EDUCATORS' VIEWS AND PRACTICES REGARDING HIGH STAKES TESTING
IN GRADE 12 MATHEMATICS CLASSROOMS

by

GLEN ERIC MACPHERSON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Curriculum Studies)

THE UNIVERSITY OF BRITISH COLUMBIA

March 2007

© Glen MacPherson, 2007
ABSTRACT

This dissertation is a case study of the views and practices of mathematics teachers in two British Columbia high schools – their pedagogical, professional, and personal responses to high stakes testing, and the factors that mediate those responses, such as administrators, department heads, other teachers, students, and external influences. There was homogeneity across the two sites in how the teachers approached the Principles of Mathematics 12 course. 'Examination-style content' was reflected in teachers’ lectures, assessment and evaluation, and classroom resources. Banks of previous examination items were a common resource at the two sites. Lecturing was the dominant mode of introducing content to students, and there were instances of the use of particular teaching practices – such as test coaching – aimed at improving examination performance. Evidence of downward examination pressure into earlier grades was found in one site. Administrators reported that provincial examinations significantly mediated administrative decisions on staffing, daily scheduling, yearly timetabling, and school-level programs. Teachers and others perceived media rankings of test scores as powerful but simplistic. Students expressed confidence in their teachers, reporting no undue levels of stress as a result of provincials examinations. These findings are consistent with other research that claims there are significant relationships between external testing and teaching practices, and supports the findings of previous examination-impact research from Anderson, Muir, Bateson, Blackmore, and Rogers (1990), and Wideen, O’Shea, Pye, and Ivany (1997).
# TABLE OF CONTENTS

Abstract ............................................................................................................................................... ii

Table of contents ................................................................................................................................. iii

List of tables ......................................................................................................................................... vi

List of figures ......................................................................................................................................... vii

Acknowledgements .............................................................................................................................. viii

Dedication ........................................................................................................................................... ix

CHAPTER 1 – Introduction and Background to the Problem ......................................................... 1

The Research Questions ...................................................................................................................... 4

Significance of the Problem Area ........................................................................................................ 5

Delimitations ........................................................................................................................................ 6

Limitations .......................................................................................................................................... 7

CHAPTER 2 – Review of the Literature ............................................................................................... 9

Introduction .......................................................................................................................................... 9

High Stakes Testing: Definition and Rationale .................................................................................... 9

International Perspectives .................................................................................................................... 11

Historical Perspectives .......................................................................................................................... 12

  Brief History of Educational Testing in BC ....................................................................................... 13

Curricular Alignment as a Model for Understanding the Relationships Between Objectives, Testing, and Teaching ........................................................................................................... 18

  Curricular Alignment .......................................................................................................................... 18

  The Word ‘Alignment’ ........................................................................................................................ 20

Research on Alignment Between External Testing and Teachers’ Views and Practices ..................... 22

  Criteria for this Literature Review ...................................................................................................... 24

  Studies Showing Strong Alignment Between Testing and Teachers’ Views and Practices ............. 24

  Studies showing Limited Alignment Between Testing and Teachers’ Views and Practices ............ 39

Questions and Issues Emerging from the Research Literature ...................................................... 46
LIST OF TABLES

Table 1. Summary of Data Collected at Greenhill School.................................74
Table 2. Summary of Data Collected at Pine River Secondary........................75
LIST OF FIGURES

Figure 1. The Curriculum Alignment Model ................................................................. 21
I would like to thank my committee members Dr. Cynthia Nicol and Dr. David Robitaille for bringing their experience and high standards to bear on the process that culminated in this dissertation. I would like to thank Dr. Gaalen Erickson for becoming my advisor, bringing progress to my doctoral program, and giving his knowledge and steady guidance throughout this process. Each committee member has challenged me to examine my perspectives and to strive for academic excellence.

I would like to thank my colleagues and participants at both research sites for their enthusiasm and for letting me learn more about our profession.

I would like to thank my father, Dr. Eric MacPherson, for his wisdom, endless encouragement, and for inspiring me to think for myself, ask my own questions, and find my own way.

I would like to thank Karina for giving me refuge after long days of writing.
DEDICATION

To my daughters, Scarlett and Desiree.
CHAPTER 1 – INTRODUCTION AND BACKGROUND TO THE PROBLEM

This study examines the relationship between high stakes mathematics examinations and the related views and practices of teachers. It does so by way of two case studies.

While the primary focus is on the views and practices of the mathematics teachers in two schools, it is important to take into consideration other important mediating factors influencing these views such as the beliefs and practices of the administrators and colleagues involved, those of their students, and other influences external to the schools.

Large-scale external achievement testing has a long history worldwide, dating back at least two millennia to China and Rome, continuing in nineteenth-century England, and flourishing most recently in the United States (Elman, 2000; Suen & Yu, 2006). Its history in Canada, as in other jurisdictions, has been characterized by periodic waves of heightened attention and uses. In each wave, their use has been justified by a shifting blend of perceived needs for accountability and, more recently, by the belief that testing and its consequences are levers for school reform as a means of pruning frivolous additions to the intended curriculum, and as an instrument for rewarding effective teachers and administrators, identifying weak teachers, and motivating lazy ones. In general, high stakes tests have been seen as “raising the standards of schooling” (Green, 2002; McDonnell, 1994; Phelps, 2005; Ravitch, 1995).

Concomitantly with the diversity of both intended purposes and sometimes-unanticipated effects of widespread testing, there have been significant variations in the perceived stakes attending these tests. In 1000 B.C. (see Judges 12:6 in the Bible), and
gain in the Sicilian Vespers of 1282\(^1\) the stakes were literally a matter of life or death. And while the stakes are not now those of that moment in history, the current North American wave of externally mandated high stakes testing entails higher stakes than has most often been the case. Those stakes can now affect students’ promotions, graduation, and university and college entrance (Huebert & Hauser, 1999), teachers’ salaries and promotions (Klein, Hamilton, McCaffrey & Stecher, 2000), tenure (Madaus, 1988b), administrators’ planning and decision-making (Houston, 2000), and school and school district funding (Langenfeld, Thurlow, & Scott, 1966).

The United States Congress’ *No Child Left Behind* Act requires that all states implement high stakes curriculum-linked testing across various grade levels and subject areas. The stakes include parental discretion in transferring students out of low-performing schools and the possibility of both school closings and the state takeover of low performing schools. Individual states have augmented the provisions of the Act with local incentives and punishments. Florida, for example, has implemented a direct linkage between class’ test results and their teachers’ professional evaluations. If the current wave has not abated by that time, it is projected that, as a result, by 2012, 72 per cent of American students will face high stakes school graduation tests.

State initiatives have generated diverse reactions from educators and educational researchers. The American Educational Research Association (AERA, 2006), the National Council of Teachers of Mathematics (NCTM, 2006) and journals such as the *Phi Delta Kappan* have repeatedly issued position statements regarding high stakes

\(^1\) So named on account of a church service incident that triggered a major Sicilian revolt against French rule in that year. A test similar to that reported in Judges was used to identify and dispatch a number of French soldiers. (Illuminato, 1982).
testing, urging cautions concerning the belief that they can help bring about desired educational reforms. Many of them have issued warnings about the possible negative impacts that they can have on what are widely agreed to be desirable qualities of schools and students' educational experiences. For example, the NCTM (2006) position statement asserts that "basing major decisions about students, teachers, schools, or instructional programs on a single test is inappropriate and inconsistent with what we know about learning and assessment" (p. 1). Taking a more extreme position, Amrein and Berliner (2002), Koretz (2005), Linn (2000), and Madaus (1988a) claim that high stakes testing invariably alter and corrupt the educational processes they aim to assess.

At the other end of the spectrum of views, there are authors who champion testing with stakes as a powerful lever for educational change and improvement of instruction (Popham, Cruse, Rankin, Sandifer, & Williams, 1985; Popham, 1987; Ravitch, 1995; Shanker, 1994). Furthermore, researchers have paid increasing attention to the ways in which external high stakes testing has affected schooling and the ways in which teachers approach their work. It is not surprising that, collectively, they have often been equivocal in their assessments. Even researchers who are known to be critical of the impacts of high stakes testing (e.g. Koretz, 2005) have identified impacts that they believe to be positive. Others have questioned whether or not high stakes tests are in fact very strong levers for classroom change at all (Firestone, Mayrowetz, and Fairman, 1998; Grant, 2000, 2001). One of the only consensuses in the research literature is that high stakes testing does have impacts on teaching practices (Cimbricz, 2002; Grant 2000; Madaus, 1988a; 1992; Koretz, 2005; Linn, 2000; Smith, 1991). It has often been suggested, though sometimes with scant empirical evidence, that teachers respond to high stakes testing with a wide
spectrum of views and practices, many of which fall under the umbrella term “teaching to the test” – a broad, ill-defined, and often pejorative label for the teaching behaviours that are commonly attributed to high stakes testing.

Several researchers have noted the array of inadequate evidence now available, and have expressed the need for deeper explorations in this area (Cimbricz, 2002; Haertel, 1999; Rex & Nelson, 2004; Zancanella, 1992).

For example, Cimbricz (2002) wrote,

“The influence state testing may or may not have on teachers and teaching expands beyond individual perceptions and actions to include the network of constructed meanings and significance extant within particular education contexts. … Studies that provide a richer, more in-depth understanding of the relationship between state-mandated testing and teaching in actual school settings, therefore, not only point toward important directions for future research in this area, but are greatly needed.” (Abstract).

Much more needs to be known about the relationships between external testing and teaching, because knowledge claims concerning that relationship have often been grounded in anecdotes and research reviews rather than empirical classroom observations. This observation captures the rationale for this study, and at the same time, suggests the approaches and methods that it requires.

The Research Questions

There are two research questions in this study:
1. What are mathematics teachers' views and practices regarding high stakes testing?
2. How are these views and practices mediated by administrators, mathematics department heads, other teachers, students, and external influences?

The stakeholders referred to in the research questions comprise those associated with the two high school research sites.

**Significance of the Problem Area**

My exploration of the two primary research questions is grounded first in my grasp of the political, professional, and administrative structures and relationships in British Columbia secondary schools, the secondary mathematics curriculum, and the influences of parents and the media; and second, in my field work in the two selected secondary schools and in study of the educators' views and practices in those two schools.

My aim is to paint pictures of classroom life in the twelfth grade mathematics classrooms of two schools, describing as thoroughly as possible aspects of observed classroom practices and teachers’ views as they relate to provincial mathematics examinations. This study will therefore provide a current, contextualized, and in-depth account of the impact of provincial testing on teachers and teaching. The commonalities in those accounts and their interactions with other sources of information (such as administrators, the media, and students) will determine the degree to which it will be possible to draw valid inferences concerning the impact of high stakes testing in British Columbia, and possibly elsewhere. The inferences that can confidently drawn in this study of the views and practices of the teachers provide a basis for making statements and
claims about the impact of provincial examinations on the views and practices of at least
British Columbia mathematics teachers.

Finally, this study aims to reveal how mathematics teachers make curricular decisions in the context of provincial examinations and the specific methods they use to prepare their students for these examinations. I have found that there is a large and eager audience for detailed information concerning teachers' practices, methods, and techniques. In my web searches, I have found literally hundreds of commercial products purporting to improve student test scores, and queries from teachers seeking out these kinds of resources. I will not attempt to synthesize either the full range of those queries or what has been distributed in response to them, but my results can at least contribute to that dialogue and, if nothing else, suggest ways to enhance students' test performances.

Delimitations

In discourse about the assessment of the quality of research studies, it has become fairly common to distinguish delimitations from limitations. Delimitations are deliberate narrowings of the ambit of a study, often linked to practical concerns over time, resources, and availability of research sites.

There are several delimitations in this study. First was my decision to carry out the study in only two schools. A study that aims to study this kind of phenomenon in depth must necessarily be so limited. But doing so raises concerns as to how the sites selected represent the population from which they are drawn. The second delimitation is that I planned to spend one month in full-time attendance at each school, and there was
the risk that this limited time might prove to be insufficient. It did not. After one month of immersion in the educational culture and operation of each school, I was satisfied that further time would have contributed little more.

Third, given the time available, I attended to only those significant influences external to the school that were identified by the primary participants in the study. For example, I could have interviewed parents, who might have helped to triangulate some of my observations, but doing so would have truncated the time available for interviewing and observing teachers, administrators, and students, the prime foci of this study. The possible benefits deriving from collecting such further data are therefore consigned to my suggestions for further research in chapter 5. Nevertheless, the reader must keep in mind the diversity of the population from which these schools come. The inferences drawn here may be as important for model clarification and raising questions for further study as for making far-reaching knowledge claims concerning the effects of high stakes testing on the views and practices of all British Columbia mathematics teachers.

Limitations

Limitations are those unanticipated artifacts of the conduct of a study that may moderate and limit the intended applications or generalizations. One potential limitation of this study derives from the extent to which I relied on teachers’ and administrators’ perceptions in order to answer the research questions. The extent of my access to classrooms for observation varied considerably across teachers. My access to some classrooms was restricted more than I would have preferred. In the extreme, with one
teacher I was able to observe full lessons only twice. At the other extreme, I observed one teacher formally or on a ‘drop in’ basis more than a dozen times.
CHAPTER 2 - REVIEW OF THE LITERATURE

Introduction

This chapter reviews the research literature that is relevant to the research questions:

1. What are mathematics teachers' views and practices regarding high stakes testing?
2. How are these views and practices mediated by administrators, department heads, other teachers, students, and external influences?

In it, I first define high stakes testing and provide the rationale for its use in this study, followed by a historical perspective on high stakes testing, including a brief history of governmental testing in BC. Next, I establish curricular alignment as a model for discussing the instructional impacts from high stakes testing. Finally, I review the research literature regarding the interactions between externally mandated testing and teachers' views and practices, and discuss questions and issues emerging from it, including its implications for the design of my study and the analysis of the data.

High Stakes Testing: Definition and Rationale

Test results are termed ‘high stakes’ if they “carry serious consequences for
students or for educators" (AERA, 2006, p. 1). Defined in that way, such tests clearly have implications for grade promotion and graduation (Heubert & Hauser, 1999), salaries, promotions, and tenure for teachers and administrators (Orfield & Wald, 2005), and funding levels for school districts (Langenfeld, Thurlow, & Scott, 1996). High stakes testing for public school students is currently a major concern for educational stakeholders in the United States (Amrein & Berliner, 2002; Koretz, 2005; Linn, 2000), and increasingly in Canada (McEwan, 1995).

The common thread running through most recent American educational reform movements is testing. Widespread testing has always been an attractive option to those wanting more information about educational progress, and to those interested in causing change in the system. Testing is relatively inexpensive, quickly implemented, and its products are visible (Koretz, 2002a; Linn, 2000). Its particularly attractive feature is that testing can be implemented and measured during one political term of office (McDonnell, '1994). Testing has therefore been a centerpiece for political reforms that have attempted to slice through perceived problems and have claimed to raise the standards and accountability of public schooling.

The current wave of educational testing is rooted in concerns for accountability (Amrein & Berliner, 2002; Linn, 2000). The push for accountability seems to devolve from two overlapping inputs. The first is that the quality of American education, as perceived by the public, business people, and politicians, is often inferred from results of widespread tests. The reported relatively poor performance of American students on international achievement tests of mathematics and science has raised some concerns and has helped shape these opinions. The second input is largely ideological and political,
implying that the source of reputed malaise in education is the immense inertia and self-serving resistance to change that is inherent in the structure of public education. From this viewpoint, external testing with consequences is seen to be a quick and effective solution – much quicker than comparatively nebulous proposals such as improving teacher training or attracting more capable applicants to the profession (Linn, 2000). From both perspectives, testing is viewed as a powerful motivator for ensuring that teachers teach what they should, and that students work hard to learn.

**International Perspectives**

High stakes testing is an important component of many modern educational systems, particularly in Asia. Non-Asians have written extensively about the Japanese examination system (Orphal, 2000). Japanese students face two testing barriers, the first for high school entrance and the second for university entrance. The examination sessions, known in Japan as “Hell Week,” are national institutions. Hell Week has become a cultural constant – an inevitable yearly event. Its effects spill over into many aspects of life – obligatory cramming schools (‘juku’), long hours (the common phrase, ‘pass with four, fail with five’ refers to hours of sleep), tutoring, special examination diets, the rerouting of traffic and other sources of noise during exam time, even suicides – it is understood by everyone that the examinations often define a student’s educational and vocational future.

Most countries in the world now use high stakes testing, and while there may be some internal opposition to the level of the stakes and their perceived impacts, the testing
regime seems to command respect in Japan and in other countries. Indeed, what may seem to be odious testing effects in one culture may, elsewhere, reflect widely accepted, if not embraced, cultural and social norms (Orsini & Yi, 2006; Roper Center, 2006).

**Historical Perspectives**

Throughout Asia, high stakes testing is a tradition to the extent that it can be considered part of the cultural fabric of some countries. The Chinese civil service examinations, for example, spanned a continuous period from 606 ACE through 1905 ACE (Suen & Yu, 2006). Large numbers of people wrote these exams and the stakes attached to them were life-altering. Some men spent their entire lives trying to pass them, because success on the examination promised a better life for them and their descendents (Suen & Yu, 2006). Issues such as test design, reliability, test coaching, scorer bias, construct irrelevancy, and cheating were grappled with by the test administrators. The tests were such a fixture in Chinese life that they were the subject of satires and classic novels. Similar examination systems existed for about 1000 years in Vietnam and Korea (Suen & Yu, 2006).

Between 1753 and 1909, the Cambridge Mathematics Tripos examinations exerted tremendous influence over the British university curriculum (Forfar, 1996). The media publication of the rankings of examination performance attracted wide attention. The public interest in the Tripos was so great that jargon such as “Wranglers” (top scorers), and “Optimes” (lower scorers) entered common parlance, and questions from previous tests were printed in newspapers, like today’s chess and bridge problems. The
prestige attached to scoring in the top few positions was so great that the brightest and
most promising freshmen in mathematics at Cambridge did not attend lectures from
professors. Rather, they were sequestered with private “exam coaches,” who had
reputations for intense training methods and bringing home the best results. Exam
coaches taught time management, instant recognition of question types and pre-cooked
solutions, examination structure, and so on. The great mathematician G.H. Hardy
was revolted by the impact of the Tripos on his own education and on mathematics in
Britain, and in a major speech at Cambridge blamed them for destroying Britain’s ability
to compete with France and Germany in creating advances in pure and applied
mathematics (Forfar, 1996).

**Brief History of Educational Testing in BC**

The 130-year history of widespread testing in BC is cyclic, and the themes, issues,
and debates surrounding the tests have not changed substantially: accountability,
minimum standards, professional autonomy, and the impact on instruction and students
are still the most debated issues. The purpose of this section is to provide a brief
overview of the history of province-wide testing in BC and to reveal the issues, forces,
and research that have attended these issues.²

Province-wide testing in various forms and with various labels has been in place
continuously since 1876. The initial Departmental Examinations covered four subjects,
were competitive, and were used to determine which students were ready for high school.
The then-Superintendent of Public Schools criticized educators for the poor performance

---
of the first round of students who wrote the examinations, and he said that in the future such failure would be attributable, with few exceptions, to weak teaching. By 1900, Departmental Examinations had permeated the high school program, becoming exit examinations that were the yardstick for measuring academic performance. Students could not graduate from high school unless they passed the tests.

In 1925, a Royal Commission on Education in British Columbia requested and heard evidence regarding the impact of Departmental Examinations. Putnam and Weir (1925) had been asked in 1924 to study and report on the provincial program. They claimed that standards of teaching and teacher qualifications had been improved, but that the examinations had led teachers to train students for the tests rather than teaching them the full curriculum (Putnam & Weir, 1925). They criticized the tests for causing 'grade retardation,' and encouraging weaker students to leave school so that results would improve. They also wondered about the importance of the Departmental Examinations in perceptions of school quality.

The extent of Departmental Examinations had been reduced by 1931, by which time teachers were permitted to promote students in every year up to grade 12, but not for graduation. High school accreditation was implemented in 1937, after which students who earned a C+ or better could be recommended for admittance to university without having written the Departmental Examinations. Other students and those attending schools that had not earned accreditation still had to write the tests. The period leading up to the implementation of accreditation saw serious debate surrounding the importance of external testing and the impact of such testing on the educational system. In the end, arguments in favour of testing prevailed.
In 1941, achievement tests were developed and implemented across the other high school grades. In 1946, the province consolidated its oversight of educational testing in its new Division of Tests, Standards, and Measurements. New university-entrance curricula in 1946 were coupled with University Program Departmental Examinations in those subjects and their prerequisite courses, under the ultimate aegis of the Board of Examiners. Experienced teachers, with Division oversight, were engaged to set and mark the examinations. Teachers could still recommend students for course credit, but students applying for scholarships needed to write the examinations.

The Second Royal Commission on Education of 1960 was established in part to address prevalent claims that education in BC was internationally substandard. Widespread examinations were seen as a critical component of improving schooling, for both “diverting weaker students to more appropriate courses,” and as “an incentive for more able students to work harder” (Anderson, Muir, Bateson, Blackmore, & Rogers, 1990, p. 15). The Commission members felt that teachers could still adjust their courses to meet the needs and interests of their students while providing a thorough treatment of the curriculum. Well aware of potential testing effects, they called for more essay-type questions rather than low-level recall items.

During the 1960s, there was a considerable groundswell against widespread testing. Questions about the purpose(s) of the examinations, ongoing concerns regarding the perceived negative impacts of the testing program, and a move in other jurisdictions away from governmental examinations culminated in a 1971 announcement of the elimination of government examinations in 1973. Over the next ten years, the government introduced learning assessments and standardized achievement tests in order
to provide information regarding achievement levels over time, directions for curricular change, and indicators of areas needing funding. Two programs were introduced: The Provincial Learning Assessment Program in 1976/1977, and the Classroom Achievement Test Program in 1980. These tests were voluntary but received wide acceptance and use. University Scholarship examinations remained.

The remainder of the 1970s saw increasing demands from many quarters for a return to provincial examinations. The public and representatives of the universities both expressed concerns over what they felt was the declining quality of education and the lack of standardization in teachers’ evaluations. In 1979, a survey reported that more than half of British Columbian adults felt that educational standards had eroded. At the same time, other jurisdictions in Canada, as well as in Britain and the United States, reintroduced widespread testing. The British Columbia Teachers’ Federation (BCTF) and its members were opposed to the return of governmental examinations. They were not consulted before the Minister of Education announced in 1982 that provincial examinations would return. There was some inertia at first, but by the end of 1984, provincial examinations had become a requirement for most graduating students. Anderson et al. (1990) concluded that the 1984 reintroduction of provincial examinations was essentially political, and that, despite disagreement from the BCTF, the examinations have, over time, gained wide acceptance, both with the public and among educational stakeholders.

The Third Royal Commission began in 1987 with assessment and evaluation as one aspect of the Commission’s mandate. The Commission found no groundswell of opinion against the newly introduced provincial examinations but they heard familiar
concerns, as in earlier periods, from commentators on external testing concerning the discrepancies between teacher and examination marks, the weighting of examinations in students’ final grades, the narrowing of the received curriculum, and the devaluing of non-examinable courses.

The Committee’s Research Team initially recommended that provincial examinations be established in four subject areas, but pursuant to their analysis of teaching and learning functions, and after receiving many submissions, the Research Team recommended the complete elimination of all provincial examinations. The Royal Commission itself, however, recommended that provincial examinations be established in all subjects at a reduced weight of 30 per cent of students’ final grades. The Department of Education, to whom the Commission reported, decided to maintain provincial examinations, at a weight of 40 per cent in courses in the Humanities and Sciences. Scholarship examinations were eliminated in the mid 1990s. As of 2006, there are no announced plans for formal governmental reviews or studies of the provincial examination program.
Curricular Alignment as a Model for Understanding the Relationships Between Objectives, Testing, and Teaching

Curricular Alignment

A recurring, and sometimes bitter, complaint about high stakes testing concerns its perceived power to create a large gap between prescribed curricular goals and students' educational experiences (Haney, 2000). The research literature contains examples of components of prescribed school subjects, including science, simply not being taught because those topics were not to be found in the content covered on an externally mandated high stakes test (Haney, 2000; Jones et al., 1999; Smith et al., 1989).

This phenomenon is one of several that can be viewed under the rubric of an alignment model. Here, we are concerned with the connections between what students are expected to learn and do, and the instruments used to assess progress on those expectations. Continually, it has been argued that high stakes testing is a potentially quick and effective means of improving instruction. The notion is that if tests are appropriately generated from and match curricular objectives, then we will have “a test worth teaching to” (Millman, 1981; Popham, 1987; Popham, Cruse, Rankin, Sandifer, & Williams, 1985), and a basis for “test-driven instruction” (Cohen, 1987; Comfort, 1991; Elia, 1986; Fahey, 1986; Koczor, 1984; Smith & O’Day, 1990; Tallarico, 1984). The expectation is that if there is a strong relationship between objectives and a test, then teachers will develop classroom practices that will enable their student to both achieve good test
results and attain the desired objectives. The goal is close alignment between what is expected, what teachers teach, and how students are tested (Anderson, 2002).

It seems that the desire for such alignment has been a major force in shaping high stakes tests in the United States. Canadian high stakes tests have always been aligned to curricular standards, those in BC for more than 100 years. United States Congressional investigators, who scrutinized the BC educational context, viewed this as a particularly desirable feature of high stakes assessments in Canada (US Department of Education, 1993). Concern for alignment has largely stemmed from the No Child Left Behind Act, and all US states are now attempting to implement high stakes tests aligned to their curricula. “Curriculum alignment is central to the success of accountability programs” (Anderson, 2002, p. 9).

Examined more closely, the alignment model has three primary components: assessments (including high stakes testing), prescribed curricular standards (objectives), and classroom practices and materials (Anderson, 2002; Barnes, Clarke, & Stephens, 2000; Gilbert, 1962). These three pillars can be viewed as being at the vertices of a triangle, which models the three relationships between them (Ippolito, 1990; Pickreign & Capps, 2002): alignment between standards and tests (Buckendahl, Plake, Impara, & Irwin, 2000; Kendall, 1999; Webb, 1999), and alignment between testing and teaching (Barnes, Clarke, & Stephens, 2000; Cohen, 1987; Gamoran, Porter, Smithson, & White, 1997; Pickreign & Capps, 2000; Webb, 1999), and the alignment between teaching and standards. These relationships together comprise curricular alignment in the fullest sense of the term (Anderson, 2002; Biggs, 1999). There are few studies that mention, let alone
explore, the complete set of interrelationships suggested by the alignment model, although some commentators have done so (e.g. Berube, 1994; Drake, 2004).

The curriculum alignment model is not only a taxonomical tool for categorizing elements of the phenomena but, like any good model, is a heuristic for examining facets of the phenomenon that might otherwise be missed, and a tool for clarifying discourse. It suggests studying the interactive relationships between three variables that I and others have identified as being of the greatest import: teaching practices and materials, high stakes assessments, and curricular objectives. This model frames my research.

The Word 'Alignment'

It should be clear from the above introduction that the phrase curricular alignment has been used in several different ways. In my study, the word curriculum refers to a collection of goals, mathematical content (scope and sequence), and intended experiences for students; in the BC context this means the content of the Instructional Resource Package (IRP). This is just one usage of the term curriculum and this particular one has frequently been called the ‘intended,’ ‘mandated,’ or ‘espoused’ curriculum (Olivia, 1997).

Alignment has referred to a single relationship, such as the matching of test content with standards content in the subject area (Pedulla, Abrams, Madaus, Russell, & Ramos, 2003), to “the desired convergence between a system’s expectations, as expressed in its course documents and recommendations for teaching and learning, and
what is actually mandated for assessment” (Barnes, Clarke, & Stephens, 2000), and to the model itself (Anderson, 2002; Berube, 1994).

The alignment between teaching and testing has been called “curriculum alignment” (Levine, 1982; Niedermeyer & Yelon, 1981) and “instructional alignment” (Cohen, 1987). In this study, curricular alignment refers collectively to the six vectors of alignment and influence suggested by the model shown in Figure 1 (Anderson, 2002; Biggs, 1999).

![Diagram](image)

**Figure 1. The Curriculum Alignment Model.**

Where I refer to aligning teaching with testing, I am referring to teachers harmonizing their practice to external tests or conversely, revising tests so as to harmonize with teaching with the goal of improved test results. The alignment between testing and teaching says more about what educators do – practices – than about the views that they hold, which may or not be in harmony with the instructional methods they feel compelled
to use (Madaus et al., 1992; Wideen et al., 1997). In any case, an examination of the alignment between classroom practices and high stakes exit examinations is a major component of this study.

**Research on Alignment Between External Testing and Teachers’ Views and Practices**

Most prescriptions for good teaching practices include: sufficient flexibility to allow for enrichment, exploration, and the creation of opportunities for students to learn the full spectrum of desired aims and outcomes as specified in curricular documents. Some curricula, notably in the humanities, are stated in such general terms. They provide broad descriptions of what students should be doing and learning, and suggest considerable latitude in the selection of content and methods of teaching. Even when curricula are stated more narrowly, as in mathematics and science, in many jurisdictions, teachers have not been severely restrained in the choice of their teaching methods and approaches. They have been free to craft instruction, tuning it to the diversity of their students and to local expectations. Ideally, the “curriculum-in-use” (Wilson, 2005) has adapted the legislated curriculum in ways that attend to the individual and collective needs of students. Teaching with that broad mandate has been rooted in widely held beliefs about the nature of teachers’ professionalism. In other words, teachers have been vested with sufficient autonomy to align their instruction with larger curricular goals as mediated by students’ needs.
But when high stakes tests are involved, students' immediate needs can change, especially when their performances on a single test can have serious educational and vocational implications. Teachers, therefore, have a moral and professional obligation to see that their students do as well as they can on the tests. This pressure may generate tension between teachers' educational ideals and what they must do in classrooms. This tension has been discussed often in commentaries and in the research literature (Anderson et al., 1990; Koretz, 2005; Madaus, 1988b; Wideen et al., 1997). For example, Madaus et al. (1992) wrote of the "conflict between pressure for students to do well on tests and contradictory pressure which suggest that 'teaching to the test' is professionally and/or ethically wrong" (p. 16). When professional and school reputations are on the line, teachers can be driven to value high test scores.

High testing stakes can mediate how teachers look out for the best interests of their students, whose graduation and post secondary aspirations often hinge on provincial examination scores, and one way they can do that is by aligning their instruction with examination content. Madaus (1988a) writes, "A review of the effects of high stakes tests over many years and in a number of countries indicates that, faced with a choice between objectives which are explicit in the curriculum and a different set of objectives that are implicit in the test, teachers and students generally choose to focus on the latter" (p. 39). Learning more about the subtle effects of that tension in classrooms is one of the goals of my study.
Criteria for this Literature Review

Conflicting conclusions and a litany of instructional alignment effects can be distilled from authors' personal opinions, letters, essays, anecdotal evidence, survey research, and a few observational studies. These effects were found to varying degrees and in a wide variety of schools, classrooms, grade levels, school districts, and states. In this summary, I have followed the example of Cimbricz (2002), restricting the literature review to those papers and studies that can be categorized as "research" according to the criteria established by Howe and Eisenhardt (1990) and to widely and frequently cited papers that review this literature or theorize on it, particularly studies addressed to the teaching of mathematics (ideally at the high school level), and to two local studies that help frame the BC context.

