

Uniform Examinations in Québec Mathematics Education: An Exploratory Study of the
Perspectives of Secondary Mathematics Teachers in a Secondary School in Montréal

Dan Elbling

A Thesis
in
The Department
of
Education

Presented in Partial Fulfilment of the Requirements
for the Degree of Master of Arts in Educational Studies
at Concordia University
Montréal, Québec, Canada

September, 2018

© Daniel Elbling 2018

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Dan Elbling

Entitled: Uniform Examinations in Québec Mathematics Education: An
Exploratory Study of the Perspectives of Secondary Mathematics Teachers
in a Secondary School in Montréal

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Educational Studies)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

Professor Elsa Lo Chair

Professor Ailie Cleghorn Examiner

Professor Miranda D'Amico Examiner

Professor Arpi Hamalian Supervisor

Approved by: _____
Chair of Department or Graduate Program Director

Dean of Faculty

Date: _____

Abstract

Uniform Examinations in Québec Mathematics Education: An Exploratory Study of the Perspectives of Secondary Mathematics Teachers in a Secondary School in Montréal

Dan Elbling

Although standardized testing is an educational phenomenon which has been extensively studied, many of the existing studies are large in scale, quantitative in nature, and conducted outside of Québec. Few qualitative studies have examined teacher opinions about the impact of standardized testing on mathematics education in Québec secondary schools. The present exploratory, qualitative, phenomenological case study aims to enrich the current understanding of uniform mathematics exams in the Québec context and potentially inform future research. Teacher opinions and experiences were targeted because teachers are familiar with provincial curriculum, educational standards, desired outcomes, and the uniform exams themselves, plus they interact directly with the students writing the exams. Semi-structured interviews were conducted with seven secondary school mathematics teachers in a single Montréal-area school. Data were analyzed for emerging themes, and results suggest that secondary school math teachers find various aspects of uniform math exams problematic, including multiple-choice questions, marking of short-answer questions, and the overall validity and reliability of exams. Furthermore, results suggest that uniform math exams in Québec have some impact on instruction and learning, specifically in terms of course pacing, the use of old uniform exams as practice, increased student stress, and potentially effects of teaching and/or learning “to the test”, all of which are consistent with findings in related literature.

Keywords: qualitative, phenomenological, case study, standardized testing, uniform examinations, mathematics, secondary school, teachers, Québec

Acknowledgements

I would like to thank the members of my thesis committee, professors Apri Hamalian, Ailie Cleghorn, and Miranda D'Amico. Without you, I surely would not have been able to complete this study. Thank you Arpi for asking the right questions at the ideal times, for pointing me in the correct directions with the most appropriate references, and for giving me a little nudge and then always leaving me the time and space to figure out the rest on my own. Thank you Ailie for introducing me to the educational studies program, for showing me how to write a research paper, and for teaching me that “qualitative” isn't a bad word... even for a “math person”. Thank you Miranda for sharing your expertise. Though I never got to take a class with you, your suggestions made my work more accurate, more complete, and simply better.

I would like to thank the administration at the school where I conducted the study. Your instantaneous acceptance of the project was a huge logistical help and a kind vote of confidence. Thank you to all of my colleagues who participated in the study. I always appreciate your advice, opinions, and expertise. Taking time out of your busy schedules to share and record them, though, went well above and beyond the call of duty and I appreciate that more than you know.

Last but not least, I would like to thank my family. To my sons, Stevie and Jack, you are constant sources of joy and inspiration. To my wife Orit, thank you for always supporting me and for pulling more than your share of the weight whenever I was working on an assignment throughout this degree. Your constant love and unwavering belief in me make me better.

Table of Contents

Chapter 1: Introduction, Research Questions, and Literature Review	1
Introduction	1
Statement of the Problem	2
Research Questions	2
Literature Review	3
A brief history of standardized testing (in Québec and around the world)	3
Fundamental issues with standardized testing	6
Accountability	6
Predicting success in post-secondary education	7
Not actually measuring student proficiency	8
How standardized testing affects teaching and learning	9
Policy	9
Practice	10
Secondary effects of standardized testing	13
Standardized testing in Québec	14
Uniform mathematics examinations	15
Viewing background research-based effects in the Québec context	18
Background research on teacher research	19
Chapter 2: Study Design and Methodology	21
Study Design	21
Setting and Context	24
Methodology	26
Emergent Themes	30
Credibility and Trustworthiness	30
Chapter 3: Presentation of Data	35
Plan for Data Presentation	35
Alice’s Interview Data Table and Summary	35
Summary of Alice’s interview	39
Exams themselves (format, content, and nature)	40
Instruction	40

