive group (schools should use standardized test scores to develop programs that promote the growth of the individual), the IQ testers (the use of standardized test results to reserve higher education for the brightest), the standards imposers (imposition of minimum performance standards for admission to higher education), and the education expanders (address students’ needs so that more students could participate in educational opportunities) (Lemann, 1999). The competition among the four streams of thought about standardized testing, its purposes, and its goals erupted into an issue of national concern, as American education became nationalized in response to the demand to develop a viable work force and to modernize its economy (Lemann, 1999). Although these issues related to the goals and purposes of standardized testing and to the production of a viable and competent workforce became visible in the 1930s, these same issues populate today’s discussions about standardized testing and its relationship to the goals and purposes of education.

Henry Chauncey, dean of students at Harvard University, recognized the potential of mental testing as a “science with limitless possibilities”; it was tied to progress, modernization, and everything that seemed to be progressive (Lemann, 1999, p. 5). The notion that mental testing seemed to be scientific and that it involved measurement and statistical analysis was critical to Chauncey’s beliefs about modernization and progress (Lemann, 1999). Chauncey also envisioned that standardized mental testing would offer the opportunity to remake American society by changing the rules for participation in leadership of America (Lemann, 1999).

Henry Chauncey dean of students Harvard University (February 4, 1945):

This project [implementation of wide-scale standardized testing] requires consideration from a lot of angles but men of vision in the field of testing,
vocational guidance, government, economics, education could be consulted individually and eventually in groups and a program eventually developed. Men with whom this might be discussed might even be so high in authority as ... President Roosevelt himself. (Lemann, 1999, p. 4)

Nicholas Lemann: This is what Henry Chauncey wants to do in (for, really!) postwar America: he wants to mount a vast scientific project that will categorize, sort, and route the entire population. (Lemann, 1999, p. 5)

Henry Chauncey’s dream set the stage for the development and implementation of the School Aptitude Test (SAT), currently a common admission tool for American universities, and the establishment of the Educational Testing Service (ETS) (Lemann, 1999). ETS develops and distributes many of the high stakes testing tools used in the United States and Canada including the SAT (Lemann, 1999). Henry Chauncey’s legacy resides in his vision that society could understand the results of a student’s learning as a number produced on a standardized test and that a number could serve as the basis for high stakes decision-making about the lives of students, citizens, and society (Lemann, 1999).

While Henry Chauncey was re-visionsing America as a society ruled by intellectual elite whose membership was determined by the results of standardized testing, Milton Friedman, an economist at the University of Chicago, was promoting his anti-Keynesian theories (Klein, 2007). Friedman (1982) and the Chicago School economists advocated for small government and the privatization of the production of goods and services including the provision of educational services and health care.

Milton Friedman: [The state should limit its responsibility] to protect[ing] our freedom both from enemies outside our gates and from our fellow-citizens: to preserve law and order, to enforce private contracts, to foster the competitive market. (Friedman, 1982, p.2).

Henry Chauncey’s dream and Milton Friedman’s economic theories found fertile ground in American educational policy as evidenced in the report, A Nation at Risk
and subsequently in the *No Child Left Behind Act* (2002). A plethora of state testing programs emerged in the United States, and in Canada conversations about possibility of a national testing program emerged in the late 1980s. These Canadian conversations grew to fruition in 1991 amid an initial groundswell of oppositional discourses (Moll, 2004b).

Meaghan and Casas (2004), Gustafson (2001), McNeil (2000), and Robertson (2007) – these commentators have established their oppositional discourses in the domain of standardized testing based on their perceptions of educational harm to both programs and students. The scope of their resistive discourses has expanded to include their issues with control exerted by governments on their citizens as an outcome of regimes of standardized testing (Meaghan and Casas, 2004; Gustafson, 2001; McNeil, 2000; and Robertson, 2007). These commentators oppose Milton Friedman’s (1982) economic vision of a laissez-faire economy in which all enterprises including education are subject to market forces, and they view standardized testing as the vehicle that provides the data that fuels the statements of justification for market-based solutions (Klein, 2007). Klein (2007) references the conversion of public schools into charter schools in New Orleans to illustrate the ease with which provision of public education can be transformed into a market-driven solution. According to Klein, it just takes the right combination of social disaster and perceptions of low performing public schools based on data from standardized testing to create the appropriate environment to re-create education as a free-market enterprise.

Milton Friedman: Most of New Orleans schools are in ruins as are the homes of the children who have attended them. The children are now scattered all over the country. This is a tragedy. It is also an opportunity to radically reform the educational system. (Friedman, 2005 cited in Klein, 2007)
Friedman’s theories concerning economic reform and the implementation of economic policies that foster unfettered trade in goods and services such as education were implemented in test sites in Latin America (Klein, 2007). These school systems are characterized by regimes of standardized testing that negatively impact the teaching and learning environment by “reducing student evaluation to mere measurement...turns the learning process into an output ... [and] facilitates comparative analysis of school performance through the mandatory publication of test results” (The Civil Society Network for Public Education in the Americas, 2004, p. 98-99).

In Canada, Milton Friedman’s theorizing is embodied in the discourse and actions of the Fraser Institute that leads advocacy initiatives for school reform and market-based education solutions by investing in research designed to influence Canadians’ willingness to accept of private options for the delivery of educational services (Froese-Germain, 2004). The Fraser Institute publishes regular report cards on schools in British Columbia that feature standardized tests scores as the major data source for school comparison (Froese-Germain, 2004). These reports cards are designed to set benchmarks for improvement, serve as a communication tool for parents, and to provide public accountability for schools in British Columbia. Educators have challenged these report cards by stating that they are simplistic school rankings designed to support an ideological project disguised with scientific rigour. Despite the concerns posed by educators, the Fraser Institute continues with its mission of advocacy for privatization of public education based on the data collected from projects of standardized testing (Froese-Germain, 2004).
From the perspective of Meaghan and Casas (2004), the discourse produced by the economists and promoted by the media has transformed education from a “social good” into a vehicle for economic prosperity. The measurement of students’ ability to achieve the prescribed learning outcomes is now of economic interest, and standardized testing with its statistical capacity seems to be the most appropriate tool to both economic and media interests (Meaghan and Casas, 2004). A marketplace ideology has subsumed education and transformed students into clients and learning objectives into educational products (Meaghan & Casas, 2004).

Donald T. Campbell’s (1988) resistive discourse took the form of a challenge. From Campbell’s perspective as a sociologist, it was not reasonable to believe that any single standardized testing instrument had sufficient accuracy to support high stakes decision making. According to Campbell, for a social science such as education to be considered a science, “the measurement of social attitudes and behaviours must be valid and reliable”; thus, social scientists developed an interest in tools such as standardized testing instruments (p. xvii). However, Campbell concludes that “approximations of knowledge are all we can ever hope to achieve through measurement,” and that our quest should focus on the development of a number of tools – imperfect though they may be – rather than a single seemingly perfect one (p. 27). Thus, Henry Chauncey’s quest for certainty and a true window into the minds of human beings based on the measurements produced through standardized testing is in reality at best an approximation of what truly exists on any particular day.

Donald T. Campbell, University Professor at Lehigh University and professor emeritus of psychology at Northwestern University: (1988)
Our longing is for data that prove and certify theory, but such is not to be our lot. Thus, it is more valid using a series of imperfect measures than a singularly imperfect measure. (p. 27).

However, designers and test users often miss the biases inherent in the measurement tools (standardized testing), and many true readings are discarded because they fail to meet the expectations of the researcher, and biased readings are retained because they comply with researchers’ expectations (Campbell, 1988).

Wasserman (2001) illustrates Campbell’s concern related to the measurement, and in particular, the measurement of human behaviour, using a single measurement through her experience of assessing reading using a standardized reading test for a group of six-year olds. Wasserman describes her classroom observations of a little boy, who was a good reader. He took his time carefully considering each question and sounding out the words that he didn’t recognize; however, he did not have enough time to finish the standardized test. In contrast, a bouncy little girl, with few demonstrable reading skills, coloured one circle for each question and completed the exam easily within the allocated time (Wasserman, 2001). The little boy who laboured over questions scored poorly; however, the little girl who treated the test as a colouring activity scored well (Wasserman, 2001).

Selma Wasserman, professor emerita, Simon Fraser University: (2001)

Our obsession with numbers has led us into deep waters of quantification, and now we have rating scales for nearly everything – relying on these measures to guide our life choices...Yet the more we strive for accuracy in our numerical ratings, the more it eludes us. (p.30)

From her position as an educator, Wasserman (2001) contextualizes the comfort that she feels when she is able to transform a complex task like a student’s learning into a
number, a grade - something that comments where that student fits in relation to other
students. For many teachers, placing a grade on a student’s work/exam provides security.

Selma Wasserman: (2001)

Hearing the numbers takes us from the unknown to the known. There is a palpable
sigh as our tension is released. For once, we can be certain that we have glimpsed the truth. (p. 32)

However, Wasserman (2001) distinguishes between the importance of teacher-led
evaluation processes and the regimes of standardized testing that have recycled
periodically during the past 40 years. She notes that this recycling continues despite the
copious documentation of serious flaws and limitations of standardized testing
instruments, the narrow outcomes measured by the tests, and the potentially unhealthy
outcomes for students. Wasserman cites a number of flaws related to measurement tool
construction: the requirement for a single correct response, the reward for information
recognition as opposed to information generation, and focus on basic skills. Wasserman’s
appraisal of the limits inherent in any regime of standardized testing reinforces
Campbell’s (1988) methodological position that all knowledge produced in the social
sciences is at best an approximation. According to Campbell (1988), data collected from
multiple perspectives are critical to constructing some understanding of human
behaviour. Wasserman adds an additional caveat that we should be especially careful
with numbers - even those emanating from very sophisticated measuring devices.

Selma Wasserman: (2001)

[M]easurement is subject to both human and technological error. Moreover, in
delivering the numbers, the needs to be self-serving and to protect one’s own
interests are also critical factors in contributing to error...If quantum theory has
any validity, then it should cause us to pause and consider the kinds of numbers
that serve as indicators of student performance on standardized tests...According
to Bohr, the numbers “mean what they mean” in mathematics...It may be the very
height of arrogance to insist that a student’s numerical score on a standardized test
is free of the variables introduced by human behaviour and that the score can mean anything at all beyond the number itself. (p. 35-36)

Wasserman (2001) draws on the quantum theory and the Heisenberg principle to explain the challenges of measuring intellectual activity. The Heisenberg Principle states that “locating a particle in a small region makes the momentum of the particle uncertain; conversely, measuring the momentum of a particle precisely makes the position uncertain” (Wikipedia, n.d). Essentially, attempts to measure human behaviour through the application of standardized testing are confounded in two dimensions: As the specificity of the test increases, the ability of the individual subject to demonstrate change in his or her behaviour increases; conversely, as the generality of the test increases, the ability of the individual to demonstrate a change in his/her behaviour will decrease. In effect, the application of quantum theory restates the concerns that Campbell (1988) identified in his early theorizing about the difficulty in making generalizations about human behaviour based on a single measurement. Generalizations concerning particularly specific and narrowly defined characteristics of a locally defined population can be made on the basis of human measurements (test scores); however, generalizations about broad characteristics such as the reading ability of Canadian grade eight students are best understood as probabilities.

Meaghan and Casas (2004) confirm Wasserman’s perspective by stating that standardized tests were not designed to be precise instruments for measuring the merits of an educational program or for evaluating a student’s learning of a specific curriculum. Hilliard (2000) draws attention to what is obscured about the teaching and learning process when this complex phenomena is reduced to single numbers placed in an array on a multi-dimensional grid.
Asa Hilliard, professor of urban education, Georgia State University: (2000)
For years I have been interested in a special group of educators. They are the
ones who create powerful education environments. They are the ones who are not
puzzled about how to raise the achievement levels of children from any
backgrounds to levels of excellence. They are the ones who see the universal
genius, spirit, and humanity in all children. Things like poverty, bilingual status,
single-parent families, and even threatening neighbourhood environments present
no obstacle to the attainment of excellence for their students. These are the
educators who are lost in the aggregate, which includes unsuccessful educators
who are puzzled about how to teach for excellence. (p.293)

Hilliard (2000) identifies challenges of reductionism: the specific, the particular,
and the extra-ordinary are aggregated with the ordinary and the substandard to comprise
the performance of the “average teacher.” Thus, the opportunity to learn from the
specific, the particular, and the “extra ordinary” is lost because its existence is not made
available through the reductionist art of statistical generalization (Hilliard, 2000).

Thus, Henry Chauncy, Milton Friedman, and Donald Campbell represent three
significant streams of thought that have collided to create the standardized testing
landscape in North America. Henry Chauncey forged the path for acceptance of
standardized testing as a method to determine who should benefit from postsecondary
education; Milton Friedman advanced the economic policies that integrated standardized
testing into strategies for the privatisation of public schools; and Donald Campbell
provided the early voice of criticism about the meaningfulness of social measurement.
The Canadian educational landscape has been influenced by the operation of these forces
in the United States and is characterized by increased reliance on standardized testing
scores for educational policy construction marked by active criticism through the public
voices of educators, scholars, and concerned citizens (Council of Ministers of Education
Canada, 2007; Wasserman, 2001). Although standardized testing is not legislated in
Canada as it in the United States, the power of its surveillance is felt at all levels of
society: citizens, educators, parents, students, and teachers. Volante (2005), associate professor of education at Brock University, states that

Holding individual teachers, schools, and districts accountable for student performance continues to be a central feature of educational reform throughout the Western world. Following the lead set by the United States, Great Britain and New Zealand, Canada is increasingly adopting a system that is driven and controlled by standardized testing. (p. 1)

In Canada, standardized testing can be grouped in three large categories: low stakes, medium stakes, and high stakes. From my perspective, low stakes standardized tests are used to provide comparative feedback to students about their academic performance. The results do not impact students’ opportunities nor do they influence policy development. Medium stakes testing produces results that may provide feedback to students and to schools; however, their principal goal is to support policy development and inform budgetary priorities. The results of high stakes testing affect the opportunities and choices for students and impact policy development and budgetary cycles.

**Accountability**

Although politicians routinely suggest that standardized testing will result in greater school accountability, there is mounting evidence to suggest that test scores are used mainly for sorting and ranking children, with serious adverse impact on some – particularly low-income and minority children. (Meaghan & Casas, 2004, p. 35)

CMEC, in its role as a federal and international voice for Canadian education policy, connects the publication and distribution of standardized test results to its conceptualization of public accountability (Council of Ministers of Education Canada, 2007). CMEC attributes its need to provide accountability to the Canadian public’s vested interest in knowing how their public schools are performing; and from CMEC’s perspective, the use standardized testing that measures the common curricular elements is
the most efficient method (Council of Ministers of Education Canada, 2007). A call for school system accountability is a right asserted by CMEC on behalf of the citizens - parents, students, and taxpayers (Dolmage, 2004). CMEC is employing a strategy of governmentality in the exercise of its programs of conduct on school systems and citizens. CMEC’s strategies are based on the assumption that citizens should want school accountability as part of its taken-for-granted Canadian landscape and should want to support activities that promote accountability. The visibility provided by test score results draw attention to those jurisdictions that perform outside the boundaries of the mean and creates the opportunity for these jurisdictions to become the subject of further actions.

According Hilliard (2000), children, parents, communities, and the public have a profound responsibility to ask for accountability and demand high standards from their public school system – and in particular, they must demand instructional excellence. The implementation of regimes of high stakes standardized testing may provide a form of accountability, but this form of accountability may not reflect high educational standards.

Asa Hilliard, professor of urban education, Georgia State University: (2000)

There can be no argument about the necessity for high standards....Confusion makes some interpret the criticism of high-stakes standardized testing as opposition to high standards...High standards may not be connected to high stakes standardized testing at all...The real issue is to give children the teachers who can make them reach high standards. (p.295)

In Canada, test results from the national assessment program, SAIP, are presented in comparison to the public’s expectations (a form of standards) about what 13- and 16-year old Canadian students should know and be able to do as constructed by a focus group of select participants (Robertson, 2004). Since the focus group was limited in size and the identity of its participants protected, the relationship of the focus group’s
decisions and the real expectations of the Canadian public remain difficult to determine (Robertson, 2004).

Linn (2000) outlines four reasons that make standardized testing attractive to policy makers such as CMEC: It is inexpensive relative to more costly reforms such as increasing instructional time, it can be externally mandated by policy or regulation, it has a short implementation time, and it provides the impression that something is being done. Through these regimes of standardized testing, neoliberal governments (the Bush administration in the United States and the Harper administration in Canada) can mobilize data that can be used to call into question the outcomes of public sector services such as education and create justification for their transformation into private enterprises (Tabb, 2002). According to Donlevy (2000), standardized tests have the capacity to produce “rich displays of first-rate student performance”; however, the types of student performances highlighted by standardized testing are not always synchronous with the more valuable understandings about the purpose of education (p.331). Standardized test results do not generally support conversations about the broad goals of education, but they do support conversations about strategies to raise test scores, a school’s reputation, and a school’s ranking in the national arena (Donlevy, 2000).

While educational reforms involving standardization, competition, and accountability bring certain changes to the public schools, their benefits are largely superficial. The schools, however, have a deeper role in helping young minds and hearts to become sensitive to the world around them and knowledgeable about how their gifts may unfold in social union. (p336)

According to Kuehn (2004), school accountability strategies based on standardized testing are not focused on inputs (students and teachers) and process elements (instructional design, instructional strategies, learning activities, and
instructional resources), but only on outputs (the production of test scores). Such calls for accountability are often accompanied by reduced funding for the activities of schools and curriculum construction, and an increased investment in resources for standardized assessment. Hampton (2004) illustrates the effects of increased calls for accountability in Alberta: a 20 percent cut in the budget for education and an increase in externally imposed tests.

CMEC uses the results of standardized testing to provide comparative data to the provinces and to promote opportunities for the provinces to develop common curricular goals (Fourier, 2000). Provinces like Ontario have chosen to use the results from standardized testing as a “gatekeeper” for high school completion credentials (Meaghan & Casas, 2004). According to Meaghan and Casas (2004) standardized testing serves the role of the “gatekeeper” by controlling the progress and recognition of students’ accomplishments: The Ontario Secondary School Literacy Test (OSSLT) regulates the distribution of the Ontario Secondary School Diplomas. Data produced by OSSLT results are also used to regulate the tenure of teachers in pay-for-performance contracts and to distribute school budgets for schools whose budgets are based on the performance of their students on standardized tests (Meaghan & Casas, 2004; Brackenreed, 2006). Test results now serve as proxies for the measurement of educational effectiveness because broad measures of school effectiveness and student achievement have been abandoned in favour of a single accountability strategy: the publication of standardized test results.

Diane Meaghan, professor of Sociology at Seneca College and François Casas, undergraduate chair of the department of Economics at the University of Toronto: (2004)

Upon release of the results of standardized achievement tests, politicians at all levels are customarily making comparisons among schools, school boards, provinces, and countries. Parents worry about the chances of their children being admitted to universities, and members of school boards and education ministry
bureaucrats praise or lament test results, while media clamour to publish school rankings based on those results. The public has been brainwashed into believing that testing is a means to ensure educational standards and an objective basis on which to distribute resources – despite the lack of compelling evidence that these alleged benefits outweigh the considerable demonstrated disadvantages. (p. 36)

A generally-accepted goal of standardized testing is the measurement of student learning and the improvement of instruction; however, an additional goal, the motivation of high-quality performance by students and teachers through the offering of rewards and sanctions, is a critical element in the discussion of accountability (Ryan, Ryan, Arbuthnot & Samuels, 2007). Thus, a school’s performance and perhaps its budget and the tenure of its teachers are tied not only to the instructional outcomes achieved by its teachers and performed by its students, but also the goals and beliefs of its students.

Katherine Ryan, associate professor of educational psychology, University of Illinois; Allison Ryan, associate professor of educational psychology, University of Illinois; Keena Arbuthnot, postdoctoral fellow, Harvard Graduate School of Education; and Maurice Samuels, doctoral student, University of Illinois: (2007)

There is a fundamental assumption that test taking is a singular experience for students. That is, the assessment context (high stakes vs. low stakes) will not influence or influence in a similar way how individuals or groups of students engage the test-taking process ... Not only knowledge but individuals’ personal beliefs and goals influence performance. (p.5)

Cawthon (2007) provides a supportive discourse that details a number of positive outcomes for children with disabilities resulting from commitment of the American government to provide a fair, equal, and significant opportunity for all children to receive a high-quality education evidenced by a minimum proficiency on state-wide standards assessments. A positive a productive outcome of the introduction of state-wide standardized assessment regimes is an increased awareness about the needs of students with disabilities (Cawthon, 2007). Teachers and administrators are more willing to provide test accommodations to students with disabilities, and administrators are more
willing to include their scores in the school performance reporting. In both Canada and in the United States, educators are grappling with the meaning of an accommodated test result and its relationship to the mastery of the test objectives (Cawthon, 2007; Brackenreed, 2006). From Cawthon’s perspective, it is critical that students with disabilities are afforded the opportunity to write the standardized assessments as their educational future is dependent on the outcomes of these assessments, and it is critical to both students and school systems to include the results of students with disabilities in school-wide performance reporting. However, Brackenreed (2006) questions the impact of accommodations on the validity of test results produced on the OSSLT.

Critics of the implementation regimes of standardized testing as an accountability strategy argue that public accountability is important; however, schools and the Canadian public education system can provide public accountability through other measures that are more meaningful and less harmful to children (Meaghan & Casas, 2004). Although most educators and citizens will agree that the Canadian public education system should be accountable for the resources it consumes and the products its produces, groups of citizens often disagree about what constitutes accountability (Hampton, 2004).


My position is that it is not a question of accountability, but rather a question of what types of measures are appropriate, used by whom, on whom and when, and what the results are used for. The question must be asked, “To what extent do we want accountability measures to drive practice?” (p. 218)

Ireland (cited in Moll, 2004) argues that processes for demonstrating the accountability for teaching and learning should be established by teachers. Teachers - not board consultants or external experts - should create the assessment tools to evaluate the outcomes of their teaching, and teachers should design the reporting instruments. Ireland
is advocating a shift in the relations of power in favour of teachers. Ireland recognizes that teachers have the expert knowledge to create large scale assessments and that it should be teachers who have the responsibility to evaluate the outputs of their work (Moll, 2004). Ireland’s resistive discourse takes a different tactic by questioning the authority of so-called experts and policy makers to assess the work of teachers through the administration of externally-administered standardized assessments.

Marita Moll, guest lecturer, Carleton University speaking about David Ireland’s accountability project: (2004)

The project worked brilliantly because teachers carried it out … And when we had school and board-wide results, what mattered to Dave [Ireland] most was what we [teachers] were going to do about them. How were we going to change our teaching so that students’ performances and their knowledge and abilities improved? (p. 3)

Ireland’s focus was clearly on the productive uses of the test results gathered through the administration of this standardized test project. His test data collection and analysis focus was centred on these key questions: What changes to teaching practice would be invoked and what improvements to students’ learning would result (Moll, 2004).

The meaning of test results is often reduced to simple rankings on league tables. If test results have any meaning, in 2003 Canadian students performed relatively well on PISA – reading (second), math (fifth) and science (fifth) tests – in relation to the other 32 participating countries (Hoffman, 2004). However, the media, politicians, and business leaders demonstrated their ability to transform a “good news” story into a tale of impending doom (Hoffman, 2004). Rather than celebrate the achievements of Canadian students, the results are treated like podium finishes: one silver and two “that don’t count.” Canadian politicians and business leaders fail to recognize that there is no world-wide curriculum, no generally agreed upon standards for public education, and no
podium medals for first place PISA test score rankings (Hoffman, 2004). The Canadian public has been conditioned to expect “bad news” stories about our educational system; and the media, politicians, and business leaders have become adept at ignoring positive test results or mis-representing test outcomes to produce statements that cause alarm about the school system as a method of gathering public sympathy for educational policy reform (Hoffman, 1999).

Robertson (2004b) notes that the media, politicians, and business leaders seek out the bad news and abandon good news stories like the test score results for equity students in Saskatchewan (those students who have been identified by their teachers or principals as having a disability). Saskatchewan, a seemingly unimportant province whose students have performed consistently below the average of those in other provinces, demonstrated above average performance for equity students on the 2003 PISA assessment (Robertson, 2004b). According to Robertson, this is a good news story that has been allowed to die. Although Saskatchewan equity students performed well, it is unlikely that the CMEC will recognize the importance of the results and translate Saskatchewan’s success in generating positive performances from equity students into generalized policy statements that promote a Saskatchewan perspective for developing educational policy related to education equity (Robertson, 2004b).

Robertson (2004b) points out that the test results on different international assessments posted by students living in the same jurisdiction may contradict each other. The differences between the test results for students living in Saskatchewan earned on PISA 2003 and the International Adult Literacy Survey (IALS) appear to be considerable. The test results of equity students in Saskatchewan exceeded the performance of equity
students in any other province on PISA 2003; however, Saskatchewan First Nations and Métis adults scored below the benchmark for participation in a complex knowledge-based economy on the IALS 2007 (Cowan, 2008, January 7). These results appear to be contradictory; however, the sample construction, the definition of “equity,” the assumptions about the category of equity (does the category include visible minorities or is it limited to students with cognitive and physical disabilities?), and level of performance expected by equity students are not included in the meta discussion posted by Robertson (2004b). These differences do serve as a caution for the reader and for those who intend to use national and international test results in decision making. Students’ performances can appear to be below the level of expectation on one assessment tool and at or above the level of expectation on another tool.

Technologies of Control

[E]conomic competitiveness depends on the “academic effectiveness” of …public schools, specifically with respect to the “basics” of shared factual disciplinary knowledge and the “functional skills” of reading, writing and arithmetic. (Vinson, 1998, p. 7)

Technologies of control operate at all levels of curriculum production and implementation. These control mechanisms regulate the following functions: selecting the content authority, identifying the content that should be taught, determining how well students must learn the content, determining some predetermined level of performance, and establishing the process for reporting the results of student and school performance to pre-identified stakeholders (Graham & Neu, 2004; Vinson, 1996). Establishing a national curriculum, whether through the actions of deliberating on what students should know or surreptitiously through the administration of regimes of national testing, privileges some knowledge and devalues other knowledge and privileges some groups of individuals over
other groups (Vinson, 1998). Technologies of control operate on the bodies of individuals, the social body, and the curricular documents (Foucault, 1980). According to Graham and Neu (2004), the activity of writing a test in a standardized environment subjects the body to regimes of disciplinary power, and the activities of reporting and publishing the test scores subjects the social body to regimes of disciplinary power that operate outside the regimes of sovereign or legislative power. These regimes of disciplinary power serve as a normalizing force to produce docile bodies, useful citizens and a compliant social body that is willing to submit to being governed (Foucault, 2007).

*Curriculum-as-a-Document*

Vinson (1998) draws on Foucault’s theories concerning truth, power, and the production of knowledge. According to Vinson, national curriculum standards serve as a regime of truth that is produced and maintained by a system of power (regulations, policies, and legislation); and this regime of truth produces disciplinary effects of power on students, teachers, parents, and the general public. The construction of curriculum is a privileging process: The knowledge of some groups is valued more than that of other groups; therefore, their truth is represented in the curriculum and forms the backbone for student assessment strategies. The implementation of the curriculum produces disciplinary effects of power in those entrusted with its administration and assessment; the curriculum controls what can be taught, how it can be taught, what should be known, and what can be assessed (Vinson, 1998). Meaghan’s (1998) theorizing complements Vinson’s work on control by referring to Foucault’s theorizing on authority and its need for both a body of theory such as that which supports a particular process of curriculum development and a set of regulating practices such a regime of standardized testing and
its mechanisms for reporting the results. Meaghan (1998) states that standardized testing and its accompanying mechanisms for reporting results serve as a disciplinary mechanism that monitors the effectiveness of jurisdiction’s curriculum and the efficacy with which schools and ultimately teachers are able to deliver the curriculum. According Meaghan (1998), the Foucauldian gaze is exercised through the administration of standardized tests because students, teachers, and classroom practice are made visible.

In Canada, curriculum production is a provincial responsibility; however, through regimes of national and international testing, CMEC influences and encourages the provinces to privilege those curricular elements that are the focus of national and international assessments through collegial discussion and the formation of professional groups such as the National Science Curriculum Consultation (Fournier, 2000). Vinson (1998) encourages scepticism about production of national curriculum standards and their assessment, as the establishment of national curriculum standards legitimizes certain “knowledges as true, certain methods as appropriate to the establishment of truth…certain acts as consistent with truth, and the power of certain individuals to determine truth and what constitutes it” (p.22). In a geographically-vast and culturally-diverse country like Canada, the questions erupting from this situation concerning the national/provincial/territorial curricular relationship is this: Whose knowledge is deemed legitimate, whose knowledge is silenced, and what violence is perpetrated through disciplinary acts of power upon those whose knowledge is silenced?

Meaghan (1998) states that teachers’ knowledge is silenced in school systems that rely on externally-mandated standardized tests to assess student learning. The implementation of regimes of standardized testing delivers a message to teachers that
they do not have the necessary expertise to be involved in such an important task as the assessment of their students’ learning, and it systematically removes the right of teachers to assess their ability to instruct the curriculum (Meaghan, 1998; Runté, 1998). In contrast to this perspective, both Meaghan and Ireland report that teacher-developed and teacher-administered assessments produce fair and accurate results for students (Moll, 2004; Meaghan, 1998). By removing teachers’ authority to assess their own students’ learning, a systematic deskilling of teacher is sanctioned through the control exerted by the regime of standardized testing. By pre-packaging not only instructional materials but also assessment tools through some centralized authority, teachers are removed from both planning and assessment functions of the teaching and learning process (Runté, 1998). Runté argues that teachers’ exercise of professional judgement is erased because teachers are no longer able to adapt curriculum to meet local needs; instead they must govern their practice with the expectation of “covering the curriculum.”

Robert Runté, professor University of Lethbridge: (1998)
The right to define and evaluate the product of one’s own labour is a key professional privilege and its loss further undermines teacher’s professional status...Once teachers lose sole responsibility for evaluating the product of their labour (that is, student achievement), they also lose control over self-assessment. (p.168)

Mullinix (2001) echoes Runté’s concern about the shrinking scope of instructional practice. According to Mullinix, the generalized introduction of the externally-mandated standardized testing constitutes the severing of assessment from learning. Learning and assessment should be linked in a self-informing cycle; learning and curriculum design inform assessment; and the results of assessment should inform the design of the teaching and learning process (Mullinix, 200). In educational systems driven by standardized
testing regimes, assessment drives learning and influences curriculum development (Mullinix, 2001).

Donlevy (2000) states the intention of today’s curriculum reflects the needs of business and is marked with the words standardization, competition, and accountability although imagination and creativity are considered to be key attributes of entrepreneurship. There appears to be an inherent conflict embedded in the processes of governmentality designed to produce societal order and self-regulation, and yet at the same time, these same processes are supposed to provide support for individuals to exploit their creative and imaginative energies. Scott (2007) explores the issues of instruction and school grades as they relate to attributes of creativity and imagination. Scott defines school by quoting Michael Thompson, a University of Chicago psychologist, who describes schools as a “place where former A students teach mostly B students to work for C students” (p. 71). The current hyper-rational school environment favours the production of B students, those who can work effectively for others, become solid citizens and earn a reasonable income (Scott, 2007). These students maintain social order and exercise self-regulation and self-discipline based on the directives of programs produced by governing agencies.

Should the Canadian educational system and its schools be concerned about the production of creative and imaginative students and the subsequent impact of creativity and imagination on standardized school performances? This rhetorical question does not have an easy answer. According to Scott, imaginative and creative students, such as Dr. James Orbinski, human rights activist with Médecins sans Frontières, do not perform well in the Canadian public school environment, and their standardized test scores may have
adversely affected the Canadian norm in some manner. Bob Young, a successful Canadian business person and C student, would argue that receiving a C was beneficial; it created the impetus for solving problems, clarified alternative methods, and highlighted opportunities (Scott, 2007). According Scott (2007), “A” students and “B” students learn to work within the educational system produced by a standardized curriculum and regimes of standardized testing; these effects of disciplinary power exerted by the curriculum and the regimes of standardized testing produce them as useful citizens who can attend to regulations operating within the school environment. C students, however, produce an oppositional discourse resistive to the disciplinary effects of power and self-regulation, yet these students may create something new and sometimes something valuable to society. As Canadians concerned about our youth and our future as a nation, we must ask ourselves if the order and control gained by imposing regimes of standardization on youth is worth the possible loss of creative and imaginative outputs from those students who are resistive to regulation imposed by the regimes of standardization.

According to Robertson (1999), parents constitute a key force in the curriculum debate as parents are concerned about their children’s opportunities to learn the skills necessary to compete in the global economy. For many parents, test scores appear to be an easy to understand indicator of their child’s relative performance. Graham and Neu (2004) describe the demand of rural Alberta parents to have children involved in standardized assessment programs because they wanted to feel confident that their schools were comparable to urban schools. Robertson (1999) identifies that many parents do favour some sort of standardized testing; however, they are resistant to initiatives that
reduce local control of curriculum and introduce standardized test scores as part of student achievement reporting. Parents are becoming increasingly aware that competition in the global economy is based on their children’s ability to compete for places in postsecondary educational institutions and eventually the global labour market; therefore, their engagement in the educational process is motivated, in part, from an economic perspective (Robertson, 1999).

Heather-Jane Robertson, author/social activist: (1999)

[T]he public support high stakes testing, but only when the stakes are very low. .. [I]f I knew less about the consequences...I would think that high-stakes testing is as neutral as taking the temperature of the room – an act that changes nothing, a benign form of measurement that is inexpensive as well as innocuous. But a parent’s desire for testing isn’t just the product of innocent curiosity. .. It is a downloading of responsibility from the rulers of the global city to individuals who download failure....The rulers of McWorld encourage us to blame the education system, not the finance minister and certainly not employers or their apologists. (p. 3)

Balancing the relationship among the various curriculum pressures is difficult and complex: Parents want assurance that children are learning what they need to know to be productive and useful citizens. The educational system has offered standardized test scores as a measure of accountability to assure parents that schools are meeting their goals; however, the publication of test scores is not always accompanied by a statement of human cost that outlines the implications of the restrictions placed on curriculum possibilities when the curriculum is restricted to only that which is assessed (Robertson, 1999). To achieve the mandated test scores, teachers often respond by eliminating locally-developed curriculum; and the school administration may reduce the number of opportunities for teachers to set their own goals and demonstrate their professional competence (Wideen et al, 1997).
Curriculum-as-Taught

The regimes of standardized testing exert control and exercise authority not only over the production of curriculum-as-a-document but also over the curriculum-as-taught. According to McAdie (2004), schools have become hyper-conformist enterprises wherein those students who can conform to disciplinary effects of power exerted by the expectations of standardized testing are successful. Hansen (2007) urges students and parents to accept the reality of standardized assessment in the school environment, and she encourages students to develop strategies to excel in this environment so that their career aspirations are not limited by low tests scores. McAdie (2004) adds that testing discourages imaginative teaching and students’ opportunities to learn through exploration in response to the narrow curriculum designed to produce competitive students.

Patricia McAdie, a research officer with the Elementary Teachers’ Federation of Ontario: (2004)

Standardized tests do not improve student learning, but they do help to separate the winners from the losers. Bringing these tests into the elementary level ensures that young children are initiated into the world of markets and meritocracy. (p. 154)

Wideen et al (1998) claim that high stakes standardized testing impacts not only the instructional activities for students targeted for assessment but also students in lower grades. Teachers were more likely to narrow the scope of their teaching to examination topics and provide additional instruction on test-taking strategies, and students were more likely to employ more memorization strategies as opposed to critical thinking (Wideen et al, 1997). Although Wideen et al were unable to describe the nature of the classroom changes due to standardized testing, they did state that there is “a sense that measurement specialists exert control over instruction through tests” (p. 431).
Marvin F. Wideen, Thomas O’Shea, Ivy Pye, professors, Simon Fraser University, and George Ivany, president, University of Saskatchewan: (1997)

Teachers concerns about narrowing the curriculum covered several aspects of teaching. Those reporting that the curriculum had narrowed said they no longer taught some topics in as much depth; others noted that some of the more difficult, but important, areas had been eliminated in order to accommodate the examination. There was a generally agreed perception that much of what is taking place in the classroom amounted to teaching how to write the tests rather than teaching science. (p. 437)

According to Meaghan (1998), standardized testing exerts pressure on classrooms to reflect the world view of dominant culture as that content informs the formation of the assessment tool. Bouvier and Karlenzig (2006) have labelled the new accountability agenda within public education in Saskatchewan as reductionist and reflective of the world view of the dominant culture. Key elements of Aboriginal epistemology – relationships, patterns, processes and the interconnectivity of all living things – are silenced (Bouvier & Karlenzig, 2006). The importance of self as a “learning being” and “respect for the unexplained” are no longer considered as legitimate forms of knowing because these ways of coming to know in the world do not serve the globalization agenda (Bouvier & Karlenzig, 2006, p. 2).

Rita Bouvier, administrator, Saskatchewan Teachers’ Federation and Bruce Karlenzig, student at the University of Regina and former director of the Saskatchewan Teachers Federation: (2006)

Layered into this provincial context for Aboriginal education has been an approach to accountability that for the most part has carefully avoided the relentless zeal for standardized testing of students, ranking and rating of schools, and teacher testing that has plagued systems elsewhere in Canada or in states south of the border...Recently, however, government officials have signalled their disappointment with Saskatchewan students’ results on the tests associated with the School Achievement Indicators Program (SAIP) and the Programme for International Student Achievement (PISA). The government has announced a new set of “Three ‘Rs ’” (responsiveness, relevance, and results) are designed to make the education system more accountable. (p.4)
The accountability response is narrow and directed to specific academic aspects of curriculum-as-a-document and to the instructional strategies that facilitate its transmission. The Saskatchewan Ministry of Learning illustrates this limited world view in its refusal to acknowledge the effects of broad issues related to marginalization, poverty, and socio-economic status on the low test scores produced by Aboriginal students (Bouvier & Karlenzig, 2006). By taking this perspective, the potential for expanding our notion of what constitutes meaning-making and for cultural bridging is negated if alternative epistemologies are deemed illegitimate and irrelevant to the project of globalization (Bouvier and Karlenzig, 2006).

The Production of Useful Citizens and Docile Social Bodies

According to Graham and Neu (2004), a regime of standardized testing may serve a primary purpose of measuring student learning, but it also serves as a technique to encourage the development of co-operative and self-disciplining citizens. Essentially, standardized testing serves as a form of governance-at-a-distance through a “network of heterogeneous agents and heterogeneous sites” (Graham & Neu, 2004, p.298). Education is governed not only by the policies and regulations that governments initiate; but also through a complex web of administrators, teachers, counsellors, curriculum developers - all organized to achieve a common goal (Graham & Neu, 2004). Standardized testing provides a technology for organizing and regulating the overlapping interests of the heterogeneous sites and agents operating within education. Governments are able to use the outcomes of standardized testing for generative purposes, to produce statistics about the population which it governs (Graham & Neu, 2004). Because governments not only receive the data, they also generate statistics that can be used to serve the governance
goals of governments, justify the production of new policies, rewarding exemplary performance and sanctioning those groups whose performance is below the predetermined standard.

Cameron Graham, assistant professor York University and Dean Neu, professor University of Calgary: (2004)

Unlike the examinee, for whom the examination itself is crucial, the other participants are disciplined by the way testing results are used. In part, the participants are disciplined by the way testing results are used. In part, the participants adjust their behaviour because these tools of governmentality operate invisibly. Participants do not know, if, when, how or by whom the numbers will be used. (p.311)

According to Donlevy (2000), school administrators and teachers have responded to the potential use of standardized testing results as a regulatory apparatus by transforming their behaviour from trust and thoughtfulness to that characterized by anxiety and impulsiveness. The body is the site of inscription of discourses; therefore, the production of different somatic responses by teachers and students as a result of the transformation of the school discourse is consistent with Foucault’s theorizing (Foucault, 2006). School leaders have transformed from reflective educators to coaches in the competitive drive for supremacy in school rankings based on test results (Donlevy, 2000). The school administrators are forced to take instructional time away from teachers and students and transform it into assessment time; and for the sake of accountability, administrators list this time as an “in kind contribution” to the standardized testing project (Robertson, 1999). Teachers have been stripped of their ability to act as autonomous agents in their classrooms, and students have been regimented into performing routines and demonstrating docility (Pierce, 2006). The focus on standardized test results introduces new levels of predictability and control that restrict teachers’ and students’ opportunities for autonomy and creativity (Donlevy, 2006).
Predictability and control imposed by a rigid curriculum assessed by regimes of standardized testing magnifies the manifestation of the violence of the disciplinary effects of power on students and communities in Canada’s northern territory of Nunavut where the culture and language of the Inuit people contest the need for predictability (Berger, Epp & Møller, 2006). Inuit children are encouraged by their parents to develop independence and to make choices for themselves while the regimes of timetables, punctuality, and standardized assessments on foreign content instructed in a foreign language comprise an assault on students' self esteem. Students respond by refusing to comply with the regimes of discipline that the school system imposes on their bodies (Berger et al, 2006).

Paul Berger, Juanita Ross Epp, and Helle Møller, researchers at the University of Alberta: (2006)

An example of a specific practice that disadvantages Inuit students is the departmental exam created in Alberta, which must be passed by students for high school graduation. [Teachers] reported that the exams are culturally biased and unsuitable for ESL students. This type of assessment is incongruous with traditional Inuit assessment. Here, as in other jurisdictions, minorities are disadvantaged by standardized tests that do not take into account their context, learning style or ESL status...When access to goods and status is limited by credentials, structuring access to these credentials through high stakes tests that disadvantage Inuit students is indefensible. (p.11).

Although Aboriginal and Inuit students are often excluded from the sample of students selected to participate in the administration of PISA and other academic achievement surveys, the omission of Aboriginal and Inuit students in the sampling and measurement process introduces another form of silencing (Frenette, 2007). Their voices are not heard (Frenette, 2007).

Students and parents formulate opinions about standardized testing based on the rewards or sanctions that standardized testing have bestowed on them as parents or their
children. For students in Ontario, the implementation of a standardized testing program has reduced the budget allocation for resources and faculty while increasing the resources to support the standardized testing program (Goldenthal-Walters, 2004). Some of the students affected by the Ontario educational reform have chosen to organize and initiate strategies to inform the public about their perceptions of education reform (Goldenthal-Walters, 2004).

Hesch (2000) describes the ominous implications of mass testing projects in Winnipeg and the reproduction of children and parents as “failures.” According to Meaghan (1998) and Kohn (2004), socioeconomic indicators such as income level are just as reliable as standardized test scores in predicting school rankings and student performance. Arguments have been made that income level and neighbourhood wealth predict test scores, and conversely, intelligence scores of parents predict family income level and neighbourhood wealth (Herrnstein & Murray, 1994). Ultimately, school performance and income levels are related, and to some degree, predictive of each other. Hesch (2000) and Kohn (2004) argue that we do not need additional testing to continually re-establish the existence of this relationship; however, as a society, we need to accept that the poor have always been part of our social fabric, and we need to address the issues of poverty if we want real benefits from the activities of our educational system. Kohn (2004) states the crises is here: “The intellectual life is being squeezed out of schools as they are being transformed into giant test centres” (p. 1). However, Kohn does not encourage anarchy on the part of teachers; he encourages teachers to prepare students for the tests using inventive and innovative strategies and then return to real learning.