Studies Showing Strong Alignment Between Testing and Teachers' Views and Practices

A majority of opinions expressed in the literature take it as a given that there are significant relationships between testing and teaching. The following behaviours have been identified as teachers' strategies moves to align their instruction with high stakes tests: narrowing the instructional focus and bringing it into closer alignment with expected test content (Barnes, Clarke, & Stephens, 2000; Koretz, 2005; Koretz, Mitchell, 2000). Howe and Eisenhardt (1990) provide five general standards for qualitative and quantitative educational research, specifically, that there be: (a) a fit between research questions and data collection and analysis techniques; (b) the effective application of specific data collection and analysis techniques; (c) an alertness to and coherence of background assumptions; (d) an overall warrant; and (e) an awareness of both external and internal value constraints.
Barron, & Keith, 1996; Madaus, 1988a; Madaus et al., 1992; Smith, 1991; Wideen et al., 1997), bringing instruction into better alignment with curricular objectives (Koretz, 2002b, 2005; Smith, 1991; Stecher, 2002), shaping assessment practices so that quizzes and tests mirror external test content and timing (Center on Educational Policy, 2005), providing more instructional time (Koretz, 2005; Stecher, Baron, Chun, & Ross, 2000, 2002), working harder and more efficiently (Borko & Elliott, 1999; Koretz, 2005; Wolf & McIver, 1999), reallocating instructional time (Koretz, 2005), exhorting students to do better (Smith, 1991), coaching (Lin, 2000; Wideen et al., 1997); and cheating (Cannell, 1989; Cizek, 1999).

Madaus (1988a)

In an influential paper, Madaus aimed to establish some foundational and general relationships between testing and teaching. His overarching observation is that high stakes testing significantly affects instruction, often negatively. This view seems to be supported by many educators and educational researchers. In his largely anecdotal paper, Madaus adduced six related theories, or principles, that describe and support his position as to the consequences of test-driven instruction under high stakes.

Madaus calls attention to three artifacts of the Testing-Practice vectors of the model. First, he asserts that if important decisions are presumed to be related to test results, then teachers will teach to the test and claims that teachers’ interpretations of stakes, rather than the legislated ones, mediate instructional responses. Other researchers
support this claim (Cimbricz, 2002; Corbett & Wilson, 1991; Koretz, 2005). Previous research in BC indicates that teachers, and others, view provincial examinations as having very high stakes (Anderson et al., 1990; US Department of Education, 1993; Wideen et al., 1997). It should be noted that Amrein and Berliner (2002) claim that high school exit examinations often have some of the highest stakes, legislated and otherwise. Here “Teaching to the test” refers to the overall gearing of instruction to serve the test. He also suggests that widely-viewed media reports could be a significant aspect of the stakes. BC has such a yearly report, and it serves as a touchstone for exploring the lived stakes for teachers.

Second, Madaus claims that in any setting with a high stakes test, a tradition of past tests develops, and it eventually leads to the de facto definition of the curriculum. This effect is strong and frequently seen (e.g., Anderson et al., 1990; Koretz, 2005; Smith, 1991; Smith et al., 1989; Wideen et al., 1997). In some cases, the effect is startling, where specific multiple choice questions from, and even incidental aspects of, past tests are incorporated into instruction (Koretz, 2005; Wideen et al., 1997). This observation suggests that even though there may be sound, detailed, and comprehensive expressions of what should comprise curriculum, a high stakes test makes those standards irrelevant to teachers’ planning and instruction. Past tests are the curriculum, and, moreover, they may sample the subject area domain in such a limited and narrow way as to undermine a meaningful treatment of the content. The tradition of past tests is explored in some depth by Anderson et al. (1990) and Wideen et al., (1997), who both reported strong evidence of this effect.
Third, teachers pay particular attention to the form and format of the questions on a high stakes test (e.g., short answers essay, multiple choice) and adjust their instruction accordingly.

Madaus (1988a) could been characterized as being an attack on widespread testing, and may reflect only one researcher's prejudices. As discussed above in the section on alignment, there are those who believe that high stakes tests can drive teaching in positive ways. We saw, in fact, that alignment between instruction and assessments is an expressed goal of those who wish to use testing as a lever for educational reform. Madaus (1988) does not address alignment issues, but he does respond to advocates of test-driven instruction and coaching. "The view that we can coach for the skills apart from the tradition of test question embodies a staggeringly optimistic view of human nature" (p. 40).

Koretz (2005)

Koretz (2005) was concerned principally with the validity of gain-scores on standardized tests. The validity of increased test scores is rooted in meaningful achievement on the part of students. He asserts that some of the techniques used to increase test scores do not, on their face, cause a meaningful increase in knowledge. This weakens validity, and therefore his discussion in part addresses the methods that teachers use to respond to a high stakes test. He organized those methods along a spectrum of educational soundness. The instructional responses are: those that produce unambiguous, meaningful gains (teachers working harder, more effectively, or increasing remedial instruction time out of class); cheating in various forms; and the middle ground, which
includes reallocation, alignment, and coaching. From this perspective, there is no need to use language such as “teaching to the test” and “teaching the test.” Rather, instructional behaviours are linked to test score increases along a spectrum of validity.

Koretz’s model codifies testing impacts, and allows for positive impacts from externally mandated, high stakes testing – some of the impacts promised by those who believe in improving instruction through testing (e.g., Shanker, 1994). However, Koretz argued that alignment does not insure against invalid inflation of scores through such techniques as coaching, as forewarned by Madaus (1988), a claim that seems to be increasingly accepted (Amrein & Berliner, 2002; Linn, 1983, 2000).

The meaning of test score gains is mostly beyond the scope of this dissertation, but Koretz’ (2005) comments are also connected to an intense debate in the literature regarding the ability of high stakes tests to actually increase student achievement. In a widely cited study of American high stakes testing, Amrein and Berliner (2002) found that increases on a high stakes test did not correlate with increases on a stable, nationwide low stakes National Assessment of Educational Progress (NAEP) test that drew material from the same domain of knowledge. Critics of Amrein and Berliner (2002) identified what they held to be a fundamental design flaw in their study (Raymond & Hanushek, 2003). Similar criticisms are pending (Carnoy & Loeb, 2002).

Corbett & Wilson (1991)

Corbett and Wilson (1991), in the same vein as Firestone, Mayrowetz, and
Fairman (1998), reviewed below, compared local educators’ views of externally mandated high stakes testing in two states, one with a high stakes Minimum Competency Test (MCT). In Maryland, testing began in grade nine and continued until students passed the test. Graduation hinged on – sooner or later – passing the tests in reading and mathematics. Tests in other subjects were being phased in as the data was being gathered. The other state, Pennsylvania, had a very low stakes test for students. Students in Pennsylvania were tested in grades three, five, and eight. The Pennsylvania legislature had made provisions to fund remediation efforts to increase student success, whereas Maryland did not. Corbett and Wilson (1991) also made within-state and within-school comparisons. There were no triangulating classroom observations.

Among other findings, in terms of impacts perceived by teachers, they found significant positive impacts from the high stakes test in Maryland, a better defined curriculum, more information about students, and improving student skills. However, the authors found that teachers had concerns about the tests being used as a benchmark for teacher and system effectiveness.

Even though Corbett and Wilson studied what they called an “exit test,” the stakes in Maryland were not comparable to the stakes in BC. In Maryland, at that time, the students were required simply to pass, at some point during their schooling, tests of the most basic skills. British Columbia provincial examinations sample the full range of topics in the academic, university-entrance courses as well as in more applied and essential courses. This illustrates that “exit exam,” like some other words used in this domain, has a variety of meanings.
I was curious to what degree the stakes mandated under NCLB had impacted testing in Maryland, so I checked the Maryland State Department of Education Web site (August, 2006). Testing has expanded throughout the system, which now has two major assessment tools: the Maryland School Assessment (MSA) and the High School Assessment (HSA). The stakes in these exams are still lower than those in BC because the Maryland tests are not grade 12 academic, and therefore not de facto university entrance examinations. In Maryland, some students pass the HSA as early as the end of grade 10. Examining several dozen sample items from the “Algebra/Data Analysis” assessment, I found the HSA test structure to be strongly isomorphic to the BC Principles of Mathematics 12 provincial examinations in terms of its split between multiple choice and open response items. The wording and difficulty level of the items and content match closely with the BC Principles of Mathematics 10 curriculum. Clearly, the stakes for students in Principles of Mathematics 12 in BC are higher than for grade 10 students in Maryland, who get several chances at passing, and who do not need high scores for university entrance.

Corbett and Wilson’s (1991) conclusions identify several impacts viewed to be desirable, including harder-working students and teachers, better-defined and more focused curriculum, and a sense that students are achieving more. They also reported evidence of what are commonly seen as undesirable effects, such as a sharp narrowing of curricula. This study confirms that the level of the stakes mediates responses differentially, and it also provides one benchmark for discussing the comparatively high stakes testing in BC.
Madaus et al. (1992)

Probably the most comprehensive non-observational study of the relationships between testing and the teaching of mathematics and science is that of Madaus et al. (1992). Their National Science Foundation study conducted through Boston College included, among others, surveys of 2229 mathematics and science teachers of grades 4-12 and interviews with 200 teachers and 100 administrators. Their research led to seven separate reports, in which the impact of testing on minority students was of major concern. The studies included no classroom observations.

Madaus et al. (1992) found widespread impacts of testing on teaching, and concluded that the effects of high stakes are strong and pervasive. The general effects on teachers were reported along a number of lines of inquiry, and five of them are germane here: test preparation techniques, time spent on test preparation, the alignment between the tests and their instruction, the extent of the influence of the tests on instruction, and teachers' responses to general statements about the impact of the tests. More than 80 per cent of the mathematics teachers surveyed reported some explicit test preparation. More than half reported at least one of:

- teaching test-taking skills,
- encouraging students to do better on tests, and
- teaching items known to be on the test.

Fifty-six per cent of mathematics teachers began preparing their students one month or more before the tests. Nineteen per cent of the teachers spent more than 20 hours preparing their students for the test. Forty per cent of them indicated that they
altered their curricula, both excluding and including topics, and altered their classroom assessments to match those of the external test. Teachers reported that the tests caused an increase in whole-group instruction and increased emphasis on lower-level skills. They also reported pressure from administrators for increased scores, adding to the conflict they felt existed between testing imperatives and good practice. Overall, the teachers surveyed believed that they were gearing their instruction to the tests. The authors concluded that the structure of the tests significantly shaped the instruction used in the classroom.

The researchers next analyzed all of the items on the tests in terms of content domain and the level of cognitive skills required. They found that that a very high proportion of them addressed lower-level thinking. They concluded that this led to lower-level teaching in which classroom examples and assessment often closely resembled the tests. They further found that the effects on instruction correlated highly with perceived testing stakes, as predicted by Amrein and Berliner (2002) and shown by McMillan, Myran, and Workman (1999), Pedulla et al. (2003), and Stecher, Barron, Chun, & Ross, (2000).

Anderson et al. (1990)

The aims of the Anderson et al. (1990) study were to "investigate the impacts of the provincial grade 12 examinations upon students, parents, teachers, school and school district administrators, employers of high school graduates, and Ministry of Education
officials; and upon the curriculum, teaching practice, and school administration at the class, school, district, and provincial levels at Grade 12 and lower grades” (p. 25).

Anderson et al.'s was clearly a broad survey research study supplemented with interviews. The authors conducted a pilot study of seven schools in order to interview some stakeholders. After this fine-tuning of the survey instruments and interview protocols, they proceeded with the province-wide study, receiving questionnaires back from 1833 students, 608 parents, 947 teachers, 160 principals, 137 counselors, and 35 superintendents. They followed up with focus group interviews of teachers in five locations around the province in order to explore issues raised by the questionnaire responses. The interviews were a source of further, but non-triangulated, data.

They reported numerically on a set of 30 examination-related teaching variables that, when taken with what researchers learned during interviews, led to their conclusions regarding the impact of examinations on teachers and teaching practice. They found that the Provincial Examination Program exerted a major effect on the teaching practices in grade 12 classrooms – in effect it had become the focus of instructional content through direct teaching to the test – and that the curriculum-in-use by teachers consisted primarily of examination content. There was more time spent on test-related activities, more testing in general, examination coaching, and increased multiple-choice content as a result. They reported teachers' dissatisfaction with this narrowing in the curriculum as a major impact of the examination program. Because they are forced to follow what is essentially an external description of how the course must be covered, teachers said that they had lost control of the course. They speculated that this loss of autonomy could be a source of low morale and to some degree, stress. A teacher
from the study made a clear distinction, along the lines of the alignment model, between the prescribed curriculum and the de facto curriculum created by the test:

Accountability for each teacher. Makes every teacher accountable for what they are teaching. The exam specs are the curriculum. This is good in that every student gets a fair chance and nobody gets on their pet topic and leaves out the others, but it is also bad in that it narrows the curriculum due to the type of test (focuses on trivial pursuit type knowledge and ignores processes and higher level skills). It also separates the curriculum a la curriculum guide from the curriculum a la exam specs. Despite the fact that the Ministry says that you should cover these other areas that are not examinable, that does NOT happen. The exam specs are the curriculum. It restricts what you teach since you must teach to the exam. (Anderson et al., 1990, p. 168)

Anderson et al. had a much broader mandate than mine. They aimed to survey and sample opinions from the spectrum of stakeholders in public high school education, but they also identified key areas for deeper examination. They suggested investigations of the stress on students and teachers, the extent to which there is a differential functioning between examinable and non-examinable courses and the ensuing effects, the extent to which the provincial examination program has affected the content of courses and the testing practices of the grades preceding grade twelve, and the extent to which tutorial assistance is provided to students, both within schools and commercially. A logical follow up to one aspect of Anderson et al. (1990) is to enter classrooms and spend
considerable time learning the local milieu and uncovering subtleties of the forces that mediate what people in that building do, talking with teachers about the influence of the examination in their work, watching their lessons, and learning about how the examination is assimilated into their teaching practices. This was done in the current study.

**Wideen et al. (1997)**

Wideen et al. (1997) employed a narrower research focus than Anderson et al. (1990) in describing the instructional impact of the BC provincial examinations in senior science classrooms. Their study entailed qualitative perspectives, employing such techniques as constant comparison (Glasser & Strauss, 1967; Strauss & Corbin, 1990) and triangulation (Miles & Huberman, 1984). In the first year of their study, the authors conducted case studies in two school districts: one selected to represent an urban district and the other a semi-rural one. They observed and audiotaped five science lessons by each of 24 teachers (12 in each district: 4 in each of grades 8, 10, and 12) in six schools in one district and five in the other. A pilot study led to the revision of their instruments. In the follow-up, 10 school districts were chosen at random from the 75 in the province, two schools from within each district, and three teachers from within each school. The teacher interviews included a range of questions about background, factors influencing teaching, and the basis for the curriculum. Data were obtained from 56 of these 60 teachers involved regarding the lessons they taught and their thoughts on the impact of the provincial examination.
A grade 10 teacher from the study provides a somewhat troubling description of teaching to the test:

I think those who set Grade 12 examinations are trying to upgrade teaching by providing a really severe exam. And, as a result, you have some teachers who maybe aren’t that good as teachers but spend the entire year with a bank of exams, going over and over them and the kids maybe haven’t learned anything. But, for a certain body of questions [those teachers] are great. But in terms of high-level learning and their attitude toward learning and all the other things that education is about, they really aren’t good teachers at all. (Wideen et al., 1997, pp. 438-439).

The authors concluded that the examinations have had a major negative impact on high school science teaching. Lecturing had come to dominate senior science lessons, and in some classrooms laboratory work had been eliminated altogether. They reported that a few teachers had positive comments about the examination, but larger numbers of teachers expressed serious concerns with the testing program.

The Wideen et al. study brought additional depth to qualitative research in this area, and provides some insights concerning the interactions between the examinations and what teachers were observed to be doing in their classrooms. One unique aspect of their study is that it assessed the instructional impact on teaching by comparing what was seen in classrooms with an external statement of standards regarding what ‘desirable practice’ in science classrooms should be like. The standards for best practice in science
they used were said to be widely agreeable to those in the science education community. They thereby called attention to the dilemma of how teachers implement their personal beliefs about education under high stakes tests. Under this definition of good practice, it would be possible to teach a 'good' course, but have all of one's students perform much lower than they otherwise could on a provincial examination.

Wideen et al. (1997) interviewed two teachers who seemed to accept this 'realpolitik' and yet believe that they could accomplish their educational goals. The authors used their interviews with those two teachers to exemplify that position:

Although neither of these teachers reported that the examinations had affected their teaching, both told us that Grade 12 examinations have had a negative impact both on science teaching in general and on student attitudes. Each teacher viewed himself as a "loner," as somehow outside the system. Both claimed to produce students who achieve excellent results. (p. 439).

No evidence was provided to suggest what the above teachers saw to be best practice, or that they produced good class results. The examples only suggest that there are teachers who believe that they have successfully integrated the examination into their practice. One aspect of this study touches on how they might do that.
Cimbricz (2002) reviews the research attending the relationships between externally mandated state testing and teachers' views and practices. She selected studies that met minimum requirements for research as defined by Howe & Eisenhardt (1990). Cimbricz concluded that when the literature was restricted to empirical studies, evidence for positive impacts vanished. She divided the remaining studies into two groups: the first comprising studies that find strong relationships (positive or negative) between external testing and teaching, the second comprising studies that claim that the relationships are neutral or that they have been overstated. Very few of the studies she cites were observational. Cimbricz found:

The studies reviewed suggest that while state testing does matter and influences what teachers say and do, so do other things, such as teachers' knowledge of subject matter, their approaches to teaching, their views of learning, and the amalgam of experience and status they possess in the school organization. As a result, the influence of state-mandated testing has or does not have on teachers and teaching would seem to depend on how teachers interpret state testing and use it to guide their actions. (Conclusion section)
Studies showing Limited Alignment Between Testing and Teachers’ Views and Practices

Where researchers have found limited alignment relationships between testing and teaching, they have pointed to other factors that (more strongly) mediate teachers’ pedagogy. A theme in several of these studies is that high stakes drives teaching content more than it does pedagogical styles. Several of these authors claim that teachers’ own constructions of the content area, their educational milieus, and their students’ needs, collectively have the greatest influence on classroom instruction.

Firestone, Mayrowetz, and Fairman (1998)

Firestone, Mayrowetz, and Fairman (1998) aimed to compare and contrast pedagogy in Maine (with low stakes) and Maryland (with moderate stakes). Economically disparate school districts in each state (three from Maryland and two from Maine) generated the data set. Forty-one teachers from each state were interviewed, and each site visit comprised observations of three grade 8 teachers, each for two classes, with interviews following each class. This study is one of very few with classroom observations. The authors analyzed teaching practice by problem size (large and small), student activity (practice or non-practice), and teacher activity (tell procedure or develop procedure), and found no significant differences between the states. However, teachers reported, and researchers observed, considerable evidence that the content and focus of lessons in Maryland had been affected by the external test. The authors contrasted this
data with results from Japanese classrooms in the Third International Mathematics and Science Study TIMSS video studies. The authors identified other factors, most notably teachers' personal understanding of and experiences in mathematics as mediating responses to external testing.

This study has been cited for supporting the position that there is not a strong relationship between high stakes testing and teachers' practices (Cimbricz, 2002). More precisely, the authors found that the teaching of grade 8 mathematics in Maine (which did not have high stakes testing at the time of study) was largely the same as the teaching of grade 8 mathematics in Maryland, which had moderate stakes. Their analysis separates what was taught from how it was taught. This paper is one of a number that claim that high stakes affect content rather than pedagogy. That is, a high stakes test can generate a large amount of activity surrounding it but, according to the authors, that activity is not instructional reform.

Firestone, et al. (1998) raises several issues. First, the testing stakes in Maryland are moderate. The authors studied teaching in a context where students' promotion or graduation were not affected by the external test. Maryland does indeed have exit tests, but they were not referred to in the study. The greatest testing consequences concerning grade 8 testing in Maryland involved school-level operation and control, and did not target specific teachers. The authors concluded that these stakes were "only a moderate threat" (p. 108). As one principal put it, "I don't think anybody in this county has to worry about [reconstitution of the school]" (p. 108).

A close examination of the stakes is required in order to theorize about possible relationships involving them. Firestone et al. (1998) may have seen similar teaching in
the two states, but one might question their conclusions regarding the degree to which the stakes were (or were not) involved in that. In fact, the authors seem to conflate high stakes with low/moderate stakes in their conclusions: “our observations suggest that high stakes assessments may encourage some people to think about how to change practice” (p. 111). And although Cimbricz (2002) categorizes Firestone et al. (1998) as an observational study, two classroom visits is inadequate for triangulation of a teacher’s expressed views and reported practice. More extensive observations along the lines of my study are called for.

Grant (2000, 2001)

Grant (2000) used high stakes testing in New York State as an arena for exploring how teachers learn from testing and use it as a basis for changing their pedagogy. New York State testing certainly involves high stakes for students and teachers. Testing in grades five and eight now generates scores for individual students, with multiple-choice questions comprising 55 per cent of test marks. For many years, students had the option of earning a Regents Diploma by writing challenging academic examinations, or of opting for the Regents Competency Examination (RCT), which could lead to a local diploma. Ninth graders, beginning in 2001, no longer have that option; all students need to write the academic Regents Exam in five subject areas in order to graduate.

Grant sought the views of focus groups comprising teachers of varying subject areas and grade levels. There were two focus groups in first year (seven elementary teachers and 12 secondary teachers respectively), and two in the second year (the first
comprising eight teachers representing mathematics, English, and social studies, and the second comprising five elementary teachers). There were no classroom observations in the 2000 study. Moreover, the primary focus of the study was teachers’ reactions to an impending test, not an established one. Grant described his 2000 study as “largely exploratory.” In the 2001 study, Grant conducted classroom observations of two teachers who worked in the same building and taught the same course leading to the same Regents examination.

He was particularly interested in these two teachers because their teaching styles were reported to be sharply different. He interviewed and observed each teacher, one for two lessons, the other for eight. He confirmed that their approaches to the civil rights unit being taught were radically different, and he speculated that these teachers’ individual constructions of the subject matter would play a large role in how the tests would play in the classroom.

Also, like many commentators, Grant suggests that testing impacts are a “mixed bag” (2000, p. 8), but “the overwhelming sentiment, however, was that the new tests could produce undesirable effects.” (2000, p. 17). One pedagogical concern expressed by teachers was that, rather than generate a higher standard of teaching, the new tests would instead compress and narrow teaching toward anticipated test content. Grant unpacks one aspect of the testing stakes by connecting teachers’ instructional choices to their perceptions of administrators’ desires for better test scores and the types of instruction that would accomplish that.

He mentions, but does not discuss, the implications raised by the fact that the Regents Examinations (pre-2001) are marked by the classroom teachers who administer
them, with limited state oversight. Also, the Regents Examinations have limited weight in students’ final marks, and passing them was not required for university entrance. This is highly significant because it profoundly affects the stakes. It means that teachers’ judgments are still in play and they still hold power over the results. This is important because one research goal in this area is to posit relationships about the level and nature of stakes to instructional impacts. Knowledge claims concerning that impact are suspect without a clear understanding of what stakes are involved.

The research literature frequently uses the term “externally-mandated” when referring to high stakes tests. Perhaps the qualifier “externally marked” should be used as well. External marking adds uncertainty and completely removes classroom teachers’ judgments from the testing process. In BC, for example, grade 12 teachers whose students are writing a provincial examination are not even allowed to be in the same room as their students, and teachers are forbidden to see the examination, or discuss it if they do. Many would dispute Grant’s 2000 claim that there is thin evidence that “tests drive instruction” (On Tests and Teaching section). I read the literature as saying that tests do drive instruction, but that we have a poor understanding of which aspects of it, why, and how.

Glasnapp, Poggio, and Miller (1991)

Glasnapp, Poggio, and Miller (1991) surveyed stakeholders, including 1358 teachers of grades two through eleven, regarding Kansas’ low-stakes minimum competency testing (MCT) program in 1982, 1983, and 1987. The authors reported that one effect –
increased emphasis on stated curricular goals – was identified by more than half the teachers in 1987. Teachers at grade levels with MCT tests had higher response rates than teachers who taught at grade levels at which no state tests were used. Few teachers reported a narrowing of the curriculum, or altered instructional methods, but did report a 40 per cent increase in activities such as test coaching, drill and practice, and test review. They did not perceive significant changes in their overall classroom practices.

Two important features distinguish Glasnapp et al. (1991) from my study. First, MCT can be used as a high stakes test, but derives from quite a different tradition of testing than provincial examinations in BC. Aligning tests to objectives and the testing of higher-level skills probably alters the nature of the pedagogical impacts, generating a focus on specific content. Second, the stakes in the study reviewed by them are much lower than those in BC, and hence one would assume that they are less likely to affect instruction, a conclusion reached by the authors as well.

Zancanella (1992) conducted in-depth, multi-year, observationally grounded case studies of the views and practices of three Missouri middle school – junior high teachers of literature. The original study was not designed to look at testing, but in the first interview data in the study, teachers spoke often of testing and therefore Zancanella modified his design to incorporate both those views and testing practices.

Each teacher was interviewed eight times, four times with respect to testing. Each teacher was observed for eight full lessons. Observers compiled written descriptions and
audiotapes. Zancanella also created student focus group interviews. Overall, he found that teachers' belief systems, rather than the external tests, were the primary agent of instructional change. The teachers in the study felt there was a strong contradiction between their pre-existing beliefs and practices and those they felt were implied or encouraged by the external tests.

Zancanella raises several important issues, first, regarding the stakes, which were officially low in importance. Teachers viewed the stakes as being relatively high because of the publication of school-level results. One of the teachers makes a revealing statement: “Even if they couldn’t pass the state test in the spring, if I can keep them reading and discovering some things that they are interested in, I would feel like I had succeeded” (p. 287). It seems that the testing stakes allow for this view and practice, because the teacher is not driven by concerns about the immediate impact of the test results on her students. In situations where there are high stakes for students, it seems less likely that a teacher could so easily resolve this situation in favour of preferred teaching approaches. The teachers in the study felt there was a strong contradiction between their own beliefs and practices and those they felt were implied or encouraged by the external tests, along the lines of Wideen et al. (1997).

Like most other commentators in this area, Zancanella (1992) addresses the overall research needs in this area “If, as many believe, the role testing plays in education will only increase in intensity over the years to come, then understanding the specific dynamics of interactions among teachers, learners, content, and tests has become a necessity” (p. 294). This is the intent of my study.
Questions and Issues Emerging from the Research Literature

The Perceived Implications of High Stakes Tests for Students and Teachers

One pattern emerging from the studies reviewed in this chapter is that, in cases where authors claim to have detected limited impacts of testing, the stakes have been relatively low, even when the authors have stated otherwise. This adds weight to the conclusion that, overall, testing most often does cause changes in teaching behaviour. The strongest instructional effects are seen where the stakes are perceived to be high, as in Texas (Haney, 2000). This implies that relating testing to instructional behaviour requires a good understanding of the stakes, which entails detailed observations and descriptions of the legislated and lived stakes for stakeholders. For example, Corbett and Wilson (1991) stated that Maryland had high stakes MCT “exit” examinations. Their purpose and qualities, however, are starkly different from New York Regents Examinations or BC Provincial Examinations. More nuanced differences are found in Grant (2000, 2001), who mentions in passing that the high stakes Regents Tests are marked by the same teachers who administer them, and does not mention that at that time, the Regents exam had little overall weight in students’ final marks and were not required for entrance to university (Center for Public Education, 200). Columbia University, for example, makes no reference to Regents Exams on its application Web pages. In BC, the FSA has no stakes for students, but it is externally marked and the results are the source of public media rankings and are connected to decisions concerning accountability. Firestone et al. (1998) categorized Maryland as having “moderate” stakes,
yet neither teachers nor students were reported as feeling threatened by them. This motivates a close examination of the reputed effects of the legislated and other stakes known to attend BC Provincial Examinations.

In BC, the stakes for students and their parents include a 40 per cent weighting in overall course marks, graduation, scholarships, and de facto university entrance criteria. No American studies that I have seen report such stakes. In a 2003 survey of US college admissions officers, scores on state tests were considered important by only 18 per cent of respondents, the SAT by 25 per cent of them, and final high school marks by 87 per cent of them (Center for Public Education, 2005). Because BC provincial examinations are given heavy weight in overall marks, and because of their reported close correlation with classroom marks, they effectively serve as university entrance examinations, not just to universities and colleges in BC, but to wherever else they may wish to attend college or university.

The stakes for teachers can include directives from principals, school rankings in newspapers, assessments of classroom-level performances generated by the Ministry of Education, local embarrassment or shame over low class marks, and perceptions of teacher and school quality on the part of students, parents, and principals. The stakes for principals include public prestige or embarrassment over rankings, competition with other schools, and other overall perceptions of school quality based on rankings.

In some North American jurisdictions, district-level and school-level accountability contracts are now tied to test results. As Rouk (2000) points out, "Reporting school and district test scores to the public has become the major tool by which states demonstrate accountability." (p. 1). McDonnell and Choisser reported that
"probably the most potent leverage the assessment system has over the behavior of teachers is the widespread perception that local newspapers plan to report test scores, not just by individual schools, which has been done traditionally, but also by specific grade levels and even by classroom" (1997, p. 16).

The BC Ministry of Education produces school-level results, but they generate far less publicity than the privately funded research that generates the widely discussed Fraser Institute Report Card for BC Secondary Schools, which includes a ranking of schools.

It seems the term high stakes has little meaning unless it is used in a context where the stakes are understood, for all the stakeholders and the multiple forces acting on them. It would also appear that testing stakes are not equally weighted for teachers. Where testing stakes are very high for students, this causes additional stakes for their teachers, who usually do what they can to help their students achieve on the test, and researchers have found clear and significant instructional impacts (Anderson et al., 1990; Calder, 1990; US Department of Education, 1993; Davis, 1996; Wideen et al., 1997).

**Judgments on Teaching Practices**

The research literature classifies instructional impacts as either positive or negative, along lines of conformity with some perceptions of educational soundness, or in terms of the achievement gains that are said to stem from certain teaching methods. These judgments are most often rooted in the perspectives and goals of the researchers, and in an implied value system in which some types of teaching behaviour are seen as
more meritorious than others. Judgments of educational merit can easily be circular, because they can shape what data is collected, which can in turn be used to define ‘educational merit’ and how it is interpreted. It might be possible to use my observations in that way, but it seems better to conduct this study, so far as possible, without the burden of unnecessary and possibly misleading \textit{a priori} conceptions of desired teaching practices. I was not attempting to identify poor or excellent teaching, but rather how the external tests affected the teaching practices that I observed. A more cogent concern is the impact of my preconceptions directly related to the research questions. This is discussed in Chapter 3 in the section pertaining to validity in case studies. My goal is to observe, reflect on, and report concerning a system in action, and to make knowledge claims arising from my research questions.

Even setting aside for now questions of validity, we can bring questions of the alignment interactions between those vertices in the model into the picture along lines of what the teacher is doing when that reflection is observed. That suggests examining, or at least being aware of, the following seven alignment relationships: (a) the impact of testing on teaching styles, (b) aligning instruction to past examination content, (c) school-level and classroom-level increases in instructional time, (d) end-of-course examination review, (e) aligning classroom assessments to past examination content, (f) gatekeeping, and (g) downward alignment.
The Impact of Testing on Teaching Styles

The research is not at all clear about the effects of high stakes testing on the specific modes that teachers use to introduce, develop, and explain mathematical concepts. Some of the above research refers to Minimum Competency Testing (MCT) and its connections to low-level drill-and-practice teaching methods. Such research may be useful in exposing the sequelae associated with such low-level testing, but it is questionable to what degree it can enlighten discussions of performance-based, curricularly aligned assessments. For example, MCT shares few features with BC provincial examinations.