Student learning	41
Other systemic issues	42
Bart’s Interview Data Table and Summary	42
Summary of Bart’s interview	46
Exams themselves (format, content, and nature)	46
Instruction	47
Student learning	47
Other systemic issues	48
Cleo’s Interview Data Table and Summary	48
Summary of Cleo’s interview	53
Exams themselves (format, content, and nature)	53
Instruction	54
Student learning	55
Daphne’s Interview Data Table and Summary	55
Summary of Daphne’s interview	58
Exams themselves (format, content, and nature)	59
Instruction	59
Student learning	60
Edna’s Interview Data Table and Summary	61
Summary of Edna’s interview	65
Exams themselves (format, content, and nature)	66
Instruction	66
Student learning	67
Other systemic issues	67
Flora’s Interview Data Table and Summary	68
Summary of Flora’s interview	76
Exams themselves (format, content, and nature)	76
Instruction	77
Student learning	77
Other systemic issues	78
Gidget’s Interview Data Table and Summary	78

Summary of Gidget’s interview	89
Exams themselves (format, content, and nature)	90
Instruction	91
Student learning	92
Other systemic issues	92
Researcher’s Interview Data Table	93
Chapter 4: Discussion and Conclusion	97
Discussion	97
Exams themselves (format, content & nature)	97
Multiple-choice questions	97
Short-answer questions	98
Long-answer questions	98
Existence of a uniform exam	99
Validity/reliability of the exam	99
The repetitive nature of questions	100
Moderation	101
Instructional issues	102
Using old exams to prepare for the exam	102
Timing throughout the year	102
Teaching specific test-preparatory skills	103
Teaching throughout the year	104
Learning issues	104
Student stress	104
Students not learning effectively	105
Developmental capacity of students	106
Systemic issues other than the exam	106
Contextualizing Findings in Literature	108
Fundamental issues with standardized testing	108
How standardized testing affects teaching and learning	109
Secondary effects of standardized testing	110
Possible issues unique to the Québec context	110

Strengths, Limitations, and Methodological Observations	112
Answers to Research Questions	115
Pedagogical Implications & Recommendations for Future Research	117
Conclusion	119
References	120
Appendices	124

List of Tables

Table 1: Emergent themes and sub-themes, associated codes, and frequency	32
Table 2: Interview responses contributed by Alice	35
Table 3: Interview responses contributed by Bart	42
Table 4: Interview responses contributed by Cleo	48
Table 5: Interview responses contributed by Daphne	55
Table 6: Interview responses contributed by Edna	61
Table 7: Interview responses contributed by Flora	68
Table 8: Interview responses contributed by Gidget	78
Table 9: Researcher self-interview results	93

List of Figures

Figure 1: Average June 2013 uniform math exam results by academic institution..... 25

Figure 2: Success rates on the June 2013 uniform math exams by academic institution ... 26

Chapter 1

Introduction, Research Questions, and Literature Review

Introduction

Standardized testing is generally understood to be the practice of using the same set of questions to assess an entire group of students, scoring the questions in a uniform manner, and using the results to compare the performance of individual students or sub-groups of test-takers. Typically, the questions are meant to be objective, and they are meant to be evaluated in an objective manner (Kohn, 2000).

Standardized testing has become a controversial topic in education as stakeholders are “polarized as to whether [it] is an appropriate way of evaluating students and the overall quality and effectiveness of the education system” (Després, Kuhn, Ngirumpatse, & Parent, 2013, p. 5). Consequently, it is the subject of numerous studies. Many of these studies are consistent in their findings that standardized testing has an effect on the content and nature of instruction (Abrams, Pedulla, & Madaus, 2003; Anderson, 2012; Cohen, 2015; Després et al., 2013; Graue & Johnson, 2011; Jennings & Bearak, 2014; McNeil, 2000), although those effects are varied, they have been studied in different educational settings, and there is no absolute consensus on whether these effects are to the overall benefit or detriment of the education system. Fundamental benefits of standardized testing, such as improving accountability and accurately assessing student proficiency, have not been convincingly documented, yet the practice continues in many educational settings around the world, including Québec, where successful completion of uniform examinations is required for secondary school graduation. Although the “uniform examinations” in Québec are not entirely composed of objective questions, they are composed of the same set of questions, they assess an entire grade level of students in the province, are scored in a relatively uniform manner, and their results are used to compare the performance of individual students or sub-groups of test-takers. Given that they meet many of the criteria to be considered “standardized”, they will be considered as standardized tests for the purpose of this study. A more thorough analysis and complete justification is provided under the heading ‘Standardized Testing in Québec’ in the literature review section of this chapter.