Shawn Kettner, a Winnipeg parent (in Hesch, 2000):
I do feel that the wonderful multi-aged, child-centred, activity–based program where my children are learning to be critical thinkers is in jeopardy...because these tests are part (of) a political strategy to change the way we teach our children. They are designed to weed out the winners and losers.

We know that in an inner city classroom...the tests seem nothing more than a way to say “We told you so, you cannot succeed.”...Our students are doing well. In fact, too well. We have too many children being prepared for university. We need to divide those that will get in from those that will not. We can’t all move up the ladder. So we bring in things like standards testing. (p.49)

Summary

The standardized scales developed by Binet and imported into the United States by a group of psychologists have become more than a strategy to gather information about approaches to assist children who are struggling with academic tasks. Through the initiatives of Henry Chauncey and the ETS, the influence of standardized testing initiatives is felt in almost every school in North America and in many international schools. The results of standardized testing have become a complicit partner is the economic strategies promoted and designed by Milton Friedman and his protégés to facilitate the privatization of education. Although cautionary voices of Donald Campbell, Heather-Jane Robertson, Selma Wasserman, other commentators, and concerned citizens have voiced their concern about the potential harm that these tools may inflict on our children and our society, policy makers remain immune to the din. The fascination with the potential of standardized testing increases as policy-makers distance themselves from students and the activities of the classroom (Kohn, 2004). Standardized testing serves the need of these policy-makers to govern-at-a-distance through relational acts of power that support self-regulation and self-discipline.

Alfie Kohn, a writer on human behaviour and education:

My rule of thumb is: the closer you get to real classrooms, the more people understand what a menace it is to talk about standardizing education and testing
kids constantly. ... The farther you get from real kids, the more likely you are to think that standardized testing is a fine idea. (p. 3)

According to Robertson (1999), education is not a neutral activity. Its intention is the production of useful citizens for our nation and the global economy; however, what constitutes useful citizens is often in contention, as is the means for their production. Thus, education is a political activity with an international agenda fuelled at times with issues generated by transnational corporations; its curriculum represents the official knowledge that captures the goals and aspirations for Canadian youth; and the outcomes of education determine whether the youth of Canada will be able to take up key roles on the world stage as “active citizens, flexible workers or as needy consumers” (Robertson, 1999, p.1).

Bouvier and Karlenzig (2006), Hesch (2000), and Kohn (2004) point to some imminent crises in education. This current regime of standardized, school accountability and hyper-rationalization of the education system has failed to produce the mammoth gains that its educational policy engineers have promised. Bouvier and Karlenzig (2006) and Hesch (2000) hold out a hope that Canadians as a society will come to recognize the value of our cultural and social capital and that educational improvement will come through the provision of better lives for students and teachers.

Educational accountability as it is currently practiced may be facing a crisis of legitimacy. The indigenous knowledge, values, and norms that are inherent to Aboriginal education could inform the development of meaningful alternatives that are, it is hoped, much more perceptive about human qualities of education and pluralism (Bouvier and Karlenzig, 2006, p.7)
CHAPTER 6
A GENEALOGY OF CANADA’S NATIONAL REGIME OF LARGE-SCALE STANDARDIZED TESTING

[T]hat is, what is our own actuality, what is happening around us, what is our present. It seems to me that philosophy acquired a new dimension here. Moreover, it opened up a certain task that philosophy had ignored or didn’t know existed, and that is to tell us who we are, what our present is, and what that is, today. (Foucault, 2007, p. 129-130)

My reflection on the emergence of Canada’s national testing regime and the widespread voluntary participation by all provinces and territories, leads me to ask these questions based on Foucault’s ponderings about how the present came to be: How is the actuality of our Canadian identity produced in the field of Canada’s national standardized testing project, SAIP/PCAP? How did the social, cultural, and political forces within Canadian society operate to produce the experience of the national standardized testing project? How have the forces and regimes of knowledge produced by the national standardized testing project shaped and directed the Canadian educational experience, and more broadly, the Canadian cultural, economic, and political agendas? How do the analytics of archaeology and the tactics of genealogy serve as research approaches in the construction of meaningful descriptions of the social, political, and institutional fields that comprise the large thematic topic of national standardized testing?

In Canada standardized large-scale academic assessments operate at three different jurisdictional levels: international (PISA, PIRLS and TIMSS), national (SAIP/PCAP), and provincial - standards tests in British Columbia, Alberta, Manitoba, and Ontario (Volante and Ben Jaafar, 2008). These testing systems are a relatively new mechanism designed to monitor student performance and provide stakeholder accountability about the performance of Canadian educational systems “using standards
that are congruent with those established and accepted by ‘others’”; this means that the
measures have some degree of congruence with each other (Volante and Ben Jafaar,
2008, p. 206). The Canadian educational systems operated and produced graduates prior
to the introduction of these academic assessments systems; however, the place of these
assessment systems in the Canadian educational landscape has become part of the taken-
for-granted features of the current educational systems. Of these three testing systems,
Volante and Ben Jaafar (2008) describe the participation of the Canadian provinces and
territories in SAIP/PCAP as “voluntary, as is their [provincial and territorial ministers of
education] engagement in CMEC” (p. 206). This means that the provincial ministers of
education must not only agree with unanimity but must also convince the members of
their respective legislatures that this is a “good thing to do.”

The Council of Ministers of Education, Canada, (2000) describes the
jurisdictional responsibility for the provision of educational services including the
administration of large-scale standardized assessment protocols in Canada as follows:

Canada is comprised of ten provinces and two territories, each of which, within
the federative system of shared powers, is responsible for education. The
Constitution Act, 1867, provides in part that "in and for each province, the
Legislature may exclusively make laws in relation to education." Therefore, each
of the provinces and territories has developed its own educational structures and
institutions; while these are similar in many ways, they reflect the circumstances
of regions separated by great distances and the diversity of the country’s historical
and cultural heritage. (p. 5)

This type of statement, which describes the limits placed on CMEC’s authority by the
constitution to speak as a national voice for education in Canada, is a routine message
included in most documents produced by CMEC about the status of primary and
secondary education in Canada. CMEC describes its role as providing a “forum for
education ministers to discuss matters of common concern, share information, and
represent Canadian education internationally” (p. 5). However, this discussion forum with financial support from HRSDC was able to direct the development and implementation of SAIP/PCAP, Canada’ national academic testing program and organize support for educational research on topics of national interest (Council of Ministers of Education, Canada, 2000). According to the Council of Ministers of Education, Canada (2000), projects such as SAIP

support the strengthening of accountability and building of Canadian expertise in the development of education outcome indicators. ...HRDC supports these projects because of their value in helping to ensure that Canadians have the necessary knowledge and skills required to compete in the global economy and their contribution to Canada’s knowledge base on human capital. (p. 11)

Although the Council of Ministers of Education, Canada (1996b), does not have jurisdictional authority or legislative power, it was able through a form of consensus building and the exercise of relational power among the ministers of education to set Canada on the path towards increased accountability through the implementation of a national assessment program, SAIP.

Aspects of Canadian national testing program differ from the American program in a number of significant ways. The Canadian testing program is pan-Canadian – the instrument, the sampling strategy, and the public reporting protocol were developed through the auspices of CMEC rather than by state educational agencies. CMEC relies on the relationships that the ministers of education develop through their collaborative deliberations to get things done. In the United States, the NCLB Act mandates state testing, reporting, and the expected annual yearly improvement but does prescribe the testing instrument. The American testing program relies on the rule of law to ensure that the states conduct the testing, complete the reporting protocol, and enforce consequences
on schools that fail meet the standards (Abedi, 2007). Robertson (2004a) summarizes the process of decision making employed by the CMEC as the use of consensus building to shift the constitutional powers through the activation of informal channels of influence and non-binding policy making.

For me, Robertson’s (2004a) comments trigger my memories of a conference I attended in 2006 at which Sheila Copps was the keynote speaker. Her topic was educational reform. Much of the question period was devoted to questions about the “how” of national educational reform in a system in which jurisdictional authority lies with the provinces. Her responses to the many the questions contained encouragement for us to consider the possibilities of constitutional reform mitigated through the process of consensus building – real constitutional reform is a difficult, long and arduous process. As I consider the ease with which SAIP appeared on the educational landscape – no community hall meetings or media-managed national consultations or the need to gather widespread public support or the need to commission a national report on the status of education in Canada. After consensus was reached to implement SAIP, each minister of education directed the responsible to his/her department to accommodate the new initiative and to work with their school administrators to ensure its administration. I appreciate the efficacy of making decisions in this way, but I wonder what was missed, whose voice was not heard or which child or teacher or parent was inadvertently harmed by the process? (Dagenais, April, 15, 2010)

The production of a critical and effective history of SAIP/PCAP, Canada’s national assessment program designed to measure the effectiveness of its schools, involves archaeological analysis and genealogical work as described by Foucault (1980) and elaborated by Dean (1994). This analysis is designed to reveal the ruptures that give way to the beginnings of discourses about aspects, features, and structures of the domain of standardized testing in Canada. Genealogical work involves collecting data; that is, the bringing together of documents – both scholarly work and other forms of communication related to the domain - organizing them, and following the evolving themes. Thus, genealogy, as research activity, is the tactical approach used to follow particular discursive practices, capture their twists or mutations in composition and structure, note
their extinction or disappearance from discursive practice, and follow the treacherous path of the truth - its constraints, obligations, and rituals (Foucault, 1980; Dean, 1994). The product of this work is a critical and effective history of Canada’s standardized testing project, SAIP/PCAP, that identifies the forces and lines of tension that connect the quasi-scientific field of statistical analysis, economic theorizing, social dimensions of effective schooling, learning assessment theory, and the principles of effective teaching. Complex fields of experience, like that occupied by SAIP/PCAP, operate at the margins of what society deems to be acceptable fields of scientific inquiry. Standardized testing uses the statistical and analytical tools of empirical studies mitigated by the complexities of human existence within complex and varied societal structures. According to Foucault (2007), these types of fields are most appropriate for archaeological analysis and genealogical work because they provide opportunities to contest our “sense of reason” (p. 133).

From my perspective, the production of a critical and effective history of a field like standardized testing seems to be a daunting task; however, Foucault (1980) emphasizes the importance of the article, a. For me, this means that the history that I produce is just one of many possible versions that could be constructed from the discursive evidence: documents, journal articles, newspaper clippings and other artefacts left by the social constructions of society. Another researcher may select other documents, link them in different patterns, follow different links to produce a somewhat different critical and effective history, uncover different power relationships, but it would be, nonetheless, a critical and effective history of the field of Canadian national standardized testing. (Dagenais, January 20, 2010)

Genealogy poses problems about present issues and concerns that arise from the current organization and configuration of knowledge and its interface with politics (Foucault, 1980; Dean, 1994). The objective of such histories is not the production of the truth but the grasping of the conditions that permit one to say something is true (Dean, 1994. Foucault (2007) urges us to direct our genealogical focus to problems of present
that exist in areas considered to be fragile and uncontested. SAIP/PCAP operates in the
fragile territory of extra-constitutional consensus of the provincial ministers of education
and the in-kind contributions of schools, teachers, and students that donate time and data
to its projects (Robertson, 2004a). Although the merits of the outputs of SAIP have been
the subject of some limited form of meta-analysis, its existence remains a part of the
taken-for-granted educational accountability landscape.

Thus, questions may emerge from the analysis of relations of power that concern
the social, cultural, and political circumstances that gave rise to the establishment of the
CMEC; and the nature of the production of concerns, issues, and events that opened the
space for the development and implementation of SAIP, its transformation into PCAP
with administrative complementarities to PISA. SAIP/PCAP is more than a tool for
assessing the efficacy of Canadian schools and reporting their performance by
jurisdiction; it is, in every sense, a Foucauldian apparatus of power that can provide some
form of surveillance, control, and containment of the educational systems in Canada
through voluntary and reciprocal relations of power.

The practical challenge facing the researcher is where to begin archaeological and
genealogical work as most of human cultural activities exist as a messy, layered
formation. The digging always leads to the layer placed before it and sometimes the
jagged edges of one layer push into the next layer. In researching the CMEC and
SAIP/PCAP, it is critical to recognize that the justification, organization, structure, and
restraints on the provision of educational services for Canadian children reside in the
provision of the Constitution Acts 1867 to 1982 and outside the purview of CMEC.
Forces and Lines of Tension

[I]f we would make Canada safe and secure, rich and renowned, we must liberalize – locally, sectionally, religiously, nationally. There is room enough in this country for one great free people; but there is not room enough, under the same flag and the same laws for two or three, suspicious, obstructive “nationalities.” (McGee, 1862, reprinted in Morton and Weinfled (Eds.), 1998, p.33)

McGee speaks to the question of Canadian unity and the issues of living, learning, and working together in one country and under one flag at a time of Canadian confederation. One might ask: Of what importance are issues related to confederation and Canadian unity to understanding how SAIP/PCAP came into existence, and how these issues operate in today’s reality. In Canada, unlike most other developed countries, the provision of education is not a national responsibility but a provincial or territorial responsibility (Council of Ministers of Education, 2000). The importance of a national voice for education became evident in the ‘60s as organizations like IEA were developing international studies to identify factors that contributed to effective and efficient learning (Organization for Economic Cooperation and Development, n.d.). As OECD member countries began to transform their economies, educational researchers became interested in organizing cross-national studies in the fields of school efficacy, the role of the school, the domain of education, and its role in the development of human capital and poverty reduction (Organization for Economic Cooperation and Development, n.d.). The emerging policy gaps filled by CMEC emphasized the growing importance of education in the production of resources for the national and international cultural, economic, and political markets and the need to participate as a country in large cross-national educational survey.
In response to the constitutional division of powers, the provincial governments developed their own school systems that reflected the local, social, cultural values, and beliefs of people residing in each province and/or school district (Morton, 1994). At the time of confederation, the protection of religious and linguistic interests in the delivery and organization of education superseded the economic interests of the nation (Robertson, 2007). In Canada, according to Watts (1986), the concentration of a large French-speaking minority in one province has resulted in regionally-authored political clout over education and other domains. Watts states that the form of federalism practiced in Canada was “made necessary by regional diversity, appears to have reinforced diversity by institutionalizing it” (p. 775). As a result, students’ educational experiences, as viewed from a pan-Canadian perspective, may have many elements in common, but there are also significant regional differences attributable to provincial legislation, local values and beliefs, and religious influences (Goddard, 2000). As Canadian society evolved into a complex social democracy, the provincial constitutional responsibility for education has become extremely important to Canadians’ daily lives because this allocation of responsibility ensured that education was relevant to regional needs (Watts, 1986).

The 1960s was a period of tension between the nationalist intellectuals such as Northrope Frye who wrote “extensively about modernity and its effects on Canada” and the nationalist vision authored by the Diefenbaker government which failed to recognize the need for a national economic strategy and the importance of the Quebec factor (Potter, 2000, p. xix). This turmoil created a crack for the liberal continentalist agenda of the Pearson government which Grant (2000) describes as fostering both a branch-plant
economy and a branch-plant culture (p. 32). Layered upon the turbulence in the Canadian economic and political sector, technological innovation was causing a growing upheaval in the way people in developed societies did things (Grant, 2000). This upheaval in conjunction with the growing continental sentiments led Grant to speculate that nations, such as Canada, will not be necessary as technology drives us towards a “universal and homogeneous” state; local cultures will simply become redundant.

In the 1960s, the changes to the Canadian economy drew attention to the provincial administration of education and how those local systems affected the national agenda and economic prosperity. The lines of tension increased as Canadians from other ethnic backgrounds and the First Nations people produced their perspectives on what it means to be a Canadian in their space and in their relationships with the land and with each other (A Parliamentary Report of the Royal Commission on Bilingualism and Biculturalism, 1998; George, 1998). The issues and concerns of Aboriginal people for the provision of education formed a quiet background conversation; thus, the wisdom and perspective on humanness offered by their leaders was often ignored by policy makers representing the dominant culture (George, 1998; Morton, 1994). The words of Chief Dan George (1998) offer a different but valuable approach to learning and to possible landscapes in which learning what it is to “be” and to “be in relation to the other” can happen. George (1998) also reminds us that learning and education are transactional activities; we must all give, and we must all receive to reap the gifts of our Canadian federalist experiment.

The administration of a national standardized testing project emphasizes the goals of centralism and unity and holds out the possibility of building a common understanding
in a large and diverse land where people are shaped by their environments, their histories and their relationships with each other (Chambers, 1998). Such a project may contribute to building of “one great free people” or it may fuel dissention and the production of difference” (McGee, 1862, reprinted in Morton and Weinfled (Eds.) 1998, p.33).

The Beginnings: 1964 – 1970

By forcing a centralism perhaps acceptable to some provinces but not to Quebec, and by insisting that Quebec must be like the others, we could destroy Canada. (Pearson, 1998, p.201)

Pearson’s statement highlights the complicated nature and the serious consequences related to the political desire to move toward national solutions not only for the design and delivery of education services, but also for the delivery of centralized responses in all sectors. After a brief recession in the mid ‘50s, Canada returned to a position of economic prosperity in the early ‘60s, and the post-confederation tension between the provincial governments and the centralist government in Ottawa resumed after a brief post-World War II hiatus (Morton, 1994). The Canadian economic policy in the 1960s was characterized by a form of nationalism that included protectionist policies to shore up the somewhat inefficient Canadian industries, protect employment, control foreign investment, and limit American involvement in Canadian industries (Finkel et al, 1993). However, some business leaders were showing a cautious interest in resurrecting free trade with the United States as a strategy to stimulate the Canadian economy (Finkel et al, 1993).

Education: The Forces and Lines of Tension
In 1957, the former Soviet Union launched the first rocket into space; this event stimulated concern in the United States and Canada about primacy in space exploration and efficacy of science education (Robertson, 2007). Bracey (2007), an author and noted commentator on issues related to standardized testing in the United States, offered the following comment about the Soviet launch of the Sputnik rocket: “Once the world realized that Sputnik was not a swindle, people had to explain how a technologically backward nation such as the Soviet Union could have accomplished such a feat” (p. 119). According to Bracey, the American public readily accepted the theory that “Russians beat us because they had better schools” (p. 120). The launch of Sputnik was important to the field of education, but not because of its technological importance. According to Bracey, “The U.S. could have beaten the Russians by over a year. Dwight David Eisenhower chose not to” (p. 120). The successful Sputnik launch triggered wide-sweeping changes to how the general public and policy makers valued the American public school curriculum content and American students’ proficiency in learning it. The change in American support their public school curriculum launched a series of broad curriculum changes. The revised curriculum that consisted of mathematics and science education for all students replaced the vocational-education and skills-training curriculum for students not interested in a college education.

The impact of changes to the way Americans thought about mathematics, science, and technology and their importance to the development of the American economy reached American students and influenced their perception about education, the possibilities that education may open up, and what constitutes value in the delivery of educational services. Henry Petroski, a science writer who grew up during the Sputnik
era, made these comments in a 2010 media interview about the encouragement he received from his school to study engineering because he had an aptitude for numbers and good standardized test scores: “We knew ... that as engineers, we were going to be in a position to change the world – not just study it” (Salter Reynolds, 2010). The paradigm for what counts for value had shifted away from general education to science and mathematics.

Coincidentally, in 1958 a group of international scholars, sociologists, psychologists, and psychometricians gathered to discuss and establish goals for the establishment of cross-national comparative educational research (IEA, n.d.). This group argued that effective evaluation of an educational system should not only consider the resource inputs, the conditions under which educational services are delivered but also the outcomes – the students’ knowledge and skills (IEA, n.d.). The key for these researchers was empirical school and student performance data that could be analyzed and used to isolate those variables that consistently and meaningfully impact the quality of educational outcomes (IEA, n.d.). The question remains: Did Sputnik influence the decision for these scholars to meet for purposes of designing cross-national educational effectiveness research? Or did Sputnik conveniently serve as a publicly-acceptable rationale for widespread change to the educational systems in North America as the conversations about the role of education were in play prior to Sputnik? From a Foucauldian perspective, the rules that govern what counts for value and how that value is determined changed in 1957. It was no longer acceptable to offer “data-free assertions about the relative merits of various educational systems”; data was now required (IEA, n.d., p. 1).
The Canadian federal government responded to the Sputnik event in two ways. First, it signed an agreement with NASA in 1959 to invest in one of the most progressive space programs of that era, the design and development of Alouette I communications satellite launched in 1962 (The Online Journal of Space Communication, 2003). And second, the federal government enhanced its policies for science education for Canadian children and introduced the “use of economic incentives to target and promote” these science education initiatives to the provinces (Goddard, 2000, p. 2). In 1967 the ministers of education for the provinces and territories responded to the federal government’s move into the provincial jurisdictional territory of education by organizing themselves into a small secretariat, CMEC (Goddard, 2000). Goddard describes the formation of CMEC as a “response to fears that the national government was attempting to increase its influence over education” (Goddard, 2000). Manzer (1994) describes the formation of CMEC as a proactive strategy based on the recognition that “interprovincial convergence in public education made national policy coordination conceivable, but also a growing provincial interdependence made national policy coordination desirable” (p. 196).

According to Bergen’s (n.d.) entry in the Canadian Encyclopaedia, the organizational model for CMEC was derived from the West German Kultursministerkonferenz, a standing conference of state ministers who meet to discuss issues relevant to the delivery of education and to gather consensus for federal initiatives. CMEC (2010) describes its role as one of providing “leadership in education at the pan-Canadian and international levels and contributes to the fulfilment of the constitutional responsibility for education conferred on provinces and territories” (p. 1). Robertson (2007) describes the role of the CMEC as a “quasi-national home for educational
discussions that were national in scope” (p. 94). According to Goddard (2000), “CMEC has had little direct impact on the educational landscape in Canada [because it] has no legislated authority and can only take action when there is unanimity among its members’ (p. 2). While it is true that CMEC lacks legal authority to make anyone take action, it has through a process of consensus building and the exercise of relational power facilitated the research and publication of several documents; set out its truth telling statements in documents such as Learn 2020, a pan-Canadian vision for education; and facilitated the development, implementation and revision of a pan-Canadian assessment program.

The Middle Years: 1971 - 1980

So the remarkable originality of Globalization lay elsewhere: in its assertion that all civilizations from now on were going to be led by commerce. This premise came with a purity or simplicity that went far beyond the more nuanced economic thinkers like Montesquieu, Adam Smith or even Karl Marx. The other constituent parts of human activity – from politics to social policy to culture – were going to be perceived through the prism of economics, which, once released from most government interference, would find its own natural balances. (Saul, 2005, p. 17-18)

In the 1970s Canada faced the disastrous combination of inflation and stagnation referred to as stagflation (Morton, 1994). Energy prices soared; the American appetite for Canadian goods declined, interest rates climbed, and the unemployment rate skyrocketed (Finkel et al, 1993). The advocates for the economic ideologies of globalization were gaining acceptability as Canada struggled to respond to the challenges of a stagnant economy and rising unemployment rates (Finkel et al, 1993; Saul, 2005). The Canadian government responded initially with “price and wage” policies targeted at the “demand” side of the economic equation; however, economists arguing for a monetarist approach urged the liberation of the economy from the unnatural controls of governmental policy
and regulation so that it could find its natural balance (Saul, 2005). Economists with a monetarist focus were gaining influence in the United States by arguing that free markets were symptomatic of a free and democratic state; however, free markets and democracy make poor partners (Saul, 2005). Their arguments included reductions in social spending – health and education – and eventual transitioning of these social services to private sector which would promote both efficiency and efficacy (Klein, 2007). Thus, Canadian social spending declined as politicians responded to the urgings of economists to allow the forces of globalization to determine the natural economic balance for the trade of services and goods (Klein, 2007).

While economists were engaged in re-envisioning education as a service regulated by the market, the OECD with the cooperation of CMEC conducted a review of the Canadian national education policies in 1975. The 1976 report published by OECD described the Canadian educational system, based on the vision of the Fathers of Confederation, as an economic risk to future generations of Canadians (Robertson, 2007). In addition the OECD urged Canadians to move toward the harmonization of its educational systems and the creation of a national curriculum (Robertson, 2007). Bergen (2010) highlights CMEC’s cooperation with the OECD to the complete the review of Canada’s national policies for education in 1975 as both critical and important. According to Manzer (1994), the OECD regional reports submitted by CMEC on behalf of the provinces were illustrative of “both the residual historical diversity of educational institutions and practices and an unprecedented ideological convergence of educational purposes and policies” (p. 202). However, the OECD determined that CMEC could not provide the national transparency and accountability, key features of a modern and
progressive educational system (Robertson, 2007). The OECD report fuelled the debate about the necessity for a national educational policy, and questioned the role of interprovincial co-operation facilitated through the CMEC (Robertson, 2007).

These events are symptomatic of the growing international forces that comment, criticize, and evaluate the Canadian educational system. The Canadian educational system was described as a liability by the OECD, an organization whose contributing scholars have spent more than a decade analyzing school systems from member countries to determine “what works” (Robertson, 2007). The OECD report contributed to the growing momentum for a national education policy; CMEC and its role in facilitating interprovincial cooperation did not satisfy the OECD’s growing belief that nations needed to develop the capacity to lead, assess, and evaluate the effectiveness of educational policy development in an economic climate in which human capital was critical (Robertson, 2007). Almost every other national federation was able to develop a national vehicle for educational research and policy building; yet Canada stood alone with its provincial model for the delivery of education (Robertson, 2007; Goddard, 2000).

As international criticism was mounting, national criticism also began to grow in the 1970s. Parents began to criticize their public schools as they perceived that the schools were not preparing their children for an increasingly uncertain future. Parents raised concern about the apparent lack of discipline, the seemingly chaotic classrooms, and the inability of public schools to instil the taken-for-granted basic skills – reading, writing, and mathematics (Barlow & Robertson, 1994; Morgan & Robinson, 1976). Parents perceived that the new child-centred experimental curricula and instructional strategies were resulting in lost opportunities for their children to learn the basics (Barlow
& Robertson, 1994; Morgan & Robinson, 1976). At the same time, postsecondary educational institutions in Alberta raised anecdotal concerns about students’ inability to read, although they had received high marks from their local public school (Graham & Neu, 2004). The Alberta legislature convened a committee to study the elements of student achievement; however, through the activities of this committee, the elements of student achievement transformed themselves into an official “problem ... of student achievement” that required a solution (Graham & Neu, 2004, p. 303). A made-in-Alberta standards testing program formed part of the solution to address the public perception of underachievement by Alberta’s public school graduates (Graham & Neu, 2004). The comfort with standards testing and standardized testing as a means to measure public school performance and to make the outcomes of the educational visible was growing (Graham & Neu, 2004). Parents began to formalize their perceptions about their children’s opportunities for the future and their school performance based on the results of standardized testing (Graham & Neu, 2004).

The First Steps toward a New Agenda: 1981-1990

The notion of ideology appears to me to be difficult to make use of...it always stands in virtual opposition to something else which is supposed to count as the truth. (Foucault, 1980, 118)

In the ‘80s, Canadians began to grapple with the issues related to the value and the cost of education: What type of education should my children receive? How much should the delivery of educational services cost? Are we, as taxpayers, paying too much and receiving too little? What are the alternatives, and can they provide anything better for less money? How do we, as taxpayers, know if our children are learning what they
should learn for the current and future economies? These questions emerged from the re-thinking of education as a service provided by the state to support the development of an informed and useful citizenry and reimagining education as service that supports the development of vibrant economy.

The educational agenda moved to the forefront in the minds of Canadians as the social shock of rising unemployment rates set in: Existing workers lost their jobs, and high school graduates were unable to find work (Klein, 2007; Robertson, 2007). Many citizens expressed concerns about the ability of Canada’s schools to educate its youth for the economy of tomorrow increased (Klein, 2007; Robertson, 2007). Wilkinson, a Canadian economics professor (1986), described the Canadian economic position in the 1980s as the “the most severe recession since the 1930s”; however, the costs for education continued to rise despite a declining number of students enrolled in Canadian schools (p. 536). During the 1980s, the provincial share of the education budget declined from 66.5 percent of the overall cost in 1980-81 to 58.5 percent of the cost in 1992-93 (Barlow & Robertson, 1994, p. 12). The decline in provincial funding was picked up in some instances by municipal governments that provided a greater “visibility [through that] of local taxes” (Barlow & Robertson, 1984, p. 12). The increased visibility fuelled public concerns about the cost and value of education delivered to Canadian youth.

In response to the growing concern about the state of Canada’s education systems, in 1981 the federal government assumed a more active role in education by providing funds to CMEC to establish a national testing program (Robertson, 2007). According to Barlow and Robertson (1994), this response grew out of concerns raised by organizations, such as Business Council on National Issues (BCNI), the Conference
Board of Canada, the Canadian Manufacturers’ Association and the Canadian Chambers of Commerce, whose responsibility was to speak for the needs of industry and business in Canada. Organizations like BCNI demanded the administration of school exit testing either by the schools or by industry as means to assess the degree to which public schools were preparing students for work, careers and the global marketplace (Barlow & Robertson, 1994).

The recession impacted the American economy as well, and the perceptions of low efficacy and productivity of American schools came to public attention with the release of the report, A Nation at Risk, in 1983 (National Commission on Excellence on Education, 1983). Its release signalled alarm about the lack of productivity of the American economy and the poor performance of its educational system (National Commission on Excellence on Education, 1983). The report linked the decline in the quality of American educational system and the mediocre performance of its graduates to the loss in America’s “unchallenged pre-eminence in commerce, industry, science and technological innovation” (National Commission on Excellence on Education, 1983, p. 1). Education was linked to the production of skilled and innovative citizens, a necessary resource for inevitable movement toward globalization, the fostering of unfettered markets, and the movement of intelligence around the globe (National Commission on Excellence on Education, 1983). The report states that “Knowledge, learning, information, and skilled intelligence are the new raw materials of international commerce”; thus, the national government had an interest in establishing standards for the delivery of education and the development of this resource potential to preserve the

The National Commission on Excellence in Education used America’s placement on international and national standardized tests as evidence to argue that a mediocre educational system with crumbling standards was at the heart of America’s economic decline (National Commission on Excellence on Education, 1983). The government refused to acknowledge that the national policies of Japan and Germany were providing safety and security to their internal markets and that these policies served as a viable vehicle for workforce stabilization (Finkel et al, 1993; Saul, 2005). Because Canada’s economic and cultural markets are linked to those in the United States, Canadian citizens and policy makers were also looking for the source of Canada’s economic woes; education became the accepted source of the economic hardship and the target for reform (Barlow & Robertson, 1994).

In response to the perceived weakness in the Canadian educational system, the Science Council of Canada released its influential report, Science for Every Student: Educating Canadians for Tomorrow’s World in 1984, which advocated science education for all Canadian students regardless of ability, interest, or place of residence (Council of Ministers of Education, Canada, 1996b). In addition, the report advocated the development of every citizen’s ability to participate in the political and social choices facing a technological society, additional training for those students with a special interest in science, appropriate preparation of individuals for the modern world, and programs designed to stimulate intellectual and moral growth for the development of rational and autonomous individuals (Council of Ministers of Education, Canada, 1996b). Although
the report issued by the Science Council of Canada was addressing similar challenges related to the preparation of youth and their ability to participate fully in a technological and rapidly-changing globalized economy as the American report, *A Nation at Risk*, the approach was quite different. The Canadian report addressed the need for the development of an informed citizenry capable of making rational decisions related to the adoption of new technology while the American report emphasized the perpetuation of a competitive society and the development of a competitive market advantage over morality and ethics (Council of Ministers of Education, Canada, 1996b). In *A Nation at Risk*, the sense of alarm came from the correlation between poor market performance and low test scores of American students on standardized tests. The Canadian report extended an invitation and provided encouragement to Canadians to value their engagement in a science education as a *common good* (Council of Ministers of Education, Canada, 1996b). The American report employed alarmist language that associated the economic crises of 1983 to the perceived faltering of its educational system, which the report described as plagued by low standards and a plethora of unnecessary courses (National Commission on Excellence on Education, 1983).

Following the release of *A Nation at Risk*, Wilkinson (1986) released his review of Canadian public education entitled *Elementary and Secondary Education Policy in Canada: A Survey*. Wilkinson’s survey provided a business case for the delivery of public education. While this document does not have the name recognition of its American counterpart, *A Nation at Risk*, it did describe the challenges and risks facing the educational systems in Canada and the role of standardized assessment in the provision of stakeholder accountability.
Wilkinson (1986) described the situation for delivery of educational services in the 1980s as follows: The provincial governments were facing rising educational costs due to increases to teachers’ salaries and the need to manage a system with excess capacity (beyond inflationary costs, decreased class sizes, and declining enrolments). Provincial and territorial ministers of education needed to control school budgetary expenditures; while at the same time, they felt the pressure to develop quality assurance measures and the need to identify what influenced the quality of educational outputs (Wilkinson, 1986). Although teachers in the 1980s had more formal qualifications than teachers of previous decades, the public satisfaction with the Canadian school system seemed to be falling (Wilkinson, 1986). Therefore, teacher qualifications alone did not predict the perception of a quality learning experience. From Wilkinson’s meta-analysis, the following factors were more consistently related to a quality educational experience as measured by test scores and increased grade point-averages: the ability of the teacher to interact with students both individually and as a group, the congruence between teaching methods and instructional content, and parental engagement. The difficulty with these factors from both a research and a remedial perspective was that they were difficult to quantify and to determine the threshold at which students would benefit (Wilkinson, 1986).

Wilkinson summarized the seven pressures or forces impacting the Canadian school system of the 1980s as the following: “excess capacity; rising real costs; public and academic concerns about the quality of education; individual and social returns [derived from educating a student; ...] the cost efficiency of the industry; financial constraints; and the possibility of renewed competition from private schools” (p. 536).
This review of the Canadian educational system described the transformation of the provision of education from that of satisfying a social need or mitigating a political agenda into a case study of a failing industry (Wilkinson, 1986). Wilkinson described the Canadian educational systems as the “education industry” in which parents are the consumers purchasing education through their tax dollars (p. 536). Parents, from his perspective, were not well-informed about what constituted quality in the provision of educational services and were reliant on the suppliers, the teachers, to provide product information. Wilkinson’s position like that of the IEA in 1958, argued that the quality assurance measures should shift from a focus on inputs (funding, class size, and faculty credentials) to outputs (standardized test scores, school grades, school attendance and behaviour and dropout rates). Wilkinson summarizes his findings by stating:

[Q]uality in education, as measured by the achievements and capabilities of students, does not appear to bear any consistent relationship to the dollars spent on education by different schools or school districts, or to the various direct inputs that these dollars can buy, such as teachers with more education or lower student-teacher ratios. (p. 543)

Wilkinson formally introduced the private-public school debate into his report as a mechanism to address efficacy and efficiency. By increasing government funding for private schools, Wilkinson argued, the overall quality of education would improve because there would be “greater competition for the public school system”; thereby, increasing both the quality of the product/service and efficiency by which it is produced (p. 543). However, the public debate continued to focus on identifying mechanisms for increasing parental engagement, improving the quality of children’s education, and reducing the public cost (Wilkinson, 1986). Recognizing that the private-public school debate was complicated and potential benefits of school competition not well understood, Wilkinson proposed that these “separate or competing schools are of particular interest as
they represent an opportunity for researchers and policy-makers to come to grips with the
cost, quality and efficiency issues” (p. 554).

In the 1980s school boards and provincial education departments dedicated
resources and infrastructure to the collection and analyses of indicator data as a means of
determining excellent schools, identifying the variables that contribute to school
excellence, and providing school accountability to their key stakeholders. Willms and
Raudenbush (1989) referred the collection of school indicator data by the British
Columbia government as an “objective means to hold principals and teachers
accountable”; however, they questioned the utility of this cross-sectional approach to
school performance analysis in relation to the utility of a longitudinal research approach
focused on particular schools (p, 209). Although, “objective” implies unbiased, their
analyses of these cross-sectional studies revealed several sources of bias that could enter
the data collection phase and lead to the development of policies and practices designed
to address non-existent school problems (Willms and Raudenbush, 1989).

One source of bias, according to Willms and Raudenbush (1989), was the use of
aggregate data that has the potential to erase or minimize some of the school-specific
variables that change slowly over time while emphasizing the year-to-year fluctuations in
variables that tend to remain relatively stable over a long period of time. Essentially, by
using school indicator data for policy making and school intervention, policymakers may
have been addressing issues that may have been an artefact of the data analysis.
Alternatively, Willms and Raudenbush’s longitudinal approach examined the quantifiable
nature of seemingly stable relationships between school inputs and student outcomes.
Their research on school indicators did not reveal a stable relationship between indicators
such as truancy rates, the level of school funding, and teaching practices and student
success and the ability of the school to achieve its educational outcomes. However, these
longitudinal studies did reveal positive effects of a particular teacher’s expertise in a
subject area or the positive impact of particular school policies designed to promote an
effective learning climate or the strategic funding to address a school-specific issue
(Willms & Raudenbush, 1989). Cross-sectional research based on aggregate data from
several schools would have erased these site-specific differences. According to Willms
and Raudenbush,

[T]he estimation of school effects will never be an exact science. Sampling is
often problematic especially in secondary schools ... and the measurement
characteristics of the intake, process, and outcome variables can affect estimates
of school effects. (p. 227)

Barlow and Robertson (1994) described the political nature of education and its
relationship to the policy priorities of the government: the provision of cost-effective
education with a commitment to quality assurance for stakeholders. The direction of
educational policy-making in the 1980s was derived from the politics of the economy;
however, according to Barlow and Robertson, the connection between education, politics,
and the economy was not made visible: “Education is power; power is political... its
political nature should be made more evident, so that the public can more readily identify
the intent and impact of education reforms” (p. 112).

In 1980s, commentators offered vocal but fragmented opposition to the re-
visioning of public education as a business, and they registered their opposition to
CMEC’s emphasis on the development of school-business partnerships and a school
testing program that responded to the demands of business (Barlow & Robertson, 1994).
The introduction of standardized testing reinforced business principle of product quality assurance in education.

Canadian educators have held that our schools should not operate as if they were corporations, because their mandate is to educate all children, not a selective few. Unlike business, schools don’t have the luxury of choosing among potential students those they might want to teach. ... The transnationals are blurring or erasing national sovereignty lines, and the rhetoric calling for business to rescue a failing education system remarkably similar... For big business in both the United States and Canada, creating a false crisis – that the unemployed are at fault for their status because they are not properly trained – has led to a false solution, Blame education for our economic ills, and the social and economic policies that are failing our young can be left intact. (Barlow & Robertson, 1994, p. 60)

Their words were strong and angry. Like Bracey (2007), Barlow and Robertson (1994) believed that the form of truth telling that the Canadian government and business leaders were promoting was a poor representation of the facts. Their work foreshadowed the concerns of teachers about the narrowing of the curriculum, teacher deskilling, and increasing competition between schools and school systems. However, from a Foucauldian perspective, it was the productive capacity of this truth telling activity that would lead to actions. These actions would become the focus of genealogical work - what was produced and what can be known.

During the 1980s the issues facing Canadian educational policy makers were influenced by the “[d]ecline in revenues, unfinanced provincial mandates and escalating local needs put educational philosophy in the shadow of fiscal restraint” (Barlow & Robertson, 1994, p. 18). At the end of the decade, the delivery of Canadian public education emerged as a domain produced through the application of rational management practices and principles (Manzer, 1994). The transformation from a domain largely viewed as an agent of participatory democracy into a product of rational management was difficult (Manzer, 1994). Education as a domain was reformulated within an economic
paradigm: The need to demonstrate cost effectiveness, provide quality assurance, address capacity management, create success indicators, and measure the effectiveness of school systems rose in importance (Manzer, 1994). These issues overshadowed concerns about the delivering effective programs and ensuring the ability of students to assume roles as citizens in a democracy.

The agreement of the members of the Council of Ministers of Education, Canada, to collaborate on a national school testing program in 1989 was just one of the productive outcomes of this re-thinking of the role and outcomes for education in a globalized economy. The Council of Ministers of Education, Canada (1996b) stated that

CMEC launched in 1989 the School Achievement Indicators Program (SAIP) in which a consortium of provinces worked together to produce and implement, for the first time, a national assessment instrument for elementary and secondary reading, writing, and mathematics. CMEC also formed in 1989 the joint venture with Statistics Canada, called the Canadian Education Statistics Council (CESC), to provide comparable educational statistics. (p. 13)

CMEC was silent on the forces that lead to the agreement to launch SAIP, and the literature has little specific or concrete information about the decision to move forward with SAIP. It could be that in the Canadian landscape, SAIP was of little consequence because participation was in theory voluntary (Goddard, 2000). The Council of Ministers has published little information about the history of CMEC and its funding arrangement; however, based on secondary source, Barlow and Robertson (1994) have attributed the funding of SAIP to the federal government.

Genealogical work is difficult. Locating documents, following the disparate pieces of information through a variety of puzzling and challenging thought patterns, and addressing the need to feel confident that the information is reliable cause me to feel unsettled. The Council of Ministers of Education, Canada, in particular, presents itself as an enigma. It publishes many documents and research reports but provides little information about itself or its history. Barlow and Robertson (1994) make reference to contributions from Liberal government
in 1981 to CMEC to support the development of a national testing program for Canada and an additional reference to a further endorsement from the Progressive Conservative government for the development of a national school assessment program. Their source is an unpublished paper by Kathryn Chang Barker. Should I feel confident that this information is sound? It would seem logical that the federal government would support this initiative but constitutionally difficult as the provision of education is a provincial responsibility, and the Council Of Ministers of Education Canada reminds its readers of that fact in almost every document. (Dagenais, February, 20, 2010)


We live in a world of statistics: measuring growth, productivity, height, longevity, money markets from every angle. (Saul, 2005, 135)

Governance of education in Canada emerged from the ‘80s in a somewhat scarred condition as declining enrolments and the fiscal restraint cast education in a negative light resulting in “political and policy shifting against [the form of] education” delivered in the 1980s (Manzer, 1994, p. 210). Radwanski (1987), a Canadian journalist, illustrated the shift in educational priorities in his articles, which not only emphasized the importance for high school graduates to enter the global marketplace with the job skills that met the needs of employers, but also reiterated the importance of the conceptual thinking skills as a means to maintain a free, democratic, and compassionate society.

Following the 1989 agreement to move forward with a “first-ever attempt to arrive at consensus on the elements of a national assessment,” the ministers of education for the provinces and territories signed a Memorandum of Understanding in 1991 that outlined their intent “to measure the achievement of 13-year olds and 16-year olds in reading, writing, and mathematics” (Council of Ministers of Education, Canada, 1993, p. 3). Prior to the signing of the Memorandum of Understanding (1991), CMEC had facilitated the participation of Canadian public students in several international academic assessment programs sponsored by the OECD, IEA and the International Assessment of
Educational Progress (IAEP) (Council of Ministers of Education, Canada, 1993). These international studies provided comparative analysis based on the data provided by students in particular jurisdictions – not all provinces participated in these studies (Council of Ministers of Education, Canada, 1993). Both international and national assessment studies were gaining the attention of a wider audience because education was now considered by many nations and economists to be a significant factor in economic growth as measured by the Gross Domestic Product (GDP) of nations and regions (Grineski, 2005).