Most of the research reviewed above did not address the teaching of mathematics; further complicating this is some evidence that teachers' pedagogical responses to high stakes testing are mediated by subject area (Williams, Kirst, & Haertel, 1999; Zancanella, 1992), the teachers' personal philosophy of the discipline, and the immediate environment (Siskin, 1994).

To this complexity we must add the discussion surrounding the use of the word *lecturing*. It has several shades of meaning. I and others use the word to refer to the introduction of new mathematics topics by using a teaching mode heavily dominated by teacher talk (and including students often writing notes). Some question (Anderson et al., 1990; Calder, 1990) and some decry (Wideen et al., 1997) the heavy use of lecturing in grade 12 examinable courses. A difficulty in attributing pedagogical effects to high stakes testing is that, even in jurisdictions with a very high stakes tests, such as BC, researchers
have noted significant differences between teaching styles across the grades, but it is problematic to generalize whether or not they can be attributed to high stakes testing. Stake (1995) recommends the examination of alternate theories and explanations as a way of possibly increasing the internal validity of conclusions. The posing and testing of alternate hypotheses is often informal, but with regard to lecturing, an alternate explanation is that it is widespread practice in grade 12, with or without a high stakes test (a conclusion reached in 1998 by Firestone et al., albeit at an earlier grade level and a lower stakes test). To pursue this question further, one would need an observational case study in two high schools, one with high stakes, and one without, comparing grade 12 and grade 8 mathematics instruction within and across schools and subject areas, looking for differential effects. I have not found such a study in the literature. Corbett and Wilson’s (1991) design allowed for this, but they did not pursue that line of inquiry. Wideen et al. (1997) did observe clear differences in teaching styles in grade 12 and grade 8 classrooms, but they did not have a low stakes environment for comparison purposes.

The literature makes a strong case for examining the above types of complexities, and argues that teachers’ pedagogy is more strongly rooted in the subject area, teachers’ own constructions and beliefs about the subject area, the local environment, and official curricula. In this study, I note teaching styles and their prevalence, but not sufficiently to adduce examination-related impacts on varying approaches and lesson styles.
Aligning Instruction to Past Examination Content

This category addresses the use of past test items as components of lessons, referring to the test or cheerleading for it, limiting instruction to the scope and format of examined content, coaching with specific past items, teaching the recognition of problem types, using prepared solutions, and distributing classroom assignments comprising past content.

School-level and Classroom-level Increases in Instructional Time

This category concerns impacts that are generally viewed as being positive: working harder, longer, and more efficiently (Koretz, 2005). The research literature, however, suggests that those references are often anecdotal, and that more and better evidence is needed, as well as evidence as to what influences such practices can have on non-examined courses (Anderson et al., 1990). For example, in the early 1990s the principal of the school I worked in tinkered with the senior timetable so that teachers and students in examinable courses had 130 contact hours – a 30 per cent increase. Following that adjustment, that principal was recognized by the Fraser Institute and awarded $5000 for managing a school that had achieved examination scores above and beyond what would normally be expected for a school in that socioeconomic setting. The Fraser Institute was unaware of, or unconcerned, about the likely connection between instructional time and student performance. This suggests a close look at school-level timetabling, scheduling, and any adjustments that are made as testing time approaches.
As well, it suggests examining classroom-level adjustments such as out of class tutorials, examination cramming sessions, camps, and so on, and possible obvious linkages between the examination results and those adjustments.

**End-of-Course Examination Review**

The literature contains numerous references to teachers sacrificing instructional time in order to prepare students for an external test. For example, a number of teachers in Madaus et al. (1992) spent more than a month preparing their students. Smith (1991) found that 32 per cent of the teachers she studied said they were “required” to engage in test preparation, and 28 per cent of these Arizona teachers said they begin their test preparation at least two months before the test was to be administered. Jones et al. (1999) reported that 80 per cent of the teachers in their study said that they devoted more than 20 per cent of their total instructional time to practicing for grade-end tests.

My discussions with B.C. teachers and Anderson et al.’s report (1990) suggest that, in BC, three weeks to one month is the norm for reviewing grade 12 courses and coaching for the examinations. This is a clear examination impact, and is also examined to a degree in this study.
Aligning Classroom Assessments to Past Examination Content

The research literature contains several references to the use of past test items in classroom assessments (e.g., Center for Public Education, 2005; Madaus, 1988a). To varying degrees, this leads to classroom quizzes and tests becoming proxies for subsections of the external test, an effect described in Madaus (1988a). That type of alignment could easily be classified as coaching.

Gatekeeping

Gatekeeping is a form of systemic alignment. Gates can be used to control entrance to a school, a classroom, or being permitted to write an examination. Gatekeeping to the school is often heard about in references to private schools that have entrance examinations, such as St. George’s School and Little Flower Academy in Vancouver. There are hundreds of references to gatekeeping in the broader literature, but very often those references are to some other place, person, or institution. There are better-documented examples within the research literature (Amrein & Berliner, 2002; Haladyna, Haas, & Allison, 1998). I explain in chapter 3 why it may be difficult to get a complete and accurate picture of this practice.
**Downward Alignment**

The phrase *downward alignment* is used here to refer to the appearance of examined content in earlier grades, in ways clearly not implied by curriculum documents. For example, one practice presented by a colleague of mine at a Fall Conference of the British Columbia Association of Mathematics Teachers (BCAMT) under the title, “Success in Principles of Mathematics 12,” discussed how he would identify which students in grade 11 would be continuing to Principles of Mathematics 12. Around the middle of the grade 11 course, students would receive differential assignments depending on their destination.

In my own practice, I have eyed certain grade 11 topics for elimination so as to increase grade 12 examination performance. More subtly, in Physics 11, I adopted the language and notation used in Provincial Examinations so that familiarity with them might enhance understandings in the examinable Physics 12 course the following year.

**Mixed and Unclear Results from the Research**

Zancanella (1992), Cimbricz (2002) and others reach moderately different conclusions about the strength of testing impacts, but they converge solidly with other researchers on one aspect of the research literature. They find that testing can cause (and often causes) instructional impacts, but thus far there is little empirical basis for drawing applicable conclusions. Zancanella (1992) writes, “This shortage of empirical investigations means that the hopes of policy-makers and the public that more tests will
somehow lead to better teaching or more learning rest on largely unvalidated assumptions". (p. 283). Cimbricz (2002) writes, "The influence state testing may or may not have on teachers and teaching expands beyond individual perceptions and actions to include the network of constructed meanings and significance extant within particular educational contexts" (Abstract section), and, "Studies that provide a richer, more in-depth understanding of the relationship between state-mandated testing and teaching in actual school settings, therefore, not only point toward important directions for future research in this area, but are greatly needed" (Conclusion section). Such studies that focus on high school-level mathematics are even more difficult to find in the literature. This frames the need for my research.

**Implications for the Design of the Study**

The design of any study is influenced by the type of questions one asks and the nature of the data one must subsequently collect. My research questions imply, and my review of the research literature suggests, that I need an empirical, qualitative approach for studying questions about a complex social phenomenon rather than a survey study with questions that lend themselves to specific hypotheses and experiments. The above research literature suggests paying close attention to the breadth, level, and qualities of the testing stakes. It also recommends collecting this data in classrooms, principals’ offices, and staff rooms. Therefore, I required an approach robust enough to hear from and analyze a multitude of sources, and that does so in direct observational contact with the phenomena in question. This chapter identified within-school and external influences
that affect testing stakes and instructional impacts. The instructional impacts, listed above, are used to seed the data collection and analytic processes. Chapter 3 explains in greater detail how the case study approach meets these needs.

Chapter Summary

The Curricular Alignment model addresses the mutually influential relationships between assessments, curricular objectives, and classroom practices. The alignment between testing and teaching then becomes a lens for viewing the instructional impacts of external testing. Those impacts have been found many times, and they have seemed to correlate with the level of the testing stakes. Those stakes are now high in BC. But the nature of these relationships require closer scrutiny and elaboration: much of the previous research has not had a significant observational component; there is considerable diversity in the precision with which key words and concepts have been used; and often there has been insufficient attention given to the complexity of the environments in which teaching happens. Indirectly, then, the research literature has guided me towards a methodology capable of grappling with this complex phenomenon obtaining observational data within a local culture. It also provoked me to do some model clarification in adopting the ‘three vector’ alignment model for inquiry and analysis.
CHAPTER 3 – DESIGN AND METHODS

Introduction

Chapter 2 examined the literature that surrounds the research questions and suggests methodology. This chapter lays out the methodological issues that surround my approach to answering the research questions. The first section establishes the need for a case study approach. The second section discusses theoretical aspects of case studies, including their definition, the unit of analysis, validity, and generalizability (generality). The third section discusses the conduct of the study, including the selection of sites, participants, and the sources and methods of data collection.

The Case Study Approach

The case study approach is appropriate for attempting to answer my research questions. Here we review the nature of case studies, some implications for their use, and issues of validity and generalizability.

The research literature reveals that studies of the relationships between high stakes assessments and teachers’ views and practices can benefit from two qualities: increased validity resulting from triangulating classroom observations, and an in-depth examination of the educational contexts and influences that surround instruction (Cimbricz, 2002; Zancanella, 1992). My data therefore requires richness of detail (Geertz
1983; Yin, 1984), multiple sources (Yin 1994; Stake, 1995), and confirmation (Miles & Huberman, 1984; Yin, 1984), and the picture I paint should involve tentative explanations, and continual reformulations and insights (Eisenhardt, 1989; Stake 1995). I must also bring to my research knowledge of the milieus within which teaching occurs in these schools (Feagin, Orum, Sjoberg, 1991). A case study is an appropriate and well-used approach for meeting these requirements and is therefore likely to provide some answers to my research questions.

There is no consensus as to precisely what a ‘case study’ is. Stake (1995) exhibits evidence of the diversity of views of what case studies are when he claims that to follow a case-study approach is not a choice of methodology but rather the selection of an object of study, an assertion that is at least incomplete if not problematic. Apparently he means that case studies employ specific methods that suit them to particular questions and not others. Saying that, however, does nothing to identify what those qualities of case studies are. The ‘case’ is an instance of the phenomenon of interest. In my dissertation, the mathematics teachers are the unit of analysis – that is, the object of study. They comprise two cases of the views and practices of mathematics teachers.

Other commentators offer some detail as to what they think case studies are. Hartley (1994) argues that case studies are tailor-made for investigating new processes or practices or ones that are little understood. Yin’s (1994) widely quoted definition has the most to offer: “[A case study] investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” (p. 13). Feagin, Orum, and Sjoberg, (1991) wrote that case studies aim to
provide a holistic understanding of cultural systems in action. "Richness of detail" (Geertz, 1993) is typical in case studies, and therefore the generation of and reliance upon ample field notes is standard. The data should be rich enough to allow views of many facets of the object of study, and to allow the experience of "being there" (Lincoln & Guba, 1985).

There are various rubrics for organizing case studies into types, most often based on their ostensible purposes. Yin (2003) uses a 2 x 3 grid to identify what he holds to be six types of case studies: single and multicase versions respectively of exploratory, explanatory, and descriptive approaches. Merriam (1998) groups case studies into "descriptive," "interpretive," and "evaluative" (p. 38-40). Stake (1995) categorizes case studies as intrinsic, instrumental, or collective. Such labels only help a researcher insofar as they help clarify the purpose of the study, its theoretical structure (if any), and the methods that are being used to attempt to answer the study's research questions.

A case study, in terms of the present work, aims to observe, record, explore, and understand the influences of high stakes testing on instructional practices that are generally agreed to exist, but whose workings at the classroom and school level are not well explored.

**Theoretical Considerations in Case Studies**

**Bringing Previous Research and Theory to the Design and the Data**

One fundamental, and to some degree distinguishing, feature of case studies is
that they are open to the researcher bringing existing models, theories, and research results to bear on the study. Yin (1984) writes, “Reliance on theoretical concepts to guide the design and data collection for case studies remains one of the most important strategies for completing successful case studies” (p. 3). Yet, Eisenhardt cautions, “Although early identification of the research question and possible constructs is helpful, it is equally important to recognize that both are tentative in this type of research. No construct is guaranteed a place in the final theory” (1989, p. 536).

The alignment model, outlined in Chapter 2, suggests mutual influence among what are viewed to be three important variables, and therefore initially influences decisions regarding what is and what is not relevant data. The weight of related previous research establishes widely observed instructional main effects, or alignments, which provide launch points for semi-structured interviews and observations, and serve as landmarks for analysis (Yin, 2003). Researchers themselves can bring their own experiences and perspectives to the research, but the key, according to Gummesson (1988), is for researchers to “balance on a razor’s edge using their pre-understanding without being its slave” (p. 58).

Unit of Analysis

The “unit of analysis” is the object of study (Luck, Jackson, & Usher, 2006) and an important component in a case of the phenomenon of interest. Numbers of commentators stress the importance of careful deliberation concerning the selection of the unit of study (Eisenhardt, 1989; Stake, 1995; Tellis, 1997). Stake (1988) reminds us
that a case is a ‘bounded system’ – a single actor, a single classroom, a single institution, or a single enterprise, usually observed under natural conditions. The more natural the case boundaries, the greater is its prima facie validity.

Mathematics Teachers within a School as the Unit of Analysis

The alignment model assumes that there are relationships between testing and teaching, and previous research provides further guidance in choosing an object of study that will maximize what it can show us about these dynamics. Some researchers have claimed that teachers’ departments can be the professional communities with the greatest impact on norms of practice and attitudes toward teaching and students (Siskin, 1990), and a distinct source of pedagogy (Siskin, 1994). It is not always clear whether the collaboration is professional or communal, authentic or affected, pedagogically sound or otherwise. There is also evidence that qualities of the department such as standards of collegiality and clarity of vision can significantly mediate its impact on individual teachers within the department (Ball & Bowe, 1992). The primary research question is concerned with mathematics teachers. The mathematics department has sharply defined boundaries, and because of its clear and strong connection to the research questions, it is a logical choice as the unit of analysis.

Grade 12 Mathematics Teachers as the Sub-unit of Analysis

But within the mathematics department there are boundaries. I reviewed the
research that shows how provincial examinations carry high stakes for teachers. They experience greater expectations, pressure, visibility, and scrutiny than teachers of non-examinable courses (Anderson et al., 1990; Wideen et al., 1997). The threats and imperatives stemming from the examinations tend to lead teachers to rely upon each other's knowledge and resources, particularly within a building. This sort of teamwork is more natural than contrived, and it is logical to view grade 12 teachers as a distinct group within a group, with a unique mandate, because high stakes testing has been linked to increased camaraderie among teachers of the examined course. For example, Grant (2000) writes, "The power of such informal relationships is apparent: These teachers sense that they are working with peers who hold similar goals and concerns, who are willing to share ideas and practices, and who offer a sense of belonging" (p. 12). This makes the above quoted research from Ball and Bowe (1992) and Siskin (1994) equally relevant to the grade 12 mathematics teachers as a subgroup.

The research question asks about mathematics teachers' views and practices regarding high stakes testing, and since it is the grade 12 teachers who teach the examinable course, it is natural that I sometimes focus on them and that they have a strong voice in answering the research questions, to the degree that it is sometimes useful to analyze them separately, and therefore they can be considered a "sub-unit of analysis" (Yin, 1994, p. 41). In one of the schools, all but a few mathematics courses are taught by three teachers, each of whom teaches Principles of Mathematics 12, and therefore there is no need for separate treatments.

An analysis of my own teaching career supports the above conclusion. Much of my career has involved teaching students where I was the sole teacher in some
provincially examinable courses. The examinations caused me concern, and since my teacher training made no mention of high stakes tests, I had to look outside my high school for mathematics teachers with whom I could share, discuss, evaluate, and tune classroom practice in all grades in light of the examinations. If I had had access to colleagues in the same building who taught the same courses, it is natural to expect that they would have been my principal source in learning how to respond professionally and pedagogically to a provincial examination.

In summary, the teachers of mathematics within a school are the most appropriate unit of analysis – the object of study – for answering the research questions in this case. The grade 12 teachers teach the high stakes course, and form a logical subgroup for analysis, the sub-unit of analysis in this study.

Validity of Case Studies

Triangulation in Case Studies

Triangulation is the foundation for establishing data points and for increasing internal validity in qualitative research (McMillan & Schumacher, 2001; Miles & Huberman, 1984; Stake, 1995). Published prescriptions for conducting not only case studies but all qualitative research stress the need for triangulation – viewing a phenomenon from as many different perspectives as possible so as to determine their concurrences, antagonies, and shades of interpretation. The idea of triangulation has been an important part of qualitative research in the educational community for some time.
The first explicit use of the term seems to be by Denzin (1978) and Sevigny (1978). Significant and frequently cited further commentary in this area includes Miles and Huberman (1984), Stake (1995), Tellis (1997), and Yin (1994), who stressed the basic requirement for the use of multiple sources of data.

I have been diligent in endeavouring to collect multiple sources of data that converge around particular phenomena and issues. I continually sought to know and compare the views and practices of teachers and students, individually and collectively, colleagues, administrators, the media, and, through the British Columbia Association of Mathematics Teachers (BCAMT) Web site and listserv, the views of some other teachers of mathematics in BC.

Further, as addenda to this study, I participated in extensive discussions of provincial mathematics examinations in boardrooms, principals’ offices, staffrooms, in hallways, and in classrooms. I soon concluded that the tone and content of such discussions are very much affected by who is and who is not present, and quite different pictures of administrators’ and teachers’ views and practices are sometimes painted when the same topics are discussed in public versus in private. This disparity therefore suggested a need for intensive triangulation of what researchers read, hear, and see as they study the relationships between high-stakes testing and its impact on schools. This and other threats to validity are discussed below.

**Threats to Validity in my Research**

There are two primary threats to the validity of my research that are discussed in
two sections: the viewpoint of the researcher and sensitive issues.

The viewpoint of the researcher. I have taught under BC provincial examinations in either Mathematics 12 or Physics 12 for 15 years, and to a lesser extent Biology 12 and Chemistry 12. Examination results were high stakes for students, parents, teachers, administrators, and from what I could infer, senior school board management as well. There was little doubt that, at least in the schools where I worked, provincial examinations drove many aspects of school operations. There was little doubt, at least within my own classroom practice, how past provincial examination content became an early and important guide for addressing basic content goals and for preparing students for writing provincial examinations. Dry runs and a full month of review became standard practice for me, as it was for most colleagues in most examinable subjects. Past examination content became the arbiter for what could be considered to be a fair level of difficulty. Good examination results were a source of public congratulation by the principal and some colleagues. My assessment practices, and knowledge of the subject area and the examination itself, was enhanced by participating in provincial examination marking sessions, which I found to be crucibles of content area knowledge and pedagogy in the context of a high stakes test. I was impressed by the measured reliability of the marking (in one Physics session the Pearson r correlation between marked and blind re-remarked papers was about 0.9) and the high level of scrutiny the markers gave to seeing that examination content agreed in a meaningful way with prescribed curricular goals and authorities in the discipline. There were occasions in which markers as a whole contested specific items on the grounds of the correctness of content or appropriateness of item
presentation. It was empowering to see, in those cases, how such items came to be removed from the examinations. I viewed the provincial examinations as the single most powerful force influencing what I did in grade 12 classrooms.

I have been engaged in the same instructional dilemmas and contexts as participant teachers in this study and therefore my previous experience could potentially be a strong force – positive or otherwise – in shaping the research. From years of teaching and interacting professionally with other colleagues and from my own experience as an administrator, I had come to what I thought were reasonable conclusions as to some of the kinds of things I would see and hear when I entered schools and classrooms – the typical buzz of activities that characterize schooling, including lectures, quizzes and tests, seat work, and so on. Moreover, I recognized the views and practices expressed by participants in previous research in BC, who provided details of their beliefs about provincial examinations and their perceived impacts.

I also knew from experience that no two schools or two teachers' practices are the same, and that there can be informative surprises when teachers discuss details of their classroom practices and when principals talk about their schools. My methodology allowed for, and in fact anticipated, that emerging data would refine and alter my assumptions and the conclusions and inferences in previous studies. Differences between my experiences and those of my colleagues were expected, and welcomed for enriching the data. As a teacher of grade 12, my closeness to the research phenomenon provided the benefits of an insider's understanding of the work being observed, but also carried the risk of my a priori relationships filtering out potentially important data and lines of inquiry.
During my research, I was particularly interested in the breadth of teachers’ views. I encountered several teachers whose beliefs about teaching were very different from my own and each other. These differences in viewpoints were welcomed, because they motivated deeper inquiry and offered alternate explanations and relationships (Stake, 1995). Indeed, on several occasions my observations were discordant with either my teaching experience or with previous research in this area. For example, I was momentarily taken aback when I learned that one of the principals allows his grade 12 students to choose their teacher for Principles of Mathematics 12. This is a radical break from what I thought was standard practice. It may in fact be such a break, but regardless of my or others’ views on the practice, it had meaning in the research site, played a role there, revealed aspects of the principal’s thinking, and underscored the importance of being receptive to phenomena and belief systems that are outside the viewpoint of the researcher. In the case of student stress, my findings challenge my perspective and are not in agreement with previous research in this area. Throughout the study I explored the opinions and practices of educators, and each one of them added voice and texture to the result.

Sensitive issues. Gatekeeping to courses and writing examinations are examples of practices that can be attractive to those looking for facile, quick, and effective techniques of raising class and school scores. While it can be argued that a student who has earned 27 per cent from his teacher has no business writing the government final, it becomes more difficult to defend the teacher and principal who deny a student the right to write an examination because her school mark is 45 per cent and they are concerned
that her performance will likely drag down the school’s rankings. Gatekeeping has the potential to deny students their basic right to pursue a course of study, even though they may have met the legislated minimums, or to write an examination whose results could make the difference between graduating and spending another year in school. Researchers have documented both of these situations (Kravolec & Buell, 2005). I have also seen evidence of that practice in schools in which I have worked. But even with guarantees of anonymity, there may be principals or teachers who are reluctant to expose aspects of their or others’ practices if they feel those practices could be viewed askance by authorities, commentators, or even researchers.

In recorded quotations from teachers in Anderson et al. (1990), I counted six references to gatekeeping. In five of those, the teacher referred to an action by another teacher or to school-wide practices. The sixth teacher acknowledged threatening to start “booting” students from his or her class if class averages on examinations were formally tied to teacher evaluations (p. 171). Since discussions of provincial examinations seem to be strongly affected by who is present for them, I found it necessary to be continually sensitive to potential threats to the validity of my study.

This caveat also underscores the need for triangulation, particularly concerning such matters as gatekeeping and other potentially sensitive issues. It also suggests that it was important that I not underestimate the possible benefits of reinforcing my promise of anonymity; establishing clearly that my interest was in learning about the impact of high states tests, not to judge the provincial examination program or anybody’s teaching.
Generalization in Case Studies

As qualitative research methods achieve greater currency and wider use, it becomes apparent that some initial attacks on those methods and the case study approach were clearly rooted in long established assumptions, language, and numerical techniques attending the use of inferential statistics in the social sciences. Some commentators argued that it would be impossible for case studies to generate knowledge claims with the same rigour, replicability, and generalizability as were sought in classical studies. Because qualitative inquiry was emergent and evolving, it did not yet have sufficiently agreed-upon language to distinguish the very different meanings of 'generality' in these two modes of inquiry. From a classical quantitative perspective, a generalizable result implies that what was discovered about a sample is representative of what would hopefully be found in a larger population from which the sample was drawn (Campbell & Stanley, 1963; Sharp, 1998).

Case studies are not intended to produce results that generalize to other cases. Rather, case studies assist in generating explanations of phenomena (Sharp, 1998). In other words, a case study aims to generalize to a theory, as opposed to a population (Stake, 1995). Stake (1995) quite explicitly states, “We do not study a case primarily to understand other cases” (p. 4). The utility of case studies rests not only in their ability to provoke model clarification, but also in their power to connect, to varying degrees, with readers’ lives. In an echo of the classical stance, Stake adds that a case study should be in conceptual harmony with the experience of a broad cross-section of readers and thus be a natural basis for that kind of generalization. He refers to this resonance with the reader’s
experience as "naturalistic generalization," and says that it is an intuitive process in which the reader recognizes issues and objects from his or her experience.

In the same vein, Geertz (1983) says that a case study can have the potential for what he calls *thick description* - descriptions with multiple "hooks" into readers' own experience, enabling them to make personal judgments as to its quality and generality. Cousin and Jenkins (2003) take a stance somewhat different from Geertz’s (but similar to Stake) in asserting that generalizations to other settings are the responsibility of the recipients, not the authors. The recipients are to decide upon the extent to which the thrust of the analysis and recommendations speak to their condition.

My intent, then, is that some readers will recognize the tensions reported here that may be generated by high stakes tests in their own and others’ views and practices and will be able to compare and contrast this surrogate experience (Stake, 1995) with their own. It is expected that this naturalistic generalization is likely to be strongest with grade 12 mathematics teachers and with principals, but also with researchers familiar with this field of inquiry.

**Conduct of the Study**

**Choice of Sites**

I embedded this study in two sites from the lower mainland of British Columbia, one public and one independent. I searched for two schools satisfying these criteria:
1. The administrations and mathematics teachers must be agreeable to participating in a study of this kind. Since there will no doubt be teachers who may be hesitant to offer open access to their practices, it will be stressed that it is not the purpose of the present research to make judgements of teaching practices.

2. The schools should be perceived to be middle-range academic institutions, representing neither extreme of academic concerns. This will increase the likelihood that what is seen in these schools will also be found in other high schools.

3. The mathematics departments, so far as can be determined in advance, must comprise cohesive and communicating professionals. This criterion was intended to optimize the likelihood of detecting the nature and extent of communication and co-operative actions in the departments.

I posted to the British Columbia Association of Mathematics Teachers (BCAMT) listserv, briefly describing my proposed research and asking to hear from interested parties. The mathematics head at Greenhill School replied quickly, and I met with the principal and the grade 12 mathematics teachers, who agreed to participate. The school seemed like a good candidate for research, and without a closer site to choose from at that point, Greenhill was selected out of convenience for study. In search of my second site, I posted again to the BCAMT listserv, asking for participants. There were two interested parties - because a smaller independent school might provide interesting contrasts, and because of the positive reception I experienced at the initial meeting, the independent school, Pine River, was preferred and selected for study.

In Chapter 4, I describe in greater detail the particularities and the contexts of the
research settings as I discuss the two case settings of this study.

Participants

The research question suggests the participants, who in each site are the teachers in the mathematics department, administrators, focus group students, and other teachers who may wish to participate.

Data Collection

Introduction

Luck, Jackson, and Usher (2006) provide a lengthy list of sources of data for a case study, comprising direct participation, observer participation, surveys, questionnaires, documentation, archival records, documents, interviews (both structured and unstructured), written accounts by participants, physical artifacts and researcher description of the context. Previous research in this area showed the efficacy and primacy of semi-structured and focus group interviews of the stakeholders. My analysis also rests heavily on these types of interactions with my participants, and also in how my participants interact with each other. There are semi-structured interviews with individual administrators and teachers, and focus group interviews of mathematics teachers and of grade 12 students. My study ultimately focuses on the work that teachers do in grade 12 and other classrooms. Data from lesson observations; classroom visitations; site documents such as classroom resources, quizzes and tests, homework; and examination
review material were also gathered. Tables 1 and 2 summarize the data collected at Greenhill School and Pine River Secondary, respectively.

Table 1.
Data collected at Greenhill School

<table>
<thead>
<tr>
<th>Interviews</th>
<th>Number</th>
<th>Length (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Individual</td>
<td>6</td>
<td>10-25</td>
</tr>
<tr>
<td>Teacher group</td>
<td>2</td>
<td>10, 15</td>
</tr>
<tr>
<td>Administrator Individual</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Student Focus Group</td>
<td>1</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discussion</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Individual</td>
<td>~10</td>
</tr>
<tr>
<td>Teacher group</td>
<td>2</td>
</tr>
<tr>
<td>Administrator Individual</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson Observations</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotated full lessons</td>
<td>3</td>
</tr>
<tr>
<td>Annotated Shorter Visits</td>
<td>6</td>
</tr>
<tr>
<td>Brief Visits</td>
<td>~12</td>
</tr>
</tbody>
</table>

Documents
- Classroom materials (Texts, item banks, workbooks)
- Ministry of Education Statistics
- Fraser Institute Rankings
- School District Web site
- School Web site
- Classroom test and quiz
Table 2.
Data Collected at Pine River School

<table>
<thead>
<tr>
<th>Interviews</th>
<th>Number</th>
<th>Length (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Individual</td>
<td>7</td>
<td>5-35</td>
</tr>
<tr>
<td>Teacher group</td>
<td>2</td>
<td>25, 20</td>
</tr>
<tr>
<td>Administrator Individual</td>
<td>3</td>
<td>20-45</td>
</tr>
<tr>
<td>Student Focus Group</td>
<td>1</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discussion</th>
<th>Number</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Individual</td>
<td>~30</td>
<td>During lunch and free periods</td>
</tr>
<tr>
<td>Teacher group</td>
<td>~15</td>
<td>From 2 to 6 members</td>
</tr>
<tr>
<td>Administrator Individual</td>
<td>~8</td>
<td>During lunch in staffroom</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lesson Observations</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotated full lessons</td>
<td>6</td>
</tr>
<tr>
<td>Annotated Shorter Visits</td>
<td>10</td>
</tr>
<tr>
<td>Brief Visits</td>
<td>~20</td>
</tr>
</tbody>
</table>

Documents

- Ministry of Education Statistics
- Fraser Institute Rankings
- School Web site
- School Handbook

Interviews

Interviews can embody a spectrum of structures in order to suit the intent of the study. In random telephone surveys, the interview questions are read in sequence, verbatim from a card. In unstructured interviews, both interviewer and interviewee can range spontaneously over the broadest or most minute areas. My design assumes a role for previous theory and research results by using them as launch points, which is reflected in my “semi-structured” interview approach, where structure is flexible and dynamic rather than prescribed. When exploring relatively unknown phenomenon, such
as grade 10 examinations, less structure is desirable.

This section on interviews has three subsections: starting points for semi-structured interviews with teachers, starting points for interviews with principals, and starting points for interviews with students.

Starting points for semi-structured interviews with teachers. In Chapter 2, I listed the commonly cited main instructional effects that are attributed to aligning classroom practice to high stakes testing. These served as guides for my first probes, discussions, and observations. These questions are not prescriptive, but since they stem from the shared results of previous widespread and local research, they seemed to be a logical place to begin. Answers to my initial questions served as links to other instructional effects and categories. Further, Scriven (1981, 1983) argued that an in-depth case study should be as ‘model free’ as possible in the sense that researchers must not come to it with such a firmly embedded set of questions and expectations concerning possible relationships in mind that they see only what they expect to see. In a good case study, it is impossible to know in advance exactly in what directions the initial explorations will carry the researcher and what fine structure will devolve from the first questions asked. Done well, it might not be necessary to look for Cimbricz’s ‘deeper structure’ at all; we would already have the information we came to find. Previous research should act as rough drawings on an exploratory map, rather than as a set of driving directions.

Besides the above-listed main instructional effects, the research literature connects the degree of instructional impacts to the perceived gravity of the testing stakes.
The research literature indicates that teachers’ perceptions of administrative priorities and demands can be influential on teaching and therefore were addressed as well.