Statement of the Problem

Standardized testing may not be the most appropriate method to evaluate student proficiency or the overall quality of the education system in general, nor may it be the most effective assessment method to promote the philosophical tenets of the Québec Education Program (QEP). It is also possible that practical consequences of standardized testing may have direct and undesired impacts on instruction, especially in content-based courses like mathematics. Secondary school teachers are uniquely positioned to provide insight into the effectiveness of standardized testing as well as some of its direct and/or indirect consequences. They are familiar with provincial curriculum, educational standards, desired outcomes, and the standardized tests themselves. Those who teach at the grade levels in which standardized tests are used to evaluate students' proficiency are able to observe effects on students, and they have unparalleled insight into effects of standardized testing on instruction. Though many studies have examined the educational impact of standardized testing, few qualitative studies have collected teacher opinions about the impacts of standardized testing in Québec secondary schools and analyzed them to look for emerging patterns.

Consequently, this thesis examines a small sample of Québec secondary school teachers to gain insight into their perspectives and practices surrounding standardized testing. For the purposes of this study, mathematics was selected because it is a content-based course, and, as will be seen in the literature review section of this chapter, some of the effects of standardized testing on instruction are more likely to occur, or are enhanced, in content-based courses. Furthermore, mathematics is the topic that I teach, which gives me a greater understanding of potential curricular topics, test issues, or other specifics that might arise through interviews. The present study aims to investigate what secondary school mathematics teachers view as the impacts that standardized tests have on mathematics instruction and learning.

Research Questions

In pursuit of the above-mentioned research goal, the following research questions will be addressed:

- 1) How do secondary school mathematics teachers in Québec view provincial uniform leaving exams in mathematics?

- 2) What impact, if any, does the existence of uniform examinations have on Québec secondary mathematics instruction, according to secondary school math teachers?

Literature Review

A brief history of standardized testing (in Québec and around the world).

Standardized testing in Western education has a long history, dating back to the mid-1800s in England. As state-funding for education was growing in popularity, officials sought to outline factors by which funding should be allotted. They determined that “the amount of public money paid to each elementary school” should depend on the condition of the school building, student attendance, and “the performance of the students attending the school in an oral examination” administered by a national school inspector (William, 2010, p. 108). William goes on to explain that in Europe, education for adolescents, who previously all entered the work force, was limited to students who were preparing for higher education, and thus “assessment was largely, if not entirely, based on achievement on timed, constructed-response, achievement tests based on carefully prescribed syllabuses” (2010, p. 109). The purpose of these tests was not to achieve a minimum standard of proficiency by a large number of students, but rather to identify the highest achieving students in order to provide access to higher education to a select few. “In contrast”, he writes, “in the United States, there appears to have been a widespread consensus about the value of education up to the age of 18 as a general preparation for adulthood” (William, 2010, p. 109). This meant that school was intended for all adolescents, and thus “it would have been inappropriate to assess students against standards intended for the small proportion going on to higher education” (p. 109). Rather, standardized tests in the United States had been and continue to be generally based on standards that should be achieved by all students.

Selecting students for a limited number of positions in higher education and establishing a minimum level of proficiency for completion of a given level are but two of several potential goals of standardized testing. In general, standardized tests are employed “in the name of improving educational quality” (McNeil, 2000 p. 4). In the United States, The No Child Left Behind (NCLB) Act of 2001 required states to test students yearly, and some of the goals of these tests were to “monitor student proficiency”, “track teacher qualifications”, and “essentially ensure that all schools were performing at appropriate levels” (Scogin, Kruger, Jekkals, & Steinfeldt, 2017, p. 40). Indeed, one of the desired effects of NCLB was “holding schools and

school personnel more accountable for their professional practice” (McNeil, 2000 p. 4). Testing is state-legislated and not all states use their test results the same way, but many use test results “to make highly consequential decisions at both the school and student levels” (Abrams et al., 2003, p. 19), a practice referred to as ‘high-stakes testing’.