Subsequently, the Economic Council of Canada put forth its 1992 report, A Lot to Learn, in which the positive relationship between education and economic well being was highlighted as the key to a prosperous future for all Canadians (Manzer, 1994). Layered on top of and interspersed within the ideological conversations about the goals of educations, policymakers balanced the discussions about the need to homogenize the delivery of education in Canada versus the need to respect the needs of the local community (Manzer, 1994). These conversations and discussions were further complicated by the debate about whether the Canadian priority should be placed on a child-centred or content-focused curriculum orientation (Manzer, 1994). Each of these forces acted like a large tendril yanking and pulling the conversations about education in different directions. The tendrils themselves were not uniform and were composed of seemingly disparate members; however, the debates, discussions, and conversations focused on identifying the educational approach that would be most beneficial to future economic development.
The child-centred curriculum orientation, according to Manzer (1994), was responsive to the needs of the individual, supported a democratic philosophical orientation, and recognized the importance of cultural diversity while a content-focused orientation supported an educational performance-accountability orientation and implementation of external assessment programs. Performance in an educational setting was described as the demonstration of the requisite skills and knowledge by students at various stages and grades; and accountability referred to the activities of individuals – parents, students, teachers, administrators, and policy makers - who were responsible for ensuring that students had the opportunity to perform at an appropriate level (Manzer, 1994). The performance was usually executed on a test, and accountability was provided to the stakeholders through a reporting mechanism. In Canada the debate between content-focused and child-centered curriculum orientations became complicated and messy. Manzer (1994) described most Canadian schools as the places “where cultural world-views and life-worlds are taught” and if Canadian schools “fail to incorporate ...the diverse cultural communities that comprise Canada,” they would be deemed “unjust” (p.253). These two forces – child-centered and performance-accountability - were vying for centre stage in the ‘90s. The Business lobby – Conference Board of Canada, BCNI, and the Economic Council of Canada - offered strong support for the content-centered orientation while the CTF argued for caution and reflection (Manzer, 1994). The provincial departments of education varied in their level of affinity to these ideologies: Alberta favoured a more content-focused approach with external assessments while Saskatchewan favoured a more child-centred approach with teacher-developed classroom assessments (Manzer, 1994).
In my recent conversation with teachers, about external standardized testing programs, they recognize the importance of accountability but express strong concerns about the way that they must be accountable. A British Columbia public school teacher complained bitterly about the lack of fairness: Her students were required to take the prescribed standardized tests at a time in the school year in which they would not do well, yet they were working hard to achieve the goals of the curriculum. As a public school teacher, she was required to teach the curriculum although not all the approved curriculum was assessed by the standardized test. Her colleagues at the local private school could invest more time in test preparation activities and content so their students would perform better. She felt strongly that the context of the public school - its diversity and the challenges – is not considered in the reporting of the outcomes of standardized testing. She feels that her students are not fairly represented in the process; however, she believes that educators should be accountable to their stakeholders for the outcomes of their work. (Dagenais, May 23, 2010)

In Canada the focus of content-oriented education debate centred on SAIP, the product of the 1989-policy decision by CMEC (Manzer, 1994). The quality of students’ school performance was measured by the SAIP instrument, and accountability was provided to stakeholders through the analysis of students’ performances and other pertinent school data such as dropout rates (Manzer, 1994). The CTF persisted with its argument against SAIP. The outcomes, according to the CTF, of such testing programs would result in graduates with proficient skills; however, the Canadian curricula would become progressively more homogeneous and narrow as provincial governments reorganized their curricula in response to SAIP results (Manzer, 1994). The business lobby groups argued that SAIP should reflect the mathematics, reading, and writing skills that industry was looking for in new employees and that the scores should be normed with industry requirements; however, CMEC was not persuaded that this was the appropriate direction (Manzer, 1994).

The provincial governments struggled to maintain their commitment to the project. In 1991 Ontario left the SAIP project based on curriculum concerns put forward by the CTF: “the projected tests were not based on the curricula ... a new level of testing
would inevitably have the effect of narrowing and homogenizing the curriculum in each province and teachers would be evaluated on the basis of test results rather than on the kinds of students in their classes” (Manzer, 1994, p. 245). The minister of education for Ontario agreed with the CTF and registered a further objection: “samples would not be larger enough to analyze test results in terms of their demographic composition”; hence the impact of “Ontario’s ethnic diversity” on the results would not factored into the analysis (Manzer, 1994, p. 243). Ontario rejoined the SAIP project six months later once those curriculum concerns were allayed. Subsequently, Saskatchewan withdrew from the SAIP project because the timing was not right (Manzer, 1994). Despite the last minute withdrawals from the project, the ministers of education were able to agree upon a research question that had importance to all jurisdictions: “How well are our schools preparing students for a global economy and for lifelong learning?” (Council of Ministers of Education, Canada, 1993, p. 2). The forces driving CMEC to engage in a project like SAIP were both internal and external. The project was driven by accountability and economics, framed by cultural desires to preserve some elements of Canadian cultural diversity, and designed to render some meaningful comment on the degree to which Canadian public schools were able to achieve some curricular goals.

In June 1993 following the first administration of SAIP, several key Canadian commentators gathered to debate the notion of a national role for education in Canada. While SAIP alone is not responsible for instigating the debate, it did bring to the forefront the shadowy role of the federal government in funding and shaping Canadian educational philosophy and the possible need for a national office of education or a federal ministry of education (Lupart and Nagy, 1994).
Francis Whyte (1994), representing CMEC described education as an “integral part of the social fabric of a nation,” and he categorized the problems facing education systems as those of “growing social complexity” (p. 11). Education, according to Whyte, was not a “production system’ where you analyze the quality of input and the output in relation to customer needs and make strategic adjustments” using the only the data generated by the results of standardized tests. From Whyte’s perspective “education is best delivered at the local level so that its priorities and programs recognize the diverse needs and aspirations of all Canadians,” and CMEC’s should be to further this goal through consultation and consensus building (p.19). SAIP was an example of a “true national education program, built on the basis of cooperation, consultation, and consensus, and addressing [CMEC’s] first priority: increased public understanding of how our educational systems work: input, output, process and achievement” (p. 20).

Heather-Jane Robertson (1994), representing the CTF, argued that initiatives such as SAIP should come out of a federal ministry of education that is both accountable and responsible to its citizens: “[T]he federal government has an obvious and substantive role and influence in [elementary and secondary} education” (p. 31). According to Robertson, lack of accountability, public debate, and transparency on federally-sponsored initiatives such as SAIP coupled with the existing program duplication made the current structure for the delivery of education problematic. Robertson summed up her position on federal engagement in education:

If efficiency, effectiveness and accountability favour federal involvement, what considerations favour a more limited federal role? ... [D]oes a federal presence ... impose lock-step standardized curricula which would neglect different needs and experiences? ... Not necessarily. However, in the contemporary world, it becomes ever more difficult to assert with confidence that what a person will need to know in the year 2020 ought to be shaped profoundly by the accident of the
location of her or his birth. By and large, existing variations [in curricula] are overestimated. (p. 37)

Ron Duhamel (1994), MP, Winnipeg – St. Boniface, pointed out that there was not only ambiguity about the nature of the federal role in the delivery of education, but also “about what education is all about ... [and] what does it encompass” (p. 43). From Duhamel’s perspective SAIP was a “good start ... a useful beginning” that might assist in restructuring by looking “at education from a different perspective” (p. 43). Duhamel viewed SAIP as a homogenizing tool that could generate data to assist policymakers to bring about changes that could inspire Canadians to feel “confident that what their child or young adult is receiving is comparable to what is happening in the very best and finest education systems in the world (p. 47).

Shapiro (1994), a professor at the University of Toronto, stated that if the development of SAIP and the establishment of the Statistics Council (to avoid using a federally-supported agency to analyze SAIP data) were the sole achievements of the CMEC’s 26-year history, then Canadians have not received good value for their money. According to Shapiro, lack of trust between provincial ministries of education and the federal government has limited the transparency of federal engagement in education and the development of a Canadian educational identity. Shapiro stated “we live in a globalized economy... but if there is nothing special in the Canadian content [of educational programs and delivery], the question of the federal role or the provincial role really does become moot” (p. 55).

As the Council of Ministers rolled out SAIP, the delivery of education in Canada was facing pressure from the OECD to develop national goals and standards, and the public were raising concerns the effectiveness and efficiency of the Canadian school
systems (Whyte, 1994). The CTF issued demands for increased transparency and accountability related to the federal engagement in the delivery of elementary and secondary programming and raised concerns about the preservation of the school diversity (Whyte, 1994). Could SAIP allay all these concerns?

According to Whyte (1994), SAIP has the potential to fulfill a “truly national function: presenting a picture of the achievements of individual education systems in comparison with each other” (p. 21). Essentially, SAIP was designed to fulfill the Foucauldian task of surveillance; the schools systems would become visible through the data provided by their students, and the data from one system could be compared with another system. Robertson (1994) argued that surveillance and accountability should begin with the federal government and the programs it funds in 17 different departments – federally funded programs and initiatives targeted at elementary and secondary school systems that came into existence without the benefit of public debate or discussion.

1992 marked a critical year both for SAIP and for discussions about a federal role for education. The federal and provincial ministers reached a form of constitutional agreement; the federal government agreed to abolish the Science Council and the Economic Council, “the main advocates of educational reform” at the federal level (Manzer, 1994, 253). Following this decision, the provincial and federal policymakers did congeal around these concerns: “the importance of a national policy for education” and “the need for a more content-oriented curriculum focused on basic skills (Manzer, 1994, p. 253). The discussions about the possibility of a formal national policy for the organization and structure of elementary and secondary education faded from the literature and from the critical conversations about the delivery of education in Canada;
however, the evidence of centralizing forces and a honed focus on identifying and measuring critical skills development remained as a thread throughout the development of SAIP.

The First Testing Cycle

CMEC established the goal for the first cycle of SAIP as the development of “a set of indicators that will inform the governments and the public about the performance of Canada’s education systems” (Council of Ministers of Education, Canada, 1991, p. 1). The indicators fell into three broad categories: student retention and graduation, academic performance, and students’ perception about key behavioural and attitudinal factors (Council of Ministers of Education, Canada, 1991). SAIP was designed to provide data on students’ academic performance and their perceptions about key behavioural and attitudinal factors; graduation and retention data was generated through an alternative data collection process (Council of Ministers of Education, Canada, 1991). CMEC used a consortium approach to develop the assessment instruments and a committee approach to manage the SAIP project (Council of Ministers of Education, Canada, 1991). The representatives from Alberta and Quebec provided leadership for the consortium, and representatives from Ontario were invited to join as observers once Ontario rejoined the SAIP project (Council of Ministers of Education, Canada, 1991). Deputy Ministers from the provinces and territories formed the project management committee, and the Canadian School Boards Association was invited to critique the development process (Council of Ministers of Education, Canada, 1991). Although CMEC stated that its goal was to develop its assessment instruments in “consultation with teachers, parents, trustees and other interested groups,” CMEC has relied on the ability of appointed bureaucrats
and national agencies to engage Canadians in a broad consultative process (Council of Ministers of Education, Canada, p. 1). Whyte (1994) stated that the consultative nature of the process did show that “Canadians from very diverse backgrounds can indeed work together across jurisdictions to produce state-of-the-art programs with a clearly-defined purpose” (p. 21).

The consortium team was able to draw on the large-scale assessment experience of several provincial ministries of education as several Canadian provinces had participated in international assessments (Council of Ministers of Education, Canada, p. 2). Based on their previous experience with large-scale assessments, the consortium teams framed an assessment tool that in many ways was quite unique: The development process was consultative in nature, and the product attempted to recognize the diversity of the Canadian education system.

Whyte (1994) described SAIP as “not a regular standardized test or exam, does not give grades or compare one student’s performance with another’s” (p. 21). The focus is on measuring “how well each province’s education system is doing and does not replace individual student assessment, which is the responsibility of teachers, schools, boards, and ministries of education” (Council of Ministers of Education, Canada, 1993, p. 3). Educators from Alberta and Quebec developed the SAIP assessment tool collaboratively in French and English to reduce the negative impact of translation (Whyte, 1994).

At its unveiling, SAIP’s structure was novel: “the assessment criteria have been developed on a continuum of competency along a five point scale” and each level “allowed for descriptions and reporting of students’ performances in fairly broad terms
that the public will find meaningful” (Council of Minister of Education, Canada, p. 1). Once the levels were set, the item developers used the provincial curricula as content guides; and for the SAIP Mathematics 1993, they also used the National Council of Teachers of Mathematics curriculum guide (Council of Ministers of Education, Canada, 1993). The SAIP assessments were designed so that the range of skills and knowledge that 13-year olds and 16-year olds might acquire through the schooling process were represented in the test items (Council of Ministers of Education, Canada, 1993). Both groups wrote the same test with the expectation that each group would perform at a different level (Council of Ministers of Education, Canada, 1993). Determining the mean level of performance for each group and describing the difference between the performance levels of the two groups of students was one of the research questions for the first cycle: “no standard of performance was established”; thus, “no comment can be made about how 13-year-old or 16-year-old students should be performing” on the first cycle of tests (Council of Ministers of Education, Canada, 1993, p. 13). Following the development process, all SAIP assessments were field tested on a predetermined student sample prior to their administration (Council of Ministers of Education, Canada, 1993).

SAIP assessments were administered to age-specific cohorts of students from all provinces; these cohorts were representative of the rural-urban split of the Canadian population (Council of Ministers of Education, Canada, 1993). Because the delivery of education and the curriculum varied from province to province, CMEC chose age-specific groups rather than years of schooling. According to CMEC (1991), age thirteen marks a transition for many students from one type of curriculum goals to another (in some jurisdictions, students make the transition from elementary school to secondary
school curriculum) and age sixteen marks the transition from compulsory schooling to voluntary schooling.

The administration of the SAIP Mathematics Test -1993 provided the testing ground for the assumptions that framed the structure and delivery of the first cycle of SAIP tests. While the SAIP Mathematics test – 1993 was in its final stages of revisions, the conceptualization and development of the Reading and Writing assessment and the Science assessment had begun. A similar consortium approach to assessment tool development, project management, item development and review were implemented with these assessments.

The SAIP Mathematics Assessment – 1993

The overall structure for the SAIP Mathematics test followed a five-level design with a preliminary placement test administered and marked by classroom teachers (Council of Ministers of Education, Canada, 1993). The test provided for assessment in two large thematic areas of mathematics – content and problem solving – using both multiple choice and constructed-response test items (Council of Ministers of Education, Canada, 1993). The administration procedures were standardized, and CMEC provided oversight at selected sites to ensure that the administration protocol was followed (Council of Ministers of Education, Canada, 1993).

During the three-hour testing period, both the 13-year-old and 16-year-old groups of students completed the same teacher-marked placement test booklet followed by two test booklets – one at their placement level and one at the level above their placement level (Council of Ministers of Education, Canada, 1993). The students were assigned a level based on the highest test booklet level on which they earned a score of 60 percent or
higher (Council of Ministers of Education, 1993). This testing approach was labour-intensive, but also failed to recognize that some students might have been able to complete an additional test booklet at a higher level and earn a score of 60 percent. Conversely, some students might not have been able to earn 60 percent on their first test booklet, yet they were required to attempt questions on an additional test booklet and endure the possible frustration resulting the generation of negative feelings about self. Although teachers were able to excuse students that they believed were “not capable of satisfying the criteria for Level 1,” there were many students who struggle in a standardized testing situation who did not fit into any one of these categories (Council of Ministers of Education, Canada, p. 13).

What Did the Test Results Tell Us?

The Report on Mathematics Assessment – 1993, included several key elements: sample construction; test outcomes by age group, gender, and by province; the description of curriculum resources by province; and the data captured through the behavioural and attitudinal questionnaire. Although Whyte (1994) reported that “a carefully constructed sample of 56,000 13- and 16-year-old children participated in the mathematics assessment element of SAIP” in his commentary on SAIP (p. 20), the Report on Mathematics Assessment – 1993 reported a somewhat smaller actual sample divided by test type and age group (Council of Ministers of Education, Canada, 1993). The results for SAIP Mathematics assessment were presented in a series of bar charts organized by test type that offer comparisons of the performance of 13-year olds with that of 16-year olds from a Canada-wide perspective, females with males, and provinces and
territories with the Canada-wide sample (Council of Ministers of Education, Canada, p. 13). Although the preamble stated that the data was not intended to be understood as comparative, the bar charts placed the performance of one group beside that of another; this presentation of data was designed to lead the reader into making comparisons.

The first concern posed by the CMEC was the following: Is the performance of 13-year olds on the SAIP mathematics test significantly different from that of 16-year olds? The aggregated results for the Canadian sample showed that the distribution of students’ results at Level 1, 2, and 3 are different (Council of Ministers of Education, Canada, 1993). The distribution of scores for Mathematics Content for 13-year olds showed an almost equal distribution between Levels 1, 2 and 3 (Level 1 – 28%, Level 2 – 34% and Level 3 28%) while the distribution of scores for 16 year-old olds was skewed to favour Level 3 performance (Level 1 – 15%, Level 2, 20% and Level 3 – 46%). The results for problem solving showed little difference between the performance of 13-year olds and 16-year olds at Level 2 (48% and 50% respectively); however more 13-year-olds scored at Level 1 (35% as opposed to 21% for 16 year-old olds) and more 16 year olds scored at Level 3 (21% as opposed to 8% for 13-year-olds) (Council of Ministers of Education, Canada, 1993).

The proportional distribution of scores for each level of each age group for the Canada-wide sample served as the de facto mean for the discussions about the performances by gender. Gender-based differences appeared in the results of problem solving assessment: A significantly larger proportion of 13- and 16-year-old females performed at level 2 while a significantly larger proportion of males performed at level 3.
(Council of Ministers of Education, Canada, 1993, p. 17). The commentary did not strive to explain these differences; the data was just presented.

The performance of each province was reviewed and compared with that of the Canada-wide sample. Statements such as the following described the performance of the students for each cohort relative:

The results for British Columbia students of both age groups were not significantly different from the Canada-wide results in either content or problem-solving (p. 19).

In problem-solving, there are fewer 13-year-old Manitoba English-speaking students achieving at Level 2 (43%) and Level 3 (4%) than is true Canada-wide (Level 2 - 48% and Level 3 – 8%) (p. 31).

Forty-one percent of the 13-year-old French-speaking Quebec students are achieving at Level 3 in mathematics content. These results are higher than the Canada-wide results (P. 49).

These statements used the Canada-wide results as a comparative norm – the expected result of “the larger representative group (the “norm”)” (Canadian Teachers’ Federation, 1999, p. 4). Although SAIP was developed as a criterion-referenced tool, the data reporting moved away from a discussion focused on the attainment of particular levels of skills and knowledge to a comparative discussion of the levels of achievement in each jurisdiction with the Canada-wide norm and with other jurisdictions. A criterion-referenced assessment would have provided commentary about the degree to which the students in each province achieved the predetermined criteria described by the established performance indicators. The commentary failed to elaborate on the appropriateness of Level 2 achievement as a reasonable performance outcome for 13-year-old students and Level 3 for 16 year-old students.

The sample used to construct the Canada-wide results did not appear to be representative of the Canadian student population, and the procedures for describing the
sample lacked jurisdictional consistency. However, the sampling strategy was described as random and “designed to allow for different sizes of the populations in provinces and to provide statistically valid data” (Council of Ministers of Education, Canada, 1993, p. 4). For provinces with a medium-sized population like British Columbia (1,277 formed the constructed sample for 13-year-olds) and Alberta (1,291 formed the constructed sample for 13-year-olds) about 6% and 7.25% of the student population respectively completed the content assessment (Council of Ministers of Education, Canada, pp. 19 & 26). For provinces with larger populations like Ontario and Quebec, the report provided the number of 13-year olds and 16-year olds for both the English and French speaking samples; however, the report did not translate the sample size into a proportion of the eligible student population. The total number of 13-year olds in the Ontario sample for the math content assessment was 2213; and for Quebec, was 2020 - approximately twice as large as the sample size for British Columbia; however, the size of sample does not represent a similar proportion of the population (pp. 37, 40. 45 & 48). Other provinces and territories, with much smaller populations of eligible students, assessed a much larger proportion of their populations - New Brunswick assessed 48% of its students in both age groups; Prince Edward Island, 65% of its 13-year-olds and 70% of its 16-year olds; the Yukon, 96% of its student population. The effects of the sampling strategy on the results of the assessment were not discussed; however, it is interesting to note that the Yukon, Prince Edward Island and Manitoba - provinces that were sampled more intensively - showed lower results. Ontario and Quebec, provinces whose students performed well, did not describe their samples as a proportion of the eligible students; however, based on the
number of eligible students in Ontario and Quebec, the proportional representation from these provinces was lower than that of either British Columbia or Alberta.

From a Foucauldian perspective, it is critical to consider how the truth statements were constructed and the nature of the actions that those statements provoked. The data was gathered to measure the performance of each educational system, organized to tell some truths about the performance of provincial educational systems, and to provide assistance to each province “in setting educational priorities and in planning program improvement” (Council of Ministers of Education, Canada, 1993, p. 3). Although the stated intention of the assessment was not comparative, the performance statements for jurisdictional and data subset were presented as comparative. Truth telling statements presented some jurisdictions as better performers than others, as student achievement was organized to stand in proxy for the performance of school systems.

The sampling strategy may have contributed to the types of truth statement that the data could produce. The sampling strategy was controlled for a specified level of error (+/- 3% at a 95% confidence level) and was designed to accommodate jurisdictional priorities – some jurisdictions preferred to assess a small proportion of their students while other jurisdictions preferred to assess a much larger portion (Council of Ministers of Education, Canada, 1993,). The sampling strategy did not address the differential impact of socioeconomic status (SES) and neighbourhood wealth on student performance. It was unclear if the Alberta, British Columbia, Ontario, and Quebec samples proximate the SES indicators and neighbour wealth distributions of these provinces, and the differential effects of SES the production on the jurisdictional and Canada-wide norms were not explored. The truth telling statements based on the SAIP
Mathematics test results for some provinces and territories might have been better or different if they had used a sampling strategy similar to Ontario or Quebec, or conversely Quebec and Ontario may not have fared as well in comparison to the Canada-wide norm if these provinces had constructed a sample in similar proportion to the Yukon or Prince Edward Island. Nova Scotia was the only province to express concern about the sampling protocol: According to its school administrators “several schools felt the sample selection process produced a biased sample” (p. 69).

The SAIP Mathematics assessment also included a short questionnaire on which students rated their behaviours and attitude about the learning of mathematics. These factors – the length of time in school, the numbers of hours spent watching television, the number of mathematics courses taken, and the opportunity to use a calculator or computer - did not correlate with students’ performance on the SAIP assessments. However, the number of hours spent on homework, students’ liking of mathematics, and their feelings of confidence in their ability to do mathematics related positively to the test scores (Council of Ministers of Education, Canada, 1993). CMEC, like other groups, stated that it was interested in improving the outputs of educational systems and was interested in the degree to which non academic factors influence the performance of students; therefore, the identification of the factors that correlated strongly with higher levels of academic performance was a key foundational research activity. According CMEC, the further exploration of the nature and the contribution of these factors to student success should form part of the Canadian research agenda on teaching and learning (Council of Ministers of Education, Canada, 1993). Although the students’ attitudes toward learning mathematics and their feelings of confidence in their ability to
do mathematics were measured in a Canadian context for the first time on the SAIP Mathematics – 1993 assessment, these questionnaire items have appeared consistently on other large-scale assessments.

The first round of SAIP presented an assessment structure and a results reporting strategy that would prove to be a legacy for future iterations of SAIP assessments. The struggle with criterion-based reporting versus the need to rank and compare subsets of the Canada-wide student population became evident in this first iteration of SAIP and foreshadowed this ongoing struggle. The SAIP report ascribed meaningfulness to the attainment of Levels 2 for 13-year-olds and Level 3 for 16-year-olds that resonated with attention paid to grade-point equivalencies on other standardized tests. The importance of comparing the performances of the provinces superseded the need to develop an understanding of what students’ attainment of 60% of the skills ascribed to Level 2 or 3 really meant as a measure of the performance of a school system.

*The SAIP Reading and Writing Assessment – 1994*

CMEC celebrated its success in 1994. It had not only developed a process for the creation of a bilingual reading and writing assessment, but also it had developed the protocol for its Canada-wide administration. CMEC stated that

Assessing academic achievement across nine provinces and two territories that do not have a common curriculum or set of standards was especially challenging. It is indeed remarkable that the assessment succeeded in probing the skill levels that are normally addressed in large-scale testing. (Council of Minister of Education, 1994, p. 94)

The CMEC recognized the difficulty associated with creating a collaborative environment in which educators and classroom teachers could work cooperatively to produce criteria, level descriptors, and test items; and in developing the assessment
instruments that would measure the effectiveness of school systems that used different curricula (Council of Ministers of Education, Canada, 1994).

The SAIP Reading and Writing assessment adopted similar research questions as the math assessment: “How well do students in Canada read and write?” and “Does the achievement of Canadian students change over several grades?” (Council of Ministers of Education, Canada, 1994, p. 4) Following the format established with the mathematics assessment development process, in 1990 a consortium consisting of members from Alberta and Quebec began a two-year development process that included identifying the assessment criteria and writing the test items in both French and English (Council of Ministers of Education, Canada, 1994). The assessments conformed to the five-level structure initiated in the math assessment. The following assumptions informed the development of the assessments:

A student’s reading fluency depends on the personal experience brought to the reading task, the student’s language base, the complexity of textual information, and the difficulty of the task; [whereas writing proficiency is dependent on] personal experience with written language, the degree to which student’s language base allows expression of ideas, and the complexity of the writing task. (Council of Ministers of Education, Canada, 1994, p. 7)

The assessment protocol was aligned with common classroom practices for measuring writing proficiency; therefore, students were provided with the “opportunity to read textual materials on a theme (heroes), make notes ..., discuss their ideas with peers ..., write first drafts, and edit their drafts” using reference materials and computers (Council of Ministers of Education, Canada, 1994, p. 7). The reading and writing assessment, unlike the math assessment did not use a placement test; the students were placed in the level in which they demonstrated most consistent success; however, consistent success was not defined.
The Results

The results report provided the same type of information for each participating province and territory (Saskatchewan opted out of the assessment protocol) as the SAIP Mathematics assessment report; however, the presentation of the results varied somewhat from the Mathematics assessment. The bar graphs for the Canada-wide sample presented the data for the performance of 13-year olds and 16-year olds on the same graph to facilitate comparison of the proportion of each age group at each performance level (Council of Ministers of Education, Canada, 1994). Confidence intervals were marked on the graphs to visually illustrate the degree of statistical difference in performance. For example, it was easy to note that in the Canada-wide sample 34.8 percent of the 13-year-olds and 39.1 percent of the 16-year-olds performed at Level 3. Those performances were statistically different; the bars did not overlap (Council of Ministers of Education, Canada, 1994, P. 12). The commentary focused on the differences in student performance as illustrated with this comment:

The results show that a greater percentage of older students is at the higher levels (levels 3 to 5) and a lower percentage of students is at levels 1 and 2. ... It was expected that the students with three more years of instruction would attain higher levels. The SAIP results, however, go beyond simple opinion, since they are based on the administration of a sophisticated assessment instrument to the largest sample of Canadian students ever. (Council of Ministers of Education, Canada, 1994, p. 12)

The commentary displayed confidence that the results were meaningful and illustrative of real differences and that the differences meant that the three years of additional schooling were both meaningful and valuable for 16-year olds. However, a question remains: Did the results illustrate a significant and real difference in students’ attainment educational outcomes?
The report contained additional charts that compared the performances of students from provinces and territories with the Canada-wide sample for both the reading and the writing assessment and additional analyses segmented by gender and by language. Although the presentation of the results varied from the mathematics assessment results, similar information was presented. Each chart facilitated the comparison of the performance of each jurisdiction in relation to the Canada-wide sample and to each other jurisdiction. For example, the performance of 13-year olds in the Northwest Territories in Reading at or above Level 1 was lower than the performance of the Canada-wide sample and the British Columbia sample (statistically significant), and the performance of French speaking 13-year-olds in Quebec was higher (statistically significant) than the Canada-wide sample (Council of Ministers of Education, Canada, 1994, p. 14). The sections devoted to each jurisdiction continued to affirm the role of the Canada-wide results as the performance norm. The following comment was typical of the comparative language used in explicating the results for each province:

In Ontario, the results of the English-language reading assessment for both 13-year-old and 16-year-old students, at each of the five achievement levels, do not differ from the results for all Canadian students in the English-language assessment. (Council of Ministers of Education, Canada, 1994, p. 46)

These types of comparisons were presented for the results for each province and territory. Although the introductory discussion focused on the important role of criteria in the construction of the assessment tools, the language used to discuss the results was comparative.

Results aggregated by gender reveal that “[g]irls demonstrated better reading and writing than boys” (Council of Ministers of Education, Canada, 1994, p. 27). The magnitude of differences varied in relation to the age group and the performance level.
For example, a larger percentage of 16-year-old females performed at Level 4 on the reading assessment, while a larger percentage of 16-year-old males performed at Level 2 on the writing assessment; and CMEC noted that “the results are consistent with other studies” (Council of Ministers of Education, Canada, 1994, p. 27).

The study also paid attention to the differences in performance of students who completed the assessments in French as compared with the results of those who completed the assessments in English; students completing the assessments in French performed at a higher level than those completing the assessment in English (Council of Ministers of Education, Canada, 1994). The results for the Canada-wide francophone sample and the Quebec francophone subset were higher than the results for the Canada-wide sample and most of the results for the other provinces and territories. This cautionary statement is offered “great care must be taken in comparisons between languages ... [as] several international studies have shown that exact equivalency [between versions of same test taken in different languages] is hard to achieve” (Council of Ministers of Education, Canada, 1994, p. 95).

The report provided general information about the sample size – 58,000 students – 43,000 students wrote in English and 15,000 students wrote in French, but the report does not describe the composition of the sample by jurisdiction (Council of Ministers of Education, Canada, 1994). The total number of test writers for each province was provided, but the report failed to describe the sample or the subsamples as a proportion of eligible students. The sample description did not provide information on socioeconomic status (SES), rural/urban distribution participation, or the neighbourhood wealth of the schools. Without the additional sample descriptors, it is difficult to determine if the
differences were real and significant or a product of the sampling strategy. Based on the limited information, it is difficult to determine if the samples provided by the large provinces were indeed representative of the student population and balanced for gender and for socioeconomic factors.

The Reading and Writing assessment included a background questionnaire to gather information about the characteristics of the student population and its attitudes towards behaviours that influence reading and writing proficiency. The survey gathered data on these factors: length of time the students had been in the province, grade level, availability of books at home, reading for pleasure and attitudes towards reading and writing (Council of Ministers of Education, Canada, 1994). Factors such as liking to read, availability of books in the home, reading for pleasure, and grade level – all resulted in a modest positive correlation with the test results. The relationship between length of time in the province and the test results was weak (Council of Ministers of Education, Canada, 1994). The goal of the question was to measure the influence of second language learning; however, measuring second language learning by proxy led to polluted results. The data included the results of many students who had moved between provinces or from one French-speaking or English-speaking country to Canada (Council of Ministers of Education, Canada, 1994). The number of books students own correlated positively with writing performance (Council of Ministers of Education, Canada, 1994). Although the SAIP report did not specify that SES was a factor in writing performance, the report did encourage schools to “make books available to those who do not have the means to acquire personal books” (Council of Ministers of Education, Canada, 1994, p. 89). The ability to purchase books may serve as an indicator of wealth; therefore,
CMEC’s recognition that not all students may be able to purchase books validated the potential impact of inequitable distribution of wealth on student learning as measured by SAIP. Making books available did address the issue of providing reading material; however, this remedial strategy did not address the issue of the ability to purchase books which may stand in proxy for other more complex factors related to family and neighbourhood wealth that can affect students’ performance on large-scale assessments.

*The SAIP Science Assessment – 1996*

Although the Science Council of Canada’s 1984 report, *Science for Every Student*, established the importance and the value of science education “for each of Canada’s students,” the science assessment was not part of the original memorandum of understanding (Council of Ministers of Education, Canada, 1996a, p. 6). The assessment of student learning in science was added to the assessment protocol in 1993 when the “ministers further agreed to include the assessment of achievement in science” (Council of Ministers of Education, Canada, 1996a, p. 5). The SAIP science assessment posed research questions that were similar to those posed for mathematics and reading and writing assessments: “How well do students in Canada do in science” and “Does the achievement of Canadian students in science change over several years?” (Council of Ministers of Education, Canada, 1996a, p. 5) Essentially, CMEC wanted to be able to provide feedback to its stakeholders about the level of performance of Canadian schools and Canadian students in the area of science. CMEC also added that its stakeholders would want to know “whether [Canadian] students attain similar levels of performance at about the same age” although most Canadians recognized that “school programs differ
from one part of the country to another” (Council of Ministers of Education, Canada, 1996a, p. 6). Since science curricula vary from one jurisdiction to another, CMEC adopted these guiding principles from which they developed a framework:

Science is more than a body of complex and abstract knowledge and theories about the universe. It is more than a set of processes that guide scientific inquiry and discovery. ... [For] effective learning, science must relate to the everyday life of students. (Council of Ministers of Education, Canada, 1996a, p. 6)

From these assumptions about science and the learning of science, CMEC adopted the following science domains: “knowledge and concepts of science, nature of science, relationship of science to technology and societal issues, and science inquiry skills” from TIMSS 1993 assessment (Council of Ministers of Education, Canada, 1996a, p. 7). The inclusion of these science domains was “consistent with recommendations set out in Project 2016’s first report, Science for All Americans” (Council of Ministers of Education, Canada, 1996a, p. 7). Thus, SAIP Science - 1996 represented a unique combination of influences. The construction of the SAIP Science - 1996 was shaped by the following factors: science programs in Canada were diverse, learning science was important and valuable for all students in Canada, the TIMSS assessment offered a useable and ready-made framework, and it was acceptable to borrow and use assessment frameworks developed in the United States (Council of Ministers of Education, Canada, 1996a).

The development of the SAIP Science assessment followed a similar pattern as the other assessments: the formation of a consortium in 1993, the development of the indicators for each of the five performance levels in 1993, test item development in 1994, and field testing in 1996. The fact that the consortium composed of representatives from Alberta, Saskatchewan, Ontario, and New Brunswick was significant as this marked
Saskatchewan’s first involvement with the SAIP assessment program (Council of Ministers of Education, Canada, 1996a).

The science instrument was described as a “new and different testing experience” because students were asked to “relate their understanding of science to real-life situations that were familiar to them” (Council of Ministers of Education, Canada, 1996a, p.9). The SAIP science assessment like the mathematics assessment consisted of three parts: a 12-item science placement test, a science literacy test, and a hands-on science inquiry skills test (Council of Ministers of Education, Canada, 1996a). The hands-on science inquiry test was unique for a large-scale assessment standardized assessment: “[T]he attainment of science inquiry skills is universally acknowledged to be an essential aspect of science education”; however, “the assessment of achievement in this area, particularly on a large scale, often has been seen as difficult” (Council of Ministers of Education, Canada, 1996a, p.9). The introduction of the inquiry skills test and the decision by the Saskatchewan ministry of education to participate in the SAIP science development and administration cycle were two key initiatives of this phase.

The sample size for the 1996 Science assessment – 37, 500 comprised 19,500 thirteen year-olds and 18,000 sixteen year-olds - was considerable smaller than the sample size for either the 1993-Mathematics assessment or the 1994 - Reading and Writing Assessment (Council of Ministers of Education, Canada, 1996a, p.8). About 26,500 students wrote in English while approximately 11,000 students wrote in French (Council of Ministers of Education, Canada, 1996a, p.8).
The *Report on Science Assessment SAIP: School Achievement Indicators program 1996* did not provide reasons for the steady decline in the sample size with each successive SAIP administration.

The sampling strategy failed to provide information about the socioeconomic status of the participating schools, but it did state that some jurisdictions assessed only a sample of eligible students while others assessed all eligible students (Council of Ministers of Education, Canada, 1996a). CMEC recognized the importance of sampling strategy in gathering of data and the legitimation of its findings; therefore, the 1996 Science report provided enhanced information by jurisdiction about the number, proportion, and gender of the eligible students (Council of Ministers of Education, Canada, 1996a). The following illustrates the type of enhanced information: 1805 students of both ages groups wrote the science literacy test in Saskatchewan while 1999 students wrote in British Columbia and the proportion of males and females for each province was similar (Council of Ministers of Education, Canada, 1996a, p. 110). The populations of Saskatchewan and British Columbia vary considerable in size and in the distribution of SES, linguistic diversity, and cultural diversity; therefore, it is difficult to determine if the data provided by the sample populations were representative of the real performance of each jurisdiction.

The 1996 Science assessment results were reported using the format established with the 1994 Reading and Writing assessment report: Results were reported as percentages within an appropriate confidence interval by jurisdiction, age group, and performance (Council of Ministers of Education, Canada, 1996a). The purpose of introducing the confidence interval was to reinforce the notion that “the percentages
calculated by the researchers were based on samples of students and were only estimates of the actual achievement students in a particular jurisdiction would have demonstrated if they had all taken the assessment”; however, in jurisdictions in which all students wrote, a confidence interval was not calculated (Council of Ministers of Education, Canada, 1996a, p. 21). Reported differences indicated statistically different performances that might have represented a real difference in performance attributable to an educationally grounded factor. For example, the performance of 13-year olds at level 3 (43%) is significantly different from the performance of 16-year olds (69%) at level 3 on the written assessment because the confidence intervals did not overlap (Council of Ministers of Education, Canada, 1996a, p. 23). However the comparison of 13-year-old females and males did not reveal any significant difference: “13-year-old girls performed as well as boys at all levels in the written assessment” because the confidence intervals did overlap for all levels (Council of Ministers of Education, Canada, 1996a, p. 29).

The results for each jurisdiction are reported in a similar fashion: The data was segregated by assessment (written or practical), age group, and language (Council of Ministers of Education, Canada, 1996a). For all assessments, the Canadian-wide mean served as the norm to which each jurisdiction was compared (Council of Ministers of Education, Canada, 1996a). The following were typical comparative statements: “In [British Columbia] 13-year olds performed as well as Canadians as a whole” (Council of Ministers of Education, Canada, 1996a, p. 43) or in New Brunswick, “Francophone students of both age groups showed levels of performance significantly different from those of Canadians as a whole” (Council of Ministers of Education, Canada, 1996a, p. 66) or “The percentage of 13-year-old francophone students in Quebec performing at
level 3 or above is significantly higher than that of Canada” (Council of Ministers of Education, Canada, 1996a, p. 60). These three statements were indicative of a performance below the level of Canada-wide mean, at the mean or above the mean; however, it is interesting note that the word “different” without additional qualifiers was used to indicate a performance below the mean. Students living in Alberta and Quebec performed above the Canadian mean for some levels and age groups while some age groups and levels of students living in Manitoba (francophone), Ontario, (Anglophone and francophone), Nova Scotia, Newfoundland and Labrador, and the Northwest Territories performed below the Canadian mean. Since the performances of all students writing the assessment were used to calculate the mean, it is likely that some jurisdictions would have had more students performing above or below the mean. Thus, these results may have been indicative of a trend, or they may have been a data collection artefact as the sampling strategy varied from jurisdiction to jurisdiction based on each jurisdiction’s local priorities.

CMEC determined that it was not only important to compare the performance of each jurisdiction with the Canada-wide norm but also with the expectations of the Canadian public. To facilitate the setting of Canadian expectations for student performance in science, CMEC convened an anonymous 84-member expectation-setting panel to aid in the interpretation of the SAIP science assessment results (Council of Ministers of Education, Canada, 1996a). CMEC described the rationale for establishing expectations for student performance as follows:

Publicly declared performance expectations provide a comparator, which students, parents, educators, and general public can use to gauge student outcomes. If only the assessment results are reported, the question remains, “Should we be satisfied with the results?”
Program is a criterion-referenced assessment, not a norm-referenced assessment. Criterion-referenced expectations provide another means of interpreting results, in addition to Canadian norms and interprovincial comparisons. (Council of Ministers of Education, Canada, 1996a, p. 100)

The rationale was a lofty statement based on the assumption that the listed stakeholders would both be aware of the need to set expectations and have the ability to use them to gauge or improve student performance. Many questions emerged from the introduction and establishment of the expectations-setting process: To what degree did CMEC inform its stakeholders about the need to establish expectations? What was the nature of the force that produced the need to set externally-derived expectations? Can the answers to a broad and varied group of stakeholders’ questions about satisfaction with students’ assessment results be addressed through the setting of externally-derived expectations? CMEC defined expectation-setting as the following:

an evaluative judgment of the percentage of Canadian students who should perform at or above predetermined criterion levels of scientific performance, for a specific set of assessment instruments in a specific set of circumstances or context. (Council of Ministers of Education, Canada, 1996a, p.100)

This judgement in the form of expectations was to serve as the “interpretive framework around the assessment data, and to highlight the strengths and areas for improvement in student performance”; however, the authors of the expectations were to remain anonymous (Council of Ministers of Education, Canada, 1996a, p. 100).

In addition CMEC noted that the “expectations are not fixed or static ... and will vary according to the nature or the complexity of the questions or tasks” on a particular assessment (Council of Ministers of Education, Canada, 1996a, p.100). Thus, the expectations set for this cycle of SAIP Science may not be useful in subsequent administrations.
For this assessment, the expectations-setting panel was composed of educators and non-educators who attended one of four expectations setting sessions to provide their opinions about what 13-year olds and 16-year olds should know and should be able to do in the field of science (Council of Ministers of Education, Canada, 1996a). Expectations of the educators and non-educators were collapsed into one set of expectations, yet CMEC noted that both educators and non-educators concurred on the level of their expectations (Council of Ministers of Education, Canada, 1996a).

Although the composition of the panel remained anonymous, the members included “classroom teachers, parents, university academics and science specialists, aboriginal trainers, business and industry representatives and community leaders ... from every province in Canada” (Council of Ministers of Education, Canada, 1996a, p. 39). The task of the panel was to review the “assessment procedures, materials and actual student results to determine the percentage if 13- and 16-year-old Canadians who should achieve at each ... performance level” (Council of Ministers of Education, Canada, 1996, p. 39). The aggregate of Canadian student performance for each age group and performance level was compared to the inter-quartile range in which the expectation percentage was situated (Council of Ministers of Education, Canada, 1996a). For example, the assessment results revealed that the percentage of students for both age groups and on both assessments at level 3 exceeded the expectations set by the panel; “[h]owever, insufficient numbers of Canadian students reached the top levels of excellence” (Council of Ministers of Education, Canada, 1996a, p. 39).

CMEC announced that “these realistic expectations will be used in the next three years as guidelines by the ministries of education when enhancing [science] programs
across Canada” (Council of Ministers of Education, Canada, 1996a, p. 39). Several program initiatives were influenced by the standard setting activities, and these included the development “of regional assessment instruments [for the Atlantic Provinces] that [would] be congruent with the new regional science programs” (Council of Ministers of Education, Canada, 1996a, p. 68). Thus, the establishment of the standards setting panel increased awareness about science performance in Canada, raised expectations about what science performance should be, promoted SAIP as a valid instrument to determine performance level of Canadian students, and encouraged regional cooperation, and curricular homogenization through heightened visibility made possible through the implementation of the SAIP Science – 1996 large-scale assessment and its accompanying report. The science assessment and its report served as a Foucauldian tool designed to make the “taken-for-granted” part of the visible landscape through the production of knowledge. Subsequent acts of power established programs of conduct (regional assessment instruments) designed to influence and direct the behaviour of Canadians.