My interactions with teachers were the most critical source of data as well as the most diverse, ranging across semi-structured individual and group interviews, observations, and casual conversations in the staff room and elsewhere. The linkages I made between previous research and my interactions with teachers stemmed initially from these interviews. Some of the issues or starting points that I pursued in my interviews with teachers were:

- **Course Planning and Timing** - How do teachers plan Principles of Mathematics 12?
- **Communication from Administration** – In what ways do administrators’ views of examinations manifest themselves in their discussions with other teachers?
- **Downward Alignment** – What effects do teachers perceive the twelfth grade examinations have had in earlier grades?
- **Examination Preparation** – What specific activities and timetable adjustments are intended to be for examination preparation, and when and how do they occur? It is reported that some schools practice dry runs - simulated provincial tests that are aligned as closely as possible in design, content, timing, and ambiance. Do these schools do so? If so, who organizes and administers them?
- **Alignment of Quizzes and Classroom Tests** – How have the provincial tests influenced the design, content, and timing of quizzes and in-school testing?
- **Alignment to Past Examination Content** – To what extent do the shape, style,
format, and content of questions from previous provincial examinations shape lessons, assignments, and in-class tests?

• *Gatekeeping* – Gatekeeping is generally seen to be the practice of discouraging or preventing students who are deemed to have little chance of doing well on a provincial examination test from enrolling for, completing, or writing it. Does this happen here? If so, who does it, how do they do it, and what alternatives do students have? In a closely related matter, does the school permit and/or encourage some students to take the twelfth grade course twice?

• *The Media* – How familiar are these teachers with the Fraser Institute ratings and media attention to that and other assessments? To what extent do they think that they affect school district and administrative initiatives? The literature suggests that perceptions of teaching quality inferred from media rankings of testing results can be among the most significant of the testing stakes for teachers. Every interview should therefore include a question pertaining to BC media. Such a probe can be a good starting point for an in-depth discussion of the testing stakes.

The above starting points for interviews were not used as an interview schedule or agenda. Rather, they identified areas of potential exploration, the majority of which were eventually addressed during interviews and more informal conversations. Interviews progressed naturally and, in some instances, into unexpected territory. Had I felt more inclined to impose even further on the generosity of the teachers, the interviews could have been approached in a more scripted and comprehensive manner.
Starting points for interviews with principals. Previous research guided my interviews with principals where, again, perceptions arising from rankings serve as an effective starting point for understanding principals’ views of the impact of the stakes of the provincial examinations. I enquired about:

- instructional leadership relative to examinations,
- school-level moves to increase performance,
- the dissemination of examination-related statistics to teachers and subsequent discussions of those and contextual issues such as school district priorities and initiatives.

Starting points for interviews with students. Students have limited experience and have little or no perspective on what happens before their teachers enter their classrooms – and they have no basis for comparing instructional or institutional changes over time. They can, of course, speak to the impact of provincial examinations on their school lives. Individual and collective interviews with them were therefore conducted for two purposes:

- To triangulate principal, department head, and teacher references to the nature and effects of the tests and
- To probe for references to or signs of student test-related stress.

Classroom Observations

The purpose of classroom observations is to witness and record artifacts of the
phenomenon of interest as it unfolds within its natural setting, and to reap the benefits offered by direct observation as opposed to merely collecting views. The research reviewed above indicates that the content of past tests can have a major influence on the design of the curriculum, and therefore evidence of that influence might be found in the development of concepts, lesson examples, homework assignments, homework review, and assignments during instruction. Attention was therefore given to lesson styles, in particular to the degree to which lecturing (defined and discussed in Chapter 4) dominated the presentation of new content. References to the examination were noted. Finally, to the above lengthy and in-depth probing, I appended briefer classroom visitations.

During each classroom visit I sat in the rear of the classroom. In my annotated lesson observations I recorded the flow of instruction in the classroom, taking note of the length of lesson segments, the development of concepts, examples, what I felt were key questions (from students and from teachers), references to provincial examinations, availability of resources, and homework assignments. Chapter 4 discusses the observational data in detail.

Site Documents

I examined school and class newsletters, teacher Web sites, student and staff handbooks, classroom quizzes and tests, textbooks, hallway posters and information, and other relevant classroom materials.
External Sources of Data

I examined both schools' historic performances on the Fraser Institute Report Card on BC Secondary Schools as well as the data provided by the Ministry of Education on school-wide performances across subjects. The school and district Web sites were examined to discover any aspects of a school’s culture that might otherwise have been missed and to look for data pertaining to my research questions, particularly with regard to academic matters, testing, and student achievement.
CHAPTER 4 – ANALYSIS OF THE DATA

Introduction and Organization of Chapter

These data were collected so as to address the research questions:

1. What are mathematics teachers’ views and practices regarding high stakes testing?
2. How are these views and practices mediated by administrators, department heads, other teachers, students, and external influences?

This chapter presents a summary of the design of the study, followed by the analysis of the data from two schools, “Greenhill School” and “Pine River Secondary.” The data from each school is organized by personnel identified in the research questions – grade 12 teachers, department head, other teachers, administrators, and students – and the examination-related issues that they raised.

My Analytic Approach

In this case study, as in most such studies, the collection of data and its analysis are separated in order to organize this document, but in the field that is not the case. In conducting a case study, data collection and analysis often overlap (Anderson, 2002; Esienhardt, 1989; Stake 1995). Glaser and Strauss (1967) made a case for simultaneous

2 Pseudonyms.
coding, gathering, and analysis. The alignment model suggests looking for relationships between testing and teaching, and, as a result, references to the results of previous research on the instructional impacts of high stakes testing are embedded in an iterative and recursive process in which such references interact with potentially important local themes (Creswell, 2003). Triangulation locates areas that call for attention, affects subsequent probing, and therefore creates feedback between inquiries and the data collected (Yin, 1984).

The data set is then categorized and recategorized as a result of that interaction to accommodate previously elided relationships between people, issues, and events. This reflexive process has been characterized as constant comparison (Glaser & Strauss, 1967). Eisenhardt (1989) suggests being mindful of the pitfalls accompanying leaping to conclusions based on limited data, being overly influenced by the vividness or the status of respondents, ignoring basic statistical properties, and inadvertently eliding disconfirming evidence (p. 540).

The overall structure of the observations and analysis at the two sites used in this study is very similar. The unique qualities of each site are reported in context.

**Greenhill Secondary**

**My Approach to the Site**

Greenhill Secondary is in a large school district located in the Lower Mainland of Vancouver in a mid to mid-upper socioeconomic suburban setting. The building was
about 10 years old at the time of the research and appeared to be in excellent condition. The school enrolls about 1400 students in grades eight through twelve, and employs about 80 teachers. The racial composition of the student body is tri-modal: Asian, East Indian, and Caucasian, each representing roughly one third of the student population Most of the students intend to pursue post-secondary studies. From my observations, the school is well-ordered, quiet, and clean. In this building, academic time is highly valued and protected, and nothing short of an emergency interrupts instruction.

I arrived at school every day about five to ten minutes before classes started. This school had a “big” feel to it - larger than any high school in which I have spent any significant time. I therefore found it odd that the staffroom was often empty, or nearly so, during lunch. After being in the building for a few days I learned, among many other things, that at lunch time many teachers “retreated” into small groups in various corners of the building.

I had avoided asking for any formal or written introduction to the staff, and so, until my final day in the school, many of the people I interacted with assumed that I was a substitute teacher, even after I had explained what I was doing to some of them. I wandered and observed the entire campus. I struck up conversations with teachers, assistants, and custodians and walked through the neighborhood to get a sense of the demographics of the area.

I was also able to get a broad feel for the school and perspectives from several staff members who were not directly involved in mathematic instruction. An early conversation with a teacher of computer science disclosed that we shared interests and led to good rapport and an invitation from him to view and discuss the school’s
technology classrooms. I did most of my interview transcriptions in the library, and after the librarian learned of my research intent, she gave me some important background data on the demographics of the high schools in that district. When I showed her some in-district high school examination score statistics, she discussed particular schools with me, in terms of parental education levels, perceived cultural factors, and the like. And finally, I was invited to the special education department to view aspects of their program, and discussed mathematics diagnosis and remediation, standardized testing, and related issues. One of the participating mathematics teachers was a former university student of mine.

The school district's Web site suggests that they emphasize indicators of student performance, and that it makes decisions "based on data," which in this case refers to the results from grade 12 provincial examinations, the Foundation Skills Assessment (FSA), and grade 10 provincial examinations. The school district's Web site lists projects for improving student learning that includes, "an extensive internal analysis of provincial examination results." The assistant superintendent produces and distributes detailed statistics pertaining to provincial examination performances, which are in turn shared with department heads and relevant teachers. The school district generates statistics that compare the performances of its schools according to participation rates, mean school scores on examinations, male-female comparisons, mean school teacher scores, and the like. There is evidence provided in my research that this institutional emphasis translates into overall policies, specific initiatives, and interventions in school and classroom behaviours. We see below how the principal, the mathematics department head, mathematics teachers, and students are well aware of the administrative concern for
achievement testing performances and the issues and expectations attending them, and that those perceptions have had a major impact on course planning, instruction, assessment, and course delivery.

Mathematics Teachers' Views and Practices

This section presents the data from Greenhill School pertaining to the first research question.

Introduction

During the semester of my visit, this school had four classes of what is formally called 'Principles of Mathematics 12,' or 'Grade 12 Mathematics' where the context is obvious. Two teachers each had two classes. The more experienced of the two had been teaching Grade 12 mathematics for 15 years, the other for about 6 years. From the time of my initial contact visit through the end of my stay in the school, I perceived and documented the teamwork and collaborative planning that permeated their practices. These two teachers used a common schedule and set of materials, and communicated with each regularly regarding macroscopic aspects of course planning and content and other areas of concern. They lunched together as part of what appeared to be a well-coordinated mathematics department with a high level of communication between the Head and the other members. With 12 members the department seemed large – too large in the opinion of the mathematics department head because she said too many teachers without training in mathematics were teaching junior courses. For example, a music
teacher was, for the first time in 15 years, teaching a block of grade 8 mathematics.

The views of the most senior grade 12 teacher, “Dave” (‘T12A’ in the transcripts) appeared to be quite influential with his fellow teachers. Dave had been teaching Principles of Mathematics for about 15 years. I had several opportunities to interview Dave. He was interviewed as an individual, as part of a spontaneous group discussion (Group 1 Interview), as part of the mathematics department focus group (Department Focus Group 1), and, as with other teachers, during numerous interactions during lunch, in the hallways, and the like. He was observed for two full grade 12 lessons, during which I composed a running commentary, and in a larger number of shorter visits, ranging from a brief walk-throughs to 15 minute lesson segments.

In the transcript excerpts that follow, participants are identified by position (“T” for teachers, “P” for administrators, “D” for department head, and “S” for student. Teachers are further categorized by grade level taught, followed, if necessary, by “A”, “B”, and so on. There are two grade 12 mathematics teachers at Greenhill - “T12A” and “T12B” in the transcripts. The three grade 12 mathematics teachers at Pine River are referred to as “T12C”, “T12D”, and “T12E”, respectively.

“Sally” (‘T12B’ in the transcripts), had been teaching Principles of Mathematics 12 for about six years. She was interviewed in the “Group 1 Interview,” in the “Department Focus Group 1,” and was observed for two full lessons, as well as in a number of shorter drop-in visits. I spent less time in her classroom than in Dave’s. I was unsuccessful in negotiating an individual interview with Sally. Time pressure resulted in her participating in just the two group interviews. “Leanne” (‘D1’ in the transcripts) is the mathematics department head. She was interviewed in an individual interview, in the
“Group 1 Interview,” and in the “Department Focus Group 1.” She teaches mathematics in grades 8, 9, and 10. “Chad” (‘T11A’ in the transcripts) teaches mathematics in grade 10 and 11 and was interviewed individually. “Alice” (‘T11B’ in the transcripts) teaches mathematics in grade 11 and in earlier grades. “Rob” (‘T8A’ in the transcripts) teaches one block of grade 8 mathematics.

As data collection and analysis progressed and as various coding schemes were explored, important themes emerged. Some of these reflect the main instructional alignment effects discussed in Chapter 2, whereas others reflect aspects of the research site, such as teachers’ responses to school-level or district-level initiatives.

This section is structured around those categories and themes, and comprises data and commentary on these teachers’ views and practices related to provincial testing, addressing the first research question. There are seven subsections: teachers’ perceptions of the stakes, teachers’ views on gatekeeping, aligning instruction to the content of past examinations, course planning, course modification in grade 12, downward alignment, and aligning assessment and evaluation practices.

**Teachers’ Perceptions of the Stakes**

Through the lens of my model, testing stakes can be viewed as alignment pressures. This section reports on those pressures as perceived by teachers. There are two subsections: communication from school administration, and statistics and media rankings.
Communications from school administration. The teachers responded as follows:

R: Has provincial exams or provincial exams performance, or achievement testing performance ever been a topic at staff meetings?

T8A: They actually printed out all the results of all the different teachers, that is—you know, final exam marks and all the rest of it and they showed it to the entire staff, and uh, we’ve actually had the principal call people in at various times and said, “You have too many failures, and you need to do something about it,” and we’ve actually been, hmm, encouraged to change our evaluation after we made it, and that part bothers me a lot.

R: That’s mainly at the grade 12 level or has that been pushed downward at all, or...

T8A: It’s been at pretty much every level. Various teachers have talked about it over time. (T81 Individual Interview)

One of the grade 11 teachers described administrator concern and intervention over the low stakes FSA test thus:

T11A: With the grade 10s. I think we did it one year, the kids actually laid on the gymnasium floor and wrote the math test on the gymnasium floor and the administration was concerned about the results [italics added], but if I’m taking gym class and I’m laying on the gymnasium floor and I don’t have a pen or a
pencil, I just wanna get that test over with and done with. (T11A Individual Interview)

The mathematics department head describes administrative involvement in teachers' classroom assessments this way:

D1: Well, obviously the biggest picture is to get the students doing well, passing, and . . . ensuring success. Now, if you want to look at it administratively, they go by the average school mark – compare it with the district and provincial; and the big thing in the district is comparing the average school mark with the average exam mark. If it’s over a certain value – like it needs to be within a certain range, then we need to question are we being too hard/too easy on the kids. (D1 Individual Interview)

She then went on to say:

R: . . . but is it fairly obvious, at least on a district level or on a school level, that there’s a push, or a desire, or a drive for increased student scores?

D1: Well I think that’s - that’s rather obvious.

Next she discussed communications from the administration:

D1: One of our superintendents, [name provided], is really into promoting results
now across the district; like giving us data, giving us figures, giving us a working base that we can work upon. Uh, so yeah; it’s number-driven. They’d like to say, “No, it’s not number driven,” but yes it is number driven, we’re teaching to the test. And why are we doing that? Because it’s published, and we’re held accountable, etc. (DI Individual Interview)

This point is triangulated by views expressed by “Alice” a grade 11 teacher:

R: Do you perceive any sort of moves, or uh, desire from the district level regarding achievement testing?

T11B: Yes, there is. (Group1 Interview)

Dave mentioned the focus of his examination-related conversations with the principal:

T12A: I’ve never been . . . had a discussion with our admin on provincial exams except, you know, discussions about the actual exam performance, after the fact.

However, when I asked another teacher more directly about communication from administration I heard:

R: For example, [the principal] made it clear there is discussion among departments or whatever regarding best practice or certain things that work, or certain types of things that teachers might do to increase performance, or to become more aware
of the examination program. Another example I might throw out would be that - and this is true in my school – where teachers are very strongly encouraged to go mark provincial examinations. That would be another example. Anything in that kind of category that you want to mention? And I’m just interested in what kind of communication or input as a department, or you as individuals, get from, from the school administration regarding examinations?

T11B: From admin . . . ?

(pause)

R: The answer could be ‘very little’!

T12A: The answer is ‘very little’; I don’t . . .

The principal told me:

R: Have you discussed exam results and performance specifically with the staff as a whole or with departments?

P1: Uh, with staff as a whole, with uh, and that is most probably is a bit more, uh, I wouldn’t use the word ‘superficial’ but that is a little less in depth, with the staff as a whole. With department heads on a regular basis, and with individuals that teach the courses on a regular basis. And usually we try to structure something around some discussion as a group of people who teach provincially examinable courses on what works, best practice, things like that. So it doesn’t matter if you’re teaching Chemistry 12 or Math 12, or History 12, there’s some common things that work, that we can share with each other, things that are really helpful.
So we try to get some discussion going on that basis as well, inter...

R: Cross-subject?

P1: Yeah, cross-subject.

R: Anything as a specific example come to mind, with regard to, such as general approaches or techniques that... you’ve shared with them or that they’ve shared with you?

P1: Yeah, I think, I think that we can bring as administrators - we can bring to the table um, a level of concern, a level of support, the strategies and what works. And we can support that, we can support peoples’ initiatives, but I think it has to come from within. But the – changes or things like that are best when they come from within the department.” (P1 Individual Interview)

There seems to be a strong triangulation of perception among teachers I spoke with that there is a coordinated district-wide push for increased achievement scores. There was less consensus as to how the principal communicated his desires for good and better scores to teachers, although it seemed evident overall from the teachers’ comments that their perception is that he wants better scores.

*Statistics and media rankings.* The literature suggested that media rankings could be a good starting point from which to understand the testing stakes from teachers’ viewpoints. Overall, teachers in this school view rankings as powerful but invalid measures of school and teaching quality. When I asked teachers about media rankings, none of them expressed positive comments about them.
A grade 11 teacher suggested multiple impacts resulting from the power of media rankings:

T11A: Oh, I've got a tonne of comments on those. Since they came out, you can almost see the extra stress on teachers. You're looking at the results and you can say, "yeah, that school, that's because that school has an alternate program, or there is a higher standard of living over there, so you can almost tell which schools are going to succeed and have a higher success rate than other schools. You want to tell teachers not to take it personally, because it all depends on your clientele – who you have in class. You can have this phenomenal mark one year, and then next year have a horrible mark. It doesn't have to be an indication of you as a teacher – it has everything to do with your clientele in the classroom. And now I think, I can't say this for sure, but I'm guessing that teachers actually take the publications personally, and they encourage or discourage certain students to take a class or not take a class just to make themselves look better in the newspaper. (T11A Individual Interview)

Leanne connected media rankings to high stakes for everybody in the system:

R: But is it fairly obvious, at least on a district level or school level, that there's a push, or a desire for increased student scores?
D1: Well, I think that's rather obvious.
R: Had to ask!
D1: Well, the stuff gets published in the newspapers – they rank schools on the provincials. Come on now (chuckling). One of our superintendents is really into promoting results across the district – giving us figures, giving us a working base that we can work on. Yes, it’s number-driven. They would like to say, “No, it’s not number driven,” but it is. We’re teaching to the test. And why are we doing that? Because it’s published, and we’re held accountable. (D1 Individual Interview)

Rob commented this way:

R: Have you seen those double-page spreads in the local newspapers where they rank all the schools and all the districts? Have you ever read those?

T8A: Oh yes. And unfortunately that has become a real bone of contention in a lot of schools, because everybody wants to have their school at the top, and if they see their school towards the bottom they get really offended and they want to be sure that it doesn’t happen again. So it becomes a little bit of a competition. And it gets ridiculous. (T8A Individual Interview)

From these interviews, we see that published rankings are certainly involved in competition between schools. At Greenhill School, the most intense competition appears to be between schools within the district. It is fueled in part by the in-district publication of provincial examination results for this and other districts. The following excerpt from a group interview addresses the issue of competition.
R: Let me ask you this: Do you perceive competition between schools for provincial examination results?

T11A: (Looking towards another person present) I think you have a good sense of that, having been department head and sitting through all those math department head meetings.

T12A: Yes. There is a sense of competition. (Some chuckling).

T12A: Yes. As much as we don’t want there to be, and as much as we talk about how we don’t really care, . . . (group chuckling)

T12A: The [inaudible] is, yes, there is a competition. Yes, we don’t want to be the low school on the totem pole. Yes, we want to be as high on the totem pole as we can be. (Group 1 Interview)

The above excerpt reinforces the inference that there is a direct linkage between published rankings and teachers’ motivations. Every professional in the building with whom I spoke, from the principal to the librarian and the special education teacher, had given thought to the published rankings and the drive for better results on those rankings. In-district rankings were referred to more often in the mathematics department. It can be inferred that, for some reason, in-district rankings were seen as having more significance than others.

Teachers continually referred to the impacts of such pressures. The rankings, although seen as having considerable impact, were almost consistently viewed with skepticism. In most cases, they seem to have concluded that they had a negative impact
on their work. These individual and collective analyses are professional and personal, and as such form a part of the answer to my research question.

**Teachers' Views on Gatekeeping**

The mathematics teachers were sensitive to the issues surrounding this phenomenon of 'gatekeeping,' which, through my model, can be viewed as a systemic examination alignment. Regarding access to Principles of Mathematics 12, the department head said,

D1: One thing about our school is that we have one of the highest participation rates in the district, so we don't block people from taking Math 12. If they want to take it, they can take it. And still, even with that, we're still above district and provincial averages. So . . . (D1 Individual Interview)

The following is an excerpt from a meeting with the departmental focus group:

R: I want to ask you this. In the grade 11 course, do you get any sense among kids, or do you get any feedback from them, relative to the grade 12 course? I mean, what type of communications do you have with them relative to going on to grade 12?

T12B: "How hard is it?" (Group laughter)

T12A: Funny you should mention that, because I go around every year to all the
grade 11 classes and I lay out the program and say, “This is where you are now, this is what I want you to know by the end of grade 11, and this is the kind of mark you’re going to need to be successful in 12: the difficulty levels, calculus, where things fit. I do this 15 to 20 minute thing for all the grade 11 classes. And I just heard one of the girls today say, “Man, Mr [name], you really scared us!” (Group laughter)

I have no evidence as to whether or not these grade 11 classroom visits steered any students away from the regular course. Nevertheless, the teacher who spoke creates a tone with his visits, which could act as a psychological barrier for some students.

A grade 11 teacher connected gatekeeping to media rankings:

T11A: I can’t say this for sure, but I’m guessing that teachers actually uh, actually take … the .. publications personal, and they encourage or discourage certain students to take a class or not to take a class, just to make themselves, uh, looking better in the newspaper. (T11A Individual Interview)

Like many other references to gatekeeping, the above quote is in the context of somebody else’s classroom practice. With regard to his own advice to grade 11 students he said:

T11A: Math 12, I’ll encourage it for sure; to go into the Principles if they can do it. But if they can’t, then … I made the mistake of, uh, actually this year. A couple of kids came to see me and they wanted out of the Applications math into
the Principles math, and of course I’m always siding with the students, so I made room for them in my class. I already had 34 which bumped it up to 36, just, you know, I figured I’d give them a chance, give ‘em a chance. And now that they’re in that class, they’re basically bringing down the rest of the kids, so the Applications was the right setting for those kids. (T11A Individual Interview)

My overall impression is that most students end up being matched with the option that suits their abilities. In the end, students can insist on taking the examinable course if they are minimally qualified, which means earning 50 per cent in Principles of Mathematics 11.

Aligning Teaching to the Content of Past Examinations

I witnessed several significant and clear alignments between the structures and contents of lessons to past examinations. Some incidents were less clear. For example, during one lesson a teacher moved from the concept of the sum of a geometric series to the sum of a so-called ‘infinite’ geometric series. It takes only a few minutes to build that latter formula from the former, yet the teacher presented the formula, identified the relevant variables, and went on to examples. While one could argue that time-pressure on account of examinations motivated an undeveloped use of the formula, there may be other legitimate reasons why a teacher might employ that pedagogy.

During this lesson segment, Dave and a student had an exchange over the meaning of the “infinity” symbol in the formula. The student was struggling to
understand how one adds an infinite number of terms. The formula in fact computes the
limit of an arithmetic series. But the teacher summarized the discussion by saying, "[The
result of the formula] is what you get when you add an infinite number of terms." This is
mathematically incorrect. While it seems to me that developing the formula could have
had ancillary benefits, doing so would not address this pedagogical content error. At least
in this instance it appears that an inappropriate mathematical understanding is neither
ameliorated nor exacerbated by the imperatives created by the provincial examination.

The same day, I observed Sally teach the same material – "infinite" geometric
series - to her grade 12 class. To introduce the concept, she presented one of Zeno’s
paradoxes, in which a runner repeatedly moves half-way to a destination and can
therefore never reach the destination. Sally then recalled for students the formula for the
sum of a geometric series, and explained that in cases in which the absolute value of the
common ratio is less than 1, the terms of the series become progressively smaller, and
can be made arbitrarily small by choosing the number of terms to be sufficiently large.
She then applied this notion to the formula and showed that if the number of terms is
sufficiently large, then the formula can be made arbitrarily close to what is presented as
the "infinite" geometric series formula.

Sally’s approach to this concept was different than Dave’s, and arguably better
reflects the discipline as understood by authorities in mathematics. It is interesting to note
that the provincial examination does not discriminate based on this understanding, but
rather on the mechanics of using the formula. If the provincial examination assessed finer
detail, it would be able to discern which students understood that one cannot add an
infinite number of terms. But previous examinations have not explored this area in such
detail, and therefore, at a minimum, teachers need only train students in the use of the formula.

Some teaching behaviours were more aligned to upcoming examinations, such as when a teacher provided 'tips' for handling specific question types. For example, the government examination typically asks students to compute the number of terms in a given series that is written in Sigma notation. Dave said that such questions are common on examinations, and so as to "save precious time during an examination" demonstrated a method for quickly finding an answer. The above example illustrates well the permeable boundary between teaching and examination review. The alignment model suggests that the most closely aligned teaching is examination review. To wit, I observed students being guided in learning to use exhaustion/brute force to attack multiple choice questions by substituting the possible answers into relevant formulas. Such examination preparation as this was commonly embedded in lessons, along with bits of advice on such things as time management during an examination.

Teachers at Greenhill School rely on compilations of past examination questions as components of lessons. The two teachers of Mathematics 12 have created a resource package comprising hundreds of questions from previous provincial examinations. Samplings of those questions are selected for practice during regular instruction and reportedly are used almost exclusively during the examination review portion of the course.

Previous questions are a primary guide for choosing appropriate content. At Greenhill School, those previous examination questions are grouped by topic and unit and serve as a bank of items that can be used for creating quizzes and test.
Course Planning

I explained above how the two grade 12 mathematics teachers worked with close teamwork, following a rigid schedule in which they had planned the course down to the single-lesson level and then to final review lessons just prior to the writing of the provincial examination. All of their specific content goals, practice assignments, and quiz and test dates were included in their schedule. Close adherence to such a detailed schedule was possible because in this school nothing short of an emergency is permitted to interrupt either instruction-in-progress or the daily schedule.

The coverage of the curricular content was ‘finished’ in the final class before the midwinter holidays, and students continued to work until the ringing of the final bell on that day. That left most of January for a combination of examination review and an increase in the time devoted to the pre- and post-school tutoring offered by the teachers. As Dave put it during an informal conversation, “We do the course over again.”

Aligning Grade 12 Course Structure to the Examination

Teachers of mathematics suggested the idea to offering Principles of Mathematics 12 over two semesters rather than one, an idea that had been considered for several years at the time of my visit. In essence, this would allow some students to take the course

---

4 Tutoring was regularly available during the course. As examinations neared, such sessions might be held before and after school.
twice before writing the provincial examination. When I interviewed the school’s mathematics teachers, they told me about the modifications they had so far been able to make to the grade 12 courses:

T12A: [With reference to the track created to enable weaker students to complete the 12th grade program] So we proposed it three years ago and then in the springtime said, “This would be helpful for a lot of students.”

R: This is a good example of what I am looking for in this study, the type of thing I am focusing on. My impression is . . . correct me if I’m wrong, that this is a way to address student needs in a way that is pragmatic about examinations. I say ‘pragmatic’ because my understanding is that, by going this route, the full-year thing, they arguably pull down your school percentage – by three, four, or five points perhaps.

T12A: Yes, it was. That was a thought of ours as well. Absolutely.

R: Right.

T12B: But in discussions with you, you’ve seen that this is an opportunity for these kids to have success. It’s . . . you know.

T12A: Yes

T12B: It just takes them longer to process the same information.

R: Right. And so I’m viewing this as really pragmatic, that’s my impression.

T12A: Well, yes. I have to be honest here. That was part of our discussion. If these kids were in the regular program and did that exam, our mark would be lower - our average exam mark. (Group 1 Interview)
In the above interchange the teachers, with some hesitation, called attention to the performance-related and test-related qualities of this course. The principal later confirms this. This kind of provision is not unique to this school. One of the above teachers mentioned that a major Vancouver-area district now teaches all of its grade 12 mathematics over two semesters.

**Downward Alignment**

"Downward alignment" here refers to what Anderson et al. (1990) describe as "the extent to which the provincial examination program has affected the content of courses and the testing practices of the grades preceding grade 12" (p. 88). The current Instructional Resource Package in BC for Principles of Mathematics 11 (with no provincial examination) has topics that do not appear in the Principles of Mathematics 12 course. This makes them vulnerable to downward alignment because their removal can create room for topics that will be examined in grade 12. The following excerpt shows evidence of downward alignment into Mathematics 11 and the possible price that might have been paid for not grooming students in a concept area.

T11B: [Grade 12 teachers] have enabled me to see what needs to be emphasized in Math 11 before they go on. We have even worked in topics that are not ordinarily part of the curriculum to better prepare them. Trigonometry, for example, isn't even in the Math 11 program – I think all of [naming another
school district] does that; a big trigonometry unit in grade 11. And also the equation of the circle and things like that. (Group 1 Interview)

The above comment suggests that this and other schools have changed the mathematics curriculum in grade 11 with an aim of improving grade 12 examination scores. This practice is discussed in greater detail:

R: And I know you mentioned one thing . . . that there’s occasionally items in grade 11 that get extra emphasis in order to give kids a hand to get them ready, and, any other specifics that come to mind?
T11C: Radians...
T12A: Trigonometry.
R: And that’s basically because they’re trying to eliminate it in grade 11?
T12A: Yeah, yes, and it’s such a huge unit in grade 12.
T11B: And it’s such an interesting unit; they should have it, I don’t know why they don’t.
T12B: [inaudible] . . . and transformations, especially the reciprocal transformations, the harder questions like that.
T12A: Transformations of functions. That’s right, because that again is a whole unit in Math 12, and it covers trig transformations, logarithmic transformations and all the functions.
T11B: What about ellipses? Now we no longer do any with expansions or compressions.
T12A: No we don’t, all we do is the circle. And we’ve noticed that our – last time ‘round last year, our, um, section on conics – conic sections, um, was the worst that we did.

T11B: Isn’t that interesting. (Group 1 Interview).

Another grade 11 teacher told me:

T11A: So uh, if there is a topic that isn’t going to be on the exam I’ll still try to cover it, but I won’t, I won’t emphasize it by any stretch of the imagination. Or if I know they’re not going to do it the following year, I won’t spend, uh, very much time on that, I’d rather spend time on . . . exam topics and, uh, subjects and areas where they will need . . . to succeed the following year. (T11A Individual Interview)

Aligning Assessment and Evaluation Practices

The provincial examination is reflected in the structure, content, and style of the classroom quizzes and tests used by the teachers of Mathematics 12. The following excerpt verifies that their alignment is deliberate:

T12A: We make each test look like a mini-provincial exam.

R: Oh, in terms of the layout between multiple choice and open-ended?

T12A: Yes. And how it goes from a really easy question on trig to the last one
being very challenging. We try to do that, not always successfully. We always have, it’s a one-hour and ten minutes class, which is approximately half the time they get when writing the provincial. So we always make it, we try to make it, half the size of the test, so that it’s . . .

R: Oh. Okay, so it’s . . .