Standardized tests are used extensively in the United States, but not exclusively. Phelps (2000) reviews international policy documents and educational research from 1974-1999 and finds the incidence of standardized testing to be increasing globally in this period. In his review, he also finds several potential reasons for countries, states, or provinces to use standardized testing. Among them, he lists addressing “concern about eroding standards”, assessing the competence of students applying for “places in selective secondary – or higher education – level programs or schools”, standardizing the minimum level of proficiency needed to obtain credentials (such as a secondary school diploma of completion), for policymakers to “compare their system performance”, and generally a “desire of [...] policymakers for more information” (Phelps, 2000, p. 19). He also acknowledges that variations or combinations of these reasons or entirely different ones might also exist in various settings.

Phelps notes in his literature review that Canada also makes extensive use of standardized testing, with all ten provinces using some form (2000, p. 13-14). In the mid 1990s, the Ontario Royal Commission on Learning noted the same trend that Phelps discovered, which was that “standardized testing was re-emerging as a prevalent feature of educational systems across Canada and around the world” (Després, et al., 2013, p. 6). Phelps’ results show that different provinces test students at different ages, that the primary focus of the tests differ by province (exit exams, assessment, or what he calls ‘other’), and that the impact on students (he classifies low, medium, and high-stakes) varied by province as well. Because of this, it is difficult to pinpoint a single rationale for the use of standardized tests in Canada, but Després et al. enumerate some general reasons. They write that “standardized testing is premised on the idea that measuring successes and shortcomings will result in greater accountability and drive continual improvement amongst teachers and students” (Després et al., 2013, p. 8). The authors also highlight that standardized test results “are typically intended to communicate standards, focus instruction, provide feedback on curriculum strengths and weaknesses, and motivate educators and students to improve their performance”, to “identify the strengths and weaknesses of the education system” as a whole, and “to inform and develop education policies and to

allocate resources” (Després et al., 2013, p. 8). Furthermore, the report noted that standardized tests were used because they are seen to be “objective and fair, and therefore an accurate measure of what students have learned” (Després et al., 2013, p. 6).

In Québec, the major document outlining educational policy is the Québec Education Plan (QEP). The QEP makes reference to a specific taskforce report which helped to lay “the foundations of educational policy” surrounding the reformed Québec curriculum (Ministère de l’Éducation, 2001, p. 1). The curriculum reform originally had in mind several goals, among which was to review “what the secondary school diploma must encompass to warrant admission to a college and guarantee a minimum of success” (Inchauspé, 1997, p. 18). In essence, the government taskforce wanted to raise the minimum standard of skills and abilities that a graduating secondary school student should possess.

The taskforce report echoes the trend seen by Phelps and Després et al. in which standardized testing gained popularity in the period of 1974-1999. Inchauspé writes that “before 1980, evaluation was not a major component of the curriculum in Québec” and that according to his findings, “it was not considered to be an integral part of teaching, and received little, if any, attention in teacher training programs” in Québec (1997, p. 86). In his report, Inchauspé lists several purposes of using standardized testing in Québec, including “to support progress in student learning”, “to establish a summary of progress and the extent to which learning has been mastered”, “to regulate the system”, and “to report to the population and provide information on the state of education” (1997, p. 98), which are very similar to the goals enumerated by Phelps. In addition to these desired goals, the Québec government taskforce listed additional logistical benefits of using standardized tests. Inchauspé writes that standardized tests “are inexpensive to administer”, “they can be administered to large numbers of students in similar conditions”, “they are reliable”, and “their results are comparable” (1997, p. 88).

The Québec taskforce also recognizes some potential problems with standardized testing, noting that frequently, standardized tests “put too much emphasis on the most easily measured elements of learning, such as knowledge and understanding, at the expense of more complex intellectual mechanisms such as problem-solving, analysis, synthesis and oral or written expression” (Inchauspé, 1997, p. 88). The taskforce suggests moving towards “open-ended questions” meaning “questions requiring answers that include argumentation, free responses,

responses illustrated by data, oral responses, practical activities, and portfolios” (Inchauspé, 1997, p. 89).