Like its predecessors, the SAIP Science assessment -1996 included a questionnaire designed to gather data about “student characteristics, school and science course organization, attitudes and practices with regard to science” (Council of Ministers of Education, Canada, 1996a, p. 87). This 58-item questionnaire was significantly longer than its previous iterations, and the questions were organized by category. The data provided not only percentages of students who agreed or disagreed with a particular statement, but it also provided an interpretative narrative for each category of questions that linked students’ responses on particular questions with their performance on the
science assessment. In relation to questions about doing well in science, the following comment was made:

Very high percentages of students in both age groups think that to do well in science you need lots of hard work and studying at home. Those who think so perform significantly better than those who do not. The percentage of 13- and 16-year olds who think the key to success in science is to memorize the textbooks or notes are surprisingly high. (Council of Ministers of Education, Canada, 1996a, p. 88)

Although the relationship between hard work and science proficiency may be well understood by teachers, the collection of a large-scale data set that relate students’ perception about hard work and proficiency science has not been produced prior to administration of SAIP. The questionnaire explored topics such as student confidence in their ability to do science, career choice, teamwork, classroom activities related to learning science, accessing help outside the science classroom, and access to books and computers. Some interesting findings revealed that a “great majority of 13-year olds and 16-year olds feel confident about their overall abilities in science” and “there is a positive relationship with performance at higher levels” (Council of Ministers of Education, Canada, 1996a, p. 89). Although students believed that “science experiments and teamwork contribute more to their interest in science than science applications to daily life ... teamwork appears to be negatively correlated with performance level” (Council of Ministers of Education, Canada, 1996a, p. 90). Field trips, visiting museums, and watching science programs – these activities are often viewed as contributing positively to the overall science education experience; however, the SAIP results indicated that “students who engage in these activities less often show stronger performance than those who do more frequently” (Council of Ministers of Education, Canada, 1996a, p. 97). An emergent question has formed about the value of higher scores on the SAIP Science
assessment versus the opportunity to learn science through experimentation and experience. Did assessment prioritize what society values as useful science learning, or did the assessment marginalize the learning that students take away from field trips and teamwork?

The questionnaire became more sophisticated; assessment and questionnaire were linked to provide an emerging profile of the Canadian student. Although classroom teachers may have already developed an experientially-derived sense of the students they teach, the CMEC assessment was beginning to construct an image of the Canadian composite student and his/her school environment; however, the integration of socioeconomic factors and neighbourhood wealth still remain outside the parameters of the 1996 questionnaire.

The SAIP science assessment was more than a re-manufacture of the existing SAIP assessments in the field of science. The introduction of the expectation setting panel, the expansion of student questionnaire, and the introduction of a science inquiry skills assessment were significant additions to the SAIP protocol. These additions increased the complexity of the assessment products and increased the visibility of the assessments. The decision by the Saskatchewan Ministry of Education to join the SAIP consortium team and to participate in the assessments made the assessment protocol truly pan-Canadian in its design and delivery.

Summary

The first SAIP testing cycle revealed that Canadians could engage in a cross-national educational project, and educators and policy-makers could arrive at some
consensus about some meaningful assessment goals and priorities. As CMEC noted in its first publications, the project, itself, was an experiment in collaboration and cooperation. The first cycle provided opportunities for Canadian educators to engage with their curricula in a different way and created a different form of visibility for Canadian educational systems, the performance of their students, and their curricula.

The product, the SAIP assessment program, was novel as it attempted to employ a criterion-referenced assessment framework to yield comparative results. The structure was complicated but consistent with the complexities of the Canadian federation and its cultural mosaic.

The assessment tool evolved throughout the first cycle. The SAIP Mathematics assessment was in many ways a first draft: It established the framework and structure for future assessments, but limited its scope to the measurement of the academic achievement of students in Canadian school systems. The measurements were presented as absolute measures rather than estimates based on sampling data, and the student questionnaire was simplistic in nature. The Reading and Writing assessment introduced the statistical concepts of confidence intervals, estimates, and statistical difference. It was more difficult to demonstrate real difference on the Reading and Writing assessment due to the introduction of ambiguity of test score estimates. The SAIP Science assessment was a more sophisticated assessment tool as it explored the capacity to measure science inquiry skills, introduced expectations setting panels, recognized the need to include public input in the setting of student expectations, and expanded student questionnaire. The SAIP science assessment introduced additional levels of complexity into the task of measuring the achievement of students in Canadian school systems because the assessment
developers now recognized that the school environment and the student’s background are critical elements in the production of learning. Through the visibility produced by the administration of the SAIP assessments, these elements can now be made available for measurement.

Contextual Canadian Conversations

During the decade, the conversations in the Canadian educational landscape focused on issues of governance, accountability, testing, curriculum development, and the role of the teacher/educator. These conversations were not necessarily focused on SAIP, but they clarified the lines of forces that shaped and influenced the production of education, curriculum, and assessment. At the beginning of the decade the forces shaping thought about the formation and structure of school systems were national in nature and favoured some form of increased centralization in the delivery of education. These discussions also focused on the importance and the mechanisms for bringing Canada’s school systems into alignment as education was viewed as the key means for developing an economically vibrant nation. Just a few years later, the forces shaping the conversations about Canada’s school systems remained focused on the need to improve the school system to ensure the productivity of the Canadian economy. However, commentators moved away from promoting the benefits of a centralized Canadian school system as a solution to Canada’s “failing school systems” toward emphasizing the positive benefits of increased parental choice and decreased bureaucratic control of the educational process.

Educators organized themselves to discuss the importance of assessment in the delivery and management of the learning process. The Joint Advisory Committee
comprised of prominent Canadian educators produced the foundational document, *Principles for Fair Student Assessment Practices for in Canada*, which articulated principles of fairness for both classroom and externally-mandated student assessment (Joint Advisory Committee, 1993). The portion of the document addressing externally-mandated assessment was informed by the American document, the *Code of Fair Testing Practices for Education* (1988); however, the Joint Advisory Committee added a specific section to address “mandated educational assessment and testing programs developed and conducted at the national, provincial, and local levels” (p. 14). The guidelines addressed six points of fairness pertaining to the development and use of externally mandated assessments that include “Inform[ing] all persons with a stake in the assessment (administrators, teachers, students, parents/guardians) of the purpose(s) of the assessment, the uses to be made of the results, and who has access to the results” and “[providing] reports and explanations of results that can be readily understood by the intended audience(s)” (p. 20). While these guidelines did not specifically target SAIP, SAIP was included in the scope and intention of these guidelines. CMEC has published administration guidelines, information about the SAIP testing protocol, and the reports describing the results of the test administrations. However, questions could be posed about the degree to which CMEC has complied with the spirit and intention of these guidelines?

Lawton (1995a) described the Canadian school system of the ‘90s as a “disaster” situated in a cumbersome bureaucracy that should be rescued through a transition to a system that operates as a series of “free-market exchanges” (p. 14). Lawton acknowledged that this thinking was not new or novel but a reinvention of the
relationships that parents negotiated with teachers at the turn of century. The new school system should encapsulate the principle of “competition, choice, alternatives, experimentation, survival and growth” that would “breathe dynamism and meaning into education” (Lawton, 1995a, p. 14). At the heart of these discussions was something more than “teaching methods”; it was philosophical purpose of education, the vision of who children should become, and who should author that vision – the parents or the school system bureaucracy (Lawton, 1995a, p. 18). Lawton’s arguments were based on broad factors: “the perceived failure of the school system,” the perceived “the failure of education to adopt new management practices,” and the growth of education beyond “the ability of society to pay” (Lawton, 1995b, p. 22). Referencing low standardized test scores for numeracy and literacy, Lawton concluded that “one invests time, money and energy educating one’s child and one’s self” (p. 27) If this investment in human capital does not result in a “benefit of both the individual and the community ... people will justly feel like they have been deceived”(Lawton, 1995b, p. 27). The solution to the perceived disaster of schools in the mid ‘90s was charter schools; these schools represented a manifestation of “public choice” (Lawton, 1995c, p. 69).

Robertson (1995) added to Lawton’s concerns about the direction of public education by stating that “education must be part of the solution to our economic problems” (p. 92). Although Robertson did not advocate the dismantling of public education, she added her voice to growing discontent about the need to “solve the problem of unresponsive secondary and post-secondary educational institutions” so that schools can engage their communities” by reducing the negative impacts of bureaucratic structures, narrow academic values and imposed ministerial directives” (p. 93). Although
Robertson did not identify standardized testing and the use of standardized test results specifically, narrow academic values and ministerial directives were often associated with these types of protocols.

While Lawton, Robertson and others were engaged in the philosophical questions surrounding the management and delivery of educational services, Traub (1994) produced a survey of standardized testing practices in Canada. Traub introduced his survey of standardized assessment in Canada by stating the following: “[It is a] propitious time to conduct such as survey as ... the good ship Public Education: Canada [is facing challenges from] increased class sizes, reduced resources, and low, and in some cases negative, salary increases – and in international studies of educational achievement, average test scores for Canadian students that fall below the averages of students in some other industrialized countries” (Taub, 1994, p. 5). In Traub’s study, the perception of negative outcomes for Canadian education was tied to relative unease of the Canadian public about the state of education. The result was call for “increased accountability of educators and the systems that they manage” through the administration of “more testing” (p. 5).

Traub (1994) placed limits on the representational value of his survey data because only 61 percent of the 200 targeted school boards responded. The survey did reveal key findings about teachers and teacher preparation: “educational testing is a neglected subject in teacher training programs in Canada” although commercial standardized test results were used in “making special education decisions” (Traub, p. 16). The survey revealed that “teachers and principals were the main source of opposition” to standardized testing (p. 20); however, the relationship between teachers’
and principals’ attitudes and their level of formal education related to the subject of standardized testing was not explored. Interestingly, the survey results for “Saskatchewan and the Yukon indicated no source of opposition” to standardized testing (Traub, p. 20).

In his concluding remarks, Traub refrained from making conclusions or offering any predictions on the future of school board standardized testing programs: “testing done by provinces and school boards appears ... to be in a transitional phase” (p. 50).

School improvement and the search for indicators that distinguish strong schools from weak schools was a focus for researchers in the 1990s. Maheu (1995) conducted a review of Quebec’s educational indicators that focused on strategies to improve the relationship between “resources invested in a school system and the results obtained” (p. 58). Maheu suggested that indicators are “often obtained through comparisons” (p. 59); comparisons made with another school system or with previous examination years if the instrument remained consistent. Maheu noted that selecting a province for comparison was difficult:

[T]he three provinces with the best results on the [IEAP study on 13-year-olds achievement in science, mathematics and geography] were Alberta, British Columbia, and Quebec. These three provinces also had the lowest graduation rates. Conversely, the province with the highest graduation rate ranked last among the Canadian provinces in science and geography, and next-to-last in mathematics. This leads to the conclusions that when requirements are higher, the diploma is harder to get, but the students acquire more knowledge. (p. 59-60)

The final factor that Maheu explored was the role of visibility through the ranking and publication of test results. Maheu expressed some concern about the publication of results especially by school boards rather than the school system as the public is first and foremost interested in results and less interested in the “factors that may influence the results,” and school administrators may want “to improve their school board’s ranking by refusing to allow their weakest students to write the ministry
examinations” nullifying the effect of the indicator as a motivator for school improvement (p.61). The Quebec ministry has expressed the hope that that school administrators will respond to a low position on the school rankings by taking positive action, “by being more demanding and by better preparing their students” (p. 61).

In the 1990s accountability in education rose to sufficient importance that Canadian Journal of Education devoted an entire issue to the topic. The topic was approached from the perspective of teachers, provincial systems, and school boards.

Bacon (1995) approached the concept of accountability from the perspective of teachers. Accountability was a controversial topic for teachers because “there is no consensus in society about what schools should be responsible for” and while national testing (SAIP) has been deemed “cure-all for those concerned about the performance of Canada’s public school systems ... there are no magic bullets or easy answers to complex questions” (p. 85). In 1995, the CTF speaking on behalf of teachers stated that the “demands for accountability through testing illustrate ... [that] other stakeholders in education do not know what is, or has been happening in our classrooms (p. 86).

According to Bacon, too often large-scale assessments, like SAIP, were comprised of items that “we can easily measure” rather those items that assess what was important. Bacon, on behalf of the CTF, called on the general public to “challenge the assumptions that preparation for competition in the marketplace lies at the heart of education and that schools necessarily perform better if placed in a competitive situation” (p, 86).

Hodgkinson (1995) approached accountability from a provincial and a definitional perspective. In British Columbia accountability has several elements which included the results of national testing (SAIP) and international assessment through participation in
OECD initiatives. According to Hodgkinson, SAIP results were just one of the “external reference points for provincial assessment and indicator development” (p. 22).

In 1995, the Alberta government positioned accountability as a public “right to know if they are getting value for their investment” and proposed an indicator system as a means to “assess the effectiveness and efficiency of the educational enterprise, to improve education, and to provide a mechanism for accountability” (McEwen, 1995a, p. 28). For Alberta, SAIP and other international assessments provided “an external point of reference: a snapshot of how well students from other provinces and countries do on specific tasks at a particular point in time” (McEwen, 1995a, p. 31). In summary, Canada and its provinces and territories have been participating in a “structured accountability” project that was part of a larger “worldwide reform agenda” which began in the 1980s (McEwen, 1995b, p. 1). McEwen (1995b) acknowledged the influence of American literature on Canadian thinking about educational accountability; however, in Canada, like the United States, the goal for education is the production of “an educated citizenry” (p. 2). In 1995 SAIP was new but held out the promise that it would contribute to the provision of accountability by evaluating the effectiveness and efficiency of public education (McEwen, 1995b, p. 5).

These commentators approached the broad philosophical questions about what the business of schools be should, who should determine what that business should be, what responsibility do educators have to communicate the outcomes of school performance to their stakeholders, and how standardized testing should fit into notions of school accountability. These questions were and continue to be large and thorny: Some commentators recognized that standardized test results like SAIP can play a role in
addressing these issues while others focused on the basic philosophical formation of elementary and secondary education.

Cycle 2

The overall goal for the second cycle of administration of the SAIP assessments was to create conditions that would support the confident construction of cycle to-cycle comparisons in response to the key question: Has student performance changed and in what way has student performance changed? Other technical goals related to reviewing the sampling procedures, refining the test instruments, clarifying the reporting process, and reviewing the administration guidelines were a lower priority.

The SAIP Mathematics Assessment – 1997

The administration of the SAIP Mathematics – 1997 assessment was described as a “milestone,” as it was “the first subject to be assessed for the second time” (Council of Ministers of Education, Canada, 1997, p. 107). The assessment instrument remained virtually the same; however, the composition of the review team varied somewhat from the development consortium, and an expectations-setting panel was struck to gather the public’s perceptions about what 13-year-old and 16-year-old students should know and be able do in the subject area of mathematics (Council of Ministers of Education, Canada, 1997). The aggregate performance of 48,000 students was reported in a variety of comparative charts: The Canada-wide sample was compared with the expectations of the panel; the performances of girls, with boys; each jurisdiction, with the Canada-wide sample; and the Canada-wide sample - 1997, with the Canada-wide sample - 1993. A student questionnaire accompanied the assessment; the results of the questionnaire were
reported as percentages accompanied by comparative comments relating students’ aggregate responses to their aggregate academic performance (Council of Ministers of Education, 1997).

CMEC noted in its introduction to its first SAIP assessment re-administration that “Canada, like many other countries, is giving increased attention to education systems and their performance” and that “achievement in school subjects is generally considered a worthwhile indicator of an education system’s performance” (Council of Ministers of Education, 1997, p. 1). The key research question for this administration of the SAIP Mathematics assessment was the following: “Has the achievement of Canadian 13- and 16-year-old students in mathematics changed since 1993?” (Council of Ministers of Education, Canada, 1997, p. 2) This research question was based on the following assumptions: School systems vary; many senior students may not have taken mathematics for two or more years; and the results should be comparable to the 1993 administration (Council of Ministers of Education, Canada, 1997).

To ensure comparability of the results, the review team from British Columbia, Ontario, Quebec, and New Brunswick (F), limited the scope of the revisions to the replacement of only four multiple choice items, the clarification of 20 other items in the mathematics content assessment, and a reduction of the number of test items from nine to six in the problem solving test (Council of Ministers of Education, Canada, 1997, p. 2). The reduction in the number of test items in the problem-solving test and the modification of two additional problem-solving test items left four test items upon which to base the cycle-to-cycle comparison (Council of Ministers of Education, Canada, 1997). Based on the reduction of test items, the testing period was reduced from three (3)
hours to two and half hours (Council of Ministers of Education, Canada, 1997). Other packaging changes to test booklets might have impacted the comparability of the test materials; however, the test developers concluded that the changes to the “administration procedures did not result in major changes in the measurement characteristics of the assessment” (Council of Ministers of Education, Canada, 1997, p. 3). Although individually, the changes appear to be small; cumulatively, their impact may have been significant on student performance.

The sample size - 48,000 - for the 1997 administration, comprised of 26,000 13-year olds and 22,000 16-year olds, was smaller than the 56,000 students who wrote the 1993 mathematics assessment (Council of Ministers of Education, Canada, 1993). The frequency distribution tables provided additional information for each jurisdiction about the proportion of students in any particular sample that scored at each level, and the sample size by jurisdiction was provided but not the proportion of students writing the assessment in relation to the number of eligible students for each age group (Council of Ministers of Education, Canada, 1997). The sampling strategy was not proportional to the population of eligible students in each province, and it may not have been representative of the mean distribution of performances in each province. The following data illustrates the lack of proportionality in the sample construction by jurisdiction: The number of students writing in British Columbia was 1047; the number in Saskatchewan, 974; the number in Prince Edward Island, 647; and the number in Ontario (French and English), 1870 (Council of Ministers of Education, Canada, 1997, p. 118). Although the data was weighted to compensate for the some of the effects of a disproportionate and unrepresentative sampling strategy, it is difficult to determine if the weighting strategy
addressed the complexity of the effects of the differential distribution of socio-economic and school-based factors.

The protocol for student selection in the sampling procedure was also modified: “In 1997, students were selected without exclusions while, in 1993, schools could exclude students before the final sample was drawn” (Council of Ministers of Education, Canada, 1997, p. 4). The consequences of the modification to selection process resulted in “more students ... not meet[ing] the criteria for level 1” and a reduction of the in the “proportion of students in the overall sample that met the criteria for each of the levels” (Council of Ministers of Education, Canada, 1997, p. 4). The effects of inclusion were evident for the Mathematics Content assessment and dramatic for the Mathematics Problem-Solving assessment. More 13-year olds and 16-year olds scored below Level 1 on the 1997 Mathematics Content assessment (13-year olds in the below Level 1 category rose from 6% in 1993 to 10% in 1997, and for 16 year olds the percentage rose from 2% in 1993 to 5% in 1997) (Council of Ministers of Education, Canada, p. 20). Significantly more 13-year olds and 16-year olds scored below Level 1 on Mathematics Problem Solving assessment (13-year olds in the below Level 1 category rose 12.9% to 24.6%, and 16-year olds in the below Level 1 category rose 9.6% to 16.5%) (Council of Ministers of Education, Canada, 1997, p. 21). The proportion of students who were deemed eligible for exclusion in 1993 was not reported, nor was the proportion of students who would have qualified for exclusion from the 1997 assessment. Thus, it was difficult to determine if the decrease in the level of performance was due solely to the revised student selection criteria or was the combined effect of changes to test items, the packaging and presentation, and/or other changes to the student/school selection criteria.
The SAIP Mathematics assessment consortium convened an 89-member expectations-setting panel comprising a similar configuration of “educators and non-educators” as the SAIP Science – 1996 assessment (Council of Ministers of Education, Canada, 1997). The expectations setting panel was asked to provide a pan-Canadian response to the following question about the performance of Canadians students in mathematics: “What percentage of Canadian students should achieve at or above each of the performance levels, as illustrated by the criteria and questions or problems?” (Council of Ministers of Education, Canada, 1997, p. 119). The process for determining the expectations for the mathematics – 1997 assessment was consistent with Science – 1996 assessment; however, the format for the presentation of the expectations differed. The levels of expectations for the mathematics assessment were presented as a range, “the interquartile set of expectations, which represents the middle 50% of panellists’ views” in which “[t]he median is the midpoint” (Council of Ministers of Education, Canada, p. 119). The introduction of interquartile ranges recognized that expectations like assessment scores were estimates that contained a measure of uncertainty.

The mathematics performance results showed significant differences between the level of expectations of the panel and the performance of 13-year olds at Level 3 in mathematics content and in problem solving for 16-year olds at Level 3 (Council of Ministers of Education, Canada, 1997). The magnitude of the discrepancies was described as follows: For mathematics content “40% of 16-year-old students are expected to achieve level 4 performance while 14.5% actually do”; ...”For problem solving ... just over 15% of Canadian [13-year-old] students reached level 3, short of the expected 40%”; and “60% [of the 16-year-olds] are expected to reach level 3, whereas 39.8%
actually did so” (Council of Ministers of Education, Canada, 1997, p. 119). The discrepancy between the expectations and the realized student performance was large; the causes for discrepancies were not part of the study. However, additional research on the relationship between public expectations and the efficacy of testing tools may provide meaningful feedback on the degree to which those testing tools provided relevant information for the public and the degree to which public expectations for the Canadian systems have been satisfied by the process of schooling.

The test results data were aggregated by age, gender, and jurisdiction; and the data were analyzed to determine if significant inter-category differences were present. In addition, cycle-to-cycle comparisons and analyses were conducted to determine if, in fact, there had been any real or substantive changes in student performance over the past four years. Overall, the differences in performance between 13-year olds and 16-year olds were significant at all levels: “[A] greater percentage of older students at the higher levels (3 and 4) and a smaller percentage of younger students at the lower levels (1 and 2)” (Council of Ministers of Education, Canada, 1997, p. 19). This outcome was expected; however, CMEC commented that “these results were obtained through the administration of validate mathematics instruments to the largest samples of Canadian students ever” (Council of Ministers of Education, Canada, 1997, p. 19).

When the data were aggregated by gender, the outcomes for males and females were not statistically different for the 13-year olds; however, the results for 16-year old males were significantly higher “percentages at levels 3, 4, and 5, and a significantly lower percentage at level 1 than females of the same age group” for both the content and problem solving assessments (Council of Ministers of Education, Canada, 1997, p. 24).
Gender-based differences have been recorded on international and national assessments; however, gender-based differences did not appear for the 13-year old group. The factors related to the production of gender-based differences seem to be varied and complex; and these effects have been attributed to differences in teaching and learning styles, the differential response of students to the assessment situation, and the content of the assessment. The data analyses revealed the nature of the problem but did not probe the data for causation.

Like other SAIP assessments, the performance for each jurisdiction and linguistic groups were discussed in relation to the performance of the Canada-wide sample; however, the performance of each jurisdiction was not discussed in relation to its 1993 performance. In general the francophone samples posted either statistically positive differences or no differences in comparison to the mean of the Canada-wide sample. A higher percentage of francophone 13-year-old students in Manitoba performed at level 1; and also for 16-year olds, at levels 1, 2, and 4 on the content assessment (Council of Ministers of Education, Canada, 1997, p. 52). The performance of francophone students in Quebec was both positive and significantly different from the Canada-wide sample for both assessments and at all levels except level 5 (Council of Ministers of Education, Canada, 1997, p. 65). Francophone students in New Brunswick performed at a significantly higher level than the mean of the Canada-wide sample for 13-year olds at level 3 and for 16-year olds at levels 2 and 5 for the mathematics content test (Council of Ministers of Education, Canada, 1997, p. 65). The Nova Scotia francophone sample yielded significant and positive differences in performance for both groups at all levels on the mathematics content assessment and positive differences for 16-year-olds on the
problem solving assessment (Council of Ministers of Education, Canada, 1997, p. 81). While percentages of francophone students in Ontario on each test and at each level were at or below the Canada-wide sample, the data analysis revealed that “For mathematics problem solving, there are significant differences between the performance of this jurisdictions’ francophone students and Canadian performance at all levels” (Council of Ministers of Education, Canada, 1997, p. 59). With the exception of Ontario, students writing in French performed better than their English speaking peers. There are several questions about curriculum, teaching strategies, and the characteristics of French-speaking schools, teachers, parents, and students that remain uncontested by this study.

The results for students writing in English-speaking jurisdictions were mixed. The performance for students in British Columbia was consistent with the Canada-wide sample for mathematics content for all age groups but below the Canada-wide sample for problem solving for 13-year olds at levels 1 and 2 (Council of Ministers of Education, Canada, 1997, p. 39). The performance for students in Alberta was consistent with the Canadian sample for both age groups at all levels with the exception of the statistically positive difference for 13-year-olds at levels 1 and 2 on the content assessment and levels 1, 2, and 3 on the problem solving assessment (Council of Ministers of Education, Canada, 1997, p. 43). The performance of students in Saskatchewan was consistent with the Canada-wide sample for both age groups for the problem solving assessment, but it fell below the Canada-wide sample for the almost all levels on the mathematics content assessment (Council of Ministers of Education, Canada, 1997, p. 46). The performance of students in Ontario fell below the Canada-wide sample for both age groups and at almost all levels. On the surface it would appear that there was less uniformity in the
performance of students from Anglophone schools: Students in British Columbia struggled more with problem solving; students in Alberta seemed to perform consistently better than their peers in other provinces while Saskatchewan students rendered a better performance on problem solving than on math concepts, and Ontario students struggled more than most other students in Canada. With reference to the comparative work of Willms and Raudenbush (1989) on the differences between cross-sectional and longitudinal studies, the jurisdictional differences produced on SAIP’s cross-sectional data collection may be real and valid, or they may be a product of annual variations in student performance data enhanced through the process of statistical analysis - an artefact of the survey techniques. A longitudinal study conducted in comparable schools in different jurisdictions may reveal similar performance differences as the SAIP test results, or the effects of other variables and trends erased by the effects of statistical manipulation of a cross sectional analysis may become evident.

Several jurisdictions reported curriculum revisions activities targeted toward improving their performances on SAIP. British Columbia had recently reviewed its curriculum and “the revisions have been incorporated into Integrated Resource Packages (IRPs)” that contain the learning outcomes, “the content standards for the provincial curriculum” (Council of Ministers of Education, Canada, 1997, p. 37). Alberta had been working with its Western partners to “address initiatives identified by the National Council of Teachers of Mathematics” (Council of Ministers of Education, Canada, 1997, p. 40). Saskatchewan “has devoted considerable effort to reforming its curricula” so that “new courses are consistent with the Curriculum Standards for School Mathematics developed by the National Council of Teachers of Mathematics. Ontario introduced a
new curriculum for mathematics in 1997. While SAIP results raised awareness about the performance of students in other jurisdictions, initiatives sponsored by CMEC to homogenize the mathematics curriculum across jurisdictions were still limited to the adoption of the standards and framework developed by the National Council of Teachers of Mathematics.

SAIP Mathematics – 1997 – Student Questionnaire

Each test booklet for 1997 SAIP mathematics administration included a student questionnaire designed to gather information about “students’ characteristics, mathematics course organization, and attitudes and practices with regard to mathematics” (Council of Ministers of Education, Canada, 1997, p. 96). The information was gathered and summarized; however, the analyses and discussions relating to students’ behaviours, characteristics, or attitudes were limited to questionnaire data provided by students who scored at level 3 or higher on one of the academic mathematics assessments (Council of Ministers of Education, Canada, 1997). Thus, the data analysis was limited to “28.4% of 13-year-olds and 59.8% of 16 year-olds performed [who] at levels 3 and above in this assessment [sic]” (Council of Ministers of Education, Canada, 1997, p. 96).

In considering the following statement: “A large portion of 13-year olds and 16-year olds feel confident when they do mathematics questions. In both cases, there was a positive relationship with performance at the higher levels”; however, the definition of higher levels was not provided (Council of Ministers of Education, Canada, 1997, p. 97). Students also reported that “they learn mathematics most easily when they listen to the teacher; ... [h]owever more students of both age groups who report that they learn
mathematics more easily when working by themselves show stronger performance at higher levels” (Council of Ministers of Education, Canada, 1997, p. 101). Thus, students who performed consistently at higher levels (level 4 or higher) formed a small component of the sample population, and this group of students was working at a level beyond the expected performance level of the assessment and probably beyond the intention of the curriculum. The relationship between the positive impact of independent study and other independent learning approaches and students’ performances at lower levels was identified but not explored.

The questionnaire also explored the degree to which students had integrated computers into completing their mathematics assignments: “13-year olds who use a computer at home and 16-year olds who never use a computer for mathematics show better performance at higher levels in this assessment” (Council of Ministers of Education, Canada, 1997, p. 106). This finding was not surprising; 13-year olds have had more exposure to computers as everyday tools than 16-year olds; no doubt, the impact of computers on learning and mathematics proficiency will change over time. The student questionnaire provided a limited snapshot of a single variable relational analysis with a limited subset of the data.

Summary

The SAIP Mathematics – 1997 assessment offered the first opportunity to consider cycle-to-cycle changes and to compare the performance of Canadian students in mathematics to expectations set by a heterogeneous panel of interested Canadians. The test sample failed to meet the panel’s performance expectations at the higher levels. In
general, the francophone cohort performed better than the Anglophone cohort. The cycle-to-cycle comparative analysis was limited to the Canada-wide samples for 1993 and 1997 although comparing aggregated performances by gender and by linguistic groups would have provided an additional useful layer of analyses. The cosmetic changes to the packaging of the test instrument, the inclusive sampling procedures, and the revisions to key test items may have contributed significantly to the differences in performance levels between the 1993 and 1997 administrations. Given the multiple changes, it is difficult to determine what, if any particular effect, each factor had on the performance outcomes. Although the student questionnaire posed interesting and potentially valuable questions, the range of analyses was limited and focused on students performing at the higher levels. Tasks of re-contextualising analyses of the questionnaire data for students performing below Level 3 was left to the prerogative of the schools and jurisdictions that were concerned about designing and developing curriculum and introducing modifications to instructional delivery to support students performing at level 2 or below.

The SAIP Reading and Writing Assessment - 1998

Although the SAIP Reading and Writing - 1998 assessment introduced some modest changes to the scoring process and the interpretation of the results, the instrument, itself, remained virtually the same as the 1994 version. The SAIP Reading and Writing - 1998 assessment retained the structure, questions, and essay topic of the 1994 assessment with the following stipulations about the relevance of the testing protocol: "School curricula differ from one part of the country to another” however, “[y]oung Canadians in
different jurisdictions ... learn many similar skills” (Council of Ministers of Education, Canada, 1998, p. 2).

Approximately 46,000 students from all provinces and territories comprised the test sample; somewhat smaller number than the 1994 administration, but a larger sample than the 1996 SAIP Science administration. The sample was split between 13-year olds (24,000 students) and 16-year olds (22,000 students); then split again based on the acquisition of their first official language: French (12,000 students) and English (34,000 students) (Council of Ministers of Education, Canada, 1998, p. 3). The two-and-half-hour assessment followed the same administration guidelines as the 1994 administration with similar provisions for student accommodations.

Key changes to the 1998 assessment protocol included the introduction of the review team, the use of bilingual scorers, and the establishment of an 85-member panel tasked with the setting of student performance expectations for reading and writing proficiency (Council of Ministers of Education, Canada, 1998). Although the assessment remained consistent with the 1994 administration, a review team comprised of representatives from “Northwest Territories, Saskatchewan, Ontario, Quebec, New Brunswick (francophone) and Nova Scotia (francophone) came together in April 1997 to review the assessments and prepare them for re-administration. The team conducted a close review of the “assessment statistics and results, advice from statisticians and scorers, and a review of student exemplars” (Council of Ministers of Education, Canada, 1998, p. 3). Although the task of the review team was to provide assurance that the assessments were ready for re-administration, the goal of the review team was to provide credence to CMEC’s publicly stated belief that the assessments were both unique and
inclusive. CMEC described the product of the review team activities as follows: “The result was a uniquely Canadian reading and writing assessments in both official languages,” and the development of the reading and writing assessment materials was the result of the contribution of “educators, students, and members of the public in every province” (Council of Ministers of Education, Canada, 1998, p. 3).

The 1998 administration of writing assessment introduced a new strategy to address “the issue of language comparability,” the assurance that the language in which students completed the writing task was not a factor in the determining their level of performance (Council of Ministers of Education, Canada, 1998, p. 4). The strategy included the use of two teams of bilingual scorers who evaluated the writing samples written in their second language; unilingual scorers, unaware that the bilingual scorers had already evaluated the writing samples, re-evaluated the writing samples (Council of Ministers of Education, Canada, 1998). The two sets of scores provided the data for the development of an inter-rater reliability index, which was not included in publicly-available documents.

To ensure comparability of test results between the 1994 and 1998 administrations, the scoring leaders and scorers from the 1994 administration were recruited and encouraged to form “scoring communities with similar contexts for scoring” (Council of Ministers of Education, Canada, 1998, p.4). Overall, the changes to scoring protocols were designed to support cycle-to-cycle comparisons and to provide greater confidence that the results produced in English and French were indeed comparable (Council of Ministers of Education, Canada, 1998). The review team’s primary “attention was paid to this factor [comparable of results produced in French and

The results were reported using the same structure and format as the 1994 assessment. Consistent with the 1994 administration, confidence intervals were used to illustrate the range of estimate of the students’ performances; the data were aggregated by gender, language, and jurisdiction; and the Canada-wide mean served as the norm (Council of Ministers of Education, Canada, 1998). Increased attention was paid to the comparability of both the French and English versions of the reading passages (Council of Ministers of Education, Canada, 1998). A primary focus on “pedagogical differences relating to differences in language structure” increased as the instrument developers recognized the capacity of these differences to “render comparisons between languages inherently more difficult” (Council of Ministers of Education, Canada, 1998, p. 18).

The Test Results

The test results indicated statistically significant differences in the reading performance of 13-year olds on the 1994 and 1998 assessments: In 1998 a lower proportion of 13-year-olds scored at levels 3 and 4 and a “higher proportion ... score[d] at level 1” (Council of Ministers of Education, Canada, 1998, p. 21). The difference in results for 16-year olds was statistically different at level 1 only; “level 1 has a very slightly larger proportion of students” (Council of Ministers of Education, Canada, 1998, p. 22).
The differences in the performance of both age groups on the writing test were statistically significantly different from the 1994 administration; the confidence intervals did not overlap for levels 1 to 4.

A larger proportion of 13-year-old students in 1998 achieve each of the levels from 1 to 4. ... Over 95% of the younger students now have a grasp of the elements of writing, even if it is uneven or uncertain (level 2). In 1994, a little under two-thirds of the 13-year-olds had at least a general control of the elements of writing (level 3). In 1998, a little over two-thirds have such control. (Council of Ministers of Education, Canada, 1998, p. 22)

The result for the 16-year-old group was similar to that of the 13-year-old group.

A greater proportion of 16-year-old students in 1998 achieve all levels of writing performance except level 5. ... Where almost three-quarters of 16-year-olds in 1994 had at least a general control of the elements of writing, in 1998 fully 85% do [level 3]. (Council of Ministers of Education, Canada, 1998, p. 23)

The question about real and substantive change remains. Students who participated in the writing assessment prepared for the assessment using CMEC-prepared test preparation materials that mirrored the SAIP writing assessment (Council of Ministers of Education, Canada, 1998). Teachers and educators who participated in the SAIP writing assessment in 1994 brought their knowledge and experience about the writing topic and SAIP expectations to the assessment preparation phase. The evaluators were experienced and had developed a refined sense of the criteria (Council of Ministers of Education, Canada, 1998). However, the jurisdictionally-determined sampling strategy determined the degree to which the data represented a particular jurisdiction or the Canadian population (Council of Ministers of Education, Canada, 1998). All these factors could have impacted on the opportunity to produced true representation of the aggregated student performance; however, the level of their impact on the results is difficult to measure.
For both age groups, a larger proportion of female students performed at the higher levels; this outcome was consistent with the 1994 administration of the reading and writing assessment. On a Canada-wide basis, the decrease in the proportion of 13-year olds performing at higher levels might have posed some concern; however, the test developers’ expectation was that more 13-year olds should perform at level 2 – an outcome consistent with the 1998 assessment results.

The introduction of pan-Canadian expectations for reading and writing performance as an interpretative framework was a new initiative for the reading and writing assessment, but an extension of the interpretative framework initiated in the SAIP science assessment. Like the science assessment, an 85-member panel of educators and non-educators was convened at one of three sites to review “all assessment readings and questions, criteria, scoring procedures, and the actual student results to determine the percentage of 13- and 16-year old students who should achieve at each ... level” (Council of Ministers of Education, Canada, 1998, p. 26). Panellists determined the percentages independently; their combined expectations formed the interpretative framework for student performance expectations for each level of each age group (Council of Ministers of Education, Canada, 1998). Generally, students did meet the expectations set for them; however, “more students should have be able to determine and explain complex meaning in sophisticated texts and demonstrate effective and confident command of the elements in writing” (Council of Ministers of Education, Canada, 1998, p. 26). Similar to the role of the science expectations, the expectations for reading and writing in relation to the students’ performance on the SAIP reading and writing assessment were designed to inform and shape curriculum reform in each jurisdiction.
The results for the SAIP Reading and Writing assessment – 1998 were aggregated by jurisdiction but did not include a comparative analysis with the 1994 results. The format of results report was consistent with the 1994 assessment. The estimate of the results for students from British Columbia, Alberta, and Saskatchewan on the SAIP writing assessment was comparable to the Canada-wide norm; results from Manitoba – francophone and Ontario – francophone were above the Canada-wide norm (Council of Ministers of Education, Canada, 1998). The results for Nova Scotia – francophone were below the Canada-wide norm and derived from a census sample of 13-year and 16-year-old performances (Council of Ministers of Education, Canada, 1998). In general the differences in performances between jurisdictions were small with the exception of the Manitoba francophone population and the Northwest Territories. The trend toward consistent test results across jurisdictions might have been a failure of the testing instrument to discriminate between student performances, a growing consistency in reading and writing curricula across Canada, an increase in student experience with writing large scale assessments, or possibly an artefact of sample construction.

*The question that continues to permeate my thought space is this: Are the test results meaningful and indicative of real differences in students’ performances or simply an artefact of the testing environment? And do these differences represent meaningful indicators of students’ future success and the degree to which school systems are able to provide useful education and meaningful learning experiences? (Dagenais, July 25, 2010)*

The SAIP Reading and Writing - 1998 assessment included a student questionnaire, a practice consistent with the 1994 administration; however, the format was different: “four different questionnaires, one in the reading assessment, and one in the writing assessment” presented in either French or English (Council of Ministers of Education, Canada, 1998, p. 102). Each questionnaire included a total of 54 questions;
the 15 background questions were designed to gather demographic data on grade level, participation in an English or French second language learning environment, and the language spoken at home (Council of Ministers of Education, Canada, 1998). The questions used different phrasing from the 1994 version; however, their intention remained the same.

The results revealed that “[f]or both age groups, a slight proportion of students liked reading less now than four year ago” and “the pattern is also similar across the two age groups” (Council of Ministers of Education, Canada, 1998, p. 103). However, a larger proportion of students at both age levels who did liked reading scored at the upper levels on either assessment (students wrote either reading or writing but not both): “More of the students who scored at level 3 and above [on either assessment] say they like reading very much than say they enjoy writing” (Council of Ministers of Education, Canada, 1998, p. 104). Generally, Canadian students demonstrated improvement on the writing assessment, but not the reading assessment, although students seemed to like writing less than reading.

The questionnaire results also revealed a positive relationship between students’ reading test performance and the amount of time they spent reading; however, the exact nature of the relationship was unclear. For example, reading for enjoyment may predict high levels of performance on the SAIP reading assessment or high scores on the SAIP reading assessment may predict more time spent reading for enjoyment; thus, “it is less clear where the cause lies” (Council of Ministers of Education, Canada, 1998, p. 107).

The relationship between the number of books in the home and reading performance was measured in 1994 and again in 1998. The findings in 1998 echoed those
of 1994: “A larger proportion of students who report having a very large number of books in their homes score in levels 3 and above” and “more students who report not having very many books score at levels 2 and below” (Council of Ministers of Education, Canada, 1998, p. 108). In 1994 jurisdictions were urged to augment their libraries as means to ameliorate the negative impact on test scores of students’ having only a few books in their homes; however, it was unclear from the 1998 report if improving the quality of school libraries positively impacted the reading performance for students in schools that performed at lower levels in 1994. Without the benefit of additional research, it is difficult to determine if schools did take any action or if increasing library assets produced a difficult outcome from increased book ownership.

High quality instruction and the role of the teacher in organizing a sound language arts program were perceived as important variables in producing strong student performances. The SAIP questionnaire results reported that “Most students consider having a good teacher to be important to their doing well”; however, “students who score at levels 3 and above are slightly more likely to agree with this statement than lower achieving students” (Council of Ministers of Education, Canada, 1998, p. 113). This phenomenon has opened up a large research space about the significance of the role of teacher and the quality of teacher-student relationship on the production of test scores on large-scale assessments. A common taken-for-granted notion is that students can recognize a “good teacher” and will automatically benefit from the opportunity to learn from a good teacher. However, the SAIP results would seem to indicate that lower performing students may feel that they don’t need or don’t value good teaching or they
may not recognize what constitutes a good teacher or they may have few opportunities to study with a good teacher.

Summary

The SAIP Reading and Writing – 1998 assessment was a re-administration of the SAIP Reading and Writing – 1994 assessment with only a few changes to some of the test items. The significant changes focused on the introduction of the review team, the bilingual scoring teams, the anonymous expectations setting panel, and cycle-to-cycle comparative analysis. The review team and the bilingual scoring teams probably enhanced the validity of the assessments; however, the impact of these initiatives was not evaluated. Although the expectations setting panel was supposed to contribute the interpretative framework and provide an opportunity for greater participation by Canadians outside the field of education, the level of comparative analysis was limited to the Canada-wide sample, and the panel remained anonymous. Thus, the contextual composition of the opinions remains obscured. The cycle-to-cycle comparisons, too, were limited to the Canada-wide sample. The opportunity to conduct a cycle-to-cycle analysis by jurisdiction was missed as well as the opportunity to conduct a relational analysis between the questionnaire data and the test scores. The products of these analyses would have had the potential to inform curriculum innovation and pedagogical change.

The SAIP Science Assessment – 1999

The SAIP Science – 1999 assessment review team composed of members from Saskatchewan, Ontario, Quebec, and Nova Scotia (francophone) introduced some
significant changes to the student questionnaire, the sampling procedures, data analysis, as well as modifying the packaging and facilitating the cycle-to-cycle editing (Council of Ministers of Education, Canada, 1999a). The SAIP Science – 1999 assessment marked the first time that Nunavut participated as a separate jurisdiction (Council of Ministers of Education, Canada, 1999a). The data analysis added cycle-to-cycle comparative analysis of students’ performances by jurisdiction.

The assumptions upon which the academic science – 1996 assessment instrument was constructed and administered remained unchanged. However, the *SAIP Science Assessment Framework and Criteria* were deemed to be reflective also of the intentions of two more recent publications: the *Common Framework of Science Learning Outcomes* and the pamphlet, *Science Literacy and the World of Work* published by the Conference Board of Canada (Council of Ministers of Education, Canada, 1999a). The *Common Framework of Science Learning Outcomes* document, a product of a CMEC-sponsored Pan-Canadian project in science curriculum development, emphasized CMEC’s growing involvement in Canadian curriculum development. The reference to Conference Board of Canada’s publication drew attention the relationship that CMEC had been fostering with business/industry.

To ensure cycle-to-cycle comparability of the results of the Canada-wide sample and jurisdictional samples, the review team focused on these concerns: “Scoring procedures and conditions as well as administration procedures were replicated as much as possible from the documentation and information provided by the previous team” (Council of Ministers of Education, Canada, 1999a, p. 7). The research questions remained consistent with the SAIP Mathematics 1997 and SAIP Reading and Writing
1998: “How well have Canadian 13- and 16-year-old students learned science in 1999?” and “Has the achievement of Canadian 13- and 16-year-old students in science changed since 1996?” (Council of Ministers of Education, Canada, 1999a, p. 2) Other revisions included minor changes to formatting, some clarifications to the wording and the repackaging of the test booklets into one unit (rather than multiple booklets) (Council of Ministers of Education, Canada, 1999a). Unlike the SAIP mathematics assessment – 1997, test items were not substantively revised or deleted.