T12A: So instead of 43 multiple choice we’ll have 22 multiple choice, and instead of eight work-out we’ll do four work-out. So all our tests are 22 multiple choice and four work-outs that resembles half of a provincial exam. (Group 1 Interview)

The provincial examination also sets the standard for handling other external achievement tests. The school, including the mathematics department, uses the provincial examination as a lever for raising the level of student seriousness with which they take other tests such as the FSA and the impending grade 10 provincials. The department head said:

D1: We have been doing FSA tests here and we push to have the FSA very much like the provincial exam so that students will take them seriously. They weren’t taking them seriously before – they were writing them on the gym floor and all that kind of stuff. So it’s not uncommon for our upcoming grade 10s to know that there’s going to be a big exam [italics added], so that they will already have had practice to have it like a provincial. When it happens in grade 10 we’re not sure 100 per cent how we’re going to set it up, but it’s going to be just like the grade 12 provincials. (D1 Individual)
The above excerpt helps to distinguish examination-related practices from what otherwise might be taken to be standard practices in mathematics classrooms. The department head clearly links the grade 12 provincial examinations to their non-high-stakes tests. The members of the mathematics department’s comments on the FSA show that the stakes attached to a test can vary across groups. The FSA is a zero-stakes test for students, yet we see the school attempting to ramp up the level of concern for it. We saw under the ‘Principal’ heading that he believes that indicators such as the FSA results will become an important parts of the school’s goals. Those goals are under the direct oversight of the School Planning Council. The impact on educators will be felt if others use those results as an indicator of the quality of teaching and the effectiveness of the school administration. This means that the FSA may have increasing stakes for teachers and administrators.

**Teacher Summary**

These teachers work within a larger educational structure in which the Assistant Superintendent and their principal transmit achievement testing statistics directly to department heads and teachers. There are system-wide imperatives for teachers’ marks to be closely aligned with provincial examination marks. The teachers sense a push for better scores, district-wide and within their own school. The grade 12 teachers have responded with a strict schedule made possible by the school environment. They align their instruction by using previous examination questions for class assignments. They
align their assessment and evaluation by creating tests and quizzes to look like “mini examinations,” in content, format, and timing. They “finish” initial coverage of the curriculum leaving at least a month for review and dry runs.

Mathematics Department Head’s Views and Practices

The mathematics department head is keenly aware of district-level and provincial external assessment as it affects her department. Leanne receives statistics from the assistant superintendent and principal as well as producing her own. The interviews with the department head were particularly productive. She gave me excellent information regarding district-level statistics and analyses. The main thrusts of her observations are provided below:

R: So with provincial exams, what is the departmental response; what’s the big picture? Give me your perspective.

D1: Well, obviously the biggest perspective is to get the students doing well; passing and ensuring success. Now, if you want to look at it administratively, they go by the average school mark – compare it with the district and provincial. And the big thing in the district is comparing the average school mark with the average exam mark – if it’s over a certain value – like if it needs to be in a certain range.

(D1 Individual Interview)

My question had been open-ended, but the department head immediately focused on the
comparisons of local and provincial grades, presumably because that issue has a high profile for her. This focus was reiterated and broadened somewhat to include her views on evaluation and accountability as she responded to my further questions.

As the mathematics department head, she is responsible for overseeing departmental results across several achievement tests, and consistently used that data as a yardstick for assessing instructional success. For example, Leanne displayed some in-district school results and called attention to several schools’ results and discrepancies between school marks and examination marks.

R: (pointing to a district summary) And I see something in at least one school, actually.

Dl: Well, look at this: 9.1 [per cent] is the difference.

R: (pointing to a different statistic) And negative 5.6 [per cent].

Dl: And that's really skewed - that's a concern.

Dl: They come down on us when there is a big difference.

R: That's an issue, is it?

Dl: A big issue. (Dl Individual Interview)

Regarding staffing of Grade 12 courses she said:

R: On a departmental level here, can you tell me about some things that you have done, or are planning or coordinating around this big picture.

Dl: I delegate to the people who are best are doing what they need to do, and
that's my grade 12 teachers; I have 100 per cent confidence in them, so I let them run with it. They report – what we do is that every mid-term, it's consistent between every grade 12, and I get diagnostics on all of those, meaning they put the Scantrons through and then we get a diagnostic form that which tells everyone what the class average is, and what the average for each class is, so I collect that data, and then I share that data with them as well, um, so we really urge consistency, but regarding the provincial exams, I let [T12A] and [T12B] take care of that because they're my specialists, and I like to have specialists. (D1 Individual Interview)

As was noted above, there is a pervasive atmosphere of high-stakes testing in the district and in its schools. It seems that provincial examination results are the benchmark for legitimizing teachers' evaluations. The district "relies on data" to inform decisions regarding student achievement. In my time in this school, I heard only one reference to any indicator of achievement other than the FSA or the grades 10 and 12 provincial examinations. Whereas the provincial mark serves to assess instructional quality, matching it with local marks serves to indicate how well teachers are assessing their students. It can be argued that the above situation points to a positive impact of provincial examinations, levelling the field for students' final evaluations. And it is possible to identify other possible consequences. That issue is addressed in the Discussion section of Chapter 5.
Principal’s Views and Practices

Introduction

Because principals are eventually held responsible for the quality of instruction provided by the teachers in their schools, and because they are directly involved in who will teach examinable courses, principals can have considerable impact on the day-to-day work of teachers. We saw above that this school district has a strong focus on the assessments of students, an assertion supported by many of the interviews and documents that follow. The principal of Greenhill is well aware of and responds to this reality and, as we will see below, is fluent in discussing achievement testing and how it affects his school. The principal receives comprehensive performance statistics on each subject, including the school’s mean score, mean teachers’ marks, participation rates, and the like. Moreover, this principal avails himself of a further analysis supplied by Edudata\(^5\) which provides charts and statistics regarding finer details of examination performances. The principal’s level of concern became evident a number of times during my interview with him, and seemed to pervade our discussions of the ‘important’ student assessments: the provincial examinations, Foundation Skills Assessments (FSA), and the impending grade 10 provincial examinations.

The principal’s perspective intersects with the second research question through school-wide alignment practices. His views do not always mesh fully with these practices. As the analysis of this section unfolded, viewed through the model of

\(^5\) Edudata is a clearinghouse for educational achievement data, based at UBC.
alignment, a set of recurring themes, forces, and issues emerged. The principal's views and practices are grouped and discussed thusly, in five sections: views on provincial examinations, the impact of media rankings, views and practices regarding gatekeeping, communicating with teachers about examinations, and alignments aimed at improving examination performance.

**Views on Provincial Examinations**

In his brief answer to my first question, the principal captures the essence of his views on provincial examinations as they relate to this school.

R: I suppose that the first thing I am hoping to talk about is: Are the provincial exams a palpable presence in your school . . . in terms of planning, or structure, or initiatives, or . . . ?

P1: I think very much so. I think they form . . . it's the end point, and so a lot of what we do to get to the end-point determines a lot of program planning, so I think that provincial exams play a large role in the program planning from grade 10 on; in terms of people's thinking, especially people that teach the senior examinable courses and the department heads, are very conscious of where they are going, based upon the provincial exams. (P1 Interview)

With these remarks, the principal underscores the district and government-level focus on achievement testing and commented on the newly-mandated 'School Planning Councils,' a more muscular and policy-focused offshoot of Parent Advisory Councils (PAC).
R: Does the School Planning Council - I think that's what they are now called, is that right?

P1: Yes.

R: Do they ever discuss these types of issues, including achievement or FSA results or that sort of thing?

P1: Yes, that's the crux of the program. (P1 Interview)

The principal then elaborated on aspects of his philosophy concerning examinations in the following lengthy quote regarding the impending (in 2004) grade 10 provincial examinations.

P1: And exams - they're only worth 20 per cent, but 20 per cent is 20 per cent, and it raises the level of concern, right? And, um, one of the first things that some of the teachers said was, if we took last year's results, and took 80 per cent of their mark, and then added in a 20 per cent final, and assume that the students getting between 40 and 60 cent are the ones that do the least well on tests that test the whole year's work, and extrapolated that and said, well, let's - let's say that uh, here's a group that had between 40 and 60 per cent, you give 80 per cent of that mark and throw in a mark and let's say that the average mark was 5 out of 20, and then 8 out of 20, or 10 out of 20, what would that mean to our, to our results, and we came up with and said well that would mean that we need to have at least one more class of math 10 repeaters, one more class of science 10 of repeaters. Well, that's not the goal of the program! I mean those are the last kids that are
gonna benefit from repeating math 10, so what are we gonna do? What are we
gonna do so that those kids, so then somebody said are you just gonna, it wouldn’t
make sense just to raise our averages by 10 per cent so that kids wouldn’t fail,
because that’s just watering down our course, and there’s some issues like this
that we’re grappling with, how do you get teachers to see that it’s important that
every kid can learn and that every kid can do these exams, and at the same time
not increase our failure rate? It’s not going to have any benefit on the school or
the kids to have more failures. That’s not, that’s not my understanding of what the
exams are being brought in for. They’re brought in as a benchmark. (P1
Interview)

The principal related the impacts of the exam weighting on program planning,
concerns over program planning, impacts on the validity of teachers’ classroom
evaluations, and impacts on the curriculum. He went on to place achievement testing
results within his larger views of assessment:

P1: Well, yeah – I guess personally as an administrative team we’re looking at
assessment for learning. Until you get to the upper grades, assessment of learning
is important in terms of kids graduating and all that but we need to find ways to
assess what’s going to improve learning; so what are the measures? One of our
goals as a school is reading, increasing reading, uh, reading skills, so we’re trying
to find some ways to measure reading so that we can say, okay, here’s where
students are now, let’s introduce a program where are we now. So, in giving kids
that feedback so that they can see their improvement. So I think assessment, assessment for assessment’s sake, is one thing, and I guess there’s some people who believe that’s where the government is at in terms of where they’re going, but as a school we’re saying that these assessment things can be really useful if they’ll help us give feedback so we know where kids are, so we know how to improve so we, so we can improve learning; that’s what we’re looking for. (P1 Interview)

The Impact of Media Rankings

The principal provides his views on rankings of examination performance:

R: Do you think that the staff is generally aware of rankings of exam performance?

P1: Yeah. It’s published, it’s out there.

R: (interjecting) Fraser?

P1: Well, it’s published in the newspaper, it’s um, the district does a subject by subject analysis, by school the district, and as well we get the, uh, we get every year we get the, uh, for every set of exams we get the breakdown by – from the Ministry of Education we get the breakdown by each subject area, uh, and they even break it down to specific topics and that goes to departments and they, uh, and as well we buy the book from um, trying to think of the organization that gets the data uh, Edu . . . Educom.
R: Oh right, Edudata?

P1: Edudata, right. And they do a really good analysis; they’re able to take, like, that data we get from the Ministry is in table form, and it’s not always as . . . user friendly and uh, and Educom/Edudata is – are able to take that information and put into . . . into tables that give you a graphic that’s really helpful and that gives you a quick picture so that you look at stuff – you look at data and get a sense of, without having to go into detail if you’re, uh, how are we doing as a school you’re looking at a whole bunch of topics; how much’ can you look at – but this gives you data that you can look at and makes good meaning from it, and for the cost of it, I mean they have some sort of a setup with the Ministry – they’re able to get the data, put into a format and, and they’ve got the computers and the expertise to put into a format so that we can use it. (P1 Interview)

Views and Practices Regarding Gatekeeping

We turn now to the issue of gatekeeping – discouraging or preventing marginally performing students from taking examined courses. Through the lens of my model, gatekeeping can be viewed as a type of systemic alignment. Whenever students’ test performances have become of sufficiently high importance to reflect significantly on teachers’ and schools’ perceived quality, gatekeeping has often occurred, and the practice is now often attributed to ‘successful’ schools. As is seen in the foreshadowing questions, an objective of this study was to separate such attributions from actual practices and to determine administrators’ views concerning gatekeeping and the degree to which they
feel comfortable sanctioning its use. When queried about gatekeeping this principal said:

P1: Well, philosophically and on a personal level, I think that academic rigour is important and that we all need it whether we’re going to university or college or not. You need a sense, by the time you leave high school, that you’ve had a course that challenged you in the sense of academic rigour and I think that because of the provincial exams that all of those courses have that. Now, there are some other courses that don’t have a provincial exam that have all sorts of academic rigour – Law 12, Comparative Civ. etc. But you can almost guarantee that if someone takes Physics 12, Math 11, or History 12, they’re going to be pushed. If they were questioning themselves and get 57 or 65 per cent, I think they’ve really benefited from having a course like that. The lower mark as compared to other kids’ isn’t going to hurt them because they may not be going on. They may be going into a different field, like going to BCIT (British Columbia Institute of Technology) and taking maybe electronics or something like that. I think that in a History course, for example, people benefit from being in a course with other kids and being challenged. They think and move ahead and approach it like an exam that counts. I think that it’s educationally sound for kids at the lower end to be doing that, and that looking just at provincial exam results is a bit dangerous, if that’s all you’re doing. [Regarding participation rates] If you look at our participation rates, a lot has to do with personnel, but our math department in particular has always encouraged kids so we have a high participation rate in our Principles of Mathematics 11 which translates into
Principles of Mathematics 12. So we're usually about ten per cent higher than the district. (P1 Interview)

Communicating with Teachers About Examinations

I asked the Principal about his communications with teachers.

R: Have you discussed exam results and performance specifically with the staff as a whole or with departments?

P1: Uh, with staff as a whole, with uh, and that is most probably is a bit more, uh, I wouldn't use the word ‘superficial’ but that is a little less in depth, with the staff as a whole. With department heads on a regular basis, and with individuals that teach the courses on a regular basis. And usually, uh, usually we uh, try to structure something around some discussion as a group of people who teach provincially examinable courses on what works, best practice things like that. So it doesn’t matter if you’re teaching Chemistry 1 or Math 12, or History 12, there’s some common things that work, that we can share with each other things that are really helpful. So we try to get some discussion going on that basis as well, inter...

R: (interjecting) Cross-subject?

P1: Yeah, cross-subject.

R: Anything as a specific example comes to mind, umm, with regard to, such as general approaches or techniques that... you’ve shared with them or that they’ve
shared with you?

P1: Yeah, I think, I think that we can bring, as administrators, we can bring to the table, um, a level of concern a level of support, the strategies and what works and we can support that, we can support peoples’ initiatives, but I think it has to come from within. But the – changes or things like that are best when they come from within the department, within a group of people with the person saying, “I wanna try this,” and then our job is to support them in that. Umm, we can suggest things that we might hear that other schools do or things but I think that most of it is uh, I think comes from the group themselves, either individually or as a group or as a department; the best ideas, they’re the ones who do it, they’re the ones who know.

(P1 Interview)

The principal was reluctant to go into specifics regarding the above comment. My data here does not show how that ‘level of concern’ is communicated to teachers. I succeeded in getting more details in later interviews with the school’s mathematics teachers. Because it bears directly on the principal’s role and practices, the following excerpt from a teacher interview is again given here:

R: Has provincial exams or provincial exams performance, or achievement testing performance ever been a topic at staff meetings?

T8A: They actually printed out all the results of all the different teachers, that is – you know, final exam marks and uh, all the rest of it and they, uh, they showed it to the entire staff, and uh, we’ve actually had the principal call people in at
various times and said, "You have too many failures," and uh, you need to do something about it, and we've actually been, hmm, encourage to uh, change our evaluation after we made it, and that part bothers me a lot. (T8A Individual Interview)

The principal's level of concern also extends to the much lower stakes FSA, where he oversaw coaching. This is reflected in the following quote, seen above, from a grade 11 teacher:

T11A: With the grade 10s, once again, I think we did it one year, the kids actually laid on the gymnasium floor and wrote the math test on the gymnasium floor and then administration was concerned about the results; but if I'm taking gym class and I'm laying on the gymnasium floor and I don't have a pen or a pencil, I just wanna get that test over with and done with. And, you know, it's not for your marks or anything like that so who cares, and after, our results we're really poor in the math area and that, so the following couple of years we tried to uh, we made it more of a classroom setting, so we actually wrote it during the Enterprise (a school course) block, um, and we actually held of couple of tutorials as well. (T11A Individual Interview)

The department head's views intersect with the above assertion:

D1: We push to have the FSA very much like the provincial exam setting so
students would take them seriously, because they weren’t taking them seriously before; they [students] were writing them [FSA] on the gym floor and all that kind of stuff. (T11A Individual Interview)

The school district’s and principal’s expectations regarding assessment and evaluation are evident in this quote from the department head:

D1: Now, if you want to look at it administratively, they go by the average school mark – compare it with the district and provincial; and the big thing in the district is comparing the average school mark with the average exam mark. If it’s over a certain value – like it needs to be within a certain range, then we need to question are we being too hard/too easy on the kids. (D1 Interview)

She follows this up with:

D1: They’d like to say, “No, it’s not number driven,” but yes it is number driven, we’re teaching to the test, and why are we doing that? Because it’s published, and we’re held accountable, and that’s why it’s so important having the exam mark, and the school mark within a range. (D1 Individual Interview)

Alignments Aimed at Improving Examination Performance

This section itself has three subsections: staffing, modified grade 12 course, and
downward alignment.

**Staffing.** The Principal discusses staffing with respect to interactions with parents:

R: Do parents ever communicate with you regarding examination performances; past, present, or future?

P1: Lots of times, and usually, if the parents perceive it as a problem, if there’s a problem occurs, uh, they will, they will communicate and sort of say, that we think, or say that my perception is that there’s a problem in this – in this area. Like, like English 12 would be a good example of a thing that – if, if the results are not there, it’s gonna affect kids’ ability to go on to university, college, all kinds of things, and think you’re gonna get . . . I think we’re more sensitive to who’s teaching that for that reason. . . . um, and parents are more sensitive. (P1 Interview)

But when I asked him directly about the connection between perceived teaching strength and examinations, he demurred somewhat:

R: Right. Speaking of personnel, um, it’s been my experience and I think um, the experience, or at least the perception of many people that the strongest teachers will be in those positions. Is that a fair statement?

P1: I think you have to have a passion for your subject area – to want to take on
the extra work involved in a provincial exam. And if extra time, or, passion for your course also means that uh, “better instructor,” the two go hand-in-hand? Well to a large extent they do because that’s where passion comes from. If I believe that math is the most important subject on the face of the Earth, then that will translate into my teaching and into the way that my students approach it. But if I also, I can also believe that Spanish, languages are the most important – our art teacher may be the most passionate person in the world and that’s why her art course is a great program, so, the people that are teaching Math 12, taking on that extra will tend to be your passionate teachers. (P1 Interview)

**Modified grade 12 mathematics course.** As is suggested above, the school offers an alternative version of the Principles of Mathematics 12 course for weaker or marginal students, covering the full year rather than just a semester. When asked about its rationale, the principal explained:

R: So you provide the option now of some students to go year-long in Math 12?

P1: Yes

R: What are the roots of that?

P1: It came from the Math 12 teachers. They said that some students, after Principles of Math 11, struggle tremendously to make the next jump. They can get there, but they’re not going to get there in the Math 12 course because it has been semestered. It’s too short. There is no time for reflection. So they came up and said, “What if we offered students a program in grade 11-12 that they can take? A
Math 12 course, and in the second semester they would take a Math 12 program. The idea was to have a separate course for those kids. But it’s hard sometimes to get 30 kids and be able to timetable them, so essentially what we are doing is taking Math 12 – they’re taking Math 12 as a regular course, but they’re getting credit for Math 12 Essentials for the first semester and so they get a credit. Some of them decide not to go on, but the majority of them do go on.

P1: Dave looked at last year’s provincial exam results. He figured that it, and I hate to use words like ‘caused us,’ brought down our provincial exam results by about 4 or 5 per cent because of throwing in 25 students who didn’t have to be there. But we encouraged them to be there. But every kid that decided to stick with it and take it was successful in Math 12. Dave figures that most of them wouldn’t have been successful if they had taken the course just once. (P1 Interview).

Like the most experienced grade 12 teacher, the principal positively but somewhat hesitantly connected this move to examination results.

**Downward alignment.** The principal’s concern has clearly influenced the school’s planning. For example, in an excerpt from a much longer answer concerning the upcoming grade 10 provincial examinations, the principal said:

P: And now they’re looking at exams, what does that mean, what that’s going to look like, there’s a sense that – you just can’t put in exams and say, well, it’s
gonna happen, and our sense is, what do we do to prepare for that? What does it look like in Grade 9 to prepare for grade 10, what does it look like in grade 8? One of things we're looking at is, is, re-looking at our grade 8 and 9 program. (P1 Interview)

The principal was thereby directly linking the provincial examinations with one of the issues referred to in the preface to this chapter, 'downward alignment': administrative and instructional moves made in grades earlier than those in which we find provincial examinations.

Summary of Principal's Views

This principal’s concern for good test results certainly influences his planning, staffing and staff assignments, and communications in his school, but is conditioned by his concomitant concern to do no harm. That observation leaves open, of course, the questions of what an administrator considers to be ‘doing harm’ and how a dedicated professional can balance those two concerns. Overall, then, this principal provides strong input into his teachers’ work. He does this primarily through his efforts to cope with the central role that achievement testing now plays in external and internal judgments of the quality of instruction in his school. It is worthy of note that while his concerns permeate planning, staffing, and communications in the school, he stops short of school-wide examination alignments that, in his view, could be educationally harmful.
Students' Views

While the imminent tenth grade provincial tests may lead to students becoming sensitive to provincial tests in earlier grades, at the time this study was conducted, students first became familiar with provincial examinations in eleventh grade. The following excerpt from a focus group of grade 12 students indicates how:

R: Anyone feel free to answer this. When the course started, did it take long to hear about the exam?
S1: No. That was the first thing they tell you.
R: And, what did you hear? What did you find out about the provincial exam?
S2: You actually hear about it in grade eleven; they talk about it.

The above quotes, repeated frequently, reflect the primacy, from students' perspectives, of the provincial examination's role in the course. As the student focus group interview continued, there was continuing evidence, triangulating my classroom observations, of how the provincial examination is woven into many facets of the course. The following excerpt is from the same focus group student interview:

R: Have you ever seen any previous provincial exams? Have they been handed out to you? (General nodding in agreement).
S1: Yeah. We have previous questions. But we haven't had our practice

---

This was a focus group of five grade 12 students, nominated by their teacher for participation in the study. They comprise a small and not necessarily representative sample.
The students spoke highly of their teachers and their ability to navigate them through the examination experience. For some students, the stakes were very high, being determiners of possible career paths. Yet beyond some generalized angst over the provincial examination, these students as a whole were not at all overwhelmed by the pressures that could accompany their ultimate provincial examination performances. As one 17-year-old put it:

R: . . . there was a previous research study, back in '97, in which a guy at S.F.U. claimed that - who claimed that grade 12 students are under significant - he didn’t say massive, but significant - and very real stress over provincial exams and I’m wondering if that was overboard, underboard, or was that a bit much.

S: I think it all comes together in grade 12. Like, that’s why everyone feels so stressed out, and she’s right . . . you know they’re thinking about next year, your whole set routine of school is just gone. Like everything – your base, your comfort zone, everything is just gone. You’re pretty much starting from scratch and, like, technically, you’re an adult now. In a way, like, somebody I know, if they didn’t go to school, they would have to pay rent. They have to pay. (Student Focus Group 1)
Site Summary

This school is embedded in a larger district culture of achievement testing, one that relies heavily on numeric indicators of results. The provincial examination is at the core of that culture and appears to be a common thread running through district-level initiatives, competition between schools, the principal’s long-term planning, school timetabling, teachers’ course planning, instruction, in-class assessment, and public perceptions of the school’s quality. Several participants, particularly the mathematics department head, spoke openly of the test-score driven nature of the district and the school’s climate. The teachers of 12th grade mathematics introduced new material using the lecture method. Previous examination questions were involved in aligning lessons, assignments, and assessments with the external test. Classroom tests were deliberately aligned to mirror the provincial examination in design, content, and even the proportions of various kinds of questions. The teachers leave three weeks to a month of a five-month semestered course for examination review and coaching. Published rankings, both in-district and province-wide, were studied closely and affected both morale and later interventions.

The impending grade ten provincial examinations are generating considerable pro-active planning for the delivery of this soon-to-be-examinable grade 10 mathematics course, much of it intended to replicate the adjustments that have been made to accommodate the twelfth grade examination.
Pine River Secondary

My Approach to the Site

Pine River Secondary is located in an urban, middle to low-middle socioeconomic setting, close to a major thoroughfare and economic activity. It enrolls about 800 students and employs about 40 teachers. The student population comprises, in order of predominance, Caucasian (with significant proportions of Italian-Canadian and Eastern-European extraction) and Asian Canadian. This school has a religious affiliation. The school’s mandate is to provide a complete education: academic, spiritual, physical, and service. This is reflected in what the school says about itself, and in what I witnessed over a period of about five weeks. Entering the school, one sees a poster that provides some information about the school and stating that it is first and foremost an academic school. I soon found myself deeply immersed in the culture of the school, and I quickly learned that things were much more complicated than I first thought.

Underpinning much of what happens here is a drive to be the best – in sports, theatrical and musical performance, and academics. This is reflected in athletic results, superior theatrical performances and, when viewed through a more sophisticated lens than simple rankings, good provincial examination performances. But all that requires an investment in time. Time is of course a valued commodity throughout our culture and certainly in most schools, but the level of concern for it is higher and explicit references to its lack are more common in this building than I have previously encountered. I soon found that thinking of time as a limited and valued resource served as one unifying thread.
running through my observations and interviews at Pine River and helped clarify them.

Pine River Secondary encourages, and has succeeded in, having a powerful set of personal relationships, close friendships, related by family or through faith, that weave the school together. Many teachers and administrators spoke openly of the interconnectedness of the staff. The staff room was often full, and people quickly struck up conversations with me and offered their opinions openly. The teachers knew why I was there and some of them sought me out to express their views. This candor seemed to permeate much of what I saw and heard there. I was given full access to the mathematics classes and teachers, and I accepted invitations to observe some religious observances. I very quickly learned about the school’s storied athletic legacy and about the influence of the athletic department in the school.

The overarching ethic of competition and desire for winning creates an urgency among teachers concerning provincially examined courses. The teachers of these courses must compete with other pursuits in the building for time, and evidence of academic time-pressure was manifested many times during my school observations and interviews with grade 12 teachers. Understanding Pine River Secondary entails becoming familiar with the dynamics of that tension and how it affects people there, and most cogently, how it mediates the mathematics teachers’ responses to the provincial examinations.
Mathematics Teachers’ Views and Practices

Introduction

The vast majority of all mathematics courses in this school are taught by three teachers, all of whom teach Principles of Mathematics 12. Unlike at Greenhill School, the terms “mathematics teachers” and “grade 12 teachers” are interchangeable here. There is considerable teamwork among the three mathematics teachers here, who are all involved in extra-curricular activities as well as offering out of class assistance to their students. I was given full access to the mathematics classes, and I spent considerable time watching lessons across several grades. As at the first research site, I took detailed notes during a number of grade 12 lessons, and several times during lessons in earlier grades. They all use the same topic sequence, classroom resources, and assessment materials. They were always within a few lessons of each other in the course schedule, and the teachers pressed hard to ‘finish’ the curriculum in time for provincial examination review activities, which were planned to start in mid-May.

They gave every indication of viewing those examinations as a formidable reality in their lives. On many occasions they linked issues within and outside their school and components of their work to pressures arising from provincial examinations.

Overall, there was considerable standardization in grade 12 lessons, across teachers and within individual classrooms: basically one or more lectures followed by homework questions, which were not checked by any grade 12 teacher during my stay. Lecturing, as defined in Chapter 2, was the sole lesson mode used to introduce new
mathematical content.

"Syl" (‘T12C’ in the transcripts) is the most senior grade 12 mathematics teacher, and from my perspective the most influential member of the mathematics department. At the time of the interviews, he had been teaching for 23 years, the previous six involving Principles of Mathematics 12. "Sam" (‘T12D’ in the transcripts) had been teaching for eight years, but this was his rookie year teaching Principles of Mathematics 12. Annie (‘T12E’ in the transcripts) taught the Honours 11 course as well as courses in earlier grades.

It did not take long for powerful themes, addressing the first research question, to emerge from my interviews with teachers and observations of their lessons. Several coding schemes were used to interpret the data, resulting in considerable analytic overlap between the teachers here and those at Greenhill School, and their parallel treatment in this document reflects that. There were some differences, however, which is also reflected in the structure presented here. This section presents the data for Pine River Secondary as it pertains to the first research question, and is organized into eight subsections: grade 12 teachers and time, teachers’ perceptions of the stakes, teachers’ views on gatekeeping, classroom resources, aligning teaching to the content of past examinations, course planning, downward alignment, and aligning assessment and evaluation practices.

Grade 12 Teachers and Time

The grade 12 teachers (of all examinable courses) had their own interpretations of
the balance between academic instructional time and other educational pursuits. I soon confirmed that teachers of senior academic courses felt tightly constrained by time limitations. Grade 12 courses are designed to be taught using just over 100 hours of instruction. Yet one of the teachers computed that his grade 12 examinable course had, because of other school activities, been reduced to about 90 hours of time. Such an impact on instructional time certainly curtails instruction and quite possibly does affect examination results. The present teachers experienced the pressures for good scores in a highly competitive context with reduced teaching hours and variable stability in the daily schedule, in addition to what many grade 12 mathematics teachers have described as a very full curriculum. Several other teachers commented on the struggle for teaching time. A long-time teacher of senior English gave a detailed response on this issue:

R: You know, basically at the heart of my study is learning how teachers strike a balance between the imperatives of achievement testing performance and our larger educational goals – the kinds of things we think should be happening in schools. And in essence, in this school, for example, I knew that – after I was in the building for about 10 minutes, for the role that [a specific sport] for example T12F: Yeah, yeah... and [another specific sport] and drama and music, and so forth.

R: And this in essence is a small arena for that question, for example, do you detect some push and shove between – with senior students on sports and academics here?

T12F: I detect it between the teachers. I mean it becomes very frustrating when
you’re trying to cover a course, and the kids are being removed several times a year. Frequently that happens for every teacher, and.. that does put you in a real bind, because you can’t punish the kids by teaching essential material when they’re absent, but you can’t hold up the course, and it’s a constant predicament. On the other hand, we generally subscribe to the philosophy, the truism, that the well-rounded student is what we’re aiming at.

R: Uh huh.

T12F: So we all see the advantages of the kids doing drama, doing soccer, doing football – we see all those advantages, and we put ourselves in this position that we’ve proliferated these activities almost to the point of insanity in this school. You have no idea how many extracurricular activities there are. And there’s no limit to the number of activities that a student can undertake. So if the student chooses to do 12 activities, he might well miss half the year! (laughter). And we never look at this issue. Um, it’s an ongoing problem. It, it presents a problem for the kids, because they’ve still got to get all their assignments in, and we’re yanking them off to a three-day thing at Whistler, or we’re yanking them off to Kamloops, or whatever. They’ve still got to get all their work done, and you know bloody well that on road trips kids don’t sleep; they come back and they usually miss another week recuperating from the road trip, and we constantly put them in this position. (T12F Individual Interview)

From another teacher:
R: Is there a lot of impact on instructional time?