Despite the seemingly well-meaning goals of standardized testing, there are a host of theoretical and practical problems with their use aside from those highlighted in the Québec taskforce. The background research that follows aims to provide an overview of some of the most pressing dangers of standardized testing and is followed by a brief analysis which examines which potential problems might apply to the education system in Québec.

Fundamental issues with standardized testing.

The problems with standardized tests include direct and indirect issues, and are as diverse as the very reasons to use standardized testing in the first place. Some of the fundamental problems with standardized testing have to do with the possibility that they may be inefficient or entirely ineffective methods of reaching the original goals they were supposed to meet.

Accountability.

One of the basic goals of standardized testing is to increase the accountability of various players in the education system. Theoretically, standardized tests are an objective benchmark to ensure that schools and teachers are following the prescribed curriculum, and that students have met or exceeded the minimum standards for progressing to the next level or obtaining some credential. One problem with this has to do with the theoretical and perceived objectivity of standardized tests. Graue and Johnson (2011) write that “much of the faith the public puts in [standardized tests] is based on [their] presumed objective nature” (p. 1832). However, they go on to explain that “the process of constructing the system [of accountability], including setting standards and choosing proficiency levels, is a very human process, filled with political negotiations and technical discussions” (Graue & Johnson, 2011, p. 1832-3). There is no perfect test to assess a given level of proficiency, certainly not when the subject itself is complex and requires complicated cognitive processes on behalf of students. That means that the “objective” test to assess proficiency is actually a human-made construct, reflecting peoples’ biases, opinions, and beliefs.

On another note, William describes, in his research, “in-school” and “between-schools” variability in standardized test results (2010). He explains that in-school variability in test results is due to instructional changes, teacher changes, or student changes. Between-schools variability

is due to the quality of the schools themselves and systemic issues like school funding and average socioeconomic status of schools' student populations. He explains that in his research, "differences between schools account for only a small proportion of the variance in student scores (in most countries, less than 10%)" (William, 2010, p. 120). Consequently, this between-schools variability "provides an upper bound on the proportion of variance that is attributable to school quality" (William, 2010, p. 111). The reason that between-schools variability provides an upper bound for school quality is that between-schools variability is due to a multitude of factors, including but not limited to school quality. Among the other factors, as mentioned above, is the average socioeconomic status of the student population, which is known to have a relatively large impact on standardized test results. Koretz (2003) explains it by saying that "non-educational factors, such as student background, predict much of the variance in performance among schools, and typical test score databases are ill-suited to disentangling their effects from the effects of educational quality" (p. 20). William posits that if the total variability attributable to between-school effects is less than 10%, then certainly school quality on its own accounts for less than 10% of standardized test variance. Given the limited test score variance for which they account, William contests that "standardized tests are rather inappropriate tools with which to hold districts, schools, and teachers accountable" (2010, p. 120).

Whether it is a question of their true objectivity, or how closely they actually represent school quality and improvements therein, it seems that standardized tests are imperfect accountability tools. Darling-Hammond (2004) writes that "ultimately, accountability is not only about measuring student learning but actually improving it" (p. 1078). As a result, he says, "genuine accountability involves supporting changes in teaching and schooling that can heighten the probability that students meet standards" (Darling-Hammond, 2004, p. 1078). Standardized tests provide standards which are negotiated through political and bureaucratic methods and they provide means of measuring proficiency, though it seems that they may not do an effective job of measuring school quality. Furthermore, standardized tests do not directly provide support to help schools or teachers improve the quality of education, which is another fundamental shortcoming.

Predicting success in post-secondary education.

Another goal of secondary school standardized tests is to assess students' ability for the purpose of determining eligibility for higher education. However, several studies suggest that standardized tests are not, in fact, the best predictors of post-secondary success (Geiser &

Santelices, 2007; Geiser & Studley, 2002; Hoffman & Lowitzki, 2005; Rothstein, 2004). High school grades (excluding standardized test results) are much more strongly correlated with post-secondary school success than standardized tests (Geiser & Santelices, 2007; Geiser & Studley, 2002; Hoffman & Lowitzki, 2005), and while adding standardized tests to the predictive model offers an increase in predictive power, the increase is minimal (Geiser & Santelices, 2007).