The sampling procedures were clarified, and the description of composition of the sample was enhanced. The total sample for both 13- and 16-year olds was 31,000 students - 6,000 fewer students than the SAIP science assessment - 1996 (Council of Ministers of Education, Canada, 1999b, p.2). The composition of the sample included 16,000 thirteen-year-olds and 15,000 sixteen-year-olds, and “[a]bout 22,000 completed the science assessment in English, 8,500 in French” (Council of Ministers of Education, Canada, 1999b, p.2). The sampling procedure was described as follows:

The sampling scheme was designed to yield representative samples of students to permit separate reporting for each population. ... It is important to note that the sampling procedure was designed to yield a representative sample of students in each of the 18 population groups identified. For large jurisdictions an initial random sample of schools was selected, and for smaller jurisdictions all schools having students in the relevant age groups were selected. The sample can therefore be said to representative in all jurisdictions. (Council of Ministers of Education, Canada, 1999b, p. 2)

In addition to highlighting the importance of sampling procedures to the validity of the survey research results, the behavioural and attitudinal survey included demographic questions about schools, teachers, and students to gather information about school environments that could useful in subsequent meta analytical studies (Council of Minister of Education, 1999b). The sampling procedures did not address a particular policy on
inclusive sampling like the SAIP mathematics – 1997 assessment; therefore, it is difficult
to determine if inclusive sampling became a given after the mathematics assessment or
was dropped due to perceived negative effect on the test results.

The results for the SAIP science academic assessment were presented in a similar
format as other SAIP assessments and consistent with SAIP science - 1996. The Canada-
wide sample was aggregated by age, gender, and language; and comparisons were
formulated within each category as well as the cycle-to-cycle comparisons with the 1996
performance data (Council of Ministers of Education, Canada, 1999a, p. 27). A
comparative analysis of the expectations set by expectations setting panels in 1996 and
1999 expectations was conducted (Council of Ministers of Education, Canada, 1999a, p.
27). The addition of cycle-to-cycle comparative analysis for both jurisdictional and
expectations data added a new dimension to the data analysis.

Results for both the written and the practical tasks assessments aggregated by age
illustrated that: “there are more students from the 16-year old population at higher levels”
... With this data, what once would only have been an expectation can now be stated with
some certainty” (Council of Ministers of Education, Canada, 1999a, p. 20). Although this
was the sixth assessment and the second science assessment to demonstrate this outcome,
CMEC has continued to reiterate the importance of this finding.

In 1999 the comparisons for the Canada-wide samples by age group revealed that,
“nearly three-quarters of 13-year olds were able to reach level 2” and “[o]ver 76% of 16-
year olds reached level 3” on the written assessment, and about 90% of 13- year olds and
95% of 16-year olds reached level 2 on the practical skills assessment (Council of
Ministers of Education, Canada, 1999a, p. 21). The difference in performance by age
group was the proportion of 16-year olds that were able to reach levels 3, 4, and 5 “where criteria require the demonstration of considerably more sophisticated skills” (Council of Ministers of Education, Canada, 1999a, p. 21).

The cycle-to-cycle analysis revealed some statistically significant increases in the proportion of both 13-year olds and 16-year olds at some of the performance levels. A significantly higher proportion of both age groups reached levels 3, 4, and 5 in 1999 written assessment (Council of Ministers of Education, Canada, 1999a). CMEC stated that “[t]his demonstrates a general increase in the sophistication of science understanding by Canadian students in the period 1996-99” (Council of Ministers of Education, Canada, 1999a, p. 21). A larger proportion of students from both age groups performed at higher levels on the practical assessment; however, the “differences may reflect changes in the scoring process” as the “criteria ... were more clearly defined” (Council of Ministers of Education, Canada, 1999a, p. 22). The minor changes to the scoring criteria illustrate how some minor changes to a large-scale assessment can possibly lead to the production of favourable or unfavourable measurement of educational outcomes, yet the preparation of the students prior to the assessment may be the same.

The 1999 Canada-wide sample for both assessments aggregated by gender revealed few statistically significant differences: “The overall message given by this data suggests that the efforts to make science education more relevant to, and more inclusive of, young women continue to have a positive influence on science education” (Council of Ministers of Education, Canada, 1999a, p. 21). Differences in performance revealed by a gender-based analysis were minimal.
The analysis of the Canada-wide sample aggregated by age and by language yielded some differences in performance on the written assessment. For the 13-year old sample, “slight differences at levels 1, 2, 4, and 5 in favour of those who wrote in English”; however, the meaningfulness of these slight differences was not probed for causation (Council of Ministers of Education, Canada, 1999a, p. 25). For the 16-year old sample, the differences were larger and significant “at levels 1, 2, and 3 in favour of those who wrote in French” (Council of Ministers of Education, Canada, 1999a, p. 25). The results on the practical assessment revealed a slight difference at level 2 favouring 16-year-old francophone students (Council of Ministers of Education, Canada, 1999a). Because the sampling strategy defined French immersion students as francophone, their test results were included in the francophone sample (Council of Ministers of Education, Canada, 1999a). The decision to comingle francophone and French immersion students into one jurisdiction may have depressed the scores of 13-year olds writing in French due to weaker vocabulary and reading skills of French immersion students.

The 1999 SAIP science assessment stuck a 93-member panel to set expectations for science knowledge and skills for Canadian students (Council of Ministers of Education, Canada, 1999a). The task, process and expected outcomes for the panel were consistent with other assessments. The performance of students on both assessments generally met the expectations of the panel with a few exceptions. The panel believed that “more students should be able to achieve at the higher levels, demonstrating relatively sophisticated science knowledge and skills”; more 16-year olds should reach level 5 and more 13-year olds should reach level 4 (Council of Ministers of Education, Canada,
1999a). Previous panels have expected more students to perform at the higher levels; the causes for this pattern of overstated expectations were not explored.

Jurisdictional analysis provided feedback on the performance of individual jurisdictions in comparison to the Canada-wide sample and with each jurisdiction’s performance on the 1996 assessment (Council of Ministers of Education, Canada, 1999a). The comments described statistically significant differences and the nature of those differences. The following statements for British Columbia illustrate typical statements for jurisdictional performance on written assessment.

[S]tudents performed as well as or better than Canadians as a whole. Slightly more 13-year-olds reached levels 1 and 3 than the Canadian average. The performance of 13-year old British Columbia students showed significant improvement between 1996 and 1999 at levels 3 and 4 while the performance of British Columbia 16-year-old students was significantly better in 1999 at level 3. (Council of Ministers of Education, Canada, 1999a, p. 39)

Saskatchewan was one of four jurisdictions that chose to administer the Practical Tasks assessment to a sufficiently large sample that would support a jurisdictional analysis of the data (Council of Ministers of Education, Canada, 1999a). The following statements for Saskatchewan and British Columbia illustrate typical statements for a jurisdictional analysis of students’ performance on the practical tasks assessment.

Saskatchewan students in both age groups performed at least as well as the students in the Canadian sample, except that slightly fewer Saskatchewan 16-year old students reached level 4. The performance of 13-year-old Saskatchewan students in practical tasks showed significant improvement between 1996 and 1999 at levels 4 and 5, while the performance of Saskatchewan 16-year-olds students was significantly better in 1999 at levels 3, 4, and 5. (Council of Ministers of Education, Canada, 1999a, p. 46)

With the creation of the new territory, Nunavut, on April 1, 1999, the provision of education for the children of the Eastern Arctic became the responsibility of the
government of Nunavut. To support the goals of adapting curriculum and developing programs to meet the needs of students in Nunavut, the Nunavut Department of Education determined that the “[d]ata were [to be] collected separately, in order to report baseline results for each territory” (Council of Ministers of Education, Canada, 1999a, p. 79). The role of English-as-a-second-language factors and cultural factors played a large role in the performance of the students in Nunavut: “At the time SAIP tests are being administered, most 13-year olds in Nunavut are completing their second year of formal instruction in English” (Council of Ministers of Education, Canada, 1999a, p. 79). The following statements for Nunavut illustrate typical statements for jurisdictional performance on written assessment.

There are significant differences between the performance of Nunavut 13-year-olds and Canadian students overall at levels 1, 2, 3 and 4 in the written assessment. Nunavut students performed as well in this category as students in the Canadian sample at level 5. There are significant differences between the performance of Nunavut 16-year-olds and Canadian students at all levels in the written assessment. (Council of Ministers of Education, Canada, 1999a, p. 82)

The jurisdictional performance statements provided an overview of each jurisdiction and its performance relative to the Canada-wide sample and with its performance in 1996. Because the overall performance of the Canada-wide sample demonstrated improvement, most jurisdictions also posted improved performance. A larger proportion of Alberta students posted performances at higher levels (Council of Ministers of Education, Canada, 1999a), and although the performance by students on the French version was strong, their performance lagged behind that of students in Alberta. The Atlantic Provinces were moving toward a common curriculum in 1999 which may or may not be responsible for stronger performance by these jurisdictions.
The remaking of the student questionnaire for the SAIP Science - 1999 assessment was much more than the introduction of a few new questions or the refining of other questions to target specific information collecting goals. The revisions recognized that “[l]earning is a complex process, affected by many factors within the student background and experience, school and classroom conditions, resources, motivation, quality of schooling and teaching, attitudes and expectations” (Council of Ministers of Education, Canada, 1999b p.1). These revisions returned SAIP its original vision as a “comprehensive indicators program” (Council of Ministers of Education, Canada, 1999b p.1). The scope of the questionnaire was increased to capture issues and factors affecting student success in science from the perspective of principals, science teachers, and students writing the SAIP science assessments (Council of Ministers of Education, Canada, 1999b).

The participant sample included the students writing the science assessment as well as 6,500 science teachers and 2,000 principals representing each of the schools in which students wrote the science assessment (Council of Ministers of Education, Canada, 1999b, p. 2). According to CMEC, the sample was large; therefore, it was “unlikely that any serious bias exists in that sample” (Council of Ministers of Education, Canada, 1999b, p. 2). The sampling procedures suffered from the complexities introduced from the need to make the samples statistically valid in each jurisdiction while at the same time not oversampling in the larger jurisdictions; thus, the samples might not have been representative of jurisdictional populations and the Canadian population. The
exclusion rates were controlled at the school level which introduced a level of basis external to the sampling methodology.

The goal of the survey was to link questionnaire analysis of the responses on three questionnaires (student, teacher, and principal) “with achievement levels of students ... to examine how contextual factors are related to achievement” (Council of Ministers of Education, Canada, 1999b, p. 2). Although the intention of the survey was to draw connections between contextual factors and student performance, the survey report fell short of this goal. Instead the report urged other researchers to use this foundational work to “stimulate discussions about important features of our schools, teachers, and students” (Council of Ministers of Education, Canada, 1999b, p. 2).

CMEC adopted the conceptual framework put forward by Wang, Haertel and Walhberg (1993) that defined learning as a complex interaction of three significant elements: inputs, process, and outcomes (Council of Ministers of Education, Canada, 1999b). Using the Wang, Haertel and Walhberg conceptual framework, the questionnaire developers framed questions for each target audience – students, teachers, and principals – that corresponded to the elements in the theoretical framework (Council of Ministers of Education, Canada, 1999b).

Principals were asked to respond to questions on school demographics, student characteristics, class size, and computers and their use. The results of survey indicated several demographic trends that seemed to follow an East-West alignment; for example, larger schools were located in Quebec, Ontario (English), and British Columbia while small schools were located in other jurisdictions (Council of Ministers of Education, Canada, 1999b). However, principals did not follow an East-West split when they
reported the degree to which external assessments such as SAIP influenced the delivery of instruction: “Saskatchewan, Nova Scotia, and Prince Edward Island ... reported little influence,” and “Alberta, Manitoba, Quebec, and New Brunswick (English) “reported a lot of influence” (Council of Ministers of Education, Canada, 1999b, p. 7).

Science teachers were asked to complete a 31-item questionnaire about their “professional background and experience, teaching assignments and duties, class sizes, interaction with parents and other teachers, lesson planning, classroom activities, resource use, constraints on teaching, homework, and student evaluation: (Council of Ministers of Education, Canada, 1999b, p. 24). Several of the categories on their questionnaire mirrored those presented to the principals. As an example, both teachers and principals were asked to indicate the amount of instruction time lost due to class cancellations (Council of Ministers of Education, Canada, 1999b). The perceptions of the teachers paralleled those of the principals and produced the same East-West trend “with more time being lost in the Eastern provinces and Nunavut” (Council of Ministers of Education, Canada, 1999b, p. 24). Teachers were also asked to respond to questions that targeted instructional practices such as classroom activities: Although several different types of learning activities were presented on the questionnaire, “more than 70% of teachers in almost all jurisdictions report[ed] frequent use [of]... note giving” (Council of Ministers of Education, Canada, 1999b, p. 25). Teachers were also asked to respond to some policy issues such as streaming in secondary science education” (Council of Ministers of Education, Canada, 1999b). Although most science teachers supported some form of streaming in the organization and delivery of science education, the strength of this belief
varies across Canada “from a low of less 60% in British Columbia and Saskatchewan to a high of more than 90% in both Ontario French and Ontario English populations.

After completing the SAIP science assessment, students completed a 27-item questionnaire about their “home backgrounds, educational and career aspirations, perceptions of school and science, out-of-school activities, attributions for success and failure, and classroom practices” (Council of Ministers of Education, Canada, 1999b, p. 41). Students responded to background demographic questions about where they were born and the congruence between the language spoken at home and at school: “[O]nly Ontario (English), Quebec (English) and British Columbia have more than 10% of their students born outside Canada”; however, the language gap is much larger among francophone populations outside Quebec (Council of Ministers of Education, Canada, 1999b, p. 41). Students were asked to reflect on the importance of others – parents, friends and teachers – in their doing well in school and in science: “Generally, a high proportion [of students] reported that parents thought it was important or very important for them to do well in school” while the “opposite belief was true for friends (Council of Ministers of Education, Canada, 1999b, p. 41). Students’ interactions with parents were explored through a series of questions: More 13-year olds reported “working with their parents on science homework a few times a month” while more 16-year olds “reported discussing their futures with their parents” (Council of Ministers of Education, Canada, 1999b, p. 44). Students were asked to provide their perspective on the types and prevalence of the classroom activities: For students, “note giving is a highly prevalent activity” (Council of Ministers of Education, Canada, 1999b, p. 44).
The introduction of the Canada Context questionnaire for students, teachers, and principals “marked another milestone in the School Achievement Indicators Program” (Council of Ministers of Education, Canada, 1999b, p. 66). The data was gathered from the administration of a theoretical-based survey instrument tailored to gather information from three key stakeholder groups engaged in different ways with education and with different education systems across Canada. The analysis of the survey remained descriptive as illustrated in the examples above and did not attempt to relate the strength of variables in the questionnaire with the level of performance on the SAIP science assessment. The report did invite researchers to use the data and conduct further analyses of the data.

The Evolution of SAIP over the Decade

At the beginning of the decade, CMEC marvelled at the possibility that most of the provinces and territories could collaborate to develop, pilot, and administer an assessment program to a large sample of 13-year-old and 16-year-old students, and report the findings to Canadian stakeholders. The premise for the initial assessment stated that the performance of Canadian students on an assessment based on content derived from commonly agreed elements of Canadian curricula could serve as a proxy for school performance. Although the initial assessment was based on an agreed upon framework, a theoretical conceptualization of learning did not undergird the statements about student proficiency. The assessment did evolve throughout the decade. By working within the structure of the initial assessment, cycle-to-cycle comparisons remained a possibility. The first cycle of testing introduced the concept of confidence intervals to reflect the notion of estimation, expectation setting panels to gather stakeholder input, and key participant
(students, teachers, and principals) questionnaires to gather information about the school environment and non-school factors. In the second cycle of testing the definition of learning expanded to acknowledge the complexity of the process. CMEC emphasized the importance of the SAIP’s capacity to provide cycle-to-cycle comparisons, and the SAIP review teams focused their attention on ensuring the comparability of test results between French and English jurisdictions. Most notably was the attention paid to the theoretical constructs that underpin the conceptualization of learning as influenced by three key elements: inputs, process, and outputs. Rather than focusing solely on the measurement of student outputs through the administration of an academic assessment, the design of the questionnaires focused on inputs from key stakeholders (teachers and principals), process (teachers and students) and outputs (students). The expectations-setting panel and the subsequent presentation of comparative data acknowledged the political nature of the assessment process and invited participation by Canadians from a variety of backgrounds who had a stake in the performance of the Canadian educational systems. Overall the definition of learning expanded throughout the decade, and the introduction of and the emphasis placed on the questionnaires illustrated the desire of CMEC to develop an indicators program that moved beyond measurement of academic outcomes of students in Canada. From a Foucauldian perspective, the statements that defined what constituted learning evolved with time. They have become more complex as the capacity to assess and measure the multiple variables that influence the learning process has increased.
Conversations as the Decade Ends

As the decade ended, conversations continued to focus on the cost-value aspect of educational services; scholars and commentators refocused their research lens to formulate meta analyses of the data produced by SAIP and other large-scale assessments; and CMEC initiated other student-related activities to measure the educational outcomes and the performance Canadian students. Commentators such as O’Sullivan (1999) were beginning to analyze the actions of CMEC in relation to current economic and educational theories. Against this backdrop of seemingly practical issues, some commentators tackled the thorny issue of the relationship between the re-construction of a Canadian world view of education in relation to educational reform and large scale assessment results.

According to O’Sullivan (1999), Canada had been in some form of educational reform for more than a century; however, in the 1990s the influence of these two paradigms shaped educational reform: the first asserted that “knowledge is a competitive asset and advantage of industrial nations” and the second asserted that knowledge should be shared as part of “interdependent global needs and responsibilities” (p. 311). O’Sullivan’s descriptions of the forces shaping educational discussions differed somewhat from Manzer’s (1994) focus on forces that influence the production of curriculum. O’Sullivan re-thought Manzer’s theoretical construct of content-focused curriculum with externally-administered evaluation into a construct that views curriculum and knowledge transfer as a competitive asset. O’Sullivan transformed Manzer’s student-centred curriculum with teacher-constructed assessment into a construct that viewed curriculum and knowledge transfer as an interdependent global need or responsibility.
Although the words seem to be quite disparate from the original dichotomy, provinces, like Alberta, that embraced content-focused curricula have also engaged in aggressive competition in the international and national testing areas (Young and Levin, 1999). These provinces have continued to support an educational reform agenda based on “choice and competition as being desirable foundations for education policy” (Young and Levin, 1999, p. 8). At the end of decade, provinces (like Saskatchewan) that had embraced a student-centred focus were less likely to participate in international assessments, viewed the delivery of education as a social responsibility, expressed concern about issues of differentiated learning, and prioritized educational spending based on issues of inclusion.

O’Sullivan (1999) described CMEC’s priorities related to education and educational policy as leaning philosophically toward the view that education was a competitive asset of industrialized nations. In the ‘90s, CMEC held discussions with national corporations, telecommunications, and high-level technology industries to “identify the kinds of knowledge and employability skills” required by high school graduates (p. 318). These types of consultations served as evidence of the “connectedness of the national [educational] achievement to global economic competitiveness and content-focused curriculum reform” (O’Sullivan, 1999, p. 318). CMEC’s national agenda was moving toward the global competitive curriculum model; however, everyday Canadians expected their educational system to be “pragmatic” and “utilitarian” and supported mostly technical reforms. By “instituting provincial and national assessment (SAIP), subscribing to national learning targets in education and centralizing educational governance and power,” CMEC was influencing Canadians’ worldview about the role of
education in the global competitive market (O’Sullivan, 1999, p. 321). The Canadian public debate on education in the ‘90s, according O’Sullivan, did not explore and analyze the potential problems that educational reform based on “global economic competitiveness” may wreak for students and society (p. 321).

Commentators towards the end of ‘90s focused on the condition of Canadian educational policy, the state of the Canadian educational systems and the need to propose solutions to “fix” the seemingly broken Canadian educational systems. Schweitzer, Crocker and Gillis (1995) presented two arguments for increasing the performance of Canadian schools: the need to develop human capital to fuel a growing and technologically-savvy economy and the need to increase the earning potential of all Canadians. These authors supported the need to improve Canada’s global competitiveness through improvements to the jurisdictional educational systems. The authors argued that keeping students in school until graduation promoted learning and that improving standardized test scores on mathematics and science would positively influence students’ earning potential and the strength of the Canadian economy. Conversely, Payette (1997) argued that the real benefit for recent graduates was much lower than for older workers entrenched within the labour market and questioned the desirability and morality of “want[ing] all, or nearly all, students to complete a mostly in-school secondary program” (p. 93). However, the in-school programs prepared students to write large-scale standardized test assessments, not alternative programs that integrated theoretical instruction with hands-on industry experience. From Schweitzer et al’s (1995) perspective, the high level of spending on education in the ‘90s gave Canadians the “right to expect unusually good performance from students”; this expectation, according to
Payette, is “both obvious and important” (p. 94). Commentators in the ‘90s were troubled with these issues: the nature of a high performing schools system, the characteristics good school system performance, the relationship of high scores on a standardized math or science assessment and/or employment in a satisfying but well-paying job to positive school system outcomes, and the relationship between high standardized test scores and well-paying and satisfying employment.

Embrouled in Schweitzer et al’s (1995) discussion of education costs and the achievement as measured by strong test scores on national and international assessments were the issues of educational de-streaming and differentiated learning. According to Schweitzer et al, the cost implications related to raising the “achievement level of the weakest 10% of students by five percentage points [might be equivalent to]... the cost of lowering the achievement of the top 10%” (p. 135). Schweitzer et al have framed the equity argument as a trade off; however, Payette (1997) reminded Canadians that “[E]quity matters in public education – and will continue to matter” (p. 96). Schweitzer et al questioned the reasonableness of the Canadian expectation that schools could have the capacity to offer programs which could be legitimately described by the phrase: “Excellence for all” (p. 104). Schweitzer et al and Payette raised several critical questions about educational priorities and their relationship to large-scale testing programs and the public’s expectations about the performance of Canadian schools. They questioned the importance of investing additional resources in the raising test scores? If raising test scores is critical, should the weakest students be the priority or should the priority be placed on the strongest students.
Following the launch of the first cycle of SAIP assessments, Willms (1997) conducted a meta analysis of National Longitudinal Study of Children and Youth – 1993 (NLSCY) in which he compared the literacy levels of 4 and 5 year-old children with parental SES gradient for each province. According to Willms, provinces differ dramatically in their SES gradients and that provinces with flatter gradients, like Quebec, provided better learning outcomes for its children while provinces, like Ontario and the Atlantic provinces, with steep gradients (large range in SES) provided poorer outcomes for their children. According Willms, children from lower SES homes were more likely to perform at a higher level if they attended schools in which the students came from a variety of SES backgrounds. Students attending Quebec schools performed better on SAIP; whereas, students in the Atlantic Provinces tended to do less well: Willms tagged this phenomenon as the “Quebec advantage” (p. 24).

The Quebec advantage as evidenced on SAIP, NLSCY, PISA and IALS seemed to be consistent and reproducible on a variety of studies; however, Canadian research studies designed to explore and describe the so-called Quebec advantage have not emerged. Researchers have not probed the causality the SES gradient in the production of the Quebec advantage. However, Willms (1997) complicated the tidiness of the relationship between the SES gradient and the Quebec advantage by introducing another variable: the school added-value factor. According to Willms, research has also shown that “large and significant differences among schools in their outcomes, even after taking into account students’ family background characteristics; ....schools differ in their ‘added value’”(p. 24). While Willms identified the importance of these two critical factors and their variable influence on student performance, he did not offer a research approach to
explore the nature and quantification of the factors operational in the Quebec advantage or the school value-added factor.

CMEC initiated the first steps towards the development of a Pan-Canadian Educational Research agenda with the call for papers on topics pertaining to “the increasing importance of education for individual and societal well-being, the need for accountability ... increasing efficiency ... and improving quality (Schultz, Clark & Crocker, 1998, p. 1). Schultz et al (1998) focused on the process of developing of learning outcomes, their role in defining assessment strategies that support broad-based indicator systems, and the political nature of the public consultation and validation process. In the ‘90s, several jurisdictions were moving away from “broad philosophical [curriculum] statements to statements expressed in outcome form” as these statements “represent the first step in defining what needs to be measured” as illustrated in the Common Framework of Science Learning Outcomes (1997) (Schultz, Clark & Crocker, 1998, p. 1). Based on the experiences of Manitoba, Quebec, and Alberta, some form of public consultation formed an “essentially political process” for identifying outcomes; however, according to Schultz et al, such processes were often “controlled by ‘insiders’ suggest[ing] that public views will not be fully represented” (p. 3). A critical issue formed around the need to identify categories of educational outcomes for which “the public would support testing,” and the development of a national indicator system (Schultz, Clark & Crocker, 1998, p. 4). The enhancements to the SAIP Science assessment – 1999 was an example of “much more comprehensive and representative data bases than had previously existed,” and based on the TIMSS conceptual model, “the inclusion of student, teacher and school questionnaires in SAIP will, for the first time,
allow the investigation of the links between science achievement and a large number of input and process variables” (Schultz, Clark & Crocker, 1998, p. 8). Clearly research questions had moved beyond students’ academic performance to the need to pinpoint those variables in the teaching and learning process that impacted on student outcomes.

Coincidental to CMEC’s efforts to frame its research agenda, Bishop (1998), an American Department Chair of the Department of Human Resources at Cornell University was researching the effect of exit exams on student achievement of “Canadian 13-year olds, their parents, teachers and school administrators” (p. 12). His meta analysis of the International Assessment of Educational Progress (IAEP) 1990-91 data for several countries including Canada yielded a comparative analyses of the performance of provinces with and without exit exams (Bishop, 1998). According to Bishop, “[p]rovincial exit exams had large effects on achievement: 19 percent of a U.S. standard deviation (about four-fifths of a U.S. grade-level equivalent) in mathematics and 13 percent of a standard deviation (about half of a grade-level equivalent) in science” favouring the students in the provinces with exit exams (p. 12). Bishop included 1,338 Canadian schools from several Canadian provinces and his experimental design measured variables such as language usage, the number of books in the home, and organization of the school; however, the design did not measure variables such as SES and rural-urban influence. Coincidentally, students in Alberta and Quebec must take provincially mandated middle-stakes exit exams, and they have consistently outperformed their peers in other Canadian provinces on national and international assessments.

Robertson (1999) stated that education is “political” because it “distributes power – or hoards it, or makes it invisible,” a key Foucauldian concerns (p. 1). For youth,
education and the value attached to particular types of education will determine whether they as future adults will participate in society or cope with it (Robertson, 1999). From Robertson’s perspective, what happens in the classroom is critical to the value of the educational experience to the student, the teacher, and society. Robertson addressed the issue of the classroom, the teacher, the curriculum, and the impacts on learning; the teaching and learning environments in Alberta and British Columbia provided the data. Although Bishop’s analysis yielded evidence that students in Alberta outperformed their peers in provinces without exit exams, Robertson’s study revealed that the curricular costs were high – the number of field trips reduced, the curriculum narrowed to follow examination topics, and the opportunities for library research reduced. The SAIP Science – 1996 assessment results revealed that field trips did not correlate positively with increased performance on standardized assessments. The results of the SAIP science – 1999 assessment, Robertson’s review, and Bishop’s study revealed a complex relationship between the mode of instruction and standardized testing activities; this relationship has implications for educators and policy makers engaged in goal setting and policy making activities.

Like Bishop, Robertson also reported that the presence of school leaving exams “affected the students in lower grades through increased school-wide tests, increased emphasis on test-taking skills, and increased attention to subject matter associated with the Grade 12 examination” (p. 3). The SAIP assessments, while less invasive on the day-to-day delivery of learning experiences, impacted curriculum and student learning through the development of Pan-Canadian frameworks to underpin the development of jurisdictionally-specific curriculum documents (Robertson, 1999). Robertson noted the
following: “The SAIP results had provided a convenient excuse to increase the space and priority that certain subjects would receive in testing and the curriculum” (p. 4).

Individual jurisdictions had begun to respond to the SAIP results. Ontario reviewed and revised its elementary and secondary mathematics curricula; the revised curricula (1997) “describes expectations for knowledge and skills, as well as assessment criteria for each grade level from Grades 1 to 8 in 5 mathematics strands” (Education Quality and Accountability Office, 1997, p. 3). Several jurisdictions participated in the development of the pan-Canadian Common Framework of Science Learning Outcomes (Fournier, 2000). Fournier, coordinator of the national testing program, described the development of the Pan-Canadian science curriculum as an “important project under the management of CMEC,” but it did not bind “jurisdictions as to what is taught in the classroom” (Fournier, 2000, p. 549). Fournier attributed the adoption of the Pan-Canadian science framework by 11 jurisdictions as “testament to the quality of the document” rather than any “impetus by SAIP to standardize curriculum” (p. 549). The curriculum revisions initiated by Ontario, the first since 1985, and the development of the Pan-Canadian framework for Science may not have been causally related to the introduction of SAIP; however, the visibility created by SAIP, the interprovincial discussions and collaboration, and the human desire to do better created an environment that supported the homogenization and standardization of acceptable best practices.

The Canadian Teachers’ federation responded harshly to the introduction of externally-derived expectations-setting and described it as a “process so bereft of validity and reliability” (Robertson, 2000, p. 549). Robertson (1999) criticized the publication of
the expectation percentages without confidence intervals for the Science - 1996 and Mathematics – 1997 assessments:

To put it mildly, this would be more than a minor glitch in most research projects; these numbers continue to be presented to the public as if they had meaning ... surely this proves that this exercise is intended to manipulate rather than to inform. (Robertson, 2000, p. 549)

An educator participant involved in the expectation setting process offered these comments about SAIP and its role in the educational process: “Do we want our testing to determine what our students should learn? ... The tests themselves do not give us an indication as to where and how we can improve the process of teaching” (Robertson, 2000, p. 549). The perspective of the classroom teacher and the student were often missing from the discussions about test scores, the meaning of the test scores, and the discussions about the nature of the really important business of schools and the role of learning in Canadian society.

The bipolarity of foundational beliefs about the purpose and role of education, the function of SAIP in policy making, and education improvement programs continued to shape the discussions and actions of the policymakers, ministries, and teachers. Concerns about growing the Canadian economy and building a skilled workforce were set against the need to take responsibility for the education and integration of the weakest students into the workforce. At the end of decade, evidence of growing uncertainty was evident about the role of locally-developed curriculum, the value of teachers as instructional managers and content assessment experts, and the importance of large-scale assessments as agents of visibility and coercion in the re-visioning of provision of education.
Cycle 3: Re-Thinking, Re-Visioning, and Re-Inventing SAIP 2001-2004

You could say that nationalism is about belonging, about a place and imagining the other. It can take a positive, civic form, one in which belonging brings obligation to reach out and to imagine the other in an inclusive, multiple way. (Saul, 2005, p. 245)

The third and final cycle of SAIP built on the existing structure of the first mathematics assessment, but it also reached out to consider learning as a complex process influenced by the mosaic of cultures that produced Canadian students as socially-interactive beings. Its design embraced that which was uniquely Canadian – our vast landscape, the complications of linguistic differences, and the messiness of multiple education systems - with the intention of developing some understanding about what was happening in these diverse systems of education and what made learning effective. The academic assessment was now just one element in a broad data collection process designed to capture some sense of what was happening in Canadian education at the beginning of the new millennia.

The SAIP Mathematics Assessment – 2001

A conceptual framework put forward by Wang, Haertel, and Walberg (1993) and enhanced by the researchers at CMEC shaped and informed the administration of the SAIP Mathematics III assessment. The conceptual model that informed the development and design of the assessment instruments and the administration practices. The statistical analysis consisted of three elements – inputs, process, and outcomes – contained within “an overall context determined by demographic features, social and economic conditions, infrastructure, and broad characteristics of society in which the enterprise operates” (Council of Ministers of Education, Canada, 2001b, p.93). The SAIP Mathematics III –
2001 academic assessment built on former assessment mathematics assessments administered in 1993 and 1997; however, the assessment reflected “changes in curriculum and teaching practice” emanating from “developments in educational research and changing public understanding of the role of education in society” (Council of Ministers of Educations, Canada, 2001, p. 3).

The sampling procedures were consistent with the SAIP Science – 1999 assessment; however, the Quebec Ministry of Education chose to limit its participation to the 13-year-old group. In addition to the administration of the academic assessment, significant attention was paid to the “collection data on the opportunities students have had to learn mathematics and on their attitudes toward mathematics, as well as other information on their interests and activities” (Council of Ministers of Education, Canada, 2001a, p.2). Questionnaires were also distributed to teachers and principals to gather their perspective on the factors that affect students’ opportunities to learn mathematics. Although students, teachers, and principals were surveyed in conjunction with the administration of the SAIP Science – 1999 assessment, the statistical analysis of the 1999 questionnaire information was limited to the description of the frequency of particular variables. The statistical analysis conducted on the information gathered on the SAIP Mathematics III questionnaires and the data from students’ academic performances were analyzed to determine the strength of the relationships between contextual factors that might have influenced student performance and students’ academic performance on SAIP Mathematics III academic assessment (Council of Ministers of Education, Canada, 2001b).
The SAIP mathematics assessment, which was funded “jointly by CMEC, ministries of education, and HRSDC,” adhered to the same framework but the “criteria that described the levels of achievement were adjusted to support consistency in the assessment of students’ performances” (Council of Ministers of Education, Canada, 2001a, p. 4). The changes reflected the increasing levels of participation of the Canadian provinces and territories in international achievement tests such as PISA and TIMSS, the collaborative work of a strong network of mathematics educators, and the recognition that the assessment focused on only “skills that can be measured by a paper-and-pencil test” (Council of Ministers of Education, Canada, 2001a, p. 5). “[T]he ability to work with manipulatives, group problem-solving skills and the exploration of complex mathematical issues” – these content areas, while important for many mathematics programs, were not included in the assessment framework (Council of Ministers of Education, Canada, 2001a, p. 5). Thus, the definition of the limitations of the assessment and that upon which it could offer feedback were becoming more clearly defined with each SAIP administration.

The mathematics assessment – 2001 posed two research questions “How well have Canadian 13- and 16-year-old students learned mathematics in 2001?” and “Has the achievement of Canadian 13- and 16-year-old students in mathematics changed since the first two assessments? (Council of Ministers of Education, Canada, 2001a, p. 3) These questions were essentially the same as other SAIP assessments; however, the instruments designed to gather the information changed as well as the capacity to analyze the data and report the findings. Although the key anchor questions remained unchanged throughout
the three administrations, the changes to the 2001-version were extensive. The
assessment developers no longer felt confident that the 1993 and 2001 administrations
could be deemed comparable; “however, care was taken to ensure that statistically sound
comparisons could continue to be made between the 1997 and 2001 results” (Council of
Ministers of Education, Canada, 2001a, p. 17).

The random sample consisted of 41,000 students drawn from all jurisdictions; the
sample was smaller than either 1993 or 1997 (Council of Ministers of Education, Canada,
2001a). The reduced sample size may in part be explained by Quebec’s decision to
withdraw its 16-year olds from the study. The sample was composed of 13-year olds
(24,000) and 16-year olds (17,000) who wrote in English (33,000) and French (8,000).
The sampling strategy was not described as inclusive like the 1997 mathematics
assessment; thus, the participant pool may have differed significantly from the 1997
administration in their overall ability to perform on the assessments.

Like previous assessments, the results for the academic assessment were
presented in a series of comparative charts: aggregated by age, gender, language, and in
relation to stakeholder-generated expectations and the 1997 results to illustrate any
significant change or difference. Approximately, “two-thirds of the 13-year-olds achieved
level 2” and “[h]alf the 16-year-olds achieved level 3” on the content assessment; the
results for the problem solving assessment were similar (Council of Ministers of
Education, Canada, 2001a, p. 22). Reflecting on the 1997 data for the content assessment,
“significantly more 13-year-old students achieved level 2,” while significantly fewer 16-
year-old students achieved level 3 (Council of Ministers of Education, Canada, 2001a, p.
24). For 16-year olds, the gap between the 2001 performance and the 1997 performance
became larger if the performance of the 16-year-old Quebec students was retained in the 1997 reference data. For the problem solving assessment, a larger percentage scored at level 2 and significantly more 16-year-olds scored at level 3; these changes, according to CMEC, were attributable to “improved student performance, rather than to any changes in the difficulty of the questions or to the scoring process” (Council of Ministers of Education, Canada, 2001a, p. 25).

Although students posted significant improvements in their performance over the 1997 results, they still failed to meet the expectations set by the 100-member expectations-setting panel (Council of Ministers of Education, Canada, 2001a). Generally, educators and non-educators were “not satisfied with the performance of Canadian students in the problem solving assessment” (p. 32), and their expectations were “higher than the achievement of 13-year-old and 16-year-old students” on the content assessment” (Council of Ministers of Education, Canada, 2001a, p. 31).

There were slight differences based on gender for 13-year olds at levels 4 and 5 for content and level 2 for problem solving, and similar slight differences for 16-year olds were present. Based on the results of SAIP Mathematics III assessment, the performance gap attributable to gender had all but vanished (Council of Ministers of Education, Canada, 2001a).

The analysis based on aggregation by language revealed that “fewer 13-year-old students who wrote in English reached levels 2 and 3” for both problem solving and content, while more 16-year-olds who wrote in English reached levels 4 and 5 (Council of Ministers of Education, Canada, 2001a, p. 28).
In 2001, the effects of gender on the performance results was minimal, students failed to meet the expectations set for them, students writing in French tended to perform at higher levels than their peers writing in English, and performance differences attributable to age grouping were still evident. The phenomenon described as the “Quebec factor” seemed not only apparent in the performances of students writing in Quebec but also in students writing in French and living in other provinces.

Results were also displayed by jurisdiction in comparative charts with both the Canada-wide 2001 norm, and the accompanying commentary described the performance relative to the 1997 results. For example, the statements pertaining to British Columbia stated that “students in both age groups performed as well as Canada as a whole” on the content assessment but “[f]ewer 16-year-old students reached level 3 in 2001 assessment than in the 1997 assessment” (Council of Ministers of Education, Canada, 2001a, p. 41). A performance trend was beginning to emerge: Students in Alberta and Quebec (French) performed at higher levels than the Canadian norm for 13-year-olds on both assessments at level 2, and the percentages of females performing at level 2 in these provinces exceeded the Canadian norm for both assessments (Council of Ministers of Education, Canada, 2001a). However, Alberta was not alone; the percentage of 13-year-olds in Ontario (English) and Quebec (English) exceeded the Canadian norm on the content assessment; the percentage of 13-year-olds in Manitoba (French), Ontario (French), Quebec (English) exceeded the Canadian norm on the problem solving assessment; and the percentage of 13-year-old females in Ontario (English) exceeded the Canadian norm on the content assessment. The trend continued for 16-year olds in Alberta as well as students writing in French in other jurisdictions: The percentage of students in Alberta,
Manitoba (French), New Brunswick (French) and Nova Scotia (French) exceeded the Canadian norm on both assessments and the percentage of females in these jurisdictions exceeded the Canadian norm on both assessments (Council of Ministers of Education, Canada, 2001a, pp 119-130). It is interesting to note that students writing in French in Manitoba, Nova Scotia and New Brunswick did not perform at a level consistent with the Canadian norm in 1993; however, in relation to their Canadian peers, their performance was ranked at a higher level in 2001.

The data revealed an interesting phenomenon related to the consistently strong performances by students writing in Alberta and in French and the strength of the performances of female students living in these jurisdictions as compared to the Canada-wide norm. The materiality of these differences and the causation for these differences has remained uncontested.

*The Canadian Context – the 2001 Contextual Questionnaire.*

In 2001, CMEC recognized that “learning is a complex process, affected by many factors within student background and experience, school and classroom conditions, resources, motivation, quality of school and teaching, attitudes and expectations” (Council of Ministers of Education, Canada, 2001b, p.3). This statement represented a shift in CMEC’s thinking about learning and the limited ability of output-focused assessments, like SAIP, to provide meaningful and useful information to educational systems. CMEC acknowledged that SAIP had been envisioned as a “comprehensive program, through which data would be collected on many factors that might influence
learning”; however, its earlier iterations were focused on measuring student performance almost exclusively (Council of Ministers of Education, Canada, 2001b, p.3).

The sampling strategy for the 2001 questionnaire was similar to that implemented for the SAIP Science - 1999 questionnaire, and “was designed to yield representative student samples of sufficient size to permit separate reporting for each population” (Council of Ministers of Education, Canada, 2001b, p. 4). The students (41,000), who wrote the SAIP Mathematics III assessment, also completed the contextual questionnaire (Council of Ministers of Education, Canada, 2001b, p.4). The principals for each of the 2,000 schools completed the school questionnaire, and the teachers who had taught mathematics to any of the 41,000 students during the 2000-01 academic year completed the questionnaires (Council of Ministers of Education, Canada, 2001b, p.4). Because the teacher sample construction did not link teachers to specific students, further correlation analysis could not be conducted (Council of Ministers of Education, Canada, 2001b). The analysis of the information generated by the teachers’ sample remained at the descriptive level.

CMEC acknowledged that “little use was made of [the] information” gathered from questionnaires administered prior to 1999; however, the 1999 questionnaire formed the foundation for the 2001 questionnaire administered to students, teachers, and principals to gather information on “student backgrounds and activities, school characteristics, decision making, resources, classroom practices, opportunity to learn, attitudes toward school and mathematics, and teacher backgrounds” (Council of Ministers of Education, Canada, 2001b, p.3). The analysis of information generated through the administration not only provided descriptive analysis but also a preliminary attempt to
link the information gathered about the variables measured through the administration of the questionnaire with the results of the SAIP Mathematics III academic assessment “in an exploratory bivariate analysis” (Council of Ministers of Education, Canada, 2001b, p. 89). The analysis was conducted by jurisdiction; however, the intention was not a comparative analysis among jurisdictions, but rather the presentation of a series of replicated trials designed to identify “relationships that are reasonably stable across jurisdictions” (Council of Ministers of Education, Canada, 2001b, p. 89). Consequently, CMEC extended an invitation to researchers to conduct a “more comprehensive analysis” and “to fully investigate these relationships, the possible cumulative effects of the various factors, and the possibility that some factors may function differently in different jurisdictions” (Council of Ministers of Education, Canada, 2001b, p. 89).

The information gathered by the questionnaires added context to the performance data generated by the SAIP Mathematics III instrument. The descriptive analysis was organized by schools (principal questionnaire), teachers, and students. The analysis of Canadian schools revealed that “school size tends to follow population size”; however, Nova Scotia and New Brunswick, both provinces with small populations have a significant number of large schools (more than 500 students) (Council of Ministers of Education, Canada, 2001b, p. 6). The analysis also revealed that “[c]ommunity size and school size are positively correlated with each other” (Council of Ministers of Education, Canada, 2001, p. 92). In addition, “large class size tends to be positively associated with higher achievement” (Council of Ministers of Education, Canada, 2001b, p. 92). These findings, CMEC noted were probably contrary to popular opinion and that in some way the findings may have been linked “in complex ways to socio-economic status of the
school, language, and other community or school variables (Council of Ministers of Education, Canada, 2001b, p. 92).