T12E: I think so. (Referring to a teaching colleague sitting beside him) thinks so.
Every now and then we would memo the office and complain. (Focus Group 2)

Syl, the most senior grade 12 mathematics teacher, made reference to this intrusion on academic time several times during my interview with him, including:

T12C: ... but I feel that Math is a much more demanding curriculum. And so there is pressure, and we know that [the principal] wants to improve marks, he's always giving us statistical information to improve, but one of the drawbacks is we tend to miss a lot of school time for a number of functions, and even if our classes are shortened rather than miss an entire period, it seems to break the flow of the learning aspect for the students, so it's easy to make excuses, but we don't have as many classes (periods) as one would anticipate or suspect that we have (Focus Group 2)

Given the above typical expressions of concern, it is easy to understand the protectionist responses toward time that many grade 12 teachers had adopted. As is mentioned above, one grade 12 teacher was so meticulous in his minute-level record keeping of lost instructional time that the principal would rely on him/her to help provide the data for the intricate and equitable timetabling adjustments that were negotiated between the principal and his grade 12 teachers. No such record keeping is needed for non-provincial courses, which are in fact set aside toward the end of the year. The vice-
principal added his perspective on this issue:

VP1⁴: . . . the only feeling that I get from talking to the math department, I always get the impression that they don’t get enough time; that they want more time. And I don’t know whether we’ve adjusted for them in recent times, in that we’ve changed our timetable, we’ve given them longer classes, they get more minutes than they did three years ago. But all our classes do. We’ve eliminated homerooms for two-thirds of the year, where we used to meet with our home classes. All that time’s gone into classroom time now, and at the end of the year here in June, we stop all our non-academic classes for Grade 12s, and we schedule more classes for our grade 12 including mathematics so I, I ask myself this question all the time: how come they can never finish!? (VP1 Individual Interview)

During my stay in this school I saw considerable evidence of the tension between academics, sports, and the performing arts. For example, about one week into my stay, I took part in a pep rally for the school’s spring musical. The following excerpt is taken from the field notes that I recorded shortly afterwards. Identifying information has been redacted:

The music/drama stream is more powerful than I had surmised. [The vice principal] called the students down by grade, to the auditorium, which is equipped

⁴ Both Vice-principals were interviewed in this school, coded as ‘VP1’ and ‘VP2’.
with an orchestra pit, professional sound and lighting equipment, and a control booth in the upper crook of an alcove of the gymnasium-sized room. When I heard the announcements I moved toward the staff room, and as [a teacher of grade 12 English] emerged from the staff room, he said, “It’s another one of those days!” meaning that the flow of teaching had been interrupted by something. I had noticed on my first day that students were being called out of class for yearbook photos.

I entered the gymnasium and noticed that the bleachers had room, but not much, and the students were well-behaved as [the drama teacher] explained the purpose of the pep rally, which I saw to be a two-scene sampler from a play. The acting and singing were excellent, on par with those in a community theatre. Busloads of elementary school children (feeder schools, I imagined) and long-term care residents lined the hallway waiting to see their matinee performance. This pep rally raised feelings of nostalgia in me, because my own high school followed a similar model wherein academics, sports, and arts were celebrated equally.

The following edited excerpt from my field notes gives a flavour of the athletic traditions of the school:

With regard to sports, I soon began to gain some depth in some of the traditions of the school. I spent at least 20 minutes examining the sports case near the gymnasium, and I was impressed by the volume and quality of evidence of
athletic accomplishments displayed in it. I saw photographs from more than 20 years ago that included a number of current staff members. I noticed that one of the mathematics teachers graduated from here and then played [a professional sport] in Germany. Two [well-known] brothers played here. I watched the staff play a student team in floor hockey and I saw some healthy competition, with both sides playing hard to win. It was a pleasure to watch both sides on occasion play very intensely, but always within the bounds of a friendly game. The staff set an excellent example for the student body, which was well represented as spectators. The staff won 2-1 as I recall.

I've since had the opportunity to talk with most of the 'coaches,' as a number of colleagues refer to the physical education teachers, who on the average are very large and athletic men – something I haven’t encountered in the public school system. I know that the term ‘coach’ is used in some American high schools and in colleges and universities, possibly to serve a symbolic purpose: to identify sports as an important and distinct entity in the building. The job of the coaches is quite demanding, even involving scouting and analyzing the competition’s games in order to learn their game plans.

Since I gathered that that there are many students in the school who participate in more than one sport/club/activity, it seems that many students attending this school, particularly in grade 12, have considerable demands beyond those associated with academic performance on their time.
Teachers’ Perceptions of the Stakes

The testing stakes can be viewed as alignment pressure. I sought here to discover what those pressures are. Most teachers, particularly during informal conversations, talked about the stakes for students attached to provincial examinations. Teachers also talked about the need to be viewed as being able to competently prepare students for provincial examinations. My conversations with them ranged across different aspects of the stakes, but two primary sources of testing stakes emerged from the analysis, and this section on teachers’ perceptions of the stakes is structured around them: communication from administration, and statistics and media rankings.

_Communication from administration._ The English teacher discusses communications from the principal regarding achievement testing results:

T12F: Um, [the principal] you know, asked us to address it. He’s asked us to make a concerted effort to improve our standing, but it’s not something that becomes a burden or something that one’s constantly reminded of and so on, and he’s certainly never come to me and said, ‘Look, uh, what are you doing about this kid or that kid?’” (T12F Individual Interview)

Sam gave his perspective:

T12D: I think part of our staff meeting at the beginning of the year is to look at
our results from the previous year, and how perhaps maybe we can improve on them, perhaps different strategies, different methods of teaching or what can you do, what can you focus on, to improve your scores.” (T12D Individual Interview)

Annie noted:

T12E: We do have meetings – academic meetings – maybe two or three times per year, where the last year’s – previous year’s exam results are compared and discussed and what we can do to improve them. And so on and so on, so um, it is brought to our attention. From that academic perspective. Those of us who are familiar with grade 12 provincial exams, the administration here keeps saying, “The academics are very important,” but, we still get them pulled in all directions. (Focus Group 2)

Annie moreover said:

T12E: [The principal] wants to improve marks, he’s always giving us statistical information to improve . . . ”

The principal’s views triangulate well with the above, in considerable depth:

P2: Yes, I asked the department heads and particularly the government exam subject teachers to consider that and are their ways – I like the binders they put
out (motioning to what appeared to be a series of Ministry of Education report binders). I’m sure you’re familiar with the analysis that the Ministry puts out, the reams of paper that they churn out. And that you can now take off their Web site. For the longest time and, again, that’s remiss on my part, since everybody calls me the ‘school statistician’ – part of my degree was statistics, the other part was hydrodynamics – but I didn’t use that statistical analysis of the results, both by topics, say within Principles of Math 12. And the skill levels – the knowledge and the higher order thinking those sorts of things. We did explain those for the benefit of the non-numerate. Department Heads and subject teachers, and they’d be say, let’s monitor these in fine detail than just averages on a year-by-year basis. And would we have now got them into the mode of looking at those when those government exams result come out Do we take account of the numbers and do we acknowledge that they drive us in some sort of fashion, and I think yes they do. (P2 Individual Interview)

Statistics and media rankings. I asked grade 12 teachers about the impact of media rankings of examination performance:

R: Is it your opinion that the staff is well aware of rankings?

T12F: What kind of rankings do you mean?

R: Of provincial examination rankings like – such as, Fraser . . .

T12F: Oh yeah, absolutely. And we are also aware of the fact that that actually has a direct impact upon us, because when it comes to choosing between Pine
River and [a rival school], the parents may want to send their kid to Pine River Secondary, but they may choose [another school] because they are ranked higher than we are. I also think that the other thing we are well aware of is that the rankings of the Fraser Institute, for example, are certainly open to a lot of question, and yet people take them – regardless, even the people who know how questionable they are, will still take those results very seriously, and to some extent, either put into a position of pride or panic, depending on where they stand [mutual laughter]. And so I think, I think you can’t ignore that, it’s out there in the public domain, and it does, it directly affects us. (T12F Individual Interview)

Sam said:

T12D: I think – from my impression. Now, it’s something that people want to focus on and want to improve, but at the same time it’s . . . many schools are at the same spot they were the year before. There’s reasons for the top-end schools, why they’re there. Clientele and all that. (T12D Individual Interview)

In what follows, Annie and Syl explain the connection between statistics and classroom practice:

R: Can you give me some indication as to how you use provincial examination results in order to fine tune or to modify or have an impact on your own classroom practices?
T12E: The provincial exam [Ministry of Education] puts out a very thorough breakdown of how various topics have been successfully handled across the program, and every now and then we can look at those statistics and say,... [drowned out by background noise]. less successfully than previous years. (Focus Group 2)

Syl had this to say:

T12C: One thing I know is that in the old curriculum, our students tended to be quite weak in the conics section, but they'd modify or change the curriculum so that conic sections are not emphasized as much. But the other areas are still relatively new, so I don't know where our strengths and weaknesses will be. I don't like, for example, the trig area, it's always been about 15 to 20 per cent of the provincial exam. Kids struggle with it, but that's because of the volume of the material that they're dealing with. But to be honest, I don't pay much attention to it because it fluctuates from one year to another. So, I don't focus in on any one topic. Depending on the scheduling for the year, will determine which topics we teach first, which topics we teach last. (T12C Individual Interview)

Teachers’ Views on Gatekeeping

It is evident in a number of interviews that this school prides itself on its participation rates in mathematics, which was evident in a number of interviews. The
view expressed by teachers was that a student in Principles of Mathematics 11 should earn an ‘A’ or a ‘B’ in order to move to Principles 12. Teachers expressed diverse views concerning this flexible boundary. Syl was not concerned:

R: . . . even though the line may be fuzzy, where do you draw the line for kids going from Math 11 to Math 12?

T12C: We encourage them to have a certain per cent - I can’t remember - 65 or something like that. I don’t think we’ve ever checked on it. Uh, but I think within the first three weeks of class, some students will realize whether they’ve misplaced themselves or not. And so they’ll drop the course and replace it with something else. (T12C Individual Interview)

Sam said the following on the issue:

T12D: We try to recommend that students who get As or Bs in Math 11, if they want to pursue a career in the sciences or in math, or if they need math for whatever they want to go into, that they have a minimum of a B, but, there are kids with C and C- that take Math 12 against our recommendation because they feel that they need that course for whatever program they’re going into. But we can only suggest, and the reasoning behind that is there’s usually a drop in mark going from Math 11 into Math 12 on the whole – I’m not sure what the average, maybe five or ten per cent. So if you score 80 per cent in Math 11 chances are you’ll score 70 per cent in Math 12, kind of thing. That’s not set in stone, but it’s
so the kids know what they're getting into. Also, we tell them the amount of time that is required to take Math 12. Like the energy and the time— they have to make sacrifices, or if they're weak to get tutors, so they know what they're up against. 

(T12D Individual Interview)

The same teacher later said:

T12D: I think it’s a good thing that because there are provincial exams, that people can see the marks; maybe teachers are hesitant that weak kids take Math 12, which is unfortunate—an unfortunate side-effect of the provincial exams. I think even if the student is weak but they think the course is interesting, they should be able to take it. The stress with the whole school performance—that it’s ranked, you’re almost hesitant that that kid writes (T12D Individual Interview)

Annie had a different viewpoint regarding the school’s high participation rate in Principles of Mathematics 12:

T12F: We have some regrets that it’s [participation rates] that high, but one of the philosophies of the school is that we are a community school, and anybody who wants to can make a go at a provincial exam course.

My impression was that the other teachers did not have regrets about the high participation rates. But Annie tilts toward the school philosophy with, “I had a kid who
was failing going in, I think he had 45 per cent on his year’s mark. Therefore he was entitled to write, and he passed the provincial – really, completely the other way around, but he passed the provincial!” (Focus Group 2). There are limits, however. Grade 12 teachers pushed the principal to decide that no student will write the examination who would go into it with less than a teacher mark of 40 per cent. The principal indicated that the teachers wanted even more stringent guidelines, but that he was reluctant to even accede to establishing a 40 per cent barrier.

The English teachers’ views triangulate on this:

T12F: Uh . . . my feeling is that we should allow any kid to take any course he or she wishes, if he or she chooses to do that with the almost inevitable result of failure. And generally speaking we have done that here, but we won’t let them write the provincial if they’ve got less than 40 [per cent].

R: Right.

T12F: Going into it. And that makes absolute sense because you’re wasting government money and everything else. But, at the same time, um, I also believe that there should be – some of these students and their parent make foolish decisions. For example, graduation’s important, period. In order to graduate you either have to have either English or Communications, some of our weaker students are not going to pass English 12. I feel that we have the right to impose upon them that they must take Communications.

R: Uh huh.

T12F: Even if their parents object, because we very often know better than they
whether the kid is going to be able to graduate or not.

R: Hmm.

T12F: And at least they can get a graduation with Communications (a basic level course), which they can’t if they fail English 12 and they haven’t taken communications. So that’s about the only condition I would put upon it, is that the school should have the right to insist that the student take Communications, even if the student also chooses to take English 12 with inevitability of failure. (T12F Individual Interview)

Classroom Resources

The grade 12 teachers have two main classroom resources: a workbook, and packages of previous questions from provincial examinations. Syl explains:

T12C: I use past provincial exams, we make up a provincial exam package that we give to the students and it consists of – what I took this year was seven past provincial exams from 2000, 2002, 2003 – the four ones that are in there; I think I held back the August one, mainly to use it as a review.

R: Right.

T12C: What I do is, after I teach a topic, I give them questions to do out of a workbook which I don’t collect because they do have the workbook in class which they do need on an ongoing basis. (T12C Individual Interview).
Aligning Teaching to the Content of Past Examinations

Syl gives an excellent glimpse into his classroom practices as they relate to alignment in this excerpt:

T12C: And then I’ve identified each examination question by topic and by section within each topic, so if there are five radian questions in the entire package I can say, “Here are the questions that you want to do.” And so I just give them that to work on and . . . and it seems to work out pretty well - I think they appreciate it. The main thing we have to do is work with our provincial exam packages, whereas in the past we never had anything. Now we have something that we can look back at and see what the types of questions are. (T12C Individual Interview)

Sam said:

T12D: Yeah, actually [Syl], a colleague of mine, has taken past provincial exams, unit by unit, to kind of put them into packages. So the trigonometry package would have for the last two years every single trigonometry question, and at the beginning of each unit, these packages are handed out to kids, and they [teachers] can use them as assignments or review for the unit test. (T12D Individual Interview)

I was told that students’ work on these questions was checked on occasion, but in a cursory way. Rather, teachers assessed students’ knowledge indirectly during lessons by
placing part of the onus on them by soliciting questions from the students who had had difficulties with the assigned past examination questions.

Syl’s grade 12 lessons did not look like his grade 8 ones. In grade 8 lessons, students received a broader range of activities and problem types. Grade 8 lessons moved at a significantly slower pace. Syl’s classroom contained many resources for teaching and learning, which he used in grade 8 lessons, such as physical models (manipulatives) and mathematics contest problem sets. Grade 8 students also used a standard text for that grade, whereas grade 12 students used a workbook and provincial examination questions. Grade 8 students wrote the low stakes FSA, but no classroom time was directed toward its preparation, and Syl did not present questions from previous administrations of that test.

Syl coached students on time management. During one of his lessons, I saw him present what appeared to be a question from a previous examination, and after a while he informed the students, “You should be budgeting no more than 90 seconds for this type of question.” The teacher’s attitude was that doing certain things will help students do better on the examination, and therefore he would present that information to them.

When providing examples of concepts taught during lessons, teachers would sometimes pose questions from previous provincial examinations. I also saw such questions used as launching points for examination coaching. A good example of this occurred during a trigonometry lesson in which Annie presented an examination-type question on the overhead, and then provided a few minutes for students to find its solution. The four possible answers were written in general form, using radians, and it seemed evident that the examination writers intended that students generate an algebraic
solution. Many students did not yet have a solid footing in this area – it also seemed that the teacher herself was unsteady on this particular question, but this was bypassed when the teacher showed the students how the answer could be found through exhaustion, by substituting the possible answers into the provided formula.

In a related incident, Annie was teaching generalized graphing techniques for trigonometric equations. She presented a provincial examination question that asked students to determine which one of four graphs matched the equation given. As in the previous example, students did not seem to have fluency in radian measures. There were two students who generated the correct graph from the equation (using radian measure on the x-axis). Most of the students seemed comfortable attempting to produce the graph using degree measure, but they did not seem able to connect it to the correct solution. Annie obviated the problem by recalling for students the conversion between degrees and radians, followed by the use of inspection to identify the correct graph. It is unclear how well this method would prepare students for answering such a question if it were not in multiple choice format.

I was in Annie’s class when a student asked about rules for decimals and rounding. I was reluctantly drawn into the discussion when Annie asked me if I knew what the most recent examination guidelines were. I replied that multiple choice answers obviated the question, and that otherwise two rounded decimals was the norm. Annie commented to the class that she would from that point standardize her instruction thusly.

The provincial examination was more explicitly referenced in Annie’s classroom practice than in others. It was common to see her link specific techniques or advice to the provincial tests, and to use the provincial examination as an authority for rationalizing
them. On several occasions she said, to considerable amusement among students, “I’ve been teaching for 120 years and I know what’s on the provincial.”

Sam was teaching Principles of Mathematics 12 for the first time. As such it seemed natural that he relied on two main authorities for determining content and timing of lessons – his colleagues (mainly Syl) and previous provincial examinations. When I watched Sam teaching grade 12, I did not observe any direct references to the provincial examination, yet his teaching, like that of his colleagues, was steeped in content from previous examinations.

The students’ comments triangulate on what I observed in this school and provided evidence in their comments on test coaching:

R: I’m also curious as to the role that the exam might play in day-to-day teaching. In [Annie’s] class when she made reference to the exam. Is that . . .

[Student interjecting] All the time.

R: Is that a regular thing? [all students nod their heads]

S1: They tell you what to look for, and how they allocate marks and stuff.

R: I see. That’s something interesting that I want to talk about, amongst other things, ... so the mark, and you talked the mark breakdowns, and how questions are marked? Any specific or general advice that’s given in regard to writing the exam?

S2: Yeah, give as much steps, show everything, and they’ll give you part marks wherever you can. (Student Focus Group)
Given the degree to which some topics of discussion and instruction are aligned to the structure and content of the examination, the provincial examination itself can be viewed as part of the curriculum – a topic that is embedded in daily instruction alongside the mathematical content of the course. This can be viewed as close alignment between instruction and external testing.

Course Planning

Syl discusses course timing:

R: Can I ask you if you have a particular target date for the completion of the curriculum?

T12C: I usually try to shoot for the middle of May and have never met that. Um, I’m a little ahead this year, but I can’t foresee finishing by the third week of May except that I’m gonna be out of here for two weeks, so that’s gonna put a little damper on it, but it’s also because, uh, we also have an accelerated math class, our Math 8/9, 90, so when they get to grade 10, they’re doing like an honours 11, and so they’re doing Math 12 with the Honours 11 students. And this year I have five or six students from last year who are repeating the course with me, for a number of reasons. Their marks were fairly good, but they just decided to repeat the course so that’s kind of accelerated the other students. They’re a little more conscientious, and a little more serious as a whole and actually I’m happy that my marks are high but it also scares me because it’s not something I’m used to, so I’m able to move a little faster with it too. (T12C Individual Interview)
Near the “end” of the course material, Syl hands out a large review package of past provincial questions.

T12C: I’ll give them – I’ll give them a review package – this year I gave them their package during our educators’ conference so they have two days off plus the weekend. So I said, “Okay, here’s some review questions to work on.” And I thought that worked quite well. (T12C Individual Interview)

Dry runs are part of course planning and standard practice here in grade 12 mathematics. Syl explains:

R: Do you have any dry runs?
T12C: In the past we have had dry runs, on occasions, we have not had them. For example, last year we did not have them because of the time factor. Our grades 12s finish their non-provincial exam subjects by May the 30th. And so what happens is when we get into, for them, regular classes and review days, if you’re going to dry run you’re going to have to do it during one of the regular classes, and because it’s uh, a two-hour provincial exam, it does create conflicts so last year was one of the first times we did not have it [dry run]. (T12C Individual Interview)

As to their daily course planning and timing, the three grade 12 mathematics teachers
perceived that they were almost always within one or two lessons of each other on an agreed-upon roster of topics. For example, Sam said:

R: So, do you get a feel from your other colleagues in terms of their timing, when they like to finish the course, or what kind of review they like to do, and those kinds of things?
T12D: In a way, like I’ve basically try to keep up with them, that’s, that’s how I know that I’m doing okay, right, I check up with them and see where they’re at, and I try to teach the units at the same time the others are, so right now we’re all pretty much at the same area, so hopefully (chuckling) they’re on pace for where they usually are. (T12D Individual Interview).

The three grade 12 teachers would end up “finishing” the curriculum topics around the beginning of June.

**Downward Alignment**

I saw little evidence of downward alignment in this school, although I did not spend enough time in grade 11 classes to be sure.

**Aligning Assessment and Evaluation Practices**

There is strong alignment between the content of past provincial examinations
and the methods used by teachers for assessing and ultimately grading their students by assigning each student with a course mark. Past provincial examinations can be viewed as having become a driving force in this teaching, assessment, and evaluation cycle. There are two subsections here: tests and quizzes, and matching school and examination marks.

Tests and quizzes. The course packages referred to above further serve as the model for classroom quizzes and tests which, as at Greenhill School, are structured proportionally to look like the provincial examination. As well, all the teachers here (and in the other site) weighted tests and quizzes far beyond any other assessment. In this light, the unit tests can be viewed as topic-specific dry runs. Sam outlined the mechanics of his assessments:

T12C: I try to give them two quizzes per chapter, and then I give them a chapter test. That’s my main form of evaluation. On a shorter chapter I might just give them one quiz and one test, so . . . so again, because they’re using their workbook in which to do homework, I don’t collect that as a straight homework assignment. I have collected certain parts of their provincial package but again, it’s a very subjective mark. I just scan through and give them a mark. (T12C Individual Interview)

Sam went on to provide finer detail about his testing practices: “Yeah, every test - [a colleague] made all the tests up that I used and involve approximately 15 multiple choice questions, and two to three written questions” (T12C Individual Interview).
This pattern mirrors the structure of the provincial examination in mathematics. Another grade 12 teacher triangulated this assessment and also underscored the professional co-dependence that I had come to expect between the grade 12 mathematics teachers in this school:

T12D: Yeah, actually [Syl], a colleague of mine, has taken past provincial exams, unit by unit, to kind of put them on to... packages. So the trigonometry package would have for the last two years every single trigonometry question, and at the beginning of each unit, these packages are handed out to kids, and they [teachers] can use them as assignments or review for the unit test. (T12D Individual Interview)

The above quote also demonstrates a connection between assessment and examination coaching because these tests and quizzes are similar in nature to the provincial examination and, therefore, as students write classroom tests and quizzes, they serve to prepare students for the upcoming examination by giving them multiple experiences in a similar setting. Also, the last quote confirms the choice of the grade 12 mathematics teachers as a sub-unit of study (Yin, 1994); classroom practices and relationships are affected when colleagues face the same high stakes test at the same grade level.

**Aligning school marks with examination marks.** At Pine River Secondary, I saw little evidence of any school-level or teacher attention to the aligningment between teacher and provincial examination marks. One grade 12 mathematics teacher said:
T12D: But my own opinion is, I think you want the kids to score similar to what they would score on the provincial exam. Why give them a false hope? They’re getting 90 per cent, they think they know what they’re doing, and then they score 60 per cent – that’s not really fair to the students. But if you make your tests too tough, harder than the provincial exam, then you haven’t been fair to the students either. But I don’t mind a little bit of inflation, I think that it’s nice for a student to have maybe a five per cent cushion. So I think it’s natural for a student’s mark to be higher than their exam mark, because they’re studying unit tests that are fresher on their minds, as opposed to a whole year’s work, I believe. (T12D Individual Interview)

The Ministry of Education does not promote teachers using a marks ‘cushion,’ but it makes it clear in many places on its Web site that teachers can expect differences between school and provincial examination marks, and that one must take into account the local factors that contribute to students’ performances. The above teacher’s comment was the sole specific reference I heard here to the matching of school and examination marks, and that might lead one to infer that school-test matching is not a significant topic at Pine River Secondary, or at least one that has not had any major effects. An examination of recent Ministry of Education examination statistics supports this inference. In two recent consecutive years, the differences between mean school marks and provincial examination marks were ten and nine per cent, respectively at Pine River.
Teacher Summary

The stakes of the provincial examination are high for these teachers, the majority of whose students aspire to post-secondary education. Under time pressure, they have responded with a strenuous effort to get the course done in time and with good examination scores. Their response to time pressure also includes negotiating with the principal over the daily and yearly schedule, as well as in regard to a fair partitioning of instructional time. The provincial examination is a constant presence, strongly influencing course planning, continual references while teaching, and assessment and evaluation. It generates the examination review phase, whose target starting date in this school is the middle of May, which leaves almost six weeks before the upcoming examination. Gatekeeping exists, but only when a student clearly has no hope of passing. Otherwise, it is not enforced. Teachers in general subscribe to the notion that any student who passes the grade 11 course should be able to take the grade 12 course.

Principal’s Views and Practices

Introduction

In my first interview with him, the principal (often spontaneously) discussed many of the issues that I wished to explore related to provincial examinations with both insight and clarity. He called particular attention to how his personal philosophy of education interacts with what he saw to be some of the risks accompanying high-stakes testing. Later interviews and discussions served to confirm and expand on that first
discussion. In all, the principal provided a great deal of data that is now woven through the analysis of this site. Several of his polices and actions clearly were aimed at improving school results on provincial examinations. My subsequent conclusion is that his remarks revealed that examination performance statistics in general, and media rankings in particular, were very high stakes for him. He communicated his desire for good and better scores to his staff. He also provided an insightful look into the power of parents in the independent school system, and how their perceptions of school quality intersect with examination results.

This summary of those interviews and discussions reflects the unique structure of this independent school, and is organized under four headings: (a) The Impact of Statistics and Media Rankings of Examination on School Prestige, Competition Between Schools, and Parental Perceptions; (b) Views and Practices Regarding Gatekeeping; (c) Communicating with Teachers About Examinations; and (d) Actions Aimed at Improving Examination Performance.

**Impact of Media Rankings on School Prestige, Competition Between Schools, and Parental Perceptions**

When asked about the role of published rankings in helping form both internal perceptions of prestige and subsequent parental perceptions and choices, the principal responded as follows:

P2: I'm not sure it comes down to anything as – on the one hand, on the one hand
anything as narrowly specific as a Fraser Institute ranking or as intangible, forget specificity. I believe parents, and probably all of us: when you buy a car, do you believe the advertising – which I call ‘Myth,’ with a capital ‘M’ – or do you thoroughly research what your options are, and dig extensively underneath? That’s the sort of thing. Parents operate on Myth, they believe, and I’ll use a university analogy – our parents being, many of them, relatively less well-educated than their own children, look at the University of British Columbia as the pinnacle – British Columbia – a province, the University of British Columbia must be the place. They then look at the University of Victoria, maybe, and they say, “That’s named after a city, that must be pretty damned good, too.” They look at Simon Fraser and say, “Some historical figure, explorer, wore moccasins and fur outfits – why would I send my kid? And that literally is – it’s a second choice. Simon Fraser is a second choice. People who live on the hill beside Simon Fraser aspire to send their kids to UBC because it’s the place to go. Bring that down a notch - people aspire to send their kids to [prestigious independent school] and [another independent school], then in the secular arena, to [independent school A], [independent school B], and [independent school C] because of the cachet, certainly in the latter three that I mentioned – the British public school type – the pristine uniform, Oxbridge grads, there are headmasters and headmistresses. I’m just a Principal teacher, not a Headmaster. That’s the kind of – it’s a level of Myth I think that people operate on. (P2 Interview)

The above quote alludes to the fact that this school indeed acts as a ‘second
choice' for some parents, a view openly acknowledged in one of the principal's newsletters. This can have long-term effects on teacher behaviour and examination scores. The principal elaborated:

P2: But we also teach everybody's child, which means unlike [independent school B], which allows upwards of 300 kids to take the entrance exam and they take the top 90. And from 91 to 300, we get some of those. So is it surprising, then, that history repeats itself five years down the road and they get better exam results than us? (P2 Interview)

The above excerpt makes explicit the principal's connection of good examination scores with school prestige: his parental 'A' list, those having greater prestige, and, using the principals' phrasing, more compelling 'Myth.' Parents of children in independent schools can ultimately exercise considerable power with their feet, and therefore their perceptions are crucial. This helps explain the rivalry between independent schools. Peter Cowley refers obliquely to this: "The [Fraser] Report Card provides an objective benchmark against which schools can improve. It is also a much-needed tool that parents and students can use to make a more informed choice of an education provider" (2002). Examination performance and parental choice are being portrayed by Cowley as being co-dependent.
Views and Practices Regarding Gatekeeping

This principal does not believe in gatekeeping, and his disdain for schools that do is apparent:

P2: I guess it’s a philosophical stance, and I know there are those who oppose me on staff and are probably dying for my retirement and hoping that someone with a particular viewpoint will come in. Within our narrow system – that’s not even looking at [Independent school A] and [Independent school B] I’m just talking about the school like us: [#3], [#4], [#5] and so on. There are practices like: if you don’t get a B in the grade 11 course, you can’t take it. There are practices like: if at this point in the year you are failing this course – this is in late April – you will not be allowed to write the exam. I wilted to a degree, under pressure, from folks and said, “I am prepared to say anyone maintaining less than 40 per cent in the school mark will not be allowed to write.” So I’ve given in to that pressure. So it’s not quite “You have to be passing,” you have to be within arm’s reach of passing. But even then they – my critics out there – know that I’m a soft touch because if you come in with a parent and sit there and say, “Joey’s had a disastrous year, he knows I’m paying huge amounts to have him tutored just before the last – he knows that he can ace this exam.” Am I going to give in to that pressure? Yes, of course I am. (P2 Interview)
Communicating with Teachers About Examinations

The principal next discussed some of his communications with teachers regarding provincial examinations:

P2: I'm much more interested in the five-year averages. Yes, I asked the department heads and particularly the government exam subject teachers to consider that and are there ways – I like the binders they put out (pointing to what appeared to be a series of Ministry of Education report binders). I'm sure you're familiar with the analysis that the Ministry puts out, the reams of paper that they churn out. And that you can now take off their Web site. ... Both by topics, say within Principles of Math 12 and the skill levels – the knowledge and the higher order thinking (the provincial examination program in Math and Sciences adopts Bloom's Taxonomy), those sorts of things. We did explain those for the benefit of the non-numerate department heads and subject teachers, and we'd say, let's monitor these in fine detail than just averages on a year-by-year basis. And we have now got them into the mode of looking at those when those government exams result come out. (P2 Interview)

He summed up his perception of this external influence by saying, "Do we take account of the numbers and do we acknowledge that they drive us in some sort of fashion? And I think yes they do" (P2 Interview).

The principal intervenes to the extent of urging the teachers of twelfth grade examinable mathematics courses to practice explicit examination coaching, including an
analysis of previous questions and strategies for answering multiple-choice questions. At the same time, he justifies the educational value of multiple choice items by bringing analysis of them back to worthy educational goals. The following excerpt from an interview with him exemplifies that view:

P2: So the fact that there is a right answer and that there are three distractors, that the very terminology used is that there are three ‘distractors,’ um, distractors in the humanities sense could mean something that sounds attractive compared to the question, but isn’t the right answer. The distractor in the mathematical sense is these; if the correct answer to the question is 2, you don’t say, here’s answers of 0, 1, and three as distractors. What you say is, the correct answer is 2; if you make this kind of major mistake, then the answer is seven and three-quarters, this kind of mistake the answer’s 0, this kind of mistake the answer’s minus 1. That’s why those distractors are there. So people need to post-factum analyze, or look at old exam papers and say, what kind of distractors are they putting in, and there again brings us right back to the notion of the cognitive difficulties that kids have.” (P2 Individual Interview)

School-wide Alignments Aimed at Examination Performance

This topic has three subsections: student choice of grade 12 teacher, altered timetable and scheduling, and grade 8 testing placement and honours streaming.
Student choice of grade 12 teacher. The principal has engineered his system so as to maximize the likelihood of matching students’ personalities and learning styles with teachers’ qualities. To that end, grade 11 students get to express their preferences and often choose which of three teachers they will have for the Principles of Mathematics 12 course. He explained:

P2: And even though we don’t publicize the fact ahead of time, they come in in August, or if [“Chris”] has her way they come in very late June, they start making choices. All they see is a matrix that has the subjects. Very occasionally, no, probably half of the time, I will have put the teachers on there so they can say, they can sell you a bill of goods and say, “I’d really like to take this course with this course and I can’t do that so you’ll have to - what the computer has churned out I’m not satisfied with, let me change. But you know what they’re really saying is, “I’ve got teacher X for Math 12 and I really want teacher Y, but I know I need to justify this.” [The] worst-case scenario is, I’ve said, all three of those teachers are going to teach math 12, and when we give kids the freedom, after the computer has created their timetable, to come in and say, “No, I really – I’ve reconsidered; I really need these changes to be made.” Two of those classes end up with 30 and one class ends up with five, then we’ve got a problem. But thank God that has never happened (P2 Individual Interview)

The ability to allow this sort of flexibility is rooted in the size of the school and in the powers of the principal. I am acquainted with schools in which offering such a choice is
rejected out of hand, either on principle or on the perceived chaos that could attend such a move in a larger school. Of course, it is impossible to determine the relationship between such student choices and later test scores, but it is significant that this practice attends to student preferences and initiatives and may even, directly or indirectly, lead to higher test scores than would not doing so.