This makes sense if you consider the perspective of Després et al., which is that “large-scale testing is unlikely to be a more fair and accurate representation of student learning than the best judgment of the well-trained teacher-assessor” (2013, p. 12). While high school grades can reflect students’ proficiency, abilities, and growth, standardized tests only “reflect student performance in a single, three-hour sitting” (Geiser & Santelices, 2007, p. 26). It is no wonder then, that high school grades are a better predictor of post-secondary success than a single snapshot assessment.

The idea that there is a single, well-defined skill-set required for post-secondary success is an overly simplistic one, and even if it were the case, there are a number of other factors that might prevent a student with that skill-set from being successful in a post-secondary setting. With that in mind, it is unlikely that any single assessment test will be able to accurately predict post-secondary success. Even if it were the case that a well-defined skill set predicted post-secondary academic success, the fact that standardized tests are not the best predictor means that they do not measure this skill-set in the most effective manner possible. More likely, though, is the possibility that there is a complex, shifting set of skills or proficiency required for post-secondary success and that these skills are only partially related to what standardized tests are measuring. This could mean that standardized tests do a *good* job of measuring *some* of the skills in the set, they do a *fair* job of measuring *many* of the skills in the set, or some combination of both. Essentially, whatever it is that standardized tests do measure, their poor predictive power of post-secondary academic success means that they are simply not measuring *all* of the right things *well*.

Not actually measuring student proficiency.

In an attempt to shed light on what exactly standardized tests *are* measuring, Amrein and Berliner (2002) analyzed standardized test results from 18 different states from the time they were introduced in each state until the time their research paper was written. They found that as time went on, standardized test results generally increased, and that where the stakes were

higher, scores on standardized tests increased more. Interestingly, they found that in no cases (high or low-stakes) did the increase in test scores correlate with increases in other related, but unofficial measures (Amrein & Berliner, 2002). In other words, results increased on the tests that counted, but not on other similar tests which did not count. Similarly, Graue and Johnson (2011) write that “increases in test scores follow repeated use of an accountability test” but that “this increase disappears if a new tool is used” (p. 1833). If the increase in standardized test results did indeed signal an increase in student proficiency or skill level, related test scores should increase as well. Amrein and Berliner (2002) tested four different but related tests and found no statistically significant increase in any of their results. The researchers acknowledge that part of the increase in standardized test results could be the result of lower-achieving students dropping out of school, but that a large part of the increase is likely due to specific preparation for those standardized tests on the part of students and teachers. “Specific preparation for standardized tests” is a deceptively complex idea. It refers to several different but somewhat related practices and has many implications in terms of what content is taught and through what teaching methods the instruction occurs. Many of these practices result in overall ineffectiveness of standardized testing in terms of measuring student proficiency, and consequently, affect accountability and the accuracy of information being reported. The next section looks at several reasons why standardized testing affects teaching and learning, including those that fall under the category of “specific test preparation”.

How standardized testing affects teaching and learning.

Generally speaking, the use of standardized tests has an impact on what is taught and how it is taught. McNeil(2002) , in her analysis of high-stakes standardized testing in the United States, finds that “standardization reduces the quality and quantity of what is taught and learned in schools” (p. 3). There are policy-linked reasons for this, as well as the practical reality based on how schools, districts, students, and other stakeholders react to knowing that a policy of standardized testing is in place.

Policy.

Ellis (2008) finds that “standardized testing of student achievement, particularly in mathematics” leads state education policymakers to “more narrowly define their expectations for what constitutes ‘proficiency’” (p. 1343). Specifically, he cites the federal Secretary of Education

who “signaled her support for aligning ‘proficiency’ on state tests with the ‘basic’ level of achievement on the National Assessment of Educational Progress” (Ellis, 2008, p. 1344). Ellis finds that states lower their expectation to the level of students’ “ability to use algorithms to calculate answers to routine problems” rather than “conceptual understanding and the ability to problem solve” (2008, p. 1344). As a direct result, he finds that “if the goal for low-achieving students is narrowly defined as improved performance on tests of basic computational skills, as opposed to more complex learning outcomes, the sort of instructional practices found to deliver such results will be correspondingly narrowly defined” (Ellis, 2008, p. 1345).