The questionnaire revealed a profile of Canadian teachers, their jobs, and their teaching preferences. In 2001 more females were employed as mathematics teachers; “Quebec stands out as having more female teachers, and Nova Scotia francophone as having more male teachers” (Council of Ministers of Education, Canada, 2001b, p. 90). Most teachers reported little involvement with parents, but teachers working in Anglophone jurisdictions reported more contact with students than those working in francophone ones (Council of Ministers of Education, Canada, 2001b). More francophone teachers identified that students’ backgrounds presented major challenges than Anglophone teachers (Council of Ministers of Education, Canada, 2001b). In general, “most teachers reported the variation of student abilities presents a major challenge to teaching” (Council of Ministers of Education, Canada, 2001, p. 90); thus, many teachers (90%) supported the streaming of students into “different programs based on their abilities (Council of Ministers of Education, Canada, 2001b, p. 33). Jurisdictional differences on the support for introduction of streaming were small; “Saskatchewan teachers and Quebec francophone teachers showed slightly lower levels of agreement than others on this issue” (Council of Ministers of Education, Canada, 2001b, p. 33). While these were just some of the examples drawing from the analysis of the teachers’ questionnaire, the results began to raise the visibility of some of the issues facing teachers: the complexity of factors influencing teaching - student ability and their backgrounds - and the interrelationship of these factors with desire to provide appropriate programming and some assurance of student success.
The information generated by the student questionnaire was presented both as a descriptive data and as a co-relational analysis with student performance by jurisdiction (Council of Ministers of Education, Canada, 2001b). Students provided their perceptions about what factors influenced their ability to get high marks: “Almost all students agreed that to do well in mathematics you need hard work and good teaching” and that “low marks can be attributed to not working hard enough” but only “half attributed poor marks to poor teaching” (Council of Ministers of Education, Canada, 2001b, p. 55). The questionnaire probed students’ beliefs about the role of luck in their ability to get either high or low marks: Interestingly, “students in Nunavut and francophone students outside Quebec were somewhat more likely to attribute marks, either high or low, to luck” (Council of Ministers of Education, Canada, 2001b, p. 55). Students were also asked to comment on how much they liked school: Students (90%) agreed that “they have lots of friends in school, that they get along with other students, and that they like to learn new things” (Council of Ministers of Education, Canada, 2001b, p. 55). By jurisdiction, more students in Nunavut reported that they felt “good about school”; however, this population also reported the highest level of absenteeism (Council of Ministers of Education, Canada, 2001b, p. 55). CMEC urged a cautious interpretation of such relationships: Did the data really imply that “students enjoy school more when they are absent?” or is this phenomenon simply indicative of some much broader interplay of factors? (Council of Ministers of Education, Canada, 2001b, p. 55).

The questionnaire revisited the nature of students’ the interactions with parents and tutors, a thematic topic from other questionnaires: “Overall more than 80% of students reported that they frequently discuss their school work with parents, with few
language or age differences” (Council of Ministers of Education, Canada, 2001b, p. 56). Although amount of time that students spend on homework was “positively associated with achievement,” the amount of time parents or tutors helped students with their homework was negatively associated with achievement (Council of Ministers of Education, Canada, 2001b, p. 91). CMEC recognized that proportionately more students who were experiencing difficulty with mathematics would seek out parental help or tutoring. However, these interactions with parents and tutors did not, in effect, produce “turn around” effects on achievement” (Council of Ministers of Education, Canada, 2001b, p. 92).

“I have spent many years working as a learning assistance instructor, worked with many students from a wide variety of backgrounds who have struggled with mathematics, and participated in their realization of some success. Personally, I believed that I offered something positive to the learning process. When I read and processed the results of this survey and also the science survey, I began to question the efficacy of such interventions. In the summer of 2010, a student with whom I had worked in past and who had a long history of difficulties with mathematics enrolled in an advanced grade 12 mathematics course which he had initially failed. I decided to back away, determined to assess the validity of the survey results – at least anecdotally. The student did well – exceptionally well – and earned a mark that would please even an exceptional student. One case does not provide definitive proof but does add more impetus to investigate how and for whom these interventions are valuable. (Dagenais, August 5, 2010)

These examples provided some sense of the context surrounding students’ learning mathematics. As illustrated by the examples, learning mathematics is complicated and many of taken-for-granted assumptions concerning the importance of the role of tutoring and parental help were challenged by the analysis of the questionnaire information and its relationship to the student performance data. Equally important was CMEC’s recognition, that these challenges to our taken-for-granted assumptions about the learning context provide sites for additional, comprehensive research to unpack the role of these variables.
and the contexts in which their expression is most relevant to student success in mathematics.

The Administration of the SAIP Mathematics III assessment was characterized by a shift in statements about assessment and learning. The language used in the 2001 reports was more cautious than other reports. The introduction of the concept of role of the learning context in the creation of opportunities for learning strengthened the theoretical underpinnings of the assessment, and the attention-paid to the social, cultural, and research environment in Canada reached a new level of awareness in 2001. Essentially, measuring student performance on an achievement test no longer constituted adequate feedback about the performance of educational jurisdictions in Canada.

The SAIP Writing Assessment – 2002

The SAIP Writing III assessment, an evolution of previous SAIP Reading and Writing assessments was jointly funded by CMEC, Canadian ministries of education, and HRSDC at a direct cost of approximately $1.5 million – about $2.25 per student (Council of Minister of Education, Canada, 2003b, p. 3). The assessment was administered to 23,680 students, a slightly smaller sample than SAIP Reading and Writing II (Council of Ministers of Education, Canada, 2003a, p. 5). The sample consisted of 12,708 thirteen-year olds and 10, 972 sixteen-year olds of which approximately two-thirds wrote in English and one-third wrote in French (Council of Ministers of Education, Canada, 2003a, p. 5).

SAIP III Writing assessment was significant for several reasons. CMEC revealed both the funding partners and the cost of assessment; the assessment framework and
criteria were modified significantly; current pedagogical practice and theory informed the redevelopment of the assessment; and the reading assessment was replaced with a novel and innovative assessment designed to measure critical thinking skills and meta-cognition (Council of Ministers of Education, Canada, 2003a). Although this was the third and last SAIP writing assessment, the developers made a significant investment in the revisions such that the comparisons with the 1994 and 1998 assessments should be approached with “considerable caution” (Council of Ministers of Education, Canada, 2003a, p. 4).

CMEC emphasized the importance of connecting educational leadership with assessment and measurement practices with the latest research when it stated:

The role of SAIP has been from its inception to provide educational leadership by producing assessments based on current innovation and the most contemporary research and practice related to student assessment. The goal has been to link innovation and research with familiar and current classroom practices. To this end, Writing III Consortium developed a new design for the assessment. ... [T]he new design embraces a more contextualized and cross-curricular framework. (Council of Ministers of Education, Canada, 2003a, p. 3).

The revisions to SAIP III writing assessment retained the five achievement levels and supported the expectation that the majority of 13-year olds would achieve level 2 while the majority of 16-year olds would achieve level 3.

Comparability between the performances of students writing in English and French has been a priority for all SAIP assessments, a priority that has been handled differently on each assessment. On the 1998 SAIP Writing assessment, the approach focused on the development of inter-rater reliability on the scoring of the writing samples; for the 2002 assessment, the focus was placed on the selection of texts for the Student Resource Booklet (Council of Ministers of Education, Canada, 2003a). The selection of the texts followed this rubric: “eight were published in both languages, three
were translated from French to English and three from English to French, and four were in English only and four in French only” (Council of Ministers of Education, Canada, 2003a, p. 4). Although the use of bilingual scorers and experienced scoring teams continued, parallel training was identified as a key component of the language parity strategy (Council of Ministers of Education, Canada, 2003a). The specific use of bilingual scorers working in their second language to provide inter-rater reliability assurance was not a component of the 2002 language parity strategy.

The framework and purpose for the SAIP writing assessment shifted considerably as the statements that governed what constituted good and useful writing had also shifted over the past decade. The conceptual framework for the writing assessment was based on the notion that “writing assessments should not simply measure whether students can produce a particular text, but whether they can apply knowledge of writing for a specific purpose in a specific context” (Council of Ministers of Education, Canada, 2003a, p. 6). Writing was no longer viewed “as an end in itself” but should create an opportunity for “writers to communicate effectively in some larger ‘real-life’ situations for reasons a classroom community might experience as authentic” (Council of Ministers of Education, Canada, 2003a, p. 6). Furthermore, CMEC adopted the following philosophical perspective pertaining to writing as it perceived writing to be “socially situated in that it is meant to be read ... social in the sense that it is intended for specific discourse communities ... [and] a means of generating ideas and information as well as a medium of communication - it is a way of creating, exploring, and refining ideas” (Council of Ministers of Education, Canada, 2003a, p. 10). This re-conceptualization of the nature of a meaningful writing task necessitated a re-thinking of the framework or strands about
which the assessment was constructed, the criteria that differentiated the levels of writing, the writing topic, and the prewriting activities and supporting documents (Council of Ministers of Education, Canada, 2003a).

The writing task was more prescriptive in nature; students were asked to use their writing to generate “public awareness about ... [an environmental] dilemma” (Council of Ministers of Education, Canada, 2003a, p. 6). The goal of the writing task was to situate student writing in the social and cultural context of real-life and to incorporate current research on writing (Council of Ministers of Education, Canada, 2003a). Thus, the writing task included a number of pre-writing activities outlined in the Student Resource Booklet, which was designed to assist with the “scaffolding” of resources,” and to provide directed-preparatory activities (Council of Ministers of Education, Canada, 2003a, p. 7). These activities included the reading of a short text and the development a written response, classroom discussions about the reading resource materials, and an opportunity to read the resource material and develop personal reflections on the material (Council of Ministers of Education, Canada, 2003a, p. 7). The students completed the preparatory process in five days and submitted the main writing task whose purpose was to “provide agency about an environmental dilemma” (Council of Ministers of Education, Canada, 2003a, p. 8).

The topic was purposefully situated in the field of environmental science and stewardship with the hope that the contextualization of the writing task might engage male students more than the previous topic on heroism (Council of Ministers of Education, Canada, 2003a). The results for the 2002 SAIP assessment confirmed an “international phenomenon in many cultures and languages, namely that girls
demonstrate writing skills at a significantly higher level than boys” (Council of Ministers of Education, Canada, 2003a, p. 28). Although the writing task was designed to address the female-male performance difference, a significantly higher percentage of females scored at levels 2 (about 10% higher for females) and 3 (about 16% higher for females) for both age groups (Council of Ministers of Education, Canada, 2003a, p. 28).

The results for the SAIP - 2002 assessment were presented as comparative charts with some discussion aggregated by age, gender, jurisdiction, and in relation to the 1998 test results and to the expectations generated by the 100-member panel (Council of Ministers of Education, Canada, 2003a). The results showed that “most of the older students write at or above level 3, compared with less than half of the younger students” (Council of Ministers of Education, Canada, 2003a, p. 27). This result is consistent with previous administrations of the SAIP writing assessment and was “not surprising” (Council of Ministers of Education, Canada, 2003a, p. 27). The students’ performance on the SAIP assessment was “at the expected levels” with exception of a lower than expected results for 16-year olds at level 3 (Council of Ministers of Education, Canada, 2003a, p. 30).

The proportion of students in several jurisdictions was at levels equal to or better than the Canadian norms for English and French. Separate Canadian norms as well as a Canadian composite norm were calculated for the SAIP III writing assessment. The performances by 13-year-old and 16 year-old students in Alberta, British Columbia, Manitoba (English), Ontario (English), Quebec (English), and Quebec (French) met or exceeded the Canada-wide norm; and the following provinces met or exceeded the Canadian norm for the 16-year-old age group only: New Brunswick (English),
Newfoundland and Labrador and Saskatchewan (Council of Ministers of Education, Canada, 2003a, p. 30). The Canadian norm for students writing in French was higher than the Canadian norm for students writing in English, but difference based on language was “less than it was in 1998” (Council of Ministers of Education, Canada, 2003a, p. 87). Quebec students writing in French have had a long tradition of performing very well on both national and international assessment; thus, the Canadian norms based on language differ.

The differences in performance based on gender were generally less than the 1998 administration, and in some jurisdictions, the percentage of male students achieving level 2 (13-year olds) and level 3 (16-year olds) exceeded the Canadian norm (Council of Ministers of Education, Canada, 2003a). Males in the following provinces performed above the Canadian norm for their age and language group: Alberta, Manitoba (English), Quebec (English and French) 16-year-olds; and Alberta, Ontario (English), Nova Scotia (English and French) 13-yr olds (Council of Ministers of Education, Canada, 2003a, p. 91 & 92). The data suggested that the these provinces may have introduced strategies that encourage male students to engage in writing tasks in a more meaningful manner, or the SAIP writing task may have been more successful in drawing out a higher level of performance from the male students in these jurisdictions. Again differential performance based on gender has been a focus of research in the past, but this data seemed suggest that either the strategies used by some jurisdictions or the SAIP III assessment alleviated some of the previous gender-based differences.
A Secondary Study: Habits of Mind

Throughout the 1980s and 1990s several Canadian jurisdictions engaged in processes of curriculum reform that included the use of performance-based learning outcomes, assessment standards, the introduction of general competencies such as those embedded in Saskatchewan’s Common Essential Learnings (Council of Ministers of Education, Canada, 2003a). Researchers and curriculum scholars recognized that these general competences such as critical thinking and tasks involving meta-cognition were valuable but difficult to evaluate (Council of Ministers of Education, Canada, 2003a).

The goal of the secondary study was to develop a task that would measure students’ disposition toward the use of critical thinking and meta-cognitive skills (Council of Ministers of Education, Canada, 2003a). The task combined a short reading task, a fable, with a short writing task based on the reading of the fable (writing proficiency was not evaluated). The goal of the task moved beyond reading comprehension and into evaluating the students’ ability to “provide meaning that goes beyond the text ... provide extended reasoning ... [and] ... reach judgment by considering aesthetic features of the text” (Council of Ministers of Education, Canada, 2003a, p. 78-79). The goal was limited, and the task was not designed to “determine the effectiveness of student thinking,” but rather to seek out and measure the “primary traits manifested in what the student did to construct meaning” (Council of Ministers of Education, Canada, 2003a, p. 79).

Although the link between reading and writing seemed to be self evident, the nature of the relationship and the strength of the linkage were the questions that interested
researchers in 2002. The following research questions guided the project and the subsequent data analysis:

Is there a direct correlation of critical habits of mind and writing skill?  
Do critical habits of mind provide a link between quality of reading and quality of writing?  
How would a successful curriculum implementation of generic skills such as critical thinking be apparent in student demonstrations of learning? (Council of Ministers of Education, Canada, 2003a, p. 80).

Essentially, CMEC’s questions were focused on curriculum development goals around the design and implementation of reading, writing and thinking skills development programs and the measurement of student learning in these domains.

The results showed that “16-year olds responded more effectively than 13-year olds” however, the de-emphasis on writing quality minimized the differences between the age groups (Council of Ministers of Education, Canada, 2003a, p. 82). The gender-based differences in performance disappeared for the 13-year-old group, and the differences in the performance percentages between males and females in the 16-year-old groups were “considerably below the current trends in literary assessments” (Council of Ministers of Education, Canada, 2003a, p. 82). The secondary study also confirmed the generally accepted thinking of cognitive scientists that have stated “students who practise higher-level habits of mind, in particular full exploration of an issue and critical judgment, also demonstrate higher levels of literacy” (Council of Ministers of Education, Canada, 2003a, p. 82). While this statement seemed to make common sense, the question that evolved from this finding was the following: Does literacy determine critical thinking and meta-cognition or does critical thinking and meta-cognition predispose an individual to become literate? These questions have had significant bearing as researchers and scholars worked
with and designed programs in conjunction with the goals and aspirations of Aboriginal Canadians.

The SAIP developers have relied on statistical difference and the notion of confidence intervals as a way of ensuring that the measures of performances were treated as estimates. However, the 2002 iteration of SAIP introduced the term, educational significance, and its importance: “Statistical significance is determined by mathematical formulas and considers issues such as sampling” where as educational significance “is a matter of interpretation” (Council of Ministers of Education, Canada, 2003a, p. 25). The importance of educational significance and its measurement have remained relatively undeveloped in previous SAIP assessments, and it was introduced as an interesting concept but was not operationalized in the SAIP 2002 assessment.

Summary

The SAIP 2002 assessment provided an innovative and contemporary writing task in both assessment activities. The rethinking of writing as a purposeful activity situated in the everyday life of people engaged in the daily tasks of a particular society constituted a significant reinvention of writing measured by large-scale assessments. The secondary task was a precursor to more broadly distributed activities related to expectations-setting around creative and critical thinking.

*As I reflect on the critical thinking and meta-cognitive activities assessment, I now understand the rationale for an invitation to sit on an expectations-setting panel to determine the expected level of critical and creative thinking for grade eight students. The task was more difficult than I had expected. Based on our review of student exemplars and the scoring rubric, the panel members were expected to focus on a particular aspect of creative or critical thinking until we could come to some level of consensus on which level of performance was most reasonable for students in grade eight. We discussed, we voted, we discussed and we voted*
again. No doubt, the 2002 SAIP assessment spurred provincial ministries to examine critical and creative thinking, its role in the curricula for all grades, and to determine strategies for measuring proficiency in these valuable but difficult to measure skill areas. (Dagenais, August 6, 2010).

The secondary assessment and the identification of jurisdictions in which the percentage of male students performed as well as or better than the Canadian norm, may provide some insight in the gender-based performance gap. In general, probing of the educational context in Quebec may also provide useful insights for other Canadian jurisdictions about curriculum making, the teaching and learning culture and the socio-economic factors that support strong performances.

*The SAIP Science Assessment – 2004*

In the spring of 2004, CMEC in partnership with Canadian ministries of education, and HRSDC launched the Science III assessment, the final assessment of the third cycle and the last assessment of the SAIP project at a direct cost of $2 million “approximately $2.50 per student” Council of Ministers of Education, Canada, 2005, p. 4). The beginning of the third cycle marked a new level of transparency concerning the partners of the SAIP project and SAIP’s direct costs. CMEC made a conscientious effort to reveal both the costs and the engagement by both federal and provincial agencies and departments.

The assessment developers reengineered a few questions on the literacy assessment, deleted the practical skills assessment based on issues of fiscal restraint, and reconfigured aspects of the questionnaire to support relational analysis between the data generated by the teacher questionnaire and the student performance data (Council of
Ministers of Education, Canada, 2005b). CMEC described the interrelationship of academic performance and the learning context as follows:

An important aspect of this assessment is the collection of contextual data on the opportunities students have had to learn science and on their attitudes toward science, as well as information on their interests and activities. (Council of Ministers of Education, Canada, 2005b, p. 1)

The statement concerning the collection of contextual data has replaced the research question in other SAIP assessments that focused on the academic performance of Canadian students in a specific subject area.

The SAIP - 2004 science written assessment remained substantially similar to the 1999 science assessment. The development team reviewed each item for linguistic parity and functionality (Council of Ministers of Education, Canada, 2005b). However, a few of the criteria on the science literacy assessment “were modified slightly and a small number of questions (12 out of a total 129 questions) were changed to reflect changes in the field of science and science education since the last administration” (Council of Ministers of Education, Canada, 2005b, p. 4). Although the theory pertaining to the teaching and learning of science has continued to evolve and change, the framework for the SAIP science assessments and most of the evaluation criteria remained relatively constant “to facilitate the comparison of results among the three assessments – an important feature of SAIP” (Council of Ministers of Education, Canada, 2005b, p. 5).

The sample size for the administration of the SAIP - 2004 science assessment was comprised of 25,700 students, similar to the 1999 administration (Council of Ministers of Education, Canada, 2005b, p. 7). The distribution of the 13-year olds (13,900) and 16-year olds (11,800) was also similar (Council of Ministers of Education, Canada, 2005b, p. 7). Approximately two-thirds of the sample wrote in English and one-third wrote in
French (Council of Ministers of Education, Canada, 2005b). All of the students completed the contextual survey. The principals for each participating school completed the school survey, and the science teachers of the students completing the academic assessment completed teacher surveys (Council of Ministers of Education, Canada, 2005b).

The SAIP Science 2004 – Academic Test Results

The reporting of the academic performance results for the SAIP - 2004 assessment used a similar format to other SAIP assessments, and the data were aggregated by age group, by cycle, gender, language, expectations, and jurisdiction. Accompanying the presentation of the results, CMEC reiterated its position related to educational difference and statistical difference and offered this cautionary warning: “There are situations where a statistically significant difference may have little educational significance,” and there are also situations where a statistically insignificant difference may have educational significance (Council of Ministers of Education, Canada, 2005b, p. 18).

The results by age group showed that more than 70% of the 13-year-olds reached level 2 “demonstrated such abilities as comparing various plant and animal adaptations and identifying technologies that influence science,” and almost two-thirds of the 16-year olds demonstrated “such abilities as using chemical properties to compare and classify substances and analyze experiments and judge their validity” (Council of Ministers of Education, Canada, 2005b, p. 19). The proportion of students for each age group reaching level 2 (13-year olds) was similar to the proportion in 1996 and 1999, but
the proportion of 16-year olds achieving level 3 was significantly lower than the proportion in 1999 (Council of Ministers of Education, Canada, 2005b). In addition, the proportion of 13-year olds achieving level 3 and the proportion of 16-year olds reaching level 4 were significantly lower than the 1999 administration (Council of Ministers of Education, Canada, 2005b, p. 21). The downward trend in performance results could be attributed to a number of factors or the interrelationship of a number of the factors. CMEC hypothesized that some, all, or none of the following factors may explain the decline in student performance:

- decreasing congruence between curricula and test design ...
- decreasing congruence between classroom practices and learning strategies and the items tested ...
- increased emphasis on literacy and numeracy in schools, sometimes at the expense of other subjects areas
- assessment fatigue, a factor for both teachers and students ...
- changes to jurisdictional policies (Council of Ministers of Education, Canada, 2005b, p. 22)

The suggestions outlined possible reasons for the decline in level of student performance, but they also recognized the complexity of the learning environment and the critical relationships between curriculum priorities; test design issues; classroom practices; and the international, political and economic pressure to increase literacy and numeracy levels of students in all schools. Commentators have raised concerns about these issues in the literature and their overall effects on the educational process (Robertson, 1999); however, CMEC has not addressed these issues in its other documentation. CMEC also introduced a new factor affecting test performance, “test fatigue,” but offered only a minimal definition of what might constitute test fatigue and how it would be produced in students and teachers. Test fatigue may be a reasonable outcome for students and teachers working in an environment designed and managed by the testing cycle.
The tests results revealed only slight differences for only a few of the performance levels when the data were aggregated by gender, language and expectations. The analysis of student performance based on gender revealed that “slightly more 13-year-old males [than females] at level 3 or above” and slightly more 16-year-old females than males at “level 1 or above” (Council of Ministers of Education, Canada, 2005b, 23). When the data were sorted and analyzed by language, “the difference in achievement at all levels for both age groups and in both languages was not significant” (Council of Ministers of Education, Canada, 2005b, 24). For the SAIP - 2004 Science assessment, like other assessments, an anonymous 100-member panel set expectations for student performance; these expectations were compared with student performance to determine the degree of congruence between the expectations of the panel members and the student performance data (Council of Ministers of Education, Canada, 2005b). Although the proportion of 13-year old students failed to meet the expectations set by the panel at levels 4 and 5 and slightly fewer 16-year olds reached level 1 and 4, the relationship of student performance with the expectations of the panel in 2004 remained consistent with those of the 1999 expectations setting panel (Council of Ministers of Education, Canada, 2005b). Overall, the analysis by gender, language, and expectations revealed that the student sample performed as expected.

The analysis by jurisdiction revealed that both age groups in Alberta performed well above the Canada-wide norm and that male and female students in Alberta for each age group performed at levels above the Canada-wide norm (Council of Ministers of Education, Canada, 2005b, pp.27, 167-170). The performance for students in both age groups in British Columbia, Ontario (English) and Quebec (French) met expectations
while 13 year olds in Manitoba (English) and Quebec (English); and 16 year olds in Newfoundland and Labrador and Yukon performed at levels consistent with the Canada-wide norm (Council of Ministers of Education, Canada, 2005b, p.27). For 13-year old group, the female students in the provinces of British Columbia, Manitoba (English), Ontario (English), Quebec (English) and Quebec (French) met or exceeded the Canada-wide norm; however, males in British Columbia, Ontario (English) and Quebec (French) achieved or surpassed the Canada-wide norm for level 2 (Council of Ministers of Education, Canada, 2005b, p. 167-168). For the 16-year-old group, the female students in the provinces of British Columbia, Ontario (English), Quebec (French) and Yukon achieved or surpassed the Canada-wide-norm; however, only the male students in Ontario (English) and Quebec (French) achieved or exceeded the Canada-wide norm (Council of Ministers of Education, Canada, 2005b, p. 169-170). Although the data analysis was limited to a simple survey of the charts and graphs, the data did suggest that provinces that were able to minimize the performance differences based on gender were able to render a stronger overall performance. This hypothesis like others emanating from the SAIP data requires further investigation to gain some understanding of what is really happening with regard to differential performance based on gender.

The analysis and documentation related to the administration of the three-pronged contextual survey occupied more than half of the SAIP 2004 report. Its importance preempted the formation of any other research questions, and its innovative nature made it stand out against the re-administration of the SAIP - 2004 academic assessment. CMEC reiterated the philosophy that governed the administration of previous contextual surveys by stating:
Most educational indicator systems are built around the concept that student learning outcomes are influenced by inputs and by the processes arising from these inputs. It is also generally recognized that education operates in an overall context determined by demographic features, social and economic conditions, infrastructure, and other broad characteristics of the society in which the enterprise operates. (Council of Ministers of Education, Canada, 2005b, p. 73)

CMEC described the ambiguity that surrounded the design and administration of the SAIP project and its journey to become a unique Canadian school indicators measurement system. In 2004 CMEC shifted its assessment priorities from the measurement of the instructional inputs, learning processes, outcomes of the learning process, and the learning environment to the difficult challenge of prioritizing those variables - at this time and in this place which variables were most critical to the production of the real and substantive change in students’ learning. CMEC described its challenge as follows: It is not always “obvious which specific context, input, or process variables are most worth investigating” at any particular time (Council of Ministers of Education, Canada, 2005b, p. 74). In 2004, learning transformed itself into an uncertain contextually-bound process that operated within a socially-constrained environment; a somewhat different conceptualization from the process whose outcomes stood in proxy for the effectiveness of the inputs and the processes dedicated to the production of students’ academic performances. Although CMEC verbalized its philosophical struggles with the capacity of its SAIP project to render the type of information it wanted, CMEC continued to administer a virtually unchanged questionnaire in 2004.

*The SAIP Science – 2004 Contextual Analysis*

The goal of the contextual questionnaire analysis was to link the information from the school, student, and teacher surveys with the performance data generated by the
students (Council of Ministers of Education, Canada, 2005b). The questionnaire was similar in structure and focus to the SAIP mathematics - 2001 version with some revisions to the questions to reflect a science focus. The analysis, itself, was predominantly descriptive and comparative rather than predictive. The following examples illustrate the type of information gathered and the type of analysis rendered.

The student questionnaire asked students to reflect on their parents’ education, their discussions with their parents, and the level of assistance that they received from parents with regard to their homework (Council of Ministers of Education, Canada, 2005b). Students in both age groups reported that about 30% of their parents had a university education; however, the proportion was much lower for parents in Newfoundland and Saskatchewan, provinces in which students have traditionally performed below the Canada-wide norm (Council of Ministers of Education, Canada, 2005b). Most students (75%) discussed daily activities with their parents; fewer 13-year-olds (15%) and 16-year-olds (10%) “worked together with their parents on science homework” (Council of Ministers of Education, Canada, 2005b, p. 77).

The relational analysis revealed that only three variables had “consistent correlations across most jurisdictions”: the frequency of assessment, frequency of field trips and outings, and students’ perceptions about their deserving the marks that they received (Council of Ministers of Education, Canada, 2005b, p. 158). Students who believed that they earned the marks they received did better; however students who “completed a quiz or test a few times a week ... [or were] ... outside the school more than few times per month” did not do as well as their peers (Council of Ministers of Education, Canada, 2005b, p. 158). For most jurisdictions, students who participated in
tutoring sessions or worked on their homework with their parents did not perform as well as their peers who did not seek additional assistance (Council of Ministers of Education, Canada, 2005b, p. 158). This relationship has been consistent for all SAIP assessment contextual questionnaires; its meaning and its implication for the design and delivery of student services has not been unpacked by CMEC researchers.

Teachers were asked to report their average class size, their perception of the relationship between class size and their choice of instructional method, and the limitations placed on their instructional methods (Council of Ministers of Education, Canada, 2005b). Teachers reported that the average class size was about 25 students with the largest classes in Quebec (francophone) and Nova Scotia (francophone) (Council of Ministers of Education, Canada, 2005b). Large classes brought more difficulties and created more limitations on the delivery of instruction especially in “jurisdictions where teachers report making frequent use of laboratories” (Council of Ministers of Education, Canada, 2005b, p. 111). Teachers also reported that the diversity of students’ background created classrooms problems: About 90% of the teachers in Nova Scotia’s francophone schools reported “problems with students’ backgrounds” and more than 80% of teachers in Canada reported a wide-spread problem related to the engagement, motivation, and management of uninterested students (Council of Ministers of Education, Canada, 2005b). CMEC offered some possible explanations about the increasing level of students’ disinterest in science; however, this issue like many others is worthy of further research (Council of Ministers of Education, Canada, 2005b).

Data from the measurement of several variables captured on the teacher questionnaire were analyzed in relation to the student performance data. However, only
two variables provided significant positive correlations: years of experience and teachers’
describing themselves as specialists (Council of Ministers of Education, Canada, 2005b).
In the majority of jurisdictions, 16-year olds achieved better performance results if their
teachers were experienced and “considered themselves to be specialists in science and
preferred to teach mainly in this area” (Council of Ministers of Education, Canada,
2005b, p. 158). This analysis was the first attempt to “establish a relationship between
teacher responses to context questions and the achievement of their students” (Council of
Ministers of Education, Canada, 2005b, p. 158). The student-teacher connections were
complex, and the analysis did not reveal cross-jurisdictional trends for other variables
(Council of Ministers of Education, Canada, 2005b).

The principals for each of the sample schools completed a school survey designed
to gather contextual data about the learning environment, the school community, and the
processes for determining the acquisition of some key inputs (teachers and staff) in the
education process. According to the data collected on the school surveys, in 2004,
Canadian schools, when grouped by size, continued to conform to the distribution
outlined on the 2001 mathematics school survey (formed three large regions: large
schools dominate the landscape in Central Canada (Council of Ministers of Education,
Canada, 2005b). Larger schools were located in British Columbia, Ontario, and Quebec
while smaller schools were located in the rest of Canada with the exception of Nova
Scotia (Council of Ministers of Education, Canada, 2005b).

The survey also gathered data about the prevalence of learning problems in
Canadian schools: The average Canadian school reported that about 17% of its student
population had “learning problems requiring special attention” (Council of Ministers of
Schools in the Northwest Territories reported a much higher prevalence of learning problems requiring specialized attention - more than 60% of their students had learning problems (Council of Ministers of Education, Canada, 2005b, p. 138). The higher incidence rate of learning problems may have been a significant contributing factor to the lower performance levels for students in the Northwest Territories.

Based on previous research studies “that have shown that children from single-parent families tend to have greater learning problems” than those from two-parent families, the survey gathered data on family structures (Council of Ministers of Education, Canada, 2005b, p. 138). On average about 25 percent of Canadian families were headed by a single parent in 2004; however, schools in Quebec (Anglophone and francophone), Yukon and Northwest Territories reported that more than 60 percent of their students lived in single parent (Council of Ministers of Education, Canada, 2005b, p. 138). Generally, a larger proportion of students in Quebec performed at a level consistent with or higher than the Canada-wide norm; whereas a larger proportion of students (both age groups) in the Northwest Territories performed below the Canada-wide norm, but only 13-year olds in the Yukon performed below the Canada-wide norm (Council of Ministers of Education, Canada, 2005b, p. 138). The variable, single-parent families, may have served as an imperfect proxy for poverty. Because women’s incomes rose in some jurisdictions and different custodial relationships were emerging, the relationship between poverty and living in a single-parent family began to erode in 2004 (Council of Ministers of Education, Canada, 2005b).
The Cycle Ends

The project has come to a close. The language used by CMEC to describe the project - what they learned and how they viewed and valued learning - shifted from an academic outcomes-focused measurement designed to provide measures of quality-assurance assessment to a relationship study focused on the development of an indicators program. In 1993, CMEC used language characterized by statements of certainty and absolute measurements of students’ academic performances. By 2004, this certainty had transformed into complexity and ambiguity. The measurement of academic performance became at best a limited estimate of some elements of commonly shared Canadian curricula, and the priority and importance of the sophisticated contextual questionnaire had risen. For SAIP – 2004, CMEC dedicated considerable resources to the development, administration, and analysis of the contextual questionnaire in its quest to determine what really impacts learning.

CMEC reconceptualised its understanding of learning and the measurement of learning process and its outcomes based on Haertel, Wang and Walberg’s theorizing about the learning process. The data gathered from the complement of contextual questionnaires rose in importance in relation to the academic assessments as CMEC researchers and assessment designers developed an appreciation for the complexity of the learning process. Consequently, the contextual questionnaires evolved into a critical element in the development of a broadly-defined learning indicators program. Thus, the certainty and singularity of academic performance measurement was replaced by the ambiguity and complexity of correlational analysis designed to determine possible relationships between environment factors and student performance in an ever-evolving
learning context. In 1993, CMEC displayed the SAIP assessment results with confidence based on sample size and the attention paid to the assessment design constructed through collaborative and cooperative inter-jurisdictional processes. However, by 2004, CMEC expressed less certainty about the outcomes of its SAIP assessment and assured its users that the measurements were just estimates. CMEC emphasized that learning environments are complex but critical to the production of learning outcomes, and these learning environments are the product of a complex interplay of multiple factors that affect student performance in some ill-defined manner.

Contextual Conversations in the First Decade of the New Millennium

The philosophical conversations at the beginning of the millennium took the form of a retrospective analysis of the consequences to the delivery of education in several Canadian provinces resulting from the embrace to neoliberal ideologies, corporatism and technopositivism. An outcome of this philosophical shift was demand for choice, options for the private delivery of educational services, and an increased accountability for both public and private providers of educational services. Against these philosophical background conversations, scholars and commentators examined the functionality of the SAIP assessment tools, and reviewed interprovincial collaborative curriculum development initiatives. CMEC supported several research projects designed to examine the gap in reading performance outcomes of rural and urban students and the academic performance of children and youth deemed as at-risk.
Philosophical and ideological conservations continued to focus on the relationship between the delivery of educational services and its interface with the economy and social technologies of choice. Riddell (2001), a Canadian economics professor, described the relationship between the economy and theory of human capital development and the provision of education. Basu (2004), a Canadian geography professor, focused on the effects of neoliberalism and the production of education policy in Ontario. Robertson (2003) offered a criticism of the growing acceptability of technopositivism, as an outgrowth of corporatism into the sphere of education. Mitchell, (2001) directed her discussion toward the evolution of the Canadian educational system and its interface with the Asian immigrant population in British Columbia. Davies (2004), an education professor, reviewed the Canadian appetite for private school and tutoring options, while Hepburn and Van Belle (2003), consultants with Fraser Institute, conducted an analysis of the availability of choice in the Canadian elementary and secondary school market.

At the beginning of the new millennium, the effects of globalization of production increased the demand for skilled workers and altered the types of skills current employees required; thus, the skill and wage gap widened (Riddell, 2001). Consequently, economists renewed their interest in the theories of human capital development and the role of human capital in the “creation of new knowledge and in the growth of living standards over time” (Riddell, 2001, p. 485). By placing emphasis on human capital development in Canada’s social policies, the Martin government reinforced the view that “education and training may ameliorate pressures for widening the inequality in economic and social outcomes” for Canadians (Riddell, 2001, p. 486). Because human capital development
was a key government priority and because publicly-funded education consumed a large portion of provincial and federal budgets, it was critical to determine how well the Canadian educational system performed in relation to other systems throughout the world (Riddell, 2001). According to Riddell (2001), Canada has invested a “substantial amount on education ... second highest among G7 countries” (Riddell, 2001, p. 488). The educational attainment of Canadian students was high, about 80 percent of Canadian students graduated with high school diplomas; however, this level of success was “substantially below the United States (87% of the population have high school diplomas) (Riddell, 2001, p. 492). Although Canada’s population was well educated and many individuals completed post-secondary educational programs, Canada’s educational system at the beginning of the millennium lagged behind “the United States in two key dimensions: the fraction of the population with completed secondary education and the proportion of the population with university degrees” (Riddell, 2001, 493). Riddell’s analysis revealed two key findings:

the human capital of college/trade school graduates who did not complete high school may be very similar to those whose highest educational attainment is a high school diploma (p. 494), [and]Canada’s provincial education systems have “forgiving” features and provide various routes to a community college diploma or trade school certificate. (Riddell, 2001, p. 495)

Riddell extended his analysis to include the performance of elementary and secondary students on international assessments such as TIMSS and concluded that “Canadian student achievement in mathematics is about average among a diverse set of countries”; however, Riddell did identify issues related to over sampling in several jurisdictions (p. 499). Oversampling, from Riddell’s perspective, may have inflated the performance of Canadian students on TIMSS. In conclusion, the role of education was critical in the
development of a skilled workforce and in sustaining prolonged high-level economic activity; however, some economists were concerned that educational inputs for the publicly-funded education systems did not result in the production of a sufficiently high rate of high school diplomas and their subsequent transformation in university degrees.

Basu (2004) moved the site of his analysis of educational ideology to the provincial level in his review of the significant changes to the delivery of education in Ontario. According to Basu, these changes were predicated on the construction of a myth of a “failing and inefficient education[al]” system (p. 621). The need to restructure the Ontario education system was motivated by economic issues: the need to cut costs and improve efficiency, raise standards and serve the economic need to “remain globally competitive” (Basu, 2004 p. 621). The application of neoliberal ideology denied the diversity of the socio-economic composition of Canadian neighbourhoods, failed to recognize the primary purpose of schools as the need to educate, and obfuscated the “essentially political nature” of schools (Basu, 2004, p. 625). The neoliberal ideology and governmental policy and regulations were used to justify and administer the centralization of the delivery of educational services, the introduction of standardized assessments at specific grades, and the release of test results to the media (Basu, 2004). Basu described the release of test results to the media as a Foucauldian “’technique of power’ [that] allowed individual schools to police themselves within the public realm” (p. 628). Although Riddell (2001) justified the economic need for schools to serve the state’s need for human capital, Basu (2004) questioned the need to incur the high social costs of educational reform justified by market-driven solutions emerging from a neoliberal ideology.
Mitchell (2001) moved the discussion of provision of educational services, the ideology of neoliberalism, and the issues of globalization to the personal level of new immigrants to Canada and their interface with the Canadian educational system. According to Mitchell, neoliberalism manifested itself in the British Columbia educational system through the manifestation of “language of choice and the push for charter schools” and the promotion of individuality (p. 61). The effects of globalization and immigration introduced immigrants to the Canadian school systems who brought the perception that “education quite strategically [was] the single most important way of achieving and retaining this competitive edge” related to status and prestige (Mitchell, 2001, p.60). A conflict waged in the newspapers centred on the expectation expressed by recent immigrant families that Canadian schools should equip their children with the skills to live and “work competitively within an increasingly competitive society” (Mitchell, 2001). From the perspective of these recent immigrants, students’ demonstration of “achievement, organization, and submission to authority” was more important than the development of assertiveness and individuality (Mitchell, 2001, 65). In particular, Chinese parents were frustrated by the “lack of standardized testing” as compared to the monthly testing administered in Hong Kong, and “they were fearful that their child might fall behind and neither the parents nor the child would be aware of it” (Mitchell, 2001, 67). Mitchell’s discussion of the expectations of the Canadian school system shed a different light on educators’ concerns about the negative effects of standardized testing on children whose first language was not English. It would seem from Mitchell’s discussion that immigrant parents wanted more testing. The test results
provided assurance that their children were learning traditional reading, mathematics, and science knowledge and skills.

Hepburn and Van Belle (2003) explored the relationship between provincial funding for education, class size, school performance based on international test scores and parental choice. According to Hepburn and Van Belle, most provincial governments have established policies that limit parents’ choice in selecting educational institutions for their children, a factor that they deem to be important in the production of educational success. Using the provision of educational services in Alberta as a case study, Hepburn and Van Belle noted a counter-intuitive relationship among the variables that prescribe the levels of inputs and the degree of parental choice with the level of performance indicators. Alberta has funded more options for educational delivery - home schooling, independent schools, charter schools, and public schools - than any other province (Hepburn and Van Belle, 2003). According to Hepburn and Van Belle’s review of PISA (2000) and SAIP (2001, 2002 and 1999) test results, students in Alberta have consistently performed at higher levels than their peers in other provinces; however, Albertans spent “considerably less per pupil than Ontario, Manitoba, Quebec, or British Columbia” on the delivery of educational services, and their schools posted larger class sizes than those of several other provinces (p. 17). However, Albertans did have the “highest personal income in Canada”, and the level of SES has been positively correlated with the production of higher test scores (Hepburn and Van Belle, 2003, p. 17). Hepburn and Van Belle hypothesized that high achievement might be linked to the “flexibility school administrators and parents have to solve the educational problems of their children” (p. 17) because parents can chose the right school for their child’s needs. However, the high
test scores and lower cost per student might be a combination of several factors: shared
public and private funding for some school alternatives; increased affordability of private
school options due to higher family incomes; and the focus of private school curriculum
on academic performance measured by standardized testing.

Davies (2004) cited the growing recognition by parents and of students that the
pursuit of higher education was necessary to avoid “the fringes of labour market” (p.
233). This expectation has driven more parents to exercise their right to choose and to
enrol their children in private schools (an increase of 20% nationally) and in tutoring
services (Davies, 2004). Although Canada lacks the high stakes testing environment of
the United States, the demand for tutoring services has risen in response to the demand
for personalized services and competitive advantage (Davies, 2004). In many ways, the
decision to purchase private tutoring is an ideological one emanating from a neoliberal
ideology prevalent at the turn of millennium. This ideology supports the availability of
private options and parental choice in delivery of public services, like education.

Robertson (2003) entered the ideological discussion from a different vantage
point and focused her comments on the purported positive relationship between the
amount of time that students spend on computers and academic achievement. According
to Robertson’s research, this relationship is questionable although corporations, like
Apple, continue to promote learning through technology as a means to raise achievement
standards. Robertson described the actions of these corporations as “bullying” because
their subtext has laid out the belief “that depriving children of maximum access to it
[technology] is tantamount to child abuse or at least education abuse” (P. 283). According
to Robertson, technopositivism has become part of the taken-for-granted subtext of how
schools get things done and that challenging technopositivism would compromise “not only the credibility of individual teachers but also the validity of schooling itself” (p. 283).

These commentators approached the effects on the neoliberal agenda from different vantage points and offered commentary on issues such as school choice; educational achievement, competition and employment; the role of technology, and the erasure of neighbourhood diversity. These conversations were both reactive to the events of the previous decade and proactive in their documentary capacity to note, record, and make visible the subtext of the organization of Canada’s schools.