**Altered scheduling and timetabling.** The school’s timetable is modified as examinations approach. The principal explained in detail:

P2: Those are two review days for 8s through 11s and for the handful of people who aren’t writing any government exams or anything. Provincial exams start there - those two weeks . . . everyone has known for many years now that grade 12s are on a special timetable, so we have [identifying information], Marketing 12, um, P.E. 12, *all of those courses stop* [italics added] and I will have circulated during mid-May, amongst the government exam folk and say, would you like double-time during that ten-day period? And I have to get my time tabling skills hat back on, and double that up. Traditionally math, physics, and chemistry, all but Biology, claim extra time. English doesn’t, lit doesn’t, French does, geography does, usually, history doesn’t. And what that means, as you know we’re on Day 1-Day 2 flip-flop timetable where it works, if you have Math 12 today you won’t have it - today being Wednesday – until Friday. *Now you can have an hour and fifteen minute class every day.*” [italics added] . . . To use an athletics-coaching analogy, that’s acknowledging that you’re two weeks away
from the marathon, so you need to go into hard training. (P2 Individual Interview)

One implication of this policy is that non-critical activities must be set aside, including instruction in non-provincial-examination courses.

On my second day in Pine River, I noted that the daily block schedule was about to be changed. Upon inquiring, I discovered that one of the senior mathematics teachers had initiated this change. Twice during my stay, when classes were interrupted for group photographs, the grade 12 teachers were clearly dismayed. Senior academic teachers in this school are quite sensitive to any intrusion on their instructional time; they insist the burden be shouldered equally among teachers of the examinable courses. The principal explained his involvement in this, in considerable detail.

P2: We know that we are, . . . we have a guest speaker in, and so these particular blocks need to be shortened. The guest speaker comes in, and then usually that means if it’s a morning thing the afternoon will be unchanged. Those can stay at one hour fifteen but these have to be reduced to 55. – and I take a look at it and say, “Golly gee, Physics 12 is there, we can’t touch that, we can’t touch the Math but Marketing 12, non-governmental exam, P.E. 12, non-governmental exam, if there are blocks where those - and to be honest the, English and [identifying course] we’d probably lump into the non-academic, where time isn’t a pressure. So that’s one driving mechanism, the other is, the [indecipherable word here] statistician on the staff, is Chris. He keeps a rigid, precise accounting of how many As, Bs, Cs, Ds, and so on [referring to block periods], he can give you a
cumulative total from September to now, he will also have a notation that will indicate, although we had an ‘F’ period at one point, it was a shortened period. So he can very nearly tell you the minutes breakdown of these things, so, in all seriousness I go to him and say, “Let’s not hit this,” and then he says, “Oh don’t worry about Physics 12; we’ve got – it’s two occurrences ahead of the ‘C’ period, so . . . but even that – it’s not as simple as that because we agreed that of you lost the next period back in September, you’ve adjusted – you’ve probably adjusted by October, and you’re where you want to be. So it’s the most recent times where you’ve had any sort of period adjustments. And then yes, quite definitely, it is the driving factor is what impact that’s going to have on the government examinations. [Another teacher] will tell you that every public school that he’s in contact with has at least 103 hours devoted to Physics 12. And he will point out to me sometime toward the end of the year, like now, that we’re operating at 92, so we are 11 hours short. But then again, that’s me. (P2 Individual Interview)

The above quote serves to show the strength of the principal’s conviction that time must be partitioned and must be seen to be done equitably. The principal constantly balances the competing forces in his school. His philosophy affects instruction, and there is considerable evidence that he stands in the way of what would otherwise be even more profound impacts of high stakes testing. He stated, “as long as I’m in this office, I’m going to say that education is with a capital ‘E’” (P2 Individual Interview). I take that to mean that he reserves the authority to supplant or alter instruction in favour of experiences that he believes appropriately serve the creed of the school.
Grade 8 testing placement and honours streaming. The principal runs a McGraw-Hill standardized placement test for students entering grade 8. Stanines are computed, and based upon those scores and commentary from grade 7 teachers, parents are invited to enroll their children in the Honours mathematics program, the regular program, or in a remedial stream. The Honours program is designed so that students complete three years of mathematics (grades 8, 9, and 10) in two years. In their fourth year of high school (chronologically grade 11) they are eligible for enrollment in Principles of Math 12. Students in the Honors program therefore have the option of writing the grade 12 provincial examination twice. One of the grade 12 teachers shared with me his understanding of the Honours program:

T12D: ... in grade 8 the kids write an entrance-kind-of-test in mathematics and the top students are selected to go into a class we call “Math 89.” Basically it packs all of grade 8 and half of grade 9 into one year – so they basically accelerate it, and the following year they take something called “Math 90,” which combines Math 9 and 10 in one year. So after two years they’ve done three years of math, so when they enter grade 11 when they’re in grade 10, they’ll take something called “Honours 11.” It’s a Math 11 course, but it’s the same group of kids who are taking it.

R: So when those kids get into grade 12, does that mean they’ll have the opportunity of writing the exam again?

T: Absolutely, yeah.
R: I see, so it helps kids meet their own academic needs as well as it boosted performance.

T12D: Yeah, and it gives kids a second chance if they don't succeed the first time on the exam result. It also allows them to take Calculus that year. (T12D Individual Interview)

Whether it was designed for the purpose or not, the Honours program clearly has a beneficial effect on examination performance, because it allows some students to take a second try at the provincial examination. As well, many of the students going through Principles of Mathematics 12 for the second time can also take calculus, a class that focuses on conceptual understandings, and with no provincial examination. The principal clarified my perception of the situation:

R: With regard to those kids who work their way up through the system, it would seem that that would have some bearing on achievement testing and exam performance. Would that be fair to say?

P2: Yes it is. (P2 Interview)

Principal Summary

This principal is guided by broad educational goals, yet he is also influenced by the imperatives and importance of examination scores. He communicates to teachers his desire for good and better scores and provides statistics and guidance to that end. He oversees an Honours program in mathematics that very likely produces increased
examination scores. The principal adjusts the year-end calendar by cancelling non-examination courses in order to increase the time needed to prepare for provincially examinable ones, and fine-tunes the daily schedule in order to fairly distribute the impacts of those adjustments on instructional time in provincially examined courses. As well, he provides a high level of student choices of grade 12 teachers.

**Students' Views**

It is explained above how some students are enabled to write the provincial examination a second time. Below I present an excerpt from the student focus group interview, in which a student comments on that experience:

S: I had a 93 term, but I messed up on the provincial, got an 81/82, something like that.

M: Anything you take away from that experience, coming at it a second time?

S: You have to practice the knowledge from old provincials pretty much. (Student Focus Group 2)

The students reported a lot of that type of practice, as well the primacy of the examination’s role in the classroom. There was considerable pressure on some of these students to do well. One student said, “Yeah. I’ve been accepted to SFU to Computer

---

5 This was a focus group interview of four grade 12 students, nominated by teachers for participation. They comprise a small and not necessarily representative sample.
Science, but I’m not sure if I’m going to go there. It depends if I keep my average up or not, to get a scholarship. Or whether I can play basketball at Cap college.” Another student told me:

M: So how many hours a week do you work?
S: I just got laid off from one of my jobs, but my other job is one or two shifts per week, but it’s an 11 hour shift.
M: So 11 hours or 22 hours kind of thing?
S: Yeah, this week will be about 20 hours.
S: I usually work about 8 hours a week now. I used to have more, but I have to take more time off now, with exams and everything.”

When asked about homework, the students again showed the importance of the examination:

M: . . . how does homework work in grade 12?
S1: I don’t have very much homework.
T: Is it mandatory, number one? (several students replied ‘optional’).
S1: Optional.
S2: Depends on which class.
S2: For math, everything’s pretty much optional.
S1: You have to do the provincial packages, everything else is optional
M: So the exam package is mandatory? [agreement] (Student Focus Group2)

In summary, the students in this school had confidence in their teachers' ability to prepare them for the examination. The students gave evidence that the provincial examination was a major component of their course. They discussed how it was continually referred to in class, and how packages of previous provincial questions were an important factor in classes, including suggested homework practice, examination coaching, and dry runs. The students seemed to handle the pressures on them with aplomb, but identified 'overload' in general as a source of stress, rather than provincial examinations in particular.

Site Summary

The alignment pressures attending Provincial examination are especially high for this highly competitive private school. Such schools occupy nine of the top 10 slots in the most recent Fraser Institute rankings and numbers of students attend this school because they were not accepted by more prestigious independent ones. This puts this school at an immediate and continuing disadvantage in the rankings. Time in general, and academic instructional time in particular, are in short supply in this building, and the mathematics teachers operate with considerably fewer teaching hours than most of their public school colleagues. Moreover, the school schedule is dynamic, and the grade 12 teachers have to adapt to this variability.

The principal plays a key role in this regard. He insists that students have a
spectrum of educational experiences, yet he is also a user and analyst of examination statistics, and communicates his desire for better scores as well as methods for achieving them.

Teachers’ lectures are fast-paced with considerable examination content in lesson examples and examination writing tips, assessment and evaluation (quizzes and tests mirror the provincial examination), course planning, and of course during the weeks left for examination review. Gatekeeping is talked about much more than it is implemented, and ultimately students have the right to take Principles of Mathematics if they meet the specified minimum requirements. The schedule for all grade 12s changes at the end of May, when non-provincial classes are cancelled, with those instructional hours partitioned among teachers who want it. The school runs an Honours program that ultimately allows a set of students to write the examination twice, thereby boosting the school’s performance.

The Impending Grade 10 Provincial Examinations in Mathematics

This section combines the grade 10 examination-related data gathered at both schools. The impending grade 10 examinations are discussed separately because they concern both the principal and the teachers. Moreover, the grade 10 examinations are an emergent phenomenon, worthy of separate treatment.

I did not sense a coordinated response to these examinations from either the principal or grade 10 teachers. The principal at Pine River Secondary said:
P2: . . . specifically in mathematics I am more than a little fearful . . . a little scared that the presence of that exam will be hurtful for the non-regular math students who we have here. So I am um, I am waiting in the weeds, standing on the sidelines, whatever metaphor you want to use. And I think that we’ve just got to play a wait-and-see game. We’re not doing anything very specific about addressing them but [teacher] is supposed to, overnight, be getting us an answer on – well, what information do we have that the Ministry has been remiss in letting us know. (P2 Interview)

One grade 10 mathematics teacher at Pine River also discussed this issue with me:

R: When you learned that [grade 10 provincial exams] were coming in, what did that trigger?

T10A: We knew then that we’d have to finish the whole course by the end of the year . . . was the first thing. Usually we leave out little bits and pieces, and we couldn’t do that this year.

R: Right.

T10A: Because you have to cover everything, because it’s all on the provincial exam.

R: So what kinds of responses did that lead into?

T10A: Quicker pace . . . basically quicker pace, um, a little bit of “If they didn’t get it, it’s too bad.” Unfortunately we had to move on.
R: Right.

T10A: We couldn’t spend a lot of time on one topic. We had to ... keep going.

R: Right. Now I know in grade 12 there’s a method that they use, particularly toward the end of the course, in reviewing and preparing for the exams.

T10A: Right.

R: Now, are you building that process for grade 10?

T10A: It’s right here on the desk! [mutual laughter] What we have done is taken the sample exams off the Internet, and then we cut it all into pieces, and then we can put chunks into chapters. And I just finished liquid-papering out all the answers. So at the end of the year I’m going to give them chapter-by-chapter – that sample without the answers on it, and then have them do the questions and then go over the answers with them.

R: Okay, I see. So did you find any, hmm, like practice tests, so-called dry-runs?

T10A: I probably will do one but I haven’t planned it yet, but I will. I want to get through these first, and then I’ll probably have one or two days left at the end for an overall dry run. (T10A Individual Interview)

These two interchanges foreshadow what the direct alignment effects of the new high stakes test in BC will be. They are likely to include increased pacing, the relinquishing of local control over areas of curricular emphasis and de-emphasis, the introduction of previous examination packages as a primary resource, dry runs, and examination review.

A grade 11 (and earlier grades) teacher at Greenhill connected the impending grade 10 examinations with downward alignment:
R: Are the coming grade 10 exams playing into your thinking or planning at all?
T11A: Uh, you’re taking about teaching more toward the exam, or,
R: Is it affecting . . .
T11A: . . . the way you teach, basically,
R: Yeah, or . . . how, like do you have any proactive kind of thinking in regards to those?
T11A: Uh, . . . We talked about, uh, adjusting the grade 9 course, we actually, all the grade 9 teachers sat down this year and talked about a time-line for the grade 9 course, and uh, talked about emphasizing more time on, say, factoring, or polynomials because that’s such a huge topic in grade 10. So, uh, yeah, it has come to . . . sitting down and talking about what to emphasize because they’re going to be seeing it later on in the final year, and we’re going to be seeing a lot more of it on the final exam.

A grade 10 teacher at Pine River said:

T10A: We knew then that we’d have to finish the whole course by the end of the year . . . was the first thing. Usually we leave out little bits and pieces, and we couldn’t do that this year.
R: Right.
T10A: Because you have to cover everything because it’s all on the provincial exam. (T10A Individual Interview)
The principal of Greenhill anticipates downward alignment:

P1: And now they’re looking at exams, what does that mean, what that’s going to look like, there’s a sense that – you just can’t put in exams and say, well, it’s gonna happen, and our sense is, what do we do to prepare for that? What does it look like in Grade 9 to prepare for grade 10, what does it look like in grade 8? One of things we’re looking at is, is, re-looking at our grade 8 and 9 program, cause we have to get students ready for the, for their graduation program which is now 10, 11, and 12, a year earlier. Well, kids don’t mature – just because you have a program doesn’t mean that they’re maturing any earlier. So what, what are we going to do to help them prepare for that? (P1 Interview)

A grade 12 teacher from Pine River Secondary said:

R: So anything to say about the impending Grade 10 examinations in mathematics regarding how you might be reacting to them or planning for them?

T12C: I have one grade 10 [class] this year and I do find um, of the 20 some-odd students, probably – out of the 25 students I’d say that 20 of them are quite weak. They are a weak group. I have a weak group, whereas [T10A] has a much stronger group, Annie’s group is ... a little bit stronger than mine, but I don’t know how the distribution came about. So, uh, I think if the provincial grade 10 exams were counted this year, I would be really concerned about my students'
marks. Uh, but in light of the fact that they’re not counted, I know it’s going to reflect but, the overall average, I think it’ll be low. Because we don’t have enough other than a sample exam to work with. They’ve changed the curriculum and I find it’s really, really overburdened. Yeah, I know I’ve taught Math 11 in the past and that there’s an entire section on functions and, and algebra which has trickled down into the grade 10, and so I think the grade 10 curriculum is now overloaded.

R: Some educators have given me the opinion that grade 10 students might not be at a level of maturity to respond appropriately to the stakes of the exam. Do you have any comments on that?

T12C: I think that perhaps more reflective of what their abilities are like, I think a conscientious student, an A student, will strive regardless of whether it’s a provincial exam or not provincial exam. A mediocre student might take it a little more lightly. They might for example say well, again, thinking perhaps how they’re thinking, well I’m not very strong in math, and this exam doesn’t count for much, so I’m not even gonna worry about it. And it’s just a cop-out, sort of. But certainly in the future, if it’s going to be counting for their graduation, yeah, I’ll think they’ll have to take it a little more seriously. And just getting back to one of the questions you’ve asked, uh, with regards to will we cut out topics? I can see myself cutting out topics. If, if I’m pressured for a provincial exam.

(T12C Individual Interview)

At both sites, teachers’ reactions to the impending grade 10 provincial examinations proved to be a valuable source of data. Some impacts of grade 10 government examinations were already being implemented. Teachers were now doing things that they
did not do before, and they identify provincial examinations as the reason for the changes.

The pedagogical changes they were making were often similar to well-tested procedures used by teachers of grade 12: increased pacing, the reliance upon previous examination packages, 'dry runs,' the expectation of losing local control over areas of emphasis and de-emphasis in the curriculum, and more tightly focused instruction.

There were differences in the level of the grade 10 examination stakes in both sites. One site was situated in a district where there is a culture of achievement testing, and district-level advance planning was already underway. References to district-level initiatives were common in my interviews. There was close to a consensus regarding the importance of achievement testing results in general, and provincial examination results in particular. That school's goals are overseen by a School Planning Council. Achievement scores are addressed in those goals and that oversight. As the principal said, the achievement scores are "the crux of the whole issue (referring to the role of School Planning Councils)."

The reactions in Pine River Secondary were more at the teacher level, with administration largely waiting to see how grade 10 examinations would settle in and what administrative responses would then be needed.

The differences between the responses at these two sites can be attributed largely to the social and administrative structures in which they are embedded. The independent school principal had discretion over many aspects of school life, and his cautious approach to grade 10 examinations may well be grounded in his belief that it will take considerable time for the grade 10 examinations to settle in before he would need
initiatives at the administrative level. The other principal worked in a district where there is a major emphasis on aligning school operations with achievement indicators, based on “data” generated by testing. Those indicators and the statistics derived from them were central to his thinking and decisions about staffing and program planning. He is addressing closely the likely program changes that will be needed at the eighth and ninth grade levels so that students will be better prepared for the grade 10 examinations.

It is clear, then, that the grade 10 provincial examinations currently have different meaning for these two administrators, but the examinations have the same meaning and probable impact on their grade 10 teachers, whose responses are described above and discussed in Chapter 5 below. Both sites have begun to use their own examination practice packages, dry runs, and the practice questions provided on the Ministry of Education Web site.

In addition to the Ministry of Education Web site, teachers have additional sources of information on the BCAMT listserv. Its postings are dominated by provincial examinations, and the looming grade 10 examinations have generated many questions and much commentary. The postings range from discussions of technical details of examination implementation to requests for information and commentary on the merits of the examination program.
CHAPTER 5 – CONCLUSIONS

This chapter presents the results of this study in four sections: Conclusions, Discussion, Implications, and Suggestions for Further Research.

Conclusions

The conclusions are structured around the research questions:

1. What are mathematics teachers’ views and practices regarding high stakes testing?
2. How are these views and practices mediated by administrators, mathematics department heads, other teachers, students, and external influences?

Overview

Overall, it is clear that the high stakes provincial examinations in British Columbia have been a significant force in shaping the twelfth grade mathematics program in these two schools, and perhaps in many other high schools in the province.

At these two sites, the examinations have influenced such administrative initiatives as program planning and staff assignments, and have encouraged staff members to explore teaching procedures that are believed to produce ‘good and better’ examination results.

From the point of view of teachers and teaching, the data provided in Chapter 4 provides details of how the provincial examinations in Principles of Mathematics 12 have become a theme in course planning, lesson planning and delivery, assessment, and review
activities. There is, in short, an overall alignment of teaching behaviour to test content. The administrators in my study accepted, or were at least resigned to accepting, high stakes testing as an established component of the high school curriculum. They expressed unease with what they saw as possibly destructive effects of high stakes testing, particularly gatekeeping, but I detected little evidence of angst. Both of them were consumers and distributors of statistics related to their schools’ performances and they communicated their desire for good and better scores to teachers in direct and indirect ways. They seemed to be aware of, but rarely went into any depth on, issues of validity – whether or not test scores are valid indicators of student understanding, and whether or not they address the aims of education.

Teachers and administrators believed that provincial examination results had had an impact on school and classroom life, shaping instruction and local assessment. They perceived that the publication of examination results had raised the stakes, on the grounds that rankings may shape public perceptions of schools’ quality. The students I met with most often said that they work hard on their courses, made it clear that they know that their performances on the provincial examinations will have a direct impact on their post-secondary careers, and said that they are undergoing no particular stress on either account. But at the same time, we must not lose sight of the fact that these data are provided to enable deeper insight into the phenomenon of high stakes testing and more subtle ways of looking at them. The intent of this study is to expose, with as much subtlety and detail as possible, how teachers, administrators, and students have coped with high stakes testing. An attempt has been made to report as many as possible of their views and practices as may bear on that objective.
Teachers’ Views and Practices Regarding High Stakes Testing

Introduction

This section addresses the first research question and organizes teachers’ views and practices in the following categories: (a) The testing stakes as perceived by teachers, (b) aligning instruction with examination content, (c) course timing and examination review, and (d) aligning classroom assessments and evaluation to the provincial examination.

The Testing Stakes as Perceived by Teachers

The alignment pressures on these teachers are considerable, although not as great as in jurisdictions where their continuing employment can hinge on students’ test scores. The teachers in my study believed that the public do use achievement indicators from test scores to judge the teachers’ professional competence. They believed that the prestige of a school and the teachers in it can hinge on these scores.

Idiosyncrasies of these sites created somewhat differing pressures on the work of grade 12 teachers. I found, however, many similarities in what happened, particularly when they entered their classrooms and began to work with students. One of their common goals, reiterated by school and district level administrators, was to produce ‘good and better’ examination scores. Probably because of the clarity and unequivocal
nature of that directive, the teachers in these schools seem to have bought into it.

Given that objective, media publications of rankings of schools provide simple numerical indicators of what is then held by some to be 'teacher effectiveness.' Public media rankings are therefore a potent means of creating alignment pressures. Teachers who regularly 'bring home good results' can be rewarded with at least considerable prestige for having done so. At the other end of the continuum, several teachers told me that they did not want to be "near the bottom of the totem pole" (Group 1 Interview).

Some care should be taken not to translate the above analysis into negative commentary on provincial examinations by the teachers of Principles of Mathematics 12 that I observed. It is not the purpose of this study to draw such inferences, and in any case, as previously reported by Anderson et al. (1990), and US Department of Education (1993), the examination program seems generally accepted.

**Aligning Instruction with Examination Content**

Overall, the two groups of mathematics teachers, in quite different administrative and social environments, were very similar in their professional practices. They taught from the same curriculum and prepared their students for the same examination. They had similar training, attended many of the same conferences, posted to the same listserv, marked provincial examinations, and maintained professional relationships with teachers in other schools. Both of these groups of teachers took the examinations as a given, and paid attention to similar teaching strategies that a teacher can employ to maximize the examination performances of their students.
Test coaching, viewed as test-taking information and skills rather than mathematical content, was observed at both sites. For example, both showed methods of parsing and of the 'brute force' approach to answering multiple choice questions when other means fail. It is not surprising that they relied upon similar approaches and methods. As was noted several times in earlier chapters, the provincial examination was a strong undercurrent in many lessons, more visible in some classrooms than in others. Several students told me that the provincial examination was the first thing they were told about in Mathematics 12. I saw a spectrum of alignments. The strongest and narrowest alignments would occur during end-of-course examination review, when previous examination content dominated what was done. Teachers presented what appeared to be intact items from past examinations, and coached their students on time management and test taking strategies. Alignment was equally evident with regard to seatwork and assignments, where packages of past items were the common source of assignments for students across the sites. Alignment was also evident in pedagogy, when teachers taught students tips for quickly producing the types of intermediate results (sub-skills) that are often tested in provincial examinations. Phenomena such as homegeneity of instructional styles and fast pacing were observed, but it cannot be inferred in this study that they were impacts of provincial testing.

Course Timing and Examination Review

I turned to my own experience for guidance in observing and posing questions about these two matters. In my experience, I had typically reserved the last two or three
weeks of a semestered course for review and examination preparation. The grade twelve mathematics teachers at these sites employed a similar practice. Further, I found on the BCAMT listserv that similar practices are widespread among those who post to the listserv. Another teacher on the listserv discussed how she administered and subsequently discussed with her class eight dry runs of previous examinations. That suggests that she devoted a good deal more time than two weeks to provincial examination preparation.

My observations confirmed that teachers at both sites also reserved a significant block of time for reviewing the course and preparing for the provincial examination. A grade 12 science teacher in one of my schools ‘finished’ the curriculum about May 1. Two grade 12 mathematics teachers followed a tight schedule that ‘finished’ the curriculum on the last teaching day before the winter holidays. The examination was scheduled for the end of January. Teachers in the site with less instructional time, where the completion date was about June 1, ‘finish’ near the middle of May.

**Aligning Classroom Assessments and Evaluations to the Provincial Examination**

Teachers in both sites relied upon two classroom resources: the textbook and examination questions in Greenhill School, a workbook and examination questions in Pine River Secondary. There was little formal assessment of assignments at either site. The completion of assignments was voluntary, and the onus was on students to raise any concerns from the previous day’s homework, which was drawn from the above mentioned sources. That practice allowed for some informal assessment.

. At both sites, tests and quizzes are structured and often administered so as to be,
as one teacher put it, "mini provincial exams." These unit tests were considered to be training grounds for the provincial examinations, not only in content and question type (proportioned to match the provincial examination), but in the testing experience. The tests are deliberately designed to mimic the time demands of the upcoming provincial examination.

The dry runs toward the end of the course are intended to polish and complete this training. For local purposes, students at both sites were evaluated principally on tests and quizzes. The tests, in turn, contain large amounts of examination-derived content. This further demonstrates how, directly and indirectly, the provincial examination is the focus of instruction (Anderson et al., 1990; Calder, 1990; Wideen et al., 1997).

As is noted earlier, I did not see homework checked during any of my classroom visits, yet most of the students in both groups of students reported doing it, from little to several hours per night. Forty-five minutes to one hour seemed to be a rough average. In any case, assessment of homework did not play a significant role in teachers' evaluations of students. All teachers assigned practice questions after lectures, drawing from either of the two main resources identified above. Given the importance of the impending examination, and the close linkage between both the contents and the format of the assignments and the tests, it seems that students did not need any further encouragement to attempt the assigned questions.

The following sections – Administrators' Views and Practices, and Students' Views – describe how the contexts, views; and practices of those groups interact with and influence the mathematics teachers, and therefore address the second research question.
Administrators' Views and Practices

At both sites, provincial examinations play a significant role in administrative initiatives and communications with teachers, many of them stemming from the phenomena identified below.

First, internal and external perceptions of the quality of these, and likely many other schools, are based, to a considerable degree, on in-school, district, and media reports of examination performances, categorized by school. Both principals saw a connection between those media reports and subsequent judgments of the quality of their schools, their administrators, and their teachers. And as those judgments become known, they create competitive atmospheres, particularly in the independent school, where parents are, by anecdotal internal and external accounts, influenced by them in their choices of schools. Both of the principals attested to the significant role of test results in generating competition in their systems, whether in a public school district or a group of independent schools drawing from the same community. They both submitted that they were driven, to a degree, by those scores.

Both principals have implemented and overseen creative initiatives and policies that address concern for students and their examination performances. There were broad similarities, but many of the initiatives and policies that I was informed of or observed directly were adjusted to take the principals' philosophical stances and their schools' milieus into account.

Both principals saw the need for 'good and better' test scores, identified specific teaching methods and coaching techniques that they believed could enhance those scores,
and discussed them with their teachers. Both saw virtues in increasing students’ contact time in Principles of Mathematics 12. In one case, that additional contact time was provided by enabling students to write the provincial examination twice, and in the other, by offering a version of Mathematics 12 in which marginal and weaker students could take the otherwise one-semester course over two full semesters.

Both principals expressed a distaste for gatekeeping – the practice of discouraging or preventing weaker students from enrolling in grade 12 examined courses, not just in mathematics. They both felt that they had been under some pressure to consider that practice. Both of them had obviously devoted considerable thought to gatekeeping, because the longest and most complex answers from to any of my questions from both of them were addressed to gatekeeping and what they saw to be its short and long-term deleterious effects. Both principals were unequivocal in their belief that even minimally qualified students ought to be permitted to enroll in Principles of Mathematics 12.

Both mentioned, one with disdain, specific schools where gatekeeping is an overtly established practice. By way of triangulation of their remarks, I can add that while triangulating my inferences in other schools, I was in a principal’s office in such a school during a principal/parent meeting and witnessed a student being denied the opportunity to write two provincial examinations on the grounds that the student had school grades between 45 and 50 per cent.

Regarding one independent school mentioned by one of these principals, I called the school in question and asked about entrance procedures and requirements. The secretary explained that there was a competitive entrance examination, and that they did not have enough places for all applicants.
A final matter that may be of some interest is conspicuous in not being addressed here. Teachers differ in their preferences and aptitudes for teaching courses at differing levels, with differing objectives, and for differing students. Long before we had high-stakes testing, any principals who did not take such matters into account in making staff assignments would be remiss. But questions concerning the degree to which high-stakes provincial examinations may have accentuated such sorting of teachers are not so easily answered as may first seem to be the case. It is a sensitive and far from simple issue and while it was occasionally referred to in passing, for good reason, neither the administrators nor the teachers in these schools volunteered any significant information concerning the assignment of teachers to courses.

**Students' Views**

The students at the two sites reported that their lessons, assignments, and local tests were well-referenced to the content and structure of the provincial examinations. For example, they reported that tips for solving particular kinds of problems were a staple in lessons; one student in a group interview qualified that it happened, “All the time.”

On account of considerable folklore and anecdotal reports, I had made an exploration of student stress a component of this study. What is remarkable is that the students in the sites I visited did not seem to exhibit any symptoms of undue examination-related stress. They so uniformly elided any such references in my open-ended interviews that I took the risk of addressing the topic directly, knowing how easy it is to suggest what I might expect to hear. But with even that opportunity, they generally
said that while they worked hard at their courses, they did not have inordinate amounts of assigned homework and suffered no particular stress on account of either homework or the provincial tests.

Both groups of students appeared to take pragmatic stances concerning their responsibilities. They lumped local and provincial examinations in with a multitude of other demands on them. They considered themselves burdened, but not stressed. One student captured well the spirit of my discussions with them:

S: I don’t think it’s just provincials, I think it’s other stuff too, ‘cause you have to meet up with requirements with university, and college, and then also, you see we have a lot going on. I think a lot of people don’t realize, but a lot of us have part-time jobs like with me. Like, I’m graduating early ‘cause I need the money for college ‘cause I don’t know if my dad would pay for it or not.

Grade 10 Provincial Examinations

This section separately deals with the impact of the impending (as of April 2004) tenth grade provincial examinations. The British Columbia Ministry of Education had recently announced what amounts to an extension of the high-stakes testing program to tenth grade. The administrators and teachers at my sites reacted somewhat differently to that announcement. In both schools, teachers have begun to do some things differently, and they identified those impending examinations as the reason. To the degree to which their alignments ran parallel, their new practices are mainly extensions of what
twelfth grade teachers have already been doing; more narrowly focused instruction, sample test items as part of instruction, fast pacing, and dry runs. Teachers uniformly seemed to foresee a loss of local control over areas of emphasis and de-emphasis in the curriculum. Otherwise, I found marked differences. Greenhill is in a district where there is a culture of achievement testing, and district-level planning for administering the tenth grade examinations is well under way. Further, the local School Planning Council, with a major role in shaping the school’s program, is now addressing the goals accompanying these tests. Previous research identifies district-level coordination as a key factor in student achievement gains (Williams, Kirst, Woody, & Levin, 2005). In Pine River Secondary, to date the reactions have been mostly at the teacher level. The principal there seems to have decided that it is prudent to wait until the new testing protocol ‘settles in,’ and then decide what administrative responses may be needed.