McNeil also finds that “aligning curricula to state tests narrows and trivializes what can be taught” in content-based courses (2000, p. 231). In his study, Anderson generally found that content-course “curriculum had become diluted by the necessity of covering long lists of unrelated topics” (2012, p. 118). The lists of standards associated with standardized tests in his study grew, forcing teachers to cover more content in the same amount of time. This ultimately meant covering those topics in less detail, which threatens to affect the quality of instruction and the level of student understanding. In mathematics in particular, McNeil finds that the prescribed content associated with standardized tests becomes “highly sequential” meaning that skills are taught and tested, and then the teacher is expected to move on. Teachers feel that this undermines their ability to have students “bring a variety of mathematical skills together to work on problems” (McNeil, 2000, p. 241). This lies in stark contrast to one of the fundamental goals of the QEP, which is to give students the “ability to integrate knowledge and transfer it to changing contexts” (Ministère de l’Éducation, 2001, p. 1). In mathematics in particular, McNeil’s findings seem to undermine the goals of the QEP, according to whom students are expected to “mobilize” their knowledge to solve complex, real-world problems called “situational problems” (Ministère de l’Éducation, 2001, p. 185). Aligning classroom instruction to prescribed curriculum standards is part of “specific test preparation”. On its own, this is not necessarily a bad thing. In fact, standardizing what is taught across many schools might be beneficial if the instruction is otherwise highly variable. However, using standardized tests to evaluate these standards has practical impacts which are more ominous.

Practice.

The implementation of standardized testing is a top-down bureaucratic approach in which decisions are made by few policymakers at the top and have broad, wide-reaching effects on

many stakeholders at the bottom. The policy effects of standardized testing occur at the very top of this process, and have extensive trickle-down effects in practice. Generally speaking, “teachers change their instruction as a result of accountability systems, with external tests shaping both the focus and the nature of instruction” (Graue & Johnson, 2011, p. 1831).

In terms of the focus of instruction, the content on “any test is a sample taken from a larger set of information” (Després et al., 2013, p. 12). If it is possible for students and/or teachers to “know ahead of time what part of the domain will be tested”, they will tend to focus on learning that part of the curriculum (Després et al., 2013, p. 12). Jennings and Bearak find that, indeed, “a small fraction of state standards is predictably tested over time” and that teachers “narrow their instruction to focus on these standards” (2014, p. 383). The reality is that standardized tests have a finite number of questions, and thus only a finite proportion of the content/skills can be evaluated. In practice, certain types of questions requiring the mobilization of certain skills are over-represented in tests compared to their relative importance in the list of standards. These skills attract increased focus of teachers compared to others. In their analysis, Jennings and Bearak find that there is a correlation between the percentage of content standards repeated year after year on standardized tests and student success on questions relating to this content (2014). The result of teachers focusing on highly tested skills is the increase in standardized test scores, but not necessarily in actual student proficiency.

This narrowed focus is part of the spectrum of instructional decisions that fall under “specific test preparation”, and is referred, along with other practices as “teaching to the test”. Jennings and Bearak find that this element of “teaching to the test”, focus on a skill or concept not taught before, or an increase in time spent on a skill or concept already taught, comes with a cost. Spending more time on highly tested content leads to “the omission of another skill not included in the standards” or not being tested (Jennings & Bearak, 2014, p. 382). Koretz refers to this process as “reallocation” and notes that not only does it happen as a phenomenon of redistributing time and focus from one within-course topic to another, but it also occurs “between subject areas”, shifting time away from courses without standardized test towards those with them (2003, p. 22).

Aside from the *focus* of instruction, standardized tests also affect the *nature* of instruction. In their study focusing on teachers’ opinions of standardized testing, Abrams et al. find that standardized testing has an effect on the nature of instruction, and it is a negative one.

Teachers “reported that their state testing program [had] led them to teach in ways that contradict their own notions of sound educational practice” (2003, p. 23). Anderson’s subjects also reported that they “no longer teach the way they think is best” (2012, p. 117). This includes “reallocating” time and focus as mentioned above, as well as other instructional choices. McNeil finds that many teachers tend to leave “personal knowledge, or their professional knowledge of their subjects” out of their classroom teaching (2000, p. 14) for fear that it will take time away from focus on content standards, and that it could promote student questions. McNeil finds that teachers tended to avoid student questions, which could shift focus away from test topics, and which teachers fear would show students how much more complex the world is than what is represented in the standardized curriculum (2000, p. 14-15). Focusing on tested content at the exclusion of encouraging students to explore areas of interest or offering enriching educational experiences through connections to the outside world is also part of “specific test preparation”, and severely limits the quality of instruction.