**Meta Research on SAIP**

Writing about SAIP emerged during this decade; commentators and researchers had begun to examine the socio-political implications of administering a large-scale test like SAIP, the functionality of the test items, and their capacity to yield the data included in the SAIP reports. Crocker (2002b), a Canadian education professor, examined both the functionality of SAIP and its utility as a tool to measure the effectiveness of the Canadian educational system as a whole. Leighton and Gokiert (2005), Canadian education professors, examined functionality of SAIP science test items, and Gierl (2003), a Canadian education professor, conducted a differential item analysis of the functionality of SAIP science test items with respect to language. Emenogu and Childs (2005), Canadian education professors, conducted differential item analysis of the functionality of the SAIP mathematics items with respect to language.
Crocker’s (2002b) analysis of the Canadian education system focused on the identification of research sites that would produce knowledge about the “nature of educational outcomes and their measurement” within a “theoretical framework of human capital” development (p.1). Crocker explored two key questions: “which outcomes are considered desirable?” and what should the basis for benchmarks be and how should targets for improvement be set? (p. 3) Crocker examined the use of expectations-setting panels in the SAIP assessments as an example of setting relatively precise targets as opposed to broader more general benchmarks. After examining PISA, TIMMS, and other provincial assessments systems, Crocker concluded that “[i]ndicator systems have not yet reached their potential for public accountability or for explaining particular observed results” and noted that most systems, like SAIP, have confined their activities to “compiling descriptive and comparative data on such matters as system structures, enrolment, finances, levels of attainment, and achievement” (p. 21). According to Crocker, there was a need for an “indicator system to begin to serve an analytical rather than descriptive and comparative purpose” (p. 21). Crocker acknowledged the criticisms levied against large-scale assessment programs related to their inability to measure “many important educational outcomes” (p. 25). However, according to Crocker, the comment has failed to recognize the potential of large-sale assessments is focused on “the current state of assessment than about any inherent features of either outcome statements or assessment techniques” (p. 25). As Crocker looked to the future of Canadian assessment programs, in particular SAIP, he noted that “a strong case can be made for redirecting SAIP toward other goals” as there was considerable overlap between SAIP and PISA.
Indicator programs, like SAIP, can serve a useful function in the effective production of human capital through the efforts of institutionalized public education systems.

Gierl (2003) examined the two-stage SAIP science assessment (a placement test followed by a performance test) from the perspective of functionality. Gierl (2003) questioned the capacity of science assessment to function equally well for students writing in English and in French. Although, “the tests were administered and scored with the implicit assumption that two-stage tests produce equivalent results across language groups,” the possibility for error due to translation could increase in two-stage tests” (Gierl, 2003, p. 3). After conducting both a statistical and differential item analysis on the stage one test followed by a review by a group of translators, “the stage-one items were [found to be] free of flagrant translations problems” (Gierl, 2003, p. 18). The subsequent review of the stage-two items revealed that these “worked well for some age and ability groups but not others” (Gierl, 2003, p. 18). Since the stage-one test determined a student’s subsequent stage-two test, the stage-one test’s discriminatory power must be high (Gierl, 2003). However, the findings revealed that “the first-stage tests did not provide strong separation between the two ability levels for either language group in either the 1996 or 1999 administrations” (Gierl, 2003, p.19). Thus, some students may have been placed in a level 3-5 test booklet (higher level) in error while others were inaccurately placed in a level 0-2 test booklet (lower level). The stronger students placed in the level 0-2 booklet would not have opportunity to place beyond level 2, and the weaker students placed in the level 3-5 booklet could not have scored below level 3. Gierl’s work did not explore the distribution of the misplaced students to determine if
they were clustered by some other factor(s); the effects of the interplay of multiple factors would have compounded the effect on the tests results of the misplaced students.

Leighton and Goikert (2005) approached the SAIP science test from the perspective of the use of words and phrases in the problem-solving items. The goal of their research was to investigate students’ comprehension of standardized test items and to use a cognitive framework to identify ambiguous test items. Their research was predicative on the notion that developing awareness of “one’s own comprehension is a powerful metacognitive skill” (Leighton & Goikert, 2005, p. 3). Thus, students who could verbalize their thoughts while they were solving problems on the SAIP science test would have had a metacognitive advantage (Leighton & Goikert, 2005, p. 3). During the research study, students struggled with the task of verbalizing their thoughts as they worked through the SAIP science questions and problems (Leighton & Goikert, 2005). The challenge with this task might have been related to the need to think out loud or it may have been “indicative of the cognitive processing elicited by the test item” (Leighton & Goikert, 2005, p. 14). Some test items drew on less sophisticated cognitive processes, and students could use one-step problem solving approaches; however, the test designers rated the complexity of the test items at a much higher level (Leighton & Goikert, 2005). The research yielded the following findings: “students had difficulty describing the objective of the problem in their own words”, although most students could “describe the specific knowledge and skills needed to answer a specific item” (Leighton & Goikert, 2005, p. 14). Leighton and Goikert concluded that both item ambiguity and item ability level predicted students’ uncertainty about comprehending the item, but only item ability level was related to students’ performance on the SAIP science test. Leighton and
Giokert’s exploration of metacognition and the quantification of student’s awareness about their thinking during a problem-solving activity dovetailed well with the Habits of Mind secondary study, part of the SAIP Writing III assessment.

In response to the call for additional research on the both the SAIP assessment tools and the contextualizing questions, Emenogu and Childs (2005) conducted a differential item analysis on the geometry questions in the SAIP III mathematics assessment to determine if translation significantly affected performance for French-speaking students living outside of Quebec. Their sample consisted of test items responses submitted by French- and English- speaking Ontario students (Emenogou & Childs, 2005). Emenogou and Childs (2005) reported that there were no statistically different response patterns for either linguistic group and that none of the test items contained structural bias.

The differential item analysis conducted by Emenogou and Childs (2005) and Gierl (2003) confirmed that the SAIP developers did exercise considerable caution in ensuring that the French and English versions of SAIP mathematics and science test items were comparable; however Gierl (2003) revealed that stage-one science test items functioned differently for different ages and abilities – a significant finding. Leighton and Goikert (2005) revealed that test items may have functioned at a level below their difficulty rating, thus, disrupting the test results and compromising the findings. These researchers have accepted CMEC’s invitation to dig beyond the surface and to disrupt the SAIP data. Their research has revealed new understandings that exist below the surface, and they have searched for meaning and produced new knowledge about learning, teaching, and assessment.
Cartwright and Allen’s (2002) work stood out because they choose the rural–urban gap in reading performance on PISA. Although PISA is a different testing instrument, the issue of differential performance based on school size has been explored on several SAIP contextual questionnaires and has revealed a similar trend – students attending larger schools in larger urban centres tended to perform at higher levels than their peers attending smaller schools in smaller centres. According to Cartwright and Allen, “rural and urban schools are much the same [and] Canadian students ranked high internationally, in part, because there are few differences”; however, Newfoundland and Labrador, Prince Edward Island, New Brunswick and Alberta showed the largest gaps in performance (p. 5). The first level of analysis examined factors attributable to the family: differences in the educational level of the parents, books, cultural possessions held by the families, and differences in participation rates in cultural activities (Cartwright & Allen, 2002). While these differences in “individual and family characteristics explain some of the rural-urban difference in student reading performance, their explanatory value is small” (Cartwright & Allen, 2002, p. 18). The secondary analysis examined the “relationship between a variety of community factors and the remaining rural-urban gap” to reveal that differences in employment rates, the average educational attainment of the adult population and the average educational status of all the parents of 15-year olds in schools” (Cartwright & Allen, 2002, p. 19). The relationship between these factors and the level of students’ reading scores on PISA was much stronger. Cartwright and Allen’s research is significant because it points to a gap in the structure of the SAIP questionnaire and the accompanying contextual analysis. The SAIP contextual analysis of factors
affecting learning has focused almost exclusively on the student, the school, and the family and has ignored the significant impact of the community on learning performance.

Pan-Canadian Assessment Program: PCAP-13 2007

The Pan-Canadian Assessment Program (PCAP) was the result of substantial rethinking and reengineering of SAIP informed by the structure and format of international assessments, PISA and TIMSS. The structure of the one-stage PCAP-13 assessment was modelled after PISA; each administration contained an assessment for reading, mathematics, and science - a substantial change from the format of SAIP (Council of Ministers of Education, Canada (2008d). Consistent with the structure of the PISA assessment, one domain was designated as the major, and the others were treated as minor domains. Reading was the major domain for PCAP-13 2007, and science and mathematics were the minor domains (Council of Ministers of Education, Canada (2008d). Assessment developers ensured that the PCAP-13 assessment mirrored the key structural elements and assessment criteria of the PISA and TIMSS assessments so researchers and policy makers could conduct cross-assessment comparisons with some degree of confidence (Council of Ministers of Education, Canada (2008d). The PCAP assessment measured the performances of 13-year olds living in the Canadian provinces and the Yukon (Northwest Territories and Nunavut did not participate).

CMEC revised the assessment title, the words that speak the assessment into existence. Just as the words are substantively different from those used to describe the former assessment program, SAIP, the structure and criteria for the PCAP assessment are also substantively different. CMEC has dropped its attempt to manage a complex, cross-
age group, criterion-referenced assessment for a streamlined, norm-referenced, single-age
group assessment. CMEC did not explain its decision to rename its assessment; however,
the assessment title more closely mirrors PISA, the tool CMEC and HRSDC have chosen
to provide metrics on the performance of Canadian 15-year olds. In every sense, the shift
from a Canadian-generated assessment tool to one informed by international assessment
protocols represents a Foucauldian rupture in the way national large-scale assessment is
thought into existence.

**PCAP-13 2007: The Academic Assessment**

PCAP-13 2007 administration included the delivery and scoring of two different
academic assessments (reading assessment and a mathematics and science assessment) to
samples of French and English students. These assessments were accompanied by three
contextual questionnaires administered to the students completing the reading
assessment, their teachers, and principals (Council of Ministers of Education, Canada,
2008c). The data analysis for the PCAP-13 2007 academic assessment and the contextual
questionnaire changed substantially. PCAP-13 2007 included an additional report
focusing solely “on particular difficulties in French literacy that have been documented in
several national and international studies” (Council of Ministers of Education, Canada,
2008c, p. 7).

PCAP-13 2007 was designed to generated data to address similar questions to
those posed on the SAIP 1993 mathematics assessment:

- how well [are Canadian] education systems ... meeting the needs of students
  and society ...
- to determine whether students across Canada reach similar levels of
  performance in these core
disciplines at about the same age” (Council of Ministers of Education, Canada, 2008a, p. 1).

CMEC recognized that the PCAP-13 2007 was not the only large-scale assessment that most students complete; therefore, the need to “complement existing assessments in each jurisdiction” was an important factor in setting the priorities for the generation of Canada-wide comparative data (Council of Ministers of Education, Canada, 2008a, p. 1). Many of the CMEC assessment goals remained consistent with its administration of SAIP; however, CMEC recognized that testing has become a burden for many schools and teachers and has identified the need to “reduce the testing burden on schools through a more streamlined administrative approach”; thus the PCAP-13 is a 90-minute assessment as opposed to the 180 – 360 minutes dedicated to the SAIP assessments (Council of Ministers of Education, Canada, 2008a, p. 1).

The development of PCAP-13 began in April 2003 and followed a similar development cycle to SAIP: the generation of a concept paper and the establishment of bilingual working groups to develop the assessment frameworks and the test items (Council of Ministers of Education, Canada, 2008a). The bilingual working groups placed a priority of the development of test items and texts “in both official languages [that were] equivalent in meaning and difficulty” (Council of Ministers of Education, Canada, 2008a, p. 2).

The sample of 30,000 students was selected from a random sample of schools; and from each selected school, the students were selected from the population of eligible students (Council of Ministers of Education, Canada, 2008a, p. 4). The student sample was subdivided into two unequal parts: those who wrote the reading assessment (15,000 in English and 5,000 in French) and those who wrote the science and mathematics
assessment (7,500 in English and 2,500 in French) (Council of Ministers of Education, Canada, 2008a, p. 4). The sample size and proportions were consistent with the total sample for each SAIP administration; however, the sample was larger than the average SAIP administration for a 13-year-old cohort. Parents could refuse to allow their children to participate in the assessment: The highest rates of refusal were in Quebec for both English (19%) and French (27.4%), and the lowest rate of refusal was in Yukon (0%) (Council of Ministers of Education, Canada, 2008a, p. 131). Schools could exempt students if “their limited proficiency in reading [was] deemed to be below level 1” (Council of Ministers of Education, Canada, 2008a, p. 24). The jurisdiction with the highest level of exemption was Prince Edward Island (French) at 12.5%, and the lowest level of exemption was Yukon at 0% (Council of Ministers of Education, Canada, 2008a, p. 131). The actual effect of level of withdrawals from the performance of individual jurisdictions and the Canada-wide sample was not determined.

About 85% of the student sample participated in the assessment, and in an effort to garner “the most positive participation and engagement in the process by teachers, students, and parents,” CMEC distributed pamphlets for parents and students and handbooks for school principals (Council of Ministers of Education, Canada, 2008a, p. 5). In previous SAIP assessments, the task of motivating and preparing students was left to teachers.

The scoring process incorporated the safeguards of other SAIP administrations to ensure fairness; however, the scores were no longer reported on a 5-level scale. The scores were converted to a standardized score with a mean of 500 and a standard deviation of 100, consistent with the presentation of PISA test scores (Council of
Ministers of Education, Canada, 2008a). The expectation was that “two-thirds of all participating students [would fall] within the range of 400 to 600 points ... a “statistically normal distribution’ of scores” (Council of Ministers of Education, Canada, 2008a, p. 11). The change in scoring practice was justified as “’scale scores’ are used to interpret more accurately the performance of students in each assessment from one administration of the assessment to another” (Council of Ministers of Education, Canada, 2008a, p. 11). CMEC used interpretative accuracy as the justification for abandoning of the criterion-referenced framework of SAIP for the norm-referenced framework of PCAP.

The anonymous expectations-setting panel composed of educators and non-educators was abandoned in favour of a bookmarking process involving a group of outside validators (Council of Ministers of Education, Canada, 2008a). The external groups set the “cut scores” or standards for levels 1 (0-379 points), level 2 (380-575 points) and level 3 (576 points and above) based on their understanding of how well 13-year-old students should be perform in reading, mathematics and science (Council of Ministers of Education, Canada, 2008a). Although the group of experts was described as from the “outside,” their relationship to the Canadian educational systems, teaching, and the typical curriculum that 13-year-olds were expected to master was unclear.

CMEC described reading a “dynamic, interactive process” in which the “interaction of reader, text, purpose, and context” would occur before, during and after the actual reading experience (Council of Ministers of Education, Canada, 2008a, p. 7). According to CMEC’s interpretation of the reading process, a reader must “make a connection with a text” (p. 7) and identify what they bring to that text; engage and make meaning from both non-fiction and fiction genres; differentiate one’s purpose from that
of the author; consider the context in which the writing was situated; and recognize that reading is a “process of continuous growth” (Council of Ministers of Education, Canada, 2008a, p. 8). CMEC expanded its definition of reading for the PCAP -13 assessment to include the consideration of how students approach reading tasks because the measurement of this additional factor might provide insight into the critical stages at which the reading process might break down. This focus was not addressed in typical Canadian reading curricula that have identified comprehension, interpretation, and response to text as their “major organizing elements” (Council of Ministers of Education, Canada, 2008a, p. 8). The design and development of PCAP-13 reading assessment was based on the assumptions that becoming a flexible reader and learning to “apply reading skills and effective strategies whenever they [students] read a text” were critical components of effective reading in everyday life (Council of Ministers of Education, Canada, 2008a, p. 10).

The results for PCAP -13 were presented in a somewhat different manner: the standardized mean score and confidence interval for each jurisdiction were presented in comparison to each other jurisdiction and the Canada-wide mean. Students in Quebec scored above the Canadian mean; students in Ontario scored at the mean; and all the other jurisdictions scored below the mean (Council of Ministers of Education, Canada, 2008a, p. 19).

When the scores for all the jurisdictions were re-aggregated by language and Canada-wide means re-calculated based on language, students in Ontario (English) performed significantly better than the Canada-wide (English) mean, but students in Quebec did not perform significantly better than the Canada-wide (French) mean;
however, francophone students in Quebec comprised the majority of the Canada-wide French sample (Council of Ministers of Education, Canada, 2008a). Generally, across all jurisdictions, approximately 81-90 percent of the students scored at level 2 or higher, and students in five jurisdictions exceeded the “Canadian expectation for this age group” (Council of Ministers of Education, Canada, 2008a, p. 24).

Particular attention was paid to gender-based differences because these differences in “assessment results in reading tend to be more significant than in the results in science and mathematics” (Council of Ministers of Education, Canada, 2008a, p. 26). The differences in the results for PCAP-13 were “consistent with international, Canadian, and provincial testing during the past decade” (Council of Ministers of Education, Canada, 2008a, p. 26). More male students performed at level 1 and more female students performed at levels 2 and 3 (Council of Ministers of Education, Canada, 2008a). The level of analysis of gender-based differences was drilled down to the level of the sub-domains: comprehension, interpretation and response to text (Council of Ministers of Education, Canada, 2008a). Students demonstrated the smallest gender-based Canada-wide discrepancy in comprehension (8 points) and the significantly larger discrepancies for interpretation (26 points) and response to text (25 points) (Council of Ministers of Education, Canada, 2008a, pp. 126-127). The discrepancies for Alberta were consistent with the Canada-wide mean while the discrepancies for Quebec were significant and much higher than the Canada-wide mean (Council of Ministers of Education, Canada, 2008a). Students in Newfoundland and Labrador demonstrated the largest gender-based discrepancies of all the jurisdictions (Council of Ministers of Education, Canada, 2008a, pp. 126-127). Like the SAIP assessments, jurisdictions –
Alberta, British Columbia, Ontario, and Quebec - in which male students performed at a level consistent with or above the Canada-wide mean, turned in better performances overall.

Providing the performances scores by sub-domain added a level of analysis that extended the utility of the scores to curriculum development and the teaching and learning process. For example students in Quebec performed equally on all three sub-domains while students in Prince Edward Island were substantially weaker in all three domains; however, they had developed much stronger reading comprehension skills than interpretation or response to text skills (Council of Ministers of Education, Canada, 2008a, p. 27). Curriculum developers and teachers who were to glean this level of understanding about their students’ reading performance in their jurisdiction could load the curriculum with learning activities that supported the development of interpretation and response to text skills and strategies.

The assessment of mathematics and science were also part of PCAP-13; however, the level of the analysis of student performances was much more limited than that for reading. The results were presented as a rank-order chart by jurisdictional mean with each confidence interval included (Council of Ministers of Education, Canada, 2008a).

Students from Quebec, Alberta, and Ontario performed at higher levels than students from other provinces on the PCAP-13. The report described the performances of these students as follows:

Considering confidence intervals, the mean score for Quebec students in mathematics [and science are] ... significantly higher than that obtained by Canadian students overall. (Council of Ministers of Education, Canada, 2008a, p. 35)
Mean scores for Ontario ... students [in mathematics and science] are not statistically different from that of Canadian students overall. (Council of Ministers of Education, Canada, 2008a, p. 35)
Mean score for ... Alberta students [in mathematics is] not significantly different from that of Canadian students overall [but] the mean score for the students in Alberta ... in science is significantly higher than that obtained by Canadian students overall. (Council of Ministers of Education, Canada, 2008a, p. 35-36)

However, when the results for Quebec students in mathematics [and science] were adjusted for language, the difference in performance was no longer statistically different from the Canada-wide mean; but “the mean score for Ontario students responding in English in mathematics is significantly higher than that obtained by Canadian students” (Council of Ministers of Education, Canada, 2008a, p. 37). The report noted that the differences were statistically different but did not comment on the degree of real educational difference

The streamlining of the PCAP-13 assessment – a one stage assessment, one age group, normative data reporting, reduced assessment length, and increased sample size - resulted in a targeted assessment directed at measuring the overall performance of 13-year olds in reading and their performance on each of the reading sub-domains. The assessment developers focused on gender-based differences in performances and how gender-based differences were produced in each jurisdiction. The improvements to data collection practices within the jurisdictions generated a number of questions related to contextual factors that led to higher or lower levels of parental and student withdrawal from the assessment protocol, encouraged more Quebec parents to withdraw their children from the study, and led to global participation in the Yukon. The study did not measure the effectiveness of CMEC’s public relations campaign designed to motivate and encourage the participation of students and parents. Although PCAP-13 2007 represented
the 13th large-scale assessment administered by CMEC, this iteration was the first to include a statement of public appreciation to the students, teachers, and administrators by noting that “the quality of your commitment has made this study possible” (Council of Minister of Education, Canada, 2008a, p. ii). This statement represents CMEC’s recognition that schools, teachers, and students have chosen to engage in this social technology because they believe that there may be some useful production of knowledge, truth, or order.

*PCAP-13 2007: The Learning Context Analysis*

School principals, language arts teachers, and students participating in the PCAP-13 academic assessment process completed a contextual questionnaire. The questionnaire - informed by those used on the PISA, SAIP, and TIMSS assessments - was reviewed and modified by the PCAP working group (Council of Ministers of Education, Canada, 2008b). The result was a streamlined questionnaire focused on reading and consistent with “predetermined theoretical constructs or important research questions” (Council of Ministers of Education, Canada, 2008b, p. 2).

The model for learning (input → process → output embedded within a context) introduced in the SAIP Science III assessments did not form part of the theoretical discussion for the PCAP-13 contextual questionnaire. The PCAP-13 assessment referred to the theoretical underpinnings of the PISA assessment to guide its questionnaire development. PISA drew on Carroll’s 1963 time-based conceptual model of learning that described the “degree of learning or achievement (L) [as] a function of the time actually spent on learning (Tsl) to the time needed to learn (Tnl), or”

\[ L = f(Tsl)/Tnl \]  

(OECD, 2010, p. 24)
The factors influencing the student’s decision about how much time to invest in a learning activity ($T_{sl}$) included these factors: “self-concept, motivation, attention and meta-cognitive processes (strategies for learning)” (OECD, 2010). According to the model’s parameters, the time needed to learn ($T_{nl}$) can be influenced by the degree that “teaching strategies match learning styles” (OECD, 2010, p. 23). The model provided the potential to develop a mathematical algorithm for the quantification of learning as a rational function of time spent (the composite of factors governing ability, aptitude, and perseverance) and time needed to learn (a composite of the individual’s cognitive ability, the effectiveness of instructional strategies and the difficulty of the content). Although the model for learning presented in the PISA documents is not a mathematical formula, it does offer a relational organizer for the variables affecting the productive outputs of the learning process.

The re-thinking of the conceptual model for learning informed the transformation of data analysis from a purely descriptive and comparative activity to bivariate and multivariate analyses to determine the loading of each variable in relationship to its effect on performance (Council of Ministers of Education, 2008b).

Demographic variables were measured and treated as covariates, as they are generally not within the control of the school system, and their influence on achievement is considered to be largely independent of educational policy or practice. That is not to say that the system should not take these variables into account. (Council of Ministers of Education, Canada, 2008b, p. 11)

The structure of analysis employed a multi-layer sampling model in which “schools [were] sampled at the first stage and students within schools at the second stage” (Council of Ministers of Education, Canada, 2008b, p. 38). Regression coefficients in this analysis could be “interpreted as representing measures of change”; both relationships
between a “single predictor and the outcome” with and without controlling for the effects of other variables (Council of Ministers of Education, Canada, 2008b, p. 38). Both the language used to describe the statistical analysis and the complexity of the data analysis served to illustrate the transformation of the questionnaire outputs from SAIP Science III to PCAP-13 2007.

The intention of the following section is not to paraphrase the results of the contextual report but to provide examples of key areas of interest related to the demographics and the analysis of factors affecting learning. The intention is to provide a sense of the language that CMEC used to speak these understandings into existence.

A major focus of the reading assessment was the effects of gender on reading performance. Although the expectation was that gender would be distributed equitably throughout the jurisdictions; this was not true; more females wrote the assessment in Yukon (55%), Quebec (53%), Newfoundland and Labrador (52%) and Alberta (51%); thus, “differences in male and female participation rates in the test may have had a small impact on the reported performance levels” (Council of Ministers of Education, Canada, 2008b, p.13).

The PCAP-13 contextual survey approached the measurement of socio-economic status through the proxy measures of “mother’s education and the number of books in the home” (Council of Ministers of Education, 2008b, p.19). The analysis yielded small differences:

Minority French populations in the western provinces and Nova Scotia show slightly higher levels of mother’s education than the Majority English populations in the same jurisdictions, while francophone population in Quebec and New Brunswick show slightly lower levels of mother’s education than their Anglophone counterparts. (Council of Ministers of Education, Canada, 2008b, p.19)
The higher levels of mother’s education for students living in the Western provinces may be consistent with the student cohort who was most likely to select French immersion programming in predominantly Anglophone communities. The relationship between reading performance and the number of books in the home was also statistically significant for students living in homes with more than 200 books (Council of Ministers of Education, Canada, 2008b). The data analysis did not evaluate the combined effect on reading scores for males who had more than 200 books in the home and a mother with a university education.

Attributes of teachers – gender, experience, and educational preparation – were considered in relation to students’ performances. While teachers’ gender did not produce a statistically significant difference; however, “higher reading scores are attained by students taught by most highly experienced teachers” (Council of Ministers of Education, Canada, 2008b, p. 26). Differences in qualifications produced significant differences for students taught by teachers who have a B.Ed. with another undergraduate degree; however, other combinations of degrees including graduate degrees did not affect the performance of students significantly (Council of Ministers of Education, Canada, 2008b).

School size and its relationship to performance were measured on previous SAIP assessments. The trend for larger schools (more than a 1000 students) to produce better results continued in the PCAP assessment (Council of Ministers of Education, Canada, 2008b, p. 32.). The questionnaire also captured data on the frequency and distribution of private schools; in Canada the number of private schools is relatively small in most jurisdictions, “with the notable exception of Quebec, where both language groups show
close to one-third private schools, and British Columbia, where the proportion exceeds 20%” (Council of Ministers of Education, Canada, 2008b, p. 32). According to the results aggregated by school type, “students in private schools [mean 523] significantly outperform those in public schools [mean 490]” (Council of Ministers of Education, Canada, 2008b, p. 33). The issue took on new meaning when the results for francophone students in Quebec were reviewed. Francophone public school students “performed at a level close to the Canadian average”; however, Quebec private school students performed at a level much higher “than any other group identified in the comparison thus far” (Council of Ministers of Education, Canada, 2008b, p. 33). The concentration of private schools, the strength of the private school performances, more females writing in Quebec, and a larger proportion of eligible students withdrawn by their parents – these factors contributed to the construction of the phenomenon labelled the Quebec factor.

Controlling for the effects of both mother’s education and the number of books present in a student’s home made “little difference to the size of the public school/private school difference” (Council of Ministers of Education, Canada, 2008b, p. 33).

The contextual survey began to put together some perspective on the differential effects of participation of Aboriginal students in the PCAP-13 assessment. Saskatchewan, Manitoba, and Yukon reported the largest proportion of schools with Aboriginal student populations exceeding 25 percent (Council of Ministers of Education, Canada, 2008b). Statistical analysis revealed the effect of the participation of Aboriginal students in PCAP-13 on school reading performance was a depression in the mean score from 502 to 451, a significant difference (Council of Ministers of Education, Canada, 2008b, p. 35). By aggregating the data for Aboriginal and non-Aboriginal students separately, the
magnitude of the difference became evident, and the magnitude of the challenge revealed itself.

These examples, while not exhaustive, provided some insight into the level of attention paid to demographics variables that affect the reading performance of 13-year olds. The analysis moved into the analytical level and shed some light on the production of the Quebec factor, and the challenges facing educators in their work with Aboriginal students and communities. As the descriptions of the contextual factors related to reading process became more interconnected, the task of isolating a particular factor and its effect became increasingly difficult.

Supplemental analysis: Performance of minority/majority language groups on PCAP -13 2007

The purpose of the minority/majority language report was to focus on “the particular difficulties in French literacy that have been documented in several national and international studies” (Council of Ministers of Education, Canada, 2008c, p. 7). Based on the findings of previous SAIP studies, students living in francophone communities outside Quebec have experienced more difficulty with “academic activities involving language processing, but fewer difficulties if any “in mathematics and science if language skills were minimally involved” (Council of Ministers of Education, Canada, 2008c, p. 7). For the purposes of the this study, students were divided into six language groups: Majority English, Quebec English, Quebec or Majority French, New Brunswick French, Ontario French and Small French Minorities (a collective of students writing in French throughout Canada) (Council of Ministers of Education, Canada,
The vast majority of students writing outside the province of Quebec were enrolled in French immersion programs and were essentially bilingual. French literacy continues to be of particular concern outside of Quebec because students often live in homes where French is used little or not at all or in communities in which there is little opportunity to use French (Council of Ministers of Education, Canada, 2008c). Students who have few opportunities to use French outside the classroom tended to have lower French literacy scores (Council of Ministers of Education, Canada, 2008c). This issue coupled with factors predicting socio-economic status were foci for this study.

The analysis conducted on the PCAP-13 results aggregated by the six minority/majority language groups revealed some interesting trends related to the number of books in the home and the mother’s level of education. While number of books in the home demonstrated “a linear effect with regard to reading performance, the effect of the level the mother’s education exhibited a similar tendency to the Canada-wide analysis but was not consistent in some of the groups” (Council of Ministers of Education, Canada, 2008c, p. 20). Because the choice to attend French language school was affected by its availability, some French-speaking students attended English language schools because they were more convenient, or they had a more vital school community (Council of Ministers of Education, Canada, 2008c). Conversely, some English speaking students chose to attend French language schools or French immersion schools because their parents had some perception about inherent value of a second-language education (Council of Ministers of Education, Canada, 2008c). These additional factors served to disrupt the expected relationship between the variables and the outcome.
In general the relationship between language use at home, language use at school, and reading performance on PCAP-13 were “only statistically significant for “the Majority English and West French groups” (Council of Ministers of Education, Canada, 2008c. P. 15). The results for the other populations were not statistically different, although there were “relatively large differences in between home language groups” and reading performance (Council of Ministers of Education, Canada, 2008c. P. 15).

The additional analysis by language group added to the understanding of how language functions in relation to reading performance in Canada. Although the literature predicted that both mother’s educational level and the language used at home would predict achievement on the PACP-13 reading test, the relationship between the variables and performance was not evident for all groups.

The administration of PCAP-13 raised the level of analysis of contextual and performance data from descriptive and comparative to analytical through the introduction of bivariate and multivariate statistical analyses; consequently, our understanding of the functionality of variables influencing the learning process and the complexity of the presentation of their effects increased also. PCAP-13 results have provided useful data for inter-jurisdictional curriculum projects and school-based teaching and learning projects. As yet the capacity and the role of PCAP-13 results in predicting particular levels of success on PISA, high school graduation or participation in university or technical school has not been determined; however, this problem could form the goal of a future longitudinal research project. CMEC has recognized the contributions of key stakeholders and agents in the educational process – parents, teachers, and school administrators - and has acknowledged the importance of their engagement in the
assessment process and with the test results. Although CMEC makes the PCAP-13 reports (and many of the SAIP reports) available to the general public, their use by principals, teachers, parents and possibly students has not been captured on any of the CMEC questionnaires.

Closing Out the Decade: Contextual Conversations – 2004 -2010

The decade is coming to a close, the first administration of PCAP-13 is completed and the second is scheduled for the fall of the 2010. Commentators and researchers are increasingly concerned about the interrelationship between accountability and assessment and the nature of the emerging Canadian philosophy about this relationship. There is a growing recognition by some commentators that these two activities may have some overlap, they are not synonymous. Other commentators have used a different tactic and reviewed the discourses used in education-related websites to determine the level harmonization of mission statements, values, and activities given that the delivery of elementary and secondary education is not coordinated through a federal agency in Canada. Other researchers engaged in the meta-analysis or secondary analysis of large-scale achievement data such as that produced by PISA, SAIP, and PCAP-13. Researchers have re-focused their attention on the concern that teachers and school administrators lack the appropriate preparation to make use of large-scale, standardized assessment results to effect changes in instructional practice in their classrooms and schools.
Dolmage (2004), a Canadian education professor, drew attention to the growing role of international assessments (PISA and TIMSS) and national assessments (SAIP/PCAP) results as evidence for a greater role for accountability in the delivery of educational services by Canadian educational systems.; however, these assessments are considered to be low-stakes – the consequences to students, teachers, and school divisions are minimal. According to Dolmage, these international assessment programs have traditionally ignored the “cultural context and sampling differences” which render the “conclusions invalid” (p. 11). Of greater concern is the increasing trend toward the use of high-stakes school leaving standardized assessments in several Canadian jurisdictions which given the variety of tests and the levels of testing strongly suggests that there is “consensus across jurisdictions concerning what constitutes worthwhile testing” (Dolmage, 2004, p.14).

Graham and Neu (2004), Canadian business professors, picked up the concern related to provincial school leaving exams and their role in the delivery and management of the Alberta education system. Graham and Neu constructed a critical and effective history of the Alberta’s testing program using Foucault’s concepts of governmentality, the construction of students as a population, and the production of student achievement as a problem that needed to be fixed. Graham and Neu raised some tough questions that echo the concerns raised by Dolmage (2004).

The evidence suggest that students from stronger socio-economic backgrounds do better on the tests than do students from poorer backgrounds or members of visible minorities. Are these results due to advantageous educational practices available only to the rich? Are they due to unfair construction of test methods and questions? Do they have the effect of exacerbating and reinforcing socio-
economic differentials, as successful students reap the rewards of successfully negotiating the barriers of higher education? (Graham & Neu, 2004, 312) Graham and Neu identified the power of standardized testing resides in its ability for creating an environment in which “the power of societal norms are internalized by the educational system participants” and the publication of the test results acts as the agent for self regulation and as a tool for surveillance (p. 313). The following example of an Alberta Education press release (2007, November 28) illustrated the actions of the Alberta government to encourage self regulation of its parents’ behaviours: “In addition, parents’ own reading habits, such as how frequently parents read, had a positive influence on their children’s reading achievement” (p.1).

Volante (2008), a Canadian education professor, described the controversial nature of student achievement results and their relationship to educational system reform. The media has played a significant role in sensationalizing the results of standardized achievement testing and their possible implications for the economic well being of Canada (Volante, 2008). Volante affirmed the concerns raised by Dolmage (2004) that the use of “testing for school improvement purposes ... can result in several unintended consequences that may have deleterious effects on teachers and students” such as narrowing the curriculum and deskilling teachers (p. 21). In addition Volante (2006) noted that test preparation strategies such as teaching to the test and engaging students in simulated test activities “do little to promote authentic learning and often provide artificially inflated scores” (p. 9). A sound large-scale assessment program should “foster, and not impede student and teacher learning ... [should] be seen as important, but not as measures for high-stakes decisions” (Volante, 2006, p. 9)
Canadian policymakers in several jurisdictions have used the results of PISA, PIRLS and TIMSS to complement the results from PCAP, and “the provincial responses to international comparison testing are largely affected by their degree of success or failure relative to other educational jurisdictions” based on simple league rankings as opposed to a careful interpretation of the complex interplay of multiple contextual and environmental factors (Volante, 2008, p. 23).

Teachers have used standardized test results with varying degrees of success: Alberta teachers’ calls for system reform “have been muffled in face of their strong showing nationally and internationally” while teachers in the Atlantic Provinces have used “low scores ... to spark renewed calls for large-scale reform” (Volante, 2008, p. 24). Volante, noted that “[i]t is bitter irony that provinces which score the highest on PISA such as Alberta and Ontario also have the lowest graduation rates, while provinces that score the lowest on PISA such as Prince Edward Island and New Brunswick have the highest graduation rates” (Volante, 2008, p. 24). What is more valuable to students over time? A high score on PISA or a grade 12 diploma.

St. Clair (2009), a consultant with ABC Canada, addressed the accountability issue from the perspective of adult literacy and the results of the International Adult Literacy Survey (IALS). Although the participants differ in age, some of the same overarching issues related to assessment data, program efficacy and efficiency, and accountability are relevant to the discussion of SAIP/PCAP. Both PISA and SAIP/PCAP have adopted definitions of reading, mathematics and science as a form of literacy rather than as a discrete set of competencies; therefore, these assessments would have experienced some of the same definitional dilemmas as the adult literacy assessment. St.
Clair stated that literacy cannot be defined as a “single set of abilities that can be placed on one scale” but rather a set of profiles describing “patterns of abilities” (St. Clair, 2009, p.1). SAIP/PCAP and the international assessments struggle to define mathematical, science and writing literacy and to develop criteria that are indicative of particular levels of performance. If programs/educational jurisdictions are to be held accountable, they should know for what they are accountable and what they can reasonably do (St. Clair, 2009). In the case of the elementary and secondary educational systems, the impact of demographic and community variables on test scores is often greater than the impact of instruction; therefore, the setting of reasonable targets based on a wide range of influencing factors is a critical first step (St. Clair, 2009).

St. Clair considered the cost of accountability programs as they are not without “associated cost in time and resources” (p. 2). The investment in accountability activities should reflect a cost-benefit analysis for the program or educational system rather than investing significant resources in the justification of what is already done well (St. Clair, 2009). Key challenges in implementing accountability systems is managing the “perspective on accountability held by different stakeholders”: Educators may believe that the goals of accountability are best served through the collection of individual student’s stories, while policy makers want numerical information and standardized test scores (St. Clair, 2009, p. 3). From St. Clair’s perspective, the assessment of student learning and accountability are two “highly specialized activities ... with different foci and different applications” that overlap in the field of evaluation (p. 6). The issue, from St. Clair’s perspective, is not whether accountability and assessment should exist, but rather how should they exist in what relation to each other.
Volante and Ben Jaafar (2008) reviewed the overall effects of administering large-scale assessment programs and the production of similar educational testing policies across the Canadian jurisdictions: “Every province and territory in Canada, with the exception of Nunavut, administers some form of mandated large-scale assessment (LSA)” (Volante & Ben Jaafar, 2008, p. 203). Although the jurisdictions offer different assessment programs, they have many similarities: “all assessment programmes test reading, writing and, mathematics ... criterion-referenced and are developed and graded by Canadian teachers” (Volante & Ben Jaafar, 2008, p. 204). In addition to the presence of provincially administered standards assessments or school leaving exams, there is also a growing trend for jurisdictions to increase participation in large scale international assessments (Volante & Ben Jaafar, 2008). This trend is curious because “Canada does not naturally adhere to the international push for standardization” (Volante & Ben Jaafar, 2008, p. 205). However, the voluntary participation by the Canadian educational jurisdictions in PCAP and several international assessments seems to reflect “an implicit approval of the adoption of the practice of collecting standard performance information as a process that holds currency in public and professional realms of education” (Volante & Ben Jaafar, 2008, p. 206). Volante and Ben Jafaar attribute the collaboration of student performance data to the need to seek public endorsement and to report “standard performance measures [that] reassure the public that the education system is monitored for effective and efficient use of public funds” (Volante & Ben Jaafar, 2008, p. 206). In Canada, assessment results are used “for both accountability and school improvement purposes”; which Volante and Ben Jaafar describe as “compatible and complementary”, but according to St. Clair (2009) not necessarily overlapping.
D’Arrisso and Lessard (2009), Canadian education professors, examined the phenomenon of convergence in the formation of educational policy for elementary and secondary education in Canada. They commented that this convergence is particularly noteworthy as Canada does not have a national agency or department of education nor does it have national acts such as the No Child Left Behind Act 2001. Their study identified the emergence of the Council of Ministers of Education, Canada, in 1967 and the priority that it gave to educational results. Other evidence of educational convergence was the development of provincial consortiums devoted to curriculum development and the participation of several Canadian jurisdictions in PISA. From these priorities emerged several pan-Canadian themes: the need for educational fiscal policies, accountability initiatives, curriculum development priorities, educational choice, social equity, and the professionalization of teachers (D’Arrisso & Lessard, 2009). D’Arrisso and Lessard found that organizations such as the association canadienne d’éducation, Canadian Association of School Administrators and fédération canadienne des enseignantes et enseignants, Conference Board of Canada, and the Fraser Institute promoted convergence through their communications and their activities. Some organizations, like the Fraser Institute, promoted competition between jurisdictions and individual schools as a means of encouraging program convergence while other organizations adopted similar mission and value statements to demonstrate convergence. In Canada, harmonization and convergence of the programming and standards governing the multiple educational systems were not and cannot be accomplished by the rule of law, but rather it has happened through the exercise of relational power and self regulation.
Although federal and provincial government invest large sums of money in large-scale assessment programs, few resources are directed to the preparation of teachers and administrators to use the data and to support the adoption assessment practices that support learning and further the goals of these programs. Both effective classroom assessment practices and the capacity to use large-scale standardized assessment data are critical to the success of educational reforms. Volante (2006) stated that “pre-service and in-service training would enhance the utilization of large-scale assessment data by classroom teachers, something that has been woefully absent from many large-scale assessment contexts” (p. 11). Volante and Fazio (2009) recognized that teachers and administrators require sophisticated assessment knowledge and skills to “promote student success” and to work effectively and efficiently in today’s assessment-driven educational environment. They also reported that many elementary school principals, unlike their secondary school counterparts were more comfortable with a variety of assessment methods, and they were able to demonstrate a higher level of competence than the teachers for whom they provide leadership. Volante and Fazio commented that “balancing out assessment methods so that students “know and “show” what they have learned remains a formidable challenge for many educators – particularly in policy contexts that emphasize summative/large-scale assessment for accountability purposes” (p. 38).

Secondary Studies and Meta Analysis

The advance of computer-enabled statistical analysis and the availability of large-sale data sets from student assessment programs have created new opportunities for
researchers to mine existing data to uncover additional relationships – both causal and co-
relational. Ungerleider (2006), a Canadian education professor, outlined conditions for 
maximizing the benefits of large scale assessments and Anderson et al (2006) conducted 
a secondary analysis of the relationship between student achievement and variables 
gathered on the SAIP contextual questionnaire. Smith, Schneider and Ruck (2005), 
Canadian psychologists working in educational institutions conducted a quantitative 
analysis of the factors affecting the performance of Black Canadian secondary students in 
Toronto and Nova Scotia. These diverse studies build on or extend the groundwork laid 
by SAIP.

Ungerleider (2006) offered a careful criticism of the repurposing of large-scale 
assessments from the survey or certification tools into tools to guide student improvement 
programs. His concerns began with the relationship between validity and purpose of the 
assessment as “there is no established procedure [to evaluate] the validity of the 
alternative uses to which large-scale assessments might be put” (Ungerleider, 2006, p. 
875). From his perspective these large-scale assessments “do not adequately assess the 
broad range of knowledge and competencies that schools are intended to develop ... [and] 
are insufficiently attentive to the breadth of knowledge in the domains they do address ”
(Ungerleider, 2006, p. 875). The narrow focus of these large-scale assessments has been 
raised by other commentators; however, Ungerleider described the limitations of these 
assessments as they relate to the repurposing.

Ungerleider raised key research questions about the differential functioning of test 
items in specific populations such as Aboriginal students, immigrants and 
minority/majority language groups. PCAP addressed the minority/majority language
issue in a special report; however, many questions remain unanswered with respect to the function of minority language effect inside and outside of Quebec. The effects of students’ use of languages other than English or French as their first or primary language on SAIP or PCAP results have not been explored.

Smith et al (2005) illustrated Ungerlieder’s (2006) concern in their exploration of the educational experiences of Black Canadian students, a group that does not fall neatly into the minority/majority language category or the immigrant student category; however, there may be students from this subcategory that fit either or both of these groups. Based on Ogbu’s cultural-deficit theory, Smith et al (2005) conducted an exploratory quantitative analysis of the differential effects of the experiences of Black Canadian secondary students living in Toronto and Halifax. The results revealed that “[most] students have high aspirations and expectations, and highly valued success”; however, female students living in Toronto obtained higher scores for academic outcomes than males students living in Toronto and all students living in Halifax (Smith, Schneider, & Ruck, 2005), p. 355). Smith et al also revealed that contrary to their expectations, “neither parental education level nor socioeconomic status had a strong influence on students’ education outcomes” (p. 356). Although Black Canadian students comprised 2 percent of the student population, their results on standardized assessments are collapsed into the larger population or other minority subgroups.