That major difference between the alignments at these two sites is likely attributable to their differing social and administrative contexts. The Pine River principal has considerable discretion over most aspects of his school’s operation and can react to new contingencies on short notice. His cautious reception of the grade ten examinations may well be grounded in his belief that since it is not certain how they will settle in, he will be able to adjust to them when the time is ripe.

The Greenhill School principal works in a quite different context. In his district, achievement indicators are based on ‘data,’ to a large degree generated by testing, and he has little difficulty anticipating the implications of these new tests for staffing and program planning. He is already addressing the program changes that will be needed at the eighth and ninth grade levels so as to best prepare students for the tenth grade
examinations. This is also direct evidence of downward alignment from the administrative level.

So, while the arrival of tenth grade tests have, to date, differing implications for the administrators involved, they have the same meaning and impact on their teachers. Teachers at both sites have begun to extend to tenth grade the common twelfth grade examination-preparation practices referred to throughout Chapter 4, supplemented by the practice questions now provided by the Ministry of Education and the mushrooming pool of resources, procedures, practice questions, and commentary on the BCAMT listserv.

Implications

Implications for Teachers

From my observations, these grade 12 mathematics teachers and teachers at other grade levels have accommodated to high stakes testing in their planning, course scheduling, teaching styles, and assessments. These data are apparent in the interviews and observations undertaken at their schools as well as on the BCAMT listserv, where teachers continually share their views and practices with their colleagues. I have indentified the following direct implications for teachers: instruction, staffing, and teachers new to the profession.
Implications for Instruction

There are likely many grade 12 mathematics teachers who want better class examination results. Given that there was strong similarity in the approaches used by teachers at both research sites, teachers who want improved class results are encouraged to examine the classroom practices described in Chapter 4.

We saw in Chapter 4, particularly at Greenhill School, that the quality of teachers' classroom assessments were judged in part by comparing classroom marks and provincial examinations scores. Insofar as teachers are concerned about such judgments from school and school district administrators, teachers should use material from past examinations in order help produce a closer correspondence between their own assessments and those on external examinations. Doing so seems like a potent and defensible approach to gauging the difficulty of one's evaluations, and for 'staying on track.'

The grade 12 teachers in both research sites planned for dry runs. Students should be familiar enough with the testing conditions – perhaps with a dry run – to minimize lost performance due to unfamiliarity with the protocols of the testing instrument (Koretz et al., 1996), thereby increasing the validity of the test results. This therefore implies that teachers should pace their instruction in order to leave enough time for dry runs and subsequent discussion.

Teachers in both sites gave students basic information about the examination marking process, so as to minimize unnecessary loss of marks and to increase the likelihood of part marks being awarded. Other teachers are encouraged to provide their
students with this knowledge as well.

**Implications for Administrators**

It is clear from the data in Chapter 4 that these schools have dedicated principals who are aware of the relationship between published rankings and perceptions of their competency by colleagues, parents and higher authorities. The need for good and better examination results continually affects their thinking, their planning, their communications with teachers, and their decision-making.

At the same time, administrators have found some conflict between the imperatives of high-stakes testing and their professional and personal values. Pointing to a Fraser Institute ranking or the high school’s average examination mark in a subject is much easier than defending the students’ poor performance on provincial examinations with a philosophical argument – what some refer to as ‘talking a good job.’ Yet, sometimes in different ways, and possibly at some personal risk, they each stand in the way of what they think could be destructive impacts of high stakes testing. Principals can consider the following implications that emerge from this study: Since there is no easy way to resolve the angst that can attend the tension between the imperatives of high-stakes testing and either professional or personal values, principals need to develop transparent and defensible strategies for managing this tension or dilemma. The dedicated principals at these sites have coped with that burden in pragmatic ways; in principle, bending to the above imperatives, but drawing a line in the sand where they see a risk of them doing significant damage to students. Higher level administrators and members of
school boards could be urged to recognize the stress that the above tension can place on principals, and try to find ways of mitigating it.

**Implications for Students**

While some bravado may have masked students' reporting of stress attributable to provincial tests, this study at least suggests that any such effect is not conspicuous among my participants. My results suggest that it may be best to wait for further studies, to determine the degree to which other factors like financial exigencies and increasing competition for university entrance are interwoven with provincial tests as a source of whatever evidence of stress is found.

Students at both sites reported the obvious method of preparing for provincial examinations; namely, by doing questions from past examinations. For students interested in doing well, this seems like an important technique.

**Discussion**

This section discusses issues raised by the study, namely: the stakes, comparisons of my results with those of previous research, 'secure' examination items, staffing issues, and the contributions my study makes to the literature on the multi-faceted relationships between high-stakes testing and teaching practices.
The Stakes

In both sites, I explored the sources of official and lived alignment pressures (stakes) and upon whom they were operating. I posit the following relationships among them as an overall description of the how alignment pressures were experienced in these schools.

All stakeholders face testing stakes in BC, but it is the students’ accomplishments that directly engage the stakes. That means teachers’ academic and personal concern for students and how teachers do their job are inextricably linked to each other and to teachers’ personal and professional interests in avoiding the possibly negative impacts of high stakes and pursuing the positive ones. This is why the combination of high stakes for both students and teachers can be a potent force. Teachers carry the additional personal and professional burden of ameliorating the consequences for their students, who, along with their parents, are counting on the results. Parental perceptions and power further raise the stakes for principals and teachers, because for some students, a small difference in performance affects their choice of university, or even whether or not they graduate.

Both principals use numerical statistics in their conversations regarding examination performance, and those statistics become a form of educational currency. Teachers feel the alignment pressures from being potentially viewed as less than competent, or highly competent, by those who officially evaluate their work. The Ministry of Education provides principals with classroom-level statistics. Because students can be so strongly affected, the grade 12 teacher now faces the potential added pressure of judgments from department heads and colleagues within the school and
school district regarding perceived teaching quality and ability to help the 'team.' Those at higher levels of administration can further focus and amplify these effects, and adding public media rankings to the mix further focuses systemic goals and ramps up the stakes for teachers and principals on account of a single rank number being used to help determine widespread public opinions.

Comparing these Results with those of Previous Research

Strength of the Overall Alignment Between Testing and Teaching

This empirical study concludes that there is a close alignment between high stakes testing and teachers' practices.

Earlier, I identified some studies in which the testing stakes were not explored in significant depth. One feature of my research is examination of the effects of high stakes testing on teachers, which are considerable. In BC, neither teachers’ salaries nor official evaluations are linked directly to test results, but many other factors influencing teaching practices have been identified in this study. The system-wide impact of the testing consequences in BC compares to most if not all of those reported in the research reviewed in Chapter 2. The teachers and students in my study were hard working and focused, and during my visits they rarely strayed beyond the curricular objectives, commonly determined by the test. The provincial examination is a powerful motivator to cover quite specific content and to do so competently.

At both sites, teachers' concern for students and for fairness permeated many of
their remarks. One quote in particular, seen earlier and worth repeating, connects teaching methods to students’ needs:

T12C: What I do is after I teach a topic I give them questions to do out of a workbook which I don’t collect because they do have the workbook in class which they do need on an ongoing basis. And then I’ve identified each exam question by topic and by section within each topic so if there are five radian questions in the entire package I can say, “Here are the questions that you want to do.” And so I just give them that to work on and... and it seems to work out pretty well – I think they appreciate it. (T12C Individual Interview).

In other words, this teacher thinks that the students appreciate his efforts in understanding the stakes, and taking action.

**Instructional Alignment**

The results of this study confirm others that claim there is a significant alignment relationship between testing and teaching content and, like them, does not suggest a strong relationship between testing and instructional styles (Corbett & Wilson, 1991; Grant, 2000, 2001; Smith et al., 1989). The teachers in this study revealed that the provincial examination has a strong influence on their work. All of the alignment effects identified in previous research, except cheating, were observed to varying degrees in my study. The content of lessons, materials, and classroom tests clearly reflect a degree of
influence by the provincial examination.

This results of this study also confirm those of Siskin (1994), who argued that departmental colleagues are a major source of mathematics teachers' pedagogy. This seems evident from my data, which indicates that department members shared resources and, to the limits their environments allowed, grade 12 teachers moved lock-step through the course. The sequencing of topics, pace, and examination review seem to have been negotiated among the grade 12 teachers. We saw how a teacher new to the grade 12 course was grateful for the solid support and direction he received from a senior colleague. This further supports the identification of grade 12 mathematics teachers as a subunit of analysis.

Unlike some prior studies, I had no intentions in this study of making judgments about the merits or validity of particular instructional or systemic alignments. The current round of provincial examinations in BC is 23 years old. It is all that most teachers have known; provincial examinations are solidly entrenched and are expanding.

Previous research, such as Anderson et al. (1990), found that high stakes testing caused teachers and students to work harder, with more focus, and with a greater sense of urgency. These effects, viewed positively by teachers, have been found in jurisdictions such as BC that use high stakes tests that are external, curriculum-based, and required for graduation (Bishop, 1999). My study could easily be interpreted to add to that corpus, not surprisingly since Bishop (1999) identified BC provincial examinations as having an overall positive impact on the quality of education in that province. The US Department of Education (1993) came to similar conclusions. My study could also be interpreted as coming to negative conclusions concerning the impact of high stakes tests on teaching
practices, along the lines of Wideen et al. (1997). The reason my study is open to differing interpretations is that the data is based primarily upon a combination of self-reported interview data along with my observations of teaching practice and I was not making any judgments of merit regarding these teaching practices. Grade 12 teachers can compare their classroom practices with those analyzed here, and use this research as a basis for making inquiries about their own practices.

Comparisons to Anderson et al. (1990) and Wideen et al. (1997).

Because the two major studies of the relationship between provincial high stakes tests and teaching practices in British Columbia were reported by Anderson et al. (1990), and Wideen et al. (1997), it is important to compare and contrast my results with their findings. Anderson’s study was intended to be a broad survey of the province-wide impact of provincial examinations. For that reason, it certainly has a broader ambit than the present study, albeit in less depth. Their overall conclusion was that when they conducted their study (1990), the twelfth grade provincial examinations had become the focus of instruction at that level. My findings concur with theirs, and with the benefit of having spent long enough at each site to visit each grade 12 teacher’s classroom several times and interview each teacher several times.

The strongest concurrence between the Anderson et al. study and mine is to found in our observations of the impact of provincial examinations on the received curriculum. A teacher in their study said:
The provincial examination separates the curriculum a la curriculum guide from curriculum a la exam specs. Despite the fact that the Ministry says you can cover the other areas that are not examinable, that does NOT happen. The exam specs are the curriculum. It restricts what you teach because you must teach to the exam. (1990, p. 168).

This quote continues to resonate with what I observed in my study. For example, a teacher in my study said: “I’ve identified each exam question by topic and by section within each topic so if there are five radian questions in the entire package I can say, ‘Here are the questions that you want to do’” (T12C Individual Interview). This teacher’s viewpoint exhibits how the curriculum has been shaped by the content of the provincial examination.

Anderson et al. (1990) gave four recommendations for further research in this area. They suggested investigations of the stress on students and teachers, the extent to which there is a differential functioning between examinable and non-examinable courses and the ensuing effects, the extent to which the provincial examination program has affected the content of courses and the testing practices of the grades preceding grade twelve, and the extent to which tutorial assistance is provided to students, both within schools and commercially. It can be seen that I have addressed a number of their suggestions in chapters one and four. As noted above, however, in one respect my findings differ from their expectation. I found that the students at these sites feel relatively confident about taking provincial examinations and reported no significant feelings of stress associated with them.
Wideen et al. (1997) collected data that was similar to my own, but from a different perspective and with a different intent. They aimed to judge teaching against what were held to be standards for ideal science instruction. The most significant overlaps between that study and mine are in comparisons of classroom observations of student work and teaching styles, the role of past examination content, and time pressures. The authors reported negative impacts on science teaching because, among other reasons, lecturing dominated instruction to the exclusion of other types of teaching styles such as laboratory work. Their conclusion that this had a negative impact on instruction was grounded in how they defined ‘best practice,’ a set of ideals regarding the teaching of science. They claimed these ideals are widely acceptable to the science education community. A different stance is that viewpoints, circumstances, and stakes often play a role in what ‘best practice’ means in a classroom. Competitive university entrance, school exit, and teachers,’ principals,’ and schools’ reputations as well as examination scores can be considered as well, and therefore even the ‘best’ teachers are at least pragmatic about examination results, especially when time is already in short supply. Concern for the best interests of students can lead to the decision that not preparing them as best as the teachers can for the exam could be a disservice to students. Best practice can therefore arguably include examination-preparation teaching behaviours. Both principals declared that teachers need these attitudes and skills as a prerequisite for teaching the examinable courses. Every teacher I saw utilized one or more such techniques. My participating teachers all mentioned that Principles of Mathematics 12 has a very full curriculum. I perceived a constant sense of hurrying during grade 12 lessons. All of the teachers talked about it. Matters outside the
examination content were not addressed because there was no time to do so. A grade 10 teacher, who was in the process of adjusting to the arrival of the government examinations in that grade, mentioned how pacing must increase in order to cover the entire curriculum to the same degree as previous to the provincial exams. The department head at Greenhill School spoke of the motivation that some students got from hearing that if they failed their current grade 10 course, then the next time around they would be facing a government examination at the end of it.

**Assessment**

The ability of many grade 12 teachers to produce close correspondence between their classroom marks and those from the examination was noted above. Using past examinations provides an effective way for teachers to maximize that correlation. At Greenhill School, this matching was mandated by the school board office. At that site, teachers' abilities to assess students was in turn assessed by their ability to produce this correlation of grades. This practice further reinforces the use of past examination material in classroom assessments.

In Pine River Secondary there was no such mandate, but one teacher mentioned being told that as much as a five per cent 'cushion' was suitable. Indeed, their provincial examination results revealed a weaker correlation with teachers marks than at Greenhill.

Participation in the marking of provincial examinations would seem to increase the ability of classroom teachers to match their classroom assessments because the grade letter cut-offs are determined collectively by the teachers who mark the examination.
These deliberations and decisions are shared openly, and it allows a large number of teachers to get a sense of where the average is, giving them the chance to adjust their own evaluation criteria. Several teachers raised the issue that excessively easy evaluation by classroom teachers is not fair to students because it may give them false hopes and it is also unfair to students from schools where teachers' evaluations are in synch with those from provincial examinations. Overly harsh evaluation, commented upon by Leanne at Greenhill School and by Sam at Pine River, is also unfair because it means that other students who may be doing no better will get higher marks and therefore have a better chance for university entrance or even, in some cases, high school graduation.

Administration

Several previous studies of testing impacts have addressed the role of administrators in mediating school-level and classroom-level responses (e.g. DeMoss, 2002), but not in a detailed and triangulated way. Much of the research does little beyond asserting that principals’ directives and information have an effect on teachers. My study was not designed to assess how effective such communications are, but it describes how they worked in these two schools. Unlike what was reported in some previous literature (e.g., Smith et al., 1989), such communications in these schools respected and were within the professional roles of teachers - no teacher was directed to do anything specific regarding provincial examination scores. The principal at Pine River took a stance that there are things to be learned from indicators of achievement, even from multiple choice instruments, and he placed the matter of assessment on the agendas of staff meetings and
relayed his statistical analyses to his teachers.

Overall, this study represents a significant attempt to explore the type of complexity asked for by Cimbricz (2002), that is, "studies that provide a richer, more in-depth understanding of the relationship between state-mandated testing and teaching in actual school settings" (Conclusion section).

Students

The testing consequences for students are more direct and well defined than for other stakeholders. Some earlier publications refer to the stress that high stakes testing can place on students. For example, Anderson et al. (1990) made reference to the "stress, pressure, and panic related to life in grade twelve with provincial examinations" (p. 63). An interviewee in that study even linked a student's suicide attempt to stress associated with the provincial examinations. Wideen et al. (1997) discussed how stressed students can and do apply direct pressure on teachers who, in their view, are drifting away from teaching them about the ways the examination will assess their knowledge of the curriculum. Eisenhardt (1989) wrote, "In replication logic, cases that confirm previous relationships enhance confidence in the validity of the relationship. Cases which disconfirm the relationships can often provide an opportunity to refine and expand the theory" (p. 544). We face here an instance of what Eisenhardt calls a disconfirming case. Both lore and some previous observations by reputable and careful researchers suggest that students may experience and report stress when faced with provincial examinations, yet students who participated in the focus group interviews at these two sites explicitly
dismissed that notion.

One reason for limited student stress is that, at these sites, both groups of students expressed confidence in their teachers' abilities to guide them through the examination process, perhaps related to the fact that both of the schools had provincial examination results better than the provincial average. But it can be argued that the apparent contradiction between Anderson et al. (1990) and my study is not a meaningful source of disagreement. Anderson et al. aimed to make general statements about province-wide views of students. My study aimed to discuss the views of the students in these sites. It is difficult to compare our conclusions because they are derived from quite different methodologies and potentially different samples of students. While the prospect of writing provincial tests might generate stress in many students, such feelings might not be generated given the atmospheres that teachers had created in these schools.

Finally, with regard to principals' stances with respect to student stress, both principals expressed unease with gatekeeping and other possibly destructive derivatives of high-stakes testing, and clearly communicated those concerns to those around them. It is possible that the atmospheres of concern for students' well-being thereby served to mitigate student stress.

In light of my research, I conclude that, in these schools, grade 12 students see the provincial examinations as being just another component of the educational process, and no more stress-inducing than any other teenage rite of passage.
"Secure" Items

One of my claims addresses the importance of past examination content in classroom assessment. This connects to a recent development from the Ministry of Education that underscores my comments about past examination material. It is planned that, in the future, provincial examination questions will be ‘secure,’ the main intent being to put barriers in place that will limit the degree to which specific examination questions will published, intentionally made available, or discussed with students. As well, the Ministry indicates that it is becoming increasingly difficult to create new questions – an indirect reference to the common phenomenon of students being taught to recognize certain ‘types’ of questions and remember how to solve them. It is further suggested that by using a recurring set of ‘secure’ questions, it will be possible to employ “sophisticated statistical techniques” (such as Item Response Theory) that can compare the relative difficulties of examinations and students’ performances on them.

There are five sessions of provincial examinations in the academic year, in January, April, June, August, and November. Starting in 2005, only one or two of those examinations or sample sets will be posted to the Internet. It will be possible, however, for teachers to view other examination questions using what is called the “principal’s secure set.” Teachers will be told not to discuss specific examination questions with students, a considerable change in practice. For example, it is now a common practice for students and teachers, following a provincial examination, to huddle and discuss the examination using the resource copy of the examination given to teachers. Such huddles would now be disallowed, not only because of the ban on discussion of specific
questions, but because there would be no examination available to discuss. This news was received with some concern on the listserv, with teachers’ comments typically falling into two categories: first, that many teachers viewed the post-examination huddles with students to be a powerful and valuable experience for students and teachers; second, a number of teachers pointed out a few obvious methods whereby examination ‘security’ could be easily circumvented. One posting mentioned mathematics teachers who might have eidetic memories. More cogently, one mathematics teacher wrote that some students had already posted questions from the examination to the Internet. Moreover, I have pointed out how hundreds of questions from previous provincial examinations have been the knowledge base for preparing for upcoming tests. Without such a resource base, many teachers who be looking elsewhere for resources. The listserv shows the keen interest in the availability of these resources and practice examinations. With illicit material potentially available on the Internet or soon to be, anyone with even the most basic Web searching skills can download it from the Internet. One listserv teacher raised the ethical question of what to do when a student asks for help with a question that s/he says is a past examination question. This hypothetical situation is a reflection of the fallout created by the American ‘Alfie Kohn Group,’ which protests high stakes testing by openly publishing the contents of American standardized high stakes tests and disrupting the administration of some tests.

Whether they are motivated by success or by protest, this suggests that students who are technologically competent or are privy to Web sites will have an advantage until such time as most people avail themselves of this knowledge base.

This information will also be quickly available to teachers who choose to look for
it. Depending on the ethical decisions that teachers make, these new provisions have the potential to create inequity in the system. In the past, teachers all worked with the same knowledge base. Apart from the few teachers involved with creating the examination, there was no underground knowledge base. With the upcoming changes, some teachers may decide to give their students this knowledge, some might choose not to. This would imbalance the ‘level playing field’ and add another dimension to whatever “best practice” might mean.

Staffing Issues

I have cited the largely anecdotal observation that the ‘strongest’ teachers now tend to be assigned to teach examined courses (e.g., Anderson et al. (1990)). It might benefit all teachers to confront that image with a close analogy from medical practice. It seems obvious that in large hospitals nurses are commonly sorted and sort themselves according to their competencies and preferences for categories of patients. Those whose taste is for comforting might end up in maternity and geriatrics. Those who prefer the tight precision and protocol of more mechanistic work might end up in the operating room. And some prefer the logistics and paper flow associated with more administrative roles. No one group of them are seen to be, or should be seen as, the ‘best’; they simply have different roles.

Rather than resent differential assignments, teachers might benefit from considering the notion that teaching examined courses is not the pinnacle of the profession – that the ‘best’ teachers in a school are to be found in all roles and at all grade levels. Those otherwise inclined can focus on students’ development, understandings,
connections to other fields, and the like.

Contributions of this Study

This study aimed to help fill a significant gap in the literature. Its overall contribution is that the research literature now has an empirical case study of the relationships between a high stakes, curricular-aligned, external exit test and the views and practices of high school mathematics teachers. I suggest the following as specific contributions to the literature:

1. This study explores the complexity of how high stakes testing is reflected in classroom and school-wide practice.
2. This study links the emergence of a new grade 10 moderate stakes test to descriptions of institutional and pedagogical changes that are accompanying it.
3. Anderson et al. (1990) called for research to investigate the extent to which there is differential functioning of the examinable and non-examinable courses, and the ensuing effects. I found evidence for this. In the sharpest example, at Pine River Secondary, non-examined grade 12 courses simply stopped at the end of May. Rather than viewing non-examined courses as unimportant, the faculty in that school were acknowledging the imperatives created by the stakes attached to the examinations. At Greenhill School, an alternate structuring of the examined course was designed and offered to students, in part to increase examination performance.
4. Anderson et al. (1990) also called for studies concerning the extent to which the provincial examination has affected the content of courses and the testing practices of the
grades preceding grade 12, or downward alignment. In the public school there was strong evidence of school-level and classroom-level alignment from provincial examinations. The impending arrival of provincial examinations in grade 10 caused the public school principal to wonder what things were going to look like as they helped students in grades 8 and 9 get ready for those examinations.

3. Finally, Anderson et al. (1990) called for study of the stress on students and teachers that is associated with the provincial examination program. My study does not seem to support earlier studies in this area, and raises questions about the sources and qualities of student stress. It was agreed among principals and teachers that teaching examined courses has higher visibility and more stress for teachers; some teachers pointed to severe time pressure, but none of them complained specifically about stress from the provincial examinations.

Suggestions for Future Research

I have identified the following areas for further research: the alignment model, dry runs and coaching, student stress, provincial examination marking sessions, pedagogy, competition among schools, parents, a broader range of schools, downward alignment, administrative behaviour, and gatekeeping.

Alignment Model

One value of the tripartheid model with which this study begins is that it suggests examinations of vectors of influence that have often been elided. For example, consider the influence of teachers' assessments on high stakes assessments. It is natural to notice
teachers’ tests when they are strikingly similar to the provincial examinations. Because they teach the course, mark the examinations, have a heavy hand in item creation and field testing of items, and dominate the committees that form curricular standards, BC teachers are at the centre of the alignment model. Teachers’ views, then, must have some influence on the shape of provincial examinations. It could even be that teachers have shaped the examination as much as it has shaped their teaching and assessments.

There is agreement in the literature that across subject area, grade level, and testing stakes, teachers have said that examination marking sessions are of great value; some teachers view it as some of the best and most potent professional development they have encountered (e.g., Anderson et al., 1990). The marking sessions are a crucible for revealing and discussing content area knowledge, pedagogical content knowledge and pedagogy in general, refining the meaning of curricular goals, the logical connection between the IRP and the examination, standards in marking, and examination preparation. As such, a study of marking sessions could be a rich source of data for exploring many of the the issues that are identified in the present study. Any of the six channels of curricular alignment could be studied at a marking session.

At the examination marking sessions, teachers are divided into groups in which they discuss, develop and test the marking key, and then mark a single question. One possible study in this area would comprise a case study of this marking process. Teachers would be interviewed individually and collectively regarding their training in mathematics and their beliefs about how it should be taught and tested, as well as their perceptions about the sources that have influenced them in that regard. Discussions at the marking table would be recorded and analyzed in order to infer how teachers’ knowledge
and beliefs are influenced by the marking process, and the degree to which those beliefs and knowledge influence the process itself.

**Dry Runs and Coaching**

There is evidence that a low level of examination coaching can increase validity, because it minimizes the risk of weaker performances caused by lack of familiarity with incidental aspects of the test, such as bubble sheets, question format, and so on. (Koretz, 2005, 2005). Another threat to validity concerns students who do not take a test seriously. I find this interesting because a great deal of research in high stakes testing addresses what is assumed to be student achievement, and much of that research uses scores on low stakes tests for ‘measures.’ I think it is obvious that students will not take a test seriously unless it has some stakes attached to it, external or otherwise. In Chapter 4 we saw how the administrators at Greenhill School tried to improve school results by creating an atmosphere of seriousness surrounding the administration of the FSA, which has no stakes for students.

Future research should examine whether or not low stakes tests indeed underestimate student achievement. I suggest the following experiment: A few weeks before the next FSA in mathematics at, say, the grade 8 level, one third of the grade 8 students are randomly chosen and given a few sessions of pizza, coaching, and cheerleading from the principal for the the FSA. The second third of the grade 8 students would be simply be told that their teachers are very interested in how well they do, and that it is hoped that they will try as hard as they can. The final third of the students would serve as control group and given no special instruction. If differential effects are seen,
they might support my theory that very low stakes (for students) assessments under-evaluate their achievement.

Dry run examinations are modelled as closely as possible after past provincial examinations in format, content, item distribution, and timing, and it is often recommended that they be written where the provincial examination will be written. It would be interesting to know how prevalent dry runs are, and equally interesting to assemble the views of teachers and administrators on their use.

**Student Stress**

As is suggested above, there is a need for a close look at the net effects of financial exigency, increasing competition for university entrance, and provincial tests on student stress. The apparent current increases in these three factors make this an apposite time for such a study.

**Competition among Schools**

Both independent school administrators at Pine River Secondary unpacked the relationships between rankings and parental perceptions of school prestige, and the subsequent choice of independent school for their children’s education. They and most of the teachers there perceived intense competition with other independent schools. The public school teachers identified parallel high stakes and keen competition with other high schools within the same school district. These aspects of high-stakes testing in these
schools may or may not be generalizable, but they are worthy of attention. Independent schools occupy the vast majority of the top 10 slots on recent Fraser Institute rankings. What methods, outside of instruction, do they use to accomplish this? Do such schools find that the results of high stakes testing have created or accentuated competition between them and other independent schools? Are public high schools more concerned about the district average performance or the provincial average performance?

Parents

When parents consider enrolling their children in independent schools, those parents are said by one principal to be strongly influenced by past students’ scores on high stakes tests. Are they?

Broader Range of Schools

As seen in Chapter 3, the criteria for the selection of sites for this study could not possibly come close to enabling broad naturalistic generalization. But selecting schools so as allow that kind of generalization would have, literally, turned it into a million-dollar project. That limitation precluded paying attention to some possibly significant qualities of other schools. For example, what is the experience of schools in low income communities? What is the experience of schools with students at the extremes of academic ability? What is the experience of schools in which, for any reasons, the administration is less concerned with pedagogical matters than is the case for the schools
in this study?

**Downward Alignment**

This study began with the intent of focusing on the impact of high stakes provincial tests on mathematics teachers and students and relevant administrators.

It was not long before teachers at one site, both directly and by inference, inserted comments about the degree to which the twelfth grade tests were affecting teachers and students in earlier grades. Accumulating sufficient data to delineate just what and how strong all those influences are would constitute a study in itself, and so far as I know there has been no in-depth study of this phenomenon. Since the results of such a study could have major implications for high-stakes testing programs, this must be considered to be a high priority topic for further research. Questions about alignment are likely to take on a higher profile in the near future. The imminent tenth grade tests are likely to exert influences on eighth and ninth grades similar to those that were, until recently, observed only in tenth and eleventh grade. This, again, provides a rare opportunity for a researcher to 'be there at the beginning.'

**Administrative Behaviour**

**Teaching Assignments**

Several times in earlier chapters I noted the common belief that teachers are assigned to examinable courses, if on no other grounds, at least on their ability to
generate high examination scores. The Center on Educational Progress (2005) refers to this reputed practice in their case studies. At Greenhill School, the principal demurred on this issue, but he did indicate that he and parents are "very aware" of who is teaching which course. From my own background in teaching, I have found that administrators and particularly teachers are hesitant to discuss the comparative skills of colleagues. I therefore decided not to pursue this line of questioning. The above quotes contain covert meanings, however. The Department Head referred to her grade 12 teachers as 'specialists' – the "best at doing what they need to do." This implies that teaching grade 12 mathematics is a distinctly different job than teaching other grades. Anderson et al. (1990) wondered whether or not teaching assignments to courses depend on the presence of a provincial examination. A future researcher may be able to find a way to obtain direct evidence of this phenomenon.

**Gatekeeping**

And as with the preceding question, it is likely on account of the sensitivity of gatekeeping, in this case with parents, that it is difficult to establish just how common the practice is. There are anecdotal references to its use in British Columbia, but attempts to determine more precisely just how prevalent it now is are complicated by the fact that there are many varieties of it, and that, near the boundary between gatekeeping and having 'high standards,' it is difficult to draw a line between them. Future research could shed light on gatekeeping and how it manifests in schools.
DISSERTATION SUMMARY

In summary, the dissertation concludes that the provincial examination is a significant force in shaping the classroom practices of teachers of grade 12 mathematics courses in this study. These teachers use the provincial examination as a source of ‘important’ course content, a generator of classroom quizzes and tests, a yardstick for their own evaluations of students, and as a source of out-of-class practice items for students. Teachers are affected by the views and practices of administrators and high levels of administration, who have communicated their desire for better scores and who have implemented strategies for improving examination results. The educators in these schools view media ranking as a significant but invalid source of testing stakes. Students do not appear to be under significant stress related to the examinations.
REFERENCES


http://www.centerforpubliceducation.org/site/c.kjJXJ5MPIwF/b.1536671/k.9B6A
/Research_review_Effects_of_highstakes_testing_on_instruction.htm#q1a

*Education Policy Analysis Archives, 10*(2). Retrieved April 10, 2006, from
http://epaa.asu.edu/epaa/v10n2.html


In G. Kulm & S. Malcom (Eds.), *Science Assessment in the Service of Reform.* 
Washington, DC: American Association for the Advancement of Science.

*Advances in program evaluation: Vol. 1. Effects of mandated assessment on teaching* (pp. 7-40). Greenwich, CT: JAI Press Ltd.

http://legacywww.coventry.ac.uk/legacy/chd/research/onthecase.htm


APPENDIX I - ETHICS APPROVAL