Jennings and Bearak explain yet another way in which standardized testing can affect the nature of instruction. Frequently, standardized test questions are repetitive in nature and can be solved in very similar ways. This “offers opportunities for educators to teach skills in ways that may improve their performance on an item without improving their understanding of a skill or concept” (Jennings & Bearak, 2014, p. 383). This is similar to, but importantly different than, teaching test-taking skills. In this case, Jennings and Bearak describe situations in which the ways that “specific *content* is introduced and taught” have been modified. Teachers would teach the content through the lens of an often repeated type of test question, which detracts from conceptual understanding, and which can reduce a rich and complex topic to a single instance of its potential applications (2014, p. 383).

The above-mentioned test-taking strategies are another potential way that standardized tests can affect the nature of instruction. Teaching test-taking strategies means taking course time not only to frame the content being taught in the context of a test, but also to spend time learning about the format and nature of the test itself with the objective of improving student test results. Cohen explains that teachers in her study

taught test-taking strategies and modeled how to eliminate ‘silly’ answers listed in multiple-choice format. They presented potentially complex information in a series of clear-cut, linear steps, acronyms, and mnemonics for the express purpose of helping

students ‘answer questions quickly’ rather than helping them make sense of academic material. In other words, they were teaching to the test (2015, p. 29)

It is not immediately clear if these findings from Cohen’s work are benefits or harms of standardized testing. These forms of “teaching to the test” seem to help students employ some higher-order thinking skills rather than simple recall. In fact, some of these skills seem to be generally useful ones for students to have, as they seem to promote understanding the context and constraints of a given assessment, developing critical test-taking skills. However, McNeil reminds us of the danger of teaching test-taking skills, which is that it has the potential to hamper students’ learning by teaching them shortcuts, keywords, and strategies which only work for a given test and which cannot be more widely generalized (2000, p. 238).

Secondary effects of standardized testing.

The problems with standardized tests discussed so far have to do with fundamental principles and direct consequences of implementation. While these alone seem serious enough to call into question the practice of standardized testing, there are many secondary effects of standardized testing as well. These effects are indirectly caused by standardized tests or the system to which they belong and they cause significant harm to various actors.

The system of standardized testing is one in which testing and test preparation replace curriculum and subject teaching, which “tends towards minimal standards and minimal effort” (McNeil, 2000, p. 11). When school requires minimal effort, students become less engaged in learning. McNeil adds that “as students disengage from enthusiastic involvement in the learning process, administrators often see the disengagement as a control problem” and they further focus their attention on student and teacher behaviour, further lowering the bar, and strengthening the cycle (2000, p. 11). Students are not the only stakeholders who become disengaged through this process. It is surely frustrating for a teacher to face a classroom of students who have lost interest in the course content forced upon them by the state. Abrams et al. write that based on their study, teachers suffer from general job dissatisfaction and want to transfer out of the grades where standardized testing occur (2003, p. 24).

Even if they are less academically engaged, students writing standardized tests still feel their effects. Students writing standardized tests are more anxious, stressed, and appeared more fatigued than those not writing, and students writing high-stakes tests more so than those writing

low-stakes tests (Abrams et al., 2003, p. 20). As one might imagine, these effects are uneven for students depending on their levels of success. Another study found “lowered self-image and increased anxiety for lower achieving students” compared to higher achieving students (William, 2010, p. 118). In some cases, the unintended outcomes of standardized tests are so strong for low-achieving students, that the tests are seen as an insurmountable barrier and may drive these students out of school (Abrams et al., 2003, p. 21).

These uneven effects are particularly disparaging when considering who they predominantly affect. As seen above, many of the secondary effects of standardized tests have more severe impacts on low-achieving students and standardized test results are known to be correlated with socioeconomic status (Geiser & Santelices, 2007, p. 2). This means that the students who are more likely to be stressed, anxious, have lower self-image, and ultimately, to drop out of school, are students of lower socioeconomic status. Unfortunately, this reinforces social stereotypes, proliferates the systemic cycle of educational inequity, and what might be worse, it all happens under the guise of being “objective”.

McNeil writes that “standardization equates sameness with equity