According to Ungerleider (2006), “the cumulative impact of school factors in student achievement is typically less than 30 per cent”; however, if schools are to impact student achievement, they need “relevant, timely, and systematically gathered quality information about schools and students prior to instruction” (p. 877). The ultimate
difficulty for school administrators and educational jurisdictions is the identification of “what information is likely to be important” and ensuring that those who need the information receive it in a timely manner and have the skills and knowledge to use it effectively (Ungerleider, 2006, p. 877).

Anderson et al (2006) explored the possibility of using the pan-Canadian assessment data for the SAIP Mathematics 2001 and Writing 2003 assessment to create a model of school performance. Although Anderson et al noted that most of the variation attributable to schools occurs at the level of the classroom, the SAIP assessments “do not permit analysis of the data at the class level,” a barrier to the creation of functioning model of school performance (p. 709). In their review of the SAIP questionnaire, only items that “were relevant and accessible to policy making [with] relatively good response rate were included in the analysis” (Anderson et al, 2006, p. 709). Incomplete responses to SAIP questionnaire items limited the usefulness of the SAIP data in secondary analysis. The results of Anderson et al’s secondary analysis revealed similar and consistent findings with the analysis of the SAIP questionnaire data: Principals’ attitudes towards their school and school environments and student performance were positively related, as well as students’ beliefs about achievement and student performance. Anderson et al stated that “policy relevance is clear: Enhancing student and principal perceptions could enhance student achievement and therefore school performance” (p. 724). However, parental assistance, and computer use were negatively related to student performance; “policy relevance here is unclear because to reduce parental involvement with a students’ work at home on school subjects would not be accepted by most educators and parents” (Anderson et al, 2006, p. 725).
In closing Riddell (2008) returns our focus to the development of human capital and the level of Canada’s investment in human capital. Over the past several decades, several commentators have raised concerns about the high level of investment in Canada’s educational systems and the limited returns on investment based on the performance of Canada’s educational systems on international standardized tests. Riddell (2008) concluded that “policy interventions that raised educational attainment had large beneficial effects on the affected individuals” as well as the production of important non-market and social benefits (p. 2). However, Canada now occupies the position near the bottom in terms of spending on early childhood education and the “reduced public expenditure [on education has resulted in] a decline in the quality of education, most evident at the university level” (p. 2). Education policy making in Canada should “narrow gaps in educational attainment and skills among key subsets of the population for equity and efficiency reasons” and encourage students to stay in secondary school (Riddell, p. 2). The results of PCAP-13 could provide valuable information to assist policymakers if questions were crafted with that purpose in mind.

The End of the Decade and the End of the Journey

SAIP began as an experiment – could the provinces really collaborate and work together to administer a Canada-wide assessment. The CMEC’s language reflected the marvel of this initiative while commentators who watched from the sidelines expressed concern about the risks to teachers and students, the constitutionality, and perceived collaboration with business. The language used to describe the SAIP results in 1993 embodied certainty which eroded as the test developers, item designers and data analysts
began to reveal the complexities of student learning. Innovations to the SAIP assessment design such as the use of confidence intervals; the introduction of an expectations-setting panels; the development of student, teacher, and school questionnaires; the introduction of increasingly complex of data analysis, and the increasing reliance on theory to frame assessment questions and explain the results changed what the assessment results could or could not contribute to the understanding of Canada’s educational systems.

In 2003 SAIP began its evolution from a criterion-referenced, curriculum-focused large-scale assessment into a norm-referenced assessment tool resembling PISA. The PCAP questionnaire retained many of questions and features of the SAIP questionnaire; however, direct references to the 2003-learning model, the conceptual basis for the SAIP 2003 questionnaire disappeared. SAIP provided comparative data for 13-year olds and 16 year olds while PCAP -13 2007 was administered as a single-stage test to only 13-year-olds. By administering PISA to a substantially larger Canadian sample - 28,000 students (as compared to approximately 5,000 students in most other countries) - and aligning PCAP to PISA, CMEC and HRSDC have constructed a workable replacement to the SAIP assessment project (Volante, 2006, p. 2).

Throughout the evolution of SAIP/PCAP, key conversations have emerged concerning the role of the federal government in the delivery of elementary and secondary education, the role of business, and the need to development human capital to fuel Canada’s economic engine. However, the implications of accountability-assessment evaluation agenda on the capacity of Canadian school systems to provide inclusive educational opportunities to all Canadian youth not as yet well understood by policy maker and educators. The language of these conversations has transformed; however,
these large thematic topics continue to permeate Canadian discussions about its educational systems and the assessment of teaching and learning.
CHAPTER SEVEN
UNDERSTANDINGS AND QUESTIONS

There is the very known of our trouble. None of us knows our history. We are a people of shadows.
Llewellyn, 1975, p. 147)

Llewellyn’s quotation speaks to me as I reflect on my journey through the archive: the copious journals, magazine articles, CMEC reports, and books that contain traces and fragments of the SAIP/PCAP story. I began this journey several years ago, working in the shadows of the standardized testing discourse, reading Foucault’s (2006) work, *Ethics: Subjectivity and Truth*, and reflecting on my personal questions about standardized testing. These questions guided my personal inquiry. How has standardized testing come to be and what is it about its nature that produces such a visceral reaction in my colleagues? Could standardized testing be understood as a field in which the researcher who is embedded in that same field stand apart and resist the urge to argue a particular view or position concerning the value of standardized large-scale assessments? Is it possible to observe, trace, and record the work of the forces shaping the field and determining what truths could be told?

It was a difficult journey; it was trying at times to remain observant of the dynamic pull and push of the forces and lines of tension and to attend to the ever-present need to record the textual voices that spoke these tales into reality. It was my job as a researcher to remain outside the ever-present argument that pits the need to provide accountability for educational outcomes, to demonstrate responsiveness to the demands for a skilled and educated workforce, and to justify educational expenditures against the need to attend to the diverse needs of children and teachers in complex teaching and
learning environments. These large-scale standardized assessment programs operate at the nexus of these emotion-filled spaces where there are no right answers – only opinions, ideologies, and theoretical perspectives travelling along lines of tension and operating in fields of opposing and complementary forces.

The Two Journeys

My dissertation is the result of two sometimes parallel journeys, but at other times intersecting journeys, that mark my growth as a scholar in a particular space – large-scale standardized testing – and my contribution to the knowledge production of that space – a critical and effective history of Canada’s national assessment program SAIP/PCAP.

My personal journey was informed by Foucault’s (2005) work, *The Hermeneutics of the Subject*, from which I incorporated the methodological concepts: memory, meditation, and method. I wrote my memories into the development of my dissertation, used my meditational moments to think about how I have come to think in this space, and used my writing as my methodological tool to develop my understandings of this space.

I used writing as a Foucauldian method of coming to know how I think about Foucault’s thoughts and concepts; and through writing and re-writing Foucault, I have come to understand how I think in this space of standardized testing. Following Foucault’s advice, I worked with one scholar/writer, Foucault, invested in reading and re-reading Foucault’s texts and the interpretations of Foucauldian scholars, made notes, re-wrote pieces of Foucault’s writing, and searched for illustrative examples. Foucault’s work formed my lens for the construction of my critical and effective history in the domain of Canada’s national standardized assessment program and the lens for my coming to understand my ontological self.
Like other scholars and writers whose paths I have crossed on this journey, I found myself re-writing and re-thinking Foucault’s conceptualizations throughout the dissertation process. Foucault, himself, encouraged this act of re-writing as a strategy to develop one’s understanding of another’s ideas and concepts and to develop one’s own tool kit or approach to understanding a domain in which one has struggled (Foucault, 2006). Throughout the dissertation process, I re-wrote Foucault, re-read Foucault’s writing, wrote some more, and re-wrote again. These writings constitute the artefacts of my attempts to produce knowledge in this domain and to record my reflections as I thought about the production of knowledge and the regimes of power that legitimate the production of truth in this field of standardized testing.

My journey in this space began with a methodological question similar to the CMEC’s question when they conceptualized SAIP. While CMEC’s methodological question focused on the possibility of developing and administering a Canada-wide assessment protocol for 13- and 16-year olds, my question focused on the possibility of actually producing a critical and effective history of this domain using Foucauldian concepts. My concerns focused on the capacity of the texts to reveal changes in statement formation that would be evident and sufficient to illustrate shifts in truth telling. I wondered if my search of the archive would reveal textual data that would be both valuable and useful to this project, and I was concerned about the gaps and the implications of those gaps in the textual record. I wondered if gathering, ordering, selecting and analyzing the texts of others was a real contribution to knowledge production. These textual acts produced disruptions of our taken-for-granted beliefs, understandings, and perhaps myths about teaching, learning and assessment. Foucault
(1982) argues that this is valuable work as it creates visibility, produces understanding and reveals aporias for future inquiry. From a Foucauldian perspective, the task of a historian is to reveal and to connect that which is not obviously connected and to disturb that which is part of the taken-for-granted reality of a society.

Research usually produces results, generalizations, causal relationships, and understandings in response to carefully crafted questions or hypotheses. A critical and effective history produces an understanding of how we have come to be in a particular space – in this case, the space of large-scale standardized assessment. Our history in this space shapes what we are able to think into existence, and our culturally-produced rules for determining the truthfulness of statements in this space are a product of our culturally-situated history.

My understandings of my personal journey appear as questions rather than statements of certainty and revelation of new understandings of the world. When I began this journey, I occupied an uncomfortable place in that taken-for-granted space of certainty surrounding standardized testing - a space shaped by strong visceral forces that either agreed or disagreed with regimes of standardized testing. My uncomfortable space was outside the argument; however, I did not know what to ask or how to ask it, but the simplicity of the polarized debate seemed to neglect a host of variables about teaching, learning, and assessment. My personal journey has led me to an understanding of my discomfort and to some comfort in forming the questions and contesting the standardized test results that are posted with such certainty.

Using Foucault’s concept of conduct on conduct embedded in his thematic of governmentality, I now feel that I understand how governments [broadly-defined]
operating-at-a-distance shape and direct the conduct of their constituents/citizens through the use of relational power. Through my archival work, I have informed myself so that I can exercise a Foucauldian care of self and shape and direct my conduct in this space through thoughtful and informed decision making.

Understandings

My understandings emerging from this space of Canadian large-scale standardized testing are organized within the framework of my guiding questions and informed by Foucault’s thematic organizer of governmentality which employs the concept of conduct on conduct as exemplified by action on action. In essence governmental organizations, like CMEC, exercise influence on the conduct of others through the outward effects of the activities related to their conduct and that of others within their sphere of influence. CMEC’s decision to implement a regime of standardized testing is an example of a government-at-a-distance conducting its actions in a manner that shape and direct the decisions and activities of its constituents. CMEC engaged in the production of truth telling statements related to standardized testing, learning, accountability and standards that have transformed over time. These truth telling statements produced knowledge about standardized testing, its outcomes, and its implications for the delivery of educational services in Canada; and these truth telling statements affected the behaviour of citizens, organizations, and CMEC, itself, through the relational effects of power.
The Production of My Understandings Based on My Research Questions

How do culturally, economically and politically derived regimes of truth telling shape the beliefs and assumption about what counts for knowledge in the intersecting domains of education, assessment and accountability?

My dissertation revolved around developing some understanding of the roles of culture, economics, and politics in shaping what we as Canadians understand to be true about the relationship of education, assessment, and accountability. My understandings have been shaped by examining the conduct of other jurisdictions, the truths produced by those jurisdictions, and the effects of their truth telling statements on students, teachers, and scholars. England, United States, and Canada have adopted national standardized testing regimes during periods of economic downturn that coincided with perceptions voiced by representatives of business/industry and the general public that schools were not producing graduates who had the skills and knowledge to transition successfully into the workplace. These sentiments were fuelled in part by the emergence and popularity of neoliberal ideologies that supported the involvement of business and market-driven solutions in spheres traditionally occupied by the public sector.

In England, the development of the national curriculum and the school accountability initiatives involving standardized testing were a response to the poor performance of the English economy in the 1980s. This poor performance was characterized by rising unemployment and the acceptance of neoliberal ideologies of minimizing government engagement in the provision of social services and maximizing the engagement of the private sector. The English national standardized testing program operates through juridical or sovereign power vested in an act of legislation, the School Reform Act (1988). Its implementation was universal, swift and public. Test scores are
published in the media; parents are offered the right to choose school options; and the
delivery of educational services is conceptualized as a market-driven solution serving the
needs of students and business/industry. The English school testing program is based on
the notion that the availability of skilled and knowledgeable graduates is essential to the
functioning of England’s economic engine.

The government provides information through the media to its citizens about
school performance in the form of standardized test scores aggregated by schools and
posted on league tables. Parents are encouraged to use the information to exercise their
right to choose the appropriate educational delivery option for their child. The
government operating-at-a-distance is directing the conduct of parents towards active
engagement and responsible decision making in their children’s futures rather than
assuming that responsibility directly. Parents have inherited the right to choose, and their
child has inherited the consequences of their choice. In terms of a Foucauldian analysis,
the government-at-a-distance is exercising its conduct on conduct by directing the
behaviour of parents and citizens through the provision of programs and regimes of
conduct that shape and direct their conduct. By providing information to parents, the
government-at-a-distance divests itself of its liability concerning the educational choices
made by parents and children.

Schools in England are conceptualized as businesses that must now compete for
students because competition should improve the quality of all schools. This regime of
conduct was based on the belief that schools would be motivated to improve the quality
of their programs to attract more students – just as businesses are motivated to improve
their products and services to attract more customers. However, school funding is
attached to student enrolment; therefore, schools that are able to attract more students receive more funding and improve their programs. The funding issue is further compounded by the provision for private endowment of schools and the corresponding provision for seats on the school’s board of directors. The government reinforces the role of competition by publishing schools’ test scores and removing school boundaries. The government believes that the parents of children attending poor performing schools would ensure that their children were able to travel to better schools; thus, encouraging poor performing schools to invest in their programming and improve their standing. However, providing information (test scores by school) to parents so they could make an informed choice and the opportunity to exercise choice are not sufficient for parents living in poor neighbours to exercise their right to choose. Meanwhile the better performing schools continue to benefit from increased resources and growing enrolments while the poorer performing schools struggle to attract students outside their catchment area. These poorer performing schools face diminishing opportunities to attract highly-qualified teachers and funding for much needed additional resources. These schools are caught in a trap of poor performance measured and bounded by standardized test scores.

The relationship between the delivery of educational services and standardized testing in the United States has a much longer but more fragmented history than England. The United States, like England, has a national department of education that regulates the transfer of funds from the federal government to the state educational agencies, but the federal government limits its juridical power to the governance of the state education agencies. Americans have a long history of resistance toward the involvement of big government in local matters. Thus, education in the United States is organized under a
single umbrella but administered by 50 agencies. Although periodic alarms about the quality of American secondary school graduates have sparked concern about the efficacy and efficiency of American school systems, the *No Child Left Behind Act, 2001 (NCLB)* was the first attempt to coordinate standardized assessment activities nationally. The *NCLB Act (2001)* mandated standardized assessment at key transition points in a student’s career, set the reporting structure for schools’ annual rates of progress, and prescribed consequences for schools failing to meet their annual rates of progress. However, the American federal government specified that the *NCLB Act (2001)* is a national, not a federal, program to be administered at the state level. Each state specifies its own assessment tool and collects the data on school performance. Standards vary from one state to another; comparisons between states are wrought with challenges. Although the target of *NCLB Act (2001)* is the improvement of student performance, the federal government has chosen to direct the conduct of state agencies of education and to assign the responsibility for prescribing the conduct of schools and teachers and the measurement of student performance to these state agencies of education.

Scholars and commentators have identified concerns about the consequences of the *NCLB Act (2001)* to American classrooms that include narrowing the curriculum, deskilling teachers, and inflicting undue anxiety on students and teachers. However, the *NCLB Act (2001)* focuses on the production of consistent outcomes for all students in all schools; this priority has refocused the American educational lens on issues of diversity and equity in the production of learning outcomes. Thus, issues of student diversity, inclusive education, and the amelioration of disadvantages due to the effects of poverty are forming part of the American education agenda. Some educators working with
students with disabilities welcome the visibility that the *NCLB Act (2001)* has brought to their educational portfolio.

Both England and the United States chose to implement national standardized testing protocols as a component of national school reform through acts of legislation. The passing of an act invites commentary, criticism and discussion. These initiatives in both countries have spawned a rich body of resistive and productive literature that describes the impacts on student learning, curriculum development and delivery, instructional strategies, the inclusive classroom, and the professionalism of teachers. Scholars and commentators have weighed in on the philosophical, ideological, and ethical implications of prescribing these types of initiatives.

In the 1980s, Canada, like England and the United States, faced a downturn in its economy coupled with rising unemployment and the perception by business and industry that the graduates of the current educational systems were not academically prepared to make the transition into employment and/or post-secondary education and training. Business and industry lobbied CMEC, the national voice for education in Canada, to establish a national standardized testing program. Without a national department of education, a legislative solution was not possible. Section 93 of the *Constitution Acts 1867 – 1982* lists education as a provincial responsibility; thus, the federal government cannot legislate or prescribe programs or curriculum for Canadian schools. In 1989, the ministers of education who comprised the CMEC secretariat signed a memorandum of agreement to proceed with the development and administration of a pan-Canadian assessment protocol. The production of the memorandum of agreement was the result of productive acts of relational power exerted by forces in Canadian society. CMEC acting
as a form of government-at-a-distance produced a regime of conduct - the memorandum
of agreement - that put into play relational acts of power that would produce the SAIP
project. The operation of these social forces illustrates a Foucauldian cycle of power:
Social power produced the conditions that led to the memorandum of agreement, CMEC
produced a regime of conduct (the SAIP project) and the SAIP project produced
knowledge in the form of test results.

The SAIP project operated as a relational act of power that produced forms of
conduct that served to guide and modify the conduct of education ministries,
departments, school boards, school administrators, teachers, students, and parents. These
relations of power produced knowledge about the functionality of the Canadian
educational systems in the form of test and survey results, created spaces for
collaboration across jurisdictions (regional consortiums on curriculum development), and
identified opportunities for research. These activities occurred outside legislative and
public debate and discussion.

The Canadian assessment instrument, SAIP, varied from the standardized
assessment tools developed in England and the United States. SAIP was criterion-
referenced, curriculum-focused, and bilingual. Although a key question posed by SAIP
cycles two and three concerned the level of comparability to the results generated by
previous iterations, the SAIP assessment tool developers refined, revised, added, and
deleted SAIP elements throughout its life cycle; therefore, its comparability was limited.
Thus, the SAIP assessment protocol evolved from an output-focused tool into a
comprehensive learning indicators assessment tool and survey instrument. Its capacity to
assess not only student learning outcomes, but also the learning environment exceeded the capacity of the instruments produced by either the English or the American projects.

Throughout the two decades of the SAIP project, the interest of CMEC drifted away from determining the proportions of the 13-year olds and 16-year olds who were able to perform at prescribed levels in mathematics, reading and writing, and science to developing of an informed understanding of the factors that affect the learning process, identifying those factors upon which schools could have an impact, and determining the factors outside the school’s realm of influence.

Although the SAIP project expanded beyond the boundaries of measuring student academic performance, CMEC continued to rationalize the importance of the SAIP project based on the theory of human capital development and the critical relationship of the productivity of Canada’s educational system to the economic well being of Canadian society. CMEC maintained its relationship with business and industry by inviting their representatives to sit on student performance expectations-setting panels and consulting with the Conference Board of Canada and OECD. In Foucauldian terms, the representatives of business and industry and the Conference Board of Canada exerted their influence through relational acts of power that resulted in the production of regimes of truth, the expectations for student performance on the SAIP assessments. The validity of these regimes of truth was assessed in the administration of the SAIP assessments. The expectations of business and industry for mathematics performance often exceeded the proportional distribution of students’ performances.

In 2007, CMEC launched PCAP, an evolution of SAIP influenced by the design and structure of PISA, an international regime of standardized testing produced by the
OECD. PCAP was the product of complex plays of relational power operating among HRSDC, the OECD and CMEC. The tangles of this web are difficult to tease out; however, HRSDC integrated PISA into its longitudinal youth survey prior to the end of the SAIP project. PISA targets 15-year olds while SAIP targeted 16-year olds; however, the SAIP and PISA assessments have significant similarities with regard to the subjects assessed, the types of test items, and the student, teacher and school questionnaires. Investing resources in the administration of PISA to a large sample of 15-year olds (about 25,000 students) and a similar investment in the administration of SAIP to an equally large sample of 16-year olds appears to be a duplication of effort and an ineffective use of scarce resources – both direct costs and in kind contributions from educational jurisdictions. To further complicate the issue, Canadian researchers in the field of assessment have contributed to HRSDC projects, the SAIP project, and the PISA project, and the role of their influence is difficult to determine. The lines of relational power are complex, unequal, and obscured through the daily activities of running these organizations. Nonetheless, PCAP-13 2007 emerged as norm-referenced, one-stage test with enhanced and targeted contextual data collection and analysis capacity based on a pre-determined research agenda.

CMEC’s decision to move forward with a norm-referenced assessment tool targeted at 13-year olds represented a significant change in Canadian philosophy about large-sale testing and a new comfort with ranking and sorting jurisdictions based on their students’ performances. Norm-referenced tools can tell different truths about student performances and place priority on rank ordering rather than skill or knowledge attainment.
CMEC through relational acts of power has pushed the boundaries on large-sale standardized assessment, influenced curriculum harmonization by making jurisdictional curricula visible, and by encouraging collaborative regional consultations on curriculum construction. As a Canadian living and working in the domain of education and testing, I recognize that the SAIP/PCAP projects have brought the Canadian educational systems out of the shadows; and through acts of visibility, Canadian jurisdictions have taken on projects that have led to harmonization and conformity in curriculum design and school administration. All this has happened with little public debate and discussion. The regimes of conduct in the form of programs and projects supported by CMEC have produced changes in the taken-for-granted everyday life in Canadian schools. On this path to harmonization, local curriculum elements have been erased and the record of what has been lost is meagre.

The English and American literature is robust and filled with commentators who write about philosophical, ideological, and ethical issues surrounding their national standardized testing enterprises and with researchers who explore the technical issues of test design and psychological implications of the testing process. The Canadian literature is limited and focuses on the functionality of SAIP test items complemented with an emerging interest in the role of standardized testing (SAIP/PACP) in educational policy construction. I believe that the necessity of keeping educational reform and the introduction of a national standardized assessment program out of the juridical sphere has stifled public debate and academic criticism about the SAIP and PCAP projects and the school reform agenda in Canada. When projects are produced as programs of relational power and operate as a type of conduct on conduct, they merge into the taken-for-granted
spaces of everyday life. When the SAIP project was launched in the ‘90s, the president of CTF expressed concern about the constitutionality of the SAIP project accompanied by a call for a national department of education and public debate about educational issues of national importance. The urgency of that need has faded. If SAIP were announced as part of a national educational strategy, it would have incited a public debate, discussion, research projects, knowledge production, and the production of statements regulating the production of the truths about education, assessment and its functionality in the Canadian social context. If the SAIP project were part of a legislative process, it would not have occupied a place in the shadows; the SAIP project would have formed a conscious element of the Canadian educational identity.

How does the knowledge production and truth telling surrounding the development, administration, and data gathering related to SAIP/PCAP inform, influence, and shape national and provincial education priorities; and conversely how do culturally- and contextually-produced factors as well as regional, provincial, national and international priorities inform and shape SAIP/PCAP?

The SAIP/PCAP projects have engaged in the production of truths about the nature of learning and its assessment, the effects of gender on the capacity of students to perform on the SAIP/PCAP assessments and the importance of language and culture on the production of student assessment results. The capacity of the SAIP/PCAP projects to produce knowledge has increased with the development of computing capacity, the emergence of sophisticated models for learning, and the recognition that learning is a complex process that operates within a context of socio-economic factors. The inception of the SAIP/PCAP projects created opportunities for teachers, educators, and statisticians to work together and to collaborate with policymakers and bureaucrats to develop new understandings about the Canadian educational systems. SAIP/PCAP projects have
provided a new level of visibility about curriculum, factors concerning students’ learning, and the role of teaching strategies in the learning process.

In the 1990s, the Canadian economy was in a downward spiral, and the focus of the public, business, and industry was on the outputs of the Canadian education systems: the skills and knowledge of their students. Thus, SAIP – 1993 was designed to measure the outputs of Canada’s educational systems: students’ skills and knowledge and students’ attitudes toward learning. The assumption underlying the construction of SAIP – 1993 was that the level of student performances reflected the level of performance of their educational system. Although the SAIP developers did little with the data from the student questionnaire data accompanying the SAIP – 1993 assessment, this data sparked an interest in the variability in attitudinal and behavioural factors that students bring to the assessment event and that the measurement of learning is a complex and uncertain activity.

The SAIP developers introduced the notion of ambiguity into their truth telling in 1994 when they acknowledged that the test results were really estimates of the true measures of student performances and not absolute measures. In 1996 the evolution of truth telling took another step away from the certainty of the statements made in the first SAIP assessment. The student questionnaire increased in length, and the statisticians conducted an analysis of relationships between students’ performances and their behaviours and attitudes. The investment in the production of this relational data shifted the truth telling about the functionality of educational systems away from measures of academic performance to the quality of relationship between students’ performances and their attitudes and behaviours.
In 1999, the definition of learning became operationalized as a relationship among these elements: inputs → processes → outputs. Since the definition of learning had expanded, collecting data on students’ performances and students’ attitudes and behaviour was insufficient. Truths about the effectiveness and efficacy of learning included the measurement of inputs (resources, school climate, and composition of the school), processes (the instructional strategies, teacher preparation, and type and quality of homework), and outputs (academic performances of students). These factors were all moderated and influenced by students’ attitudes and behaviours. The measurement of learning process and its productive outputs became increasingly complex, and understanding the differences in performance became considerably more difficult. Truth telling became a relationship activity bounded by growing uncertainty.

In the third cycle of SAIP, the definition of learning continued to expand along the same trajectory. The SAIP developers reinvented Wang, Haertel, and Walhberg theory of learning and produced it as a model with the following elements: inputs → processes → outputs – all operating within the influences of a social context. The final SAIP assessments captured academic performance, reinvented the SAIP questionnaire to capture feedback from principals, teachers, and students about the key elements affecting the learning process. Multivariate and bivariate co-relational analysis weighted the effects of individual variables on student performance. Truth telling at the end of the SAIP project was complex, ambiguous, relational, and conditional. The certainty of the truths emerging from the first SAIP assessment about the functionality of school systems was replaced by uncertainty and ambiguity about the potential of schools to influence students’ academic outcomes. The shifts in the definition of learning and what constitutes
meaningful change in academic performances was moderated by the SAIP experience, the academic environment in Canada, and shifts in the performance of the Canadian economy during the SAIP project.

The definition of learning incorporated into PCAP drew on the PISA legacy: Learning is now defined as a function of the ratio of the time taken to learn and the time needed to learn. The time taken to learn is defined as a function of variables such as motivation, attitude, and personal resources on a student’s investment in the learning process. This definition of learning presents the potential to understand learning as a mathematical function, and the net effects of modifying one or a cluster of variables could be understood as a function of this mathematical model. This definition of learning builds on the understandings derived from previous SAIP administrations, but it also recognizes the potential to understand learning as a productive output of a complex interaction of factors that could be analyzed using sophisticated statistical analyses. The re-definition of learning opens the door for different types of research and the production of new knowledge about the learning process.

The SAIP project aggregated and categorized the outcomes of learning process (student performances) by gender and language. The construction of these categories of student performance included prescriptions for generalizations about students’ behaviours, attitudes and their expected level of performance.

Thus, student performances aggregated by gender were scrutinized for differences in the degree to which their performances conformed to predetermined expectations. The student test results and questionnaires were probed for evidence of explanatory
relationships between differences in learning styles and attitudes of students based on
gender and their respective levels of performance.

Female students were expected to perform at higher levels on reading and writing
assessments and were expected to demonstrate more positive attitudes toward reading and
writing activities than males, and male students were expected to perform at higher levels
on the mathematics and science assessments and to demonstrate higher levels of
enjoyment for these subjects than female students. Initially, the SAIP assessments
produced this expected gender-based performance pattern; however, the results from the
third cycle of SAIP disrupted this expectation. Gendered-based differences for
mathematics and science test results disappeared, but gendered-based differences for
reading and writing persisted on SAIP – 2002. Despite the significant redesign of SAIP –
2002 with content specifically selected to be engaging for male students, the gender-
based differences persisted. However, jurisdictions, in which the performance of male
students approached the Canadian norm, tended to have stronger performances overall.
The issue of gender-based differences in reading and writing performance appears to be
more than just an issue of test design but may have tendrils that reach into classrooms,
families and the society in which males students live and learn.

The SAIP project had to address the issue of the measurement of learning
produced by students living and learning in French and English speaking jurisdictions.
This challenge is unique to Canadian education and is constitutionally situated. For the
SAIP project to be a truly Canadian experience, the assessment tools had to function
equally well in English and French. The SAIP test results had to inspire a sufficient level
of confidence in their comparability between the jurisdictions writing in French and those
writing in English, so that the validity of the test results would not be dismissed on the presumption of inaccurate or faulty translation. To achieve this goal, the SAIP project developers invited educators and learning specialists from both linguist groups to collaborate in the production of tests, assessments rubrics, and the marking sessions. Through their engagement in the SAIP project, educators from various jurisdictions were able to collaborate and influence the design, scope, and content of SAIP assessment tools and build rapport across jurisdictions. These activities exemplify the production of conduct by educators in response to the conduct in the form of programs and directives produced by CMEC.

As an outcome of the production of the SAIP tool in English and French, student performances were categorized, aggregated and compared on the basis of language. In terms of the categorization of student performance based on language, the results were stable and consistent with those produced on large-scale international assessments; however, these international assessment results did not form a set of comparative expectations for the SAIP results. Students writing in Quebec in French performed at consistently higher levels on SAIP than those writing in English speaking jurisdictions or writing in French outside Quebec – a result consistent with international assessment results. The relational analysis conducted on the data generated by the student, teacher and school questionnaires produced some interesting but challenging findings. The results from the teacher questionnaire revealed that teachers in Quebec spent less time engaged in after-class interactions with their students than did teachers in English-speaking jurisdictions. However, the importance of this finding is confounded by other factors: More parents in Quebec send their children to private schools, exclude their children
from participating in the testing process, and send their children to large schools which correlate positively with higher levels of student performance. The factors leading to the production of higher test scores in Quebec are complex and inter-related; however, they have relevance to the delivery of educational programs in other Canadian jurisdictions.

From a Foucauldian perspective, the transformation of the definition of learning was an expression of the cyclical effects of power. The artefacts left by the actions of power formed the stages in the evolution of the definition of learning. The SAIP developers in consultation with CMEC produced the original definition; however, its subsequent transformations informed the revisions to the SAIP assessment instrument. The results produced by the SAIP instrument informed the re-definition of the learning and informed the SAIP assessment re-development process. However, the production of categories of students with particular traits and expectations is a function of knowledge production according to Foucault (1982).

The production of knowledge about the functionality of these groups in relation to other groups is facilitated by categorization. Because these groups are constructed and not natural, the traits that distinguished one group from another can change or disappear; yet societies and cultures tend to maintain the categorization of groups and search for their distinguishing differences. Although the third cycle of SAIP test results did not produce gender-based performances differences in mathematics and science; however, the focus on gender-based differences persisted in PCAP-13 2007 mathematics and science assessment. According Foucault (1982), categories can disappear when they cease to be useful in our knowledge production projects; however, they often persist long after their utility has passed.
How do the SAIP/P-CAP results serve the normalizing agenda of accountability?

Accountability and the assessment of student learning have a complex relationship. Accountability poses large questions about who should do what, for whom, and how well. CMEC recognized that the SAIP results were only part of a much larger school accountability agenda that included graduation rates and transition rates to post-secondary education and employment. However, CMEC designed the first SAIP assessment on the assumption that the SAIP results would stand as a proxy for all other indicators of school performance and that the reporting of school performance through the use of students’ test scores was the most appropriate vehicle to provide accountability about the performance of Canada’s school systems to their stakeholders. Over the duration of the SAIP project, that assumption eroded into a belief that effects of schooling contribute no more than 30 percent of the variance in student test scores and that the test scores tell only a small part of the school performance story. CMEC shifted its truth telling statements about the quality of relationship between student assessment and school system accountability from that of certainty to that of conditionality.

Student assessment is a productive activity; it produces data, visibility, and commentary that have the potential to form the basis for the production of programs and directives that can act as a form of conduct. The SAIP assessment production activities created new venues for educators to consider, discuss, and evaluate curriculum production and forms of assessment. The SAIP results fostered a new form of visibility for Canadian school systems: The summaries of school curricula were published beside the test results for each jurisdiction to facilitate comparisons of both curricular goals and performance outcomes. In concert with the assessment production activities in 1993,
CMEC set the stage for the development of collaborative inter-jurisdictional relationships on curriculum development but sidestepped the constitutional strictures on federal engagement in elementary and secondary education by offering support but not direction for regional collaborative curriculum making activities. In response to CMEC’s initiatives, the ministers of education agreed to embark on a series of consultations on new directions for Canadian education, to support interprovincial working groups to review curriculum, and to establish regional collaborative initiatives on curriculum making and assessment (Council of Ministers of Education, Canada, 1993, September, 28). The outcomes of the collaborative activities included the Common Framework of Science Learning Outcomes (1997), the Western Protocol for curriculum collaboration (mathematics), and the Atlantic working group (Council of Ministers of Education, Canada, 1993, September, 28). The result of this ongoing collaboration is significant harmonization of Canadian elementary and secondary curricula around common learning outcomes and the production of sustainable inter-jurisdictional working groups and protocols (D’Arrisso & Lessard, 2009).

A result of signification collaboration and corporation among educators and policy makers has resulted in common curriculum for Western provinces and the Atlantic Provinces (British Columbia Ministry of Education, n.d). The introduction of common curriculum limits the negative impacts on student mobility in an era when families relocate frequently, and promotes regional standardization.

The introduction of the SAIP assessment project provided a Foucauldian gaze; the Canadian school systems, their similarities and their differences, were made visible through the production, administration and publication of the test results. Visibility
produced action in the form of consultation and collaboration. The collaboration and consultation produced harmonization and cooperative curriculum development activities. However, cooperative and collaborative teaching and learning activities can be extinguished when the standardized tests results are used as measures of teacher performance (Gustafson, 2001). Gustafson (2001) warns Canadians that the transition from measuring student performances to gather indicator data about school performance is just a few short steps away from using that data to measure teacher effectiveness. Measuring teacher effectiveness in the United States has made the easy next step to pay-for-performance schemes.

Regimes of accountability set their gaze not only on curriculum but also on teaching and learning, the primary activities of most of teachers. Instruction and the role of the teacher in the delivery of instruction have moved in and out of public scrutiny. In the 1990s and during periods of economic downturn, Manzer (1994) described the public perception about the quality of instruction and learning in Canadian classrooms as low. However, Robertson (1994) acknowledged this sentiment but stated that it is not well founded and emanated from business and industry’s desire for implementation of standardized academic testing. The issue is not the public perception about the quality of instruction in Canadian classrooms but the quality of preparation of teachers who must work in learning environments shaped to some degree by regimes of standardized testing.

Teachers, like parents, must have the requisite information and knowledge to govern their personal and professional selves. The call for academy to include standardized testing in the syllabi for teaching preparation has formed a recurrent and unanswered thread in Canadian educational literature. Wilkenson’s (1986) review of
teacher preparation and student outcomes identified the need for teachers to develop competencies in the area large-scale standardized assessment. Runté (1998) revisited the issue of teacher preparation in the field of standardized assessment in his discussion of Alberta’s school leaving exams. According to Runté, teachers who participated in Alberta’s school leaving exam preparation and marking were able to develop knowledge and skills beyond those included in their teacher preparation programs. Ungerleider (2006) added his concerns about the gaps in teacher preparation programs in the area of large-scale standardized assessment. Following the economic downturn of 2008, Zwaagstra et al (2010) argued that large-scale standardized assessment should form part of the assessment plan for elementary and secondary education; however, “many teachers’ unions and professors of education disparage standardized tests as being unfair and not authentic” (p. 32) and that “trying to improve public schools should take precedence over gaining favour with those who are the educational establishment” (p.7).

Whether teachers agree or disagree with the implementation of regimes of large-scale standardized testing, these testing regimes have become part of taken-for-granted reality of the many Canadian classrooms. Teachers should have the requisite skills and knowledge to conduct a critical review of the test results and contextual reports. According to Dean, (2010), knowledge and skills enhance one’s ability to direct one’s conduct; therefore, teachers who lack literacy about the field of large-scale standardized do not have the skills to provide the appropriate care of their selves, maintain professional and personal order, and manage professional risk. Therefore, the academy should re-evaluate its role in the development of teachers’ literacy in the field of large-scale standardized assessment.
Questions

The purpose of Foucauldian research is to not only produce some understanding through the development of a critical and effective history of how some phenomenon has come into existence but also to identify sites for further research and pose disruptive questions. Thus, questions have emerged from my years of reviewing the SAIP data and reports; some of the questions have been posed by other researchers.

The Quebec factor or the Quebec advantage emerged early as a theme in the SAIP results and was reconfirmed by the results of international testing. The SAIP questionnaires have revealed that more students in Quebec (writing in both French and English) attend private school, Quebec parents exclude their children from the testing process at a higher rate than other jurisdictions, and male students living in Quebec perform at levels equivalent to the Canadian mean on the reading and writing assessments. These same results reveal that the difference in male-female performance is greater than in any other province, the level of the student-teacher interaction is lower than in many other jurisdictions, and the teacher-class size ratio is often higher. The SAIP assessments have provided some insight into Quebec learning environment, but have not identified and described the nature of the Quebec factor. Probing the nature of the Quebec advantage has the potential to produce not only some understanding of this phenomenon, but also the potential to build and explore the construction of cross-jurisdictional working relationships.

Gender-based performance differences on large-scale standardized assessments have been tracked and followed for many years by scores of researchers in a variety of countries. The gender-based performance differences in mathematics and science have all
but disappeared on the last cycle of SAIP assessments; however, female students continue to perform at higher levels than males on reading and writing assessments. The focus of research lens should now shift to reading and writing literacy development in males. Ungerleider (2006) noted that some test items may function differently in some subpopulations; therefore, it is critical to develop a better understanding of performance of the test items in subpopulations whose performance deviates significantly from the norm. A secondary analysis of the differential functioning of the reading and writing test items may provide additional insight into the response patterns of males and females. By increasing our understanding of the differential response patterns of male and female students, educators may be able to gain some insight into the particular interests and learning patterns of males as opposed to females and into the differential functioning of test items through gender-based analysis.

CMEC introduced a student questionnaire to accompany the SAIP Science I – 1996 assessment to facilitate the gathering some preliminary data about students’ attitudes and behaviours toward the learning of science. That preliminary questionnaire included a question about parental help with homework and tutorial assistance. Results of this questionnaire and subsequent questionnaires revealed that students who performed below the level expected for their age group engaged their parents and/or tutors more frequently. These results have been consistent over time, and students with stronger performances were more likely to engage their parents in discussions about school but less likely to seek assistance from their parents or tutors with their homework. The secondary study conducted by Anderson et al (2006) confirmed that the analysis of SAIP Mathematics III data revealed the same trend; however, policy development that
introduces measures to discourage parents from assisting their children would be difficult to promote and would seem to be counter intuitive. Clearly, the effects of parental assistance and tutoring constitute a complicated and difficult issue, and the measured phenomenon may be the observed result of a complex interplay of several variables that have differential effects on the efficacy of the learning process. Additional investigation into this complex relationship would contribute to our understanding of the complex relationships between parents, tutors and students in the production of learning.

The SAIP contextual questionnaires have attempted to measure the effects of socio-economic factors through the proxies of the level of the student’s mother’s education and the number of books in the student’s home. These factors have some relationship to SES; however, the learning context is much larger than the student’s home or the student’s school environment. Cartwright and Allen (2002) examined the level of students reading performance in relation to a number of community factors including the types of employment, the educational levels of the people living in the community as well as the students’ opportunities to participate in a variety of cultural, artistic, and sporting events. Their research revealed that the overall wealth of the community (economic, social and cultural) correlated significantly with the strength of reading scores. The SAIP/PCAP project has not explored the effect of the community on the academic performance. The contribution of the community may prove to be a significant factor in small schools in rural areas which have consistently performed at levels below that of larger urban schools. This is a complex issue with a large number of interrelated variables, but in an era when we as Canadians are more concerned about creating better opportunities for more Canadians, this issue should merit investigation.
CMEC has been quiet on the issue of differential performance by Aboriginal Canadian students until the release of PCAP -13 2007. For several years, CMEC has captured demographic data for each jurisdiction on the participation of Aboriginal students. The 1999 SAIP Science questionnaire began to gather data about the perceived incidence of learning problems: The Northwest Territories, whose population is composed predominantly of people of Aboriginal ancestry, reported a much higher rate of learning problems than other jurisdictions. PCAP -13 2007 disaggregated performance data by the proportion of Aboriginal students enrolled in the school. Saskatchewan has the largest number of schools in which the proportion of Aboriginal students exceeds 25 percent. The performance level of the school correlates negatively with the proportion of Aboriginal students enrolled in the school. However, the interaction and loading of the personal socio-economic factors relating to the student’s mother’s level of education, the number of books in the home, and socio-economic factors relating to community wealth have not been isolated. Given that PCAP-13 2007 data is available for secondary analysis, it would seem to be both critical and urgent to conduct secondary studies on the data from these schools to determine which factors impact most significantly on schools success. Based on the outcomes of this analysis, school boards, school administrators, teachers and parents can explore options for targeted programming.

Closing Remarks

My understandings and questions are not exhaustive, but they do serve as a summary of my journey through the decades of SAIP and PCAP. Not only does this journey reveal something about SAIP, PCAP and our Canadian perspective on testing, it
also tracks the evolution of the Canadian educational systems, issues related to teaching and learning, and the interface of our educational systems with their school administrators, teachers, students, and parents. This has been journey about the production of our Canadian educational identity.
References


446


Demetriou, A. (2007). The g-factor is alive and well but what shall we do about it? European Journal of Personality, 21, 720-722. Published online in Wiley InterScience (www.interscience.wiley.com)


Hoffman, J. (2004). How does Canada’s education stack up against other countries? In M. Moll (Ed.) *Passing the test: The false promise of standardized testing* (pp. 51-56). Ottawa, ON: Canadian Centre for Policy Alternatives.


Joint Advisory Committee (1993). *Principles for Fair Student Assessment Practices*
in Canada. Edmonton, AB: Joint Advisory Committee, Centre for Research in Applied Measurement and Evaluation, 3-104 Education Building North, University of Alberta.


McAdie, P. (2004). Testing the limits. In M. Moll (Ed.) Passing the test: The false promise of standardized testing (pp. 149-155). Ottawa, ON: Canadian Centre for Policy Alternatives.


Organization for Economic Cooperation and Development (n.d.). History. Retrieved on November 12, 2009, from http://www.oecd.org/pages/0,3417,en_36734052_36761863_1_1_1_1_1_00.html


Riley, K. & Torrance, H. (2003). As national policy-makers seek to find solutions to national education issues, do international comparisons such as TIMMS and PISA create a wider understanding, or do they serve to promote the orthodoxies of international agencies? Journal of Educational Change, 4, 419-425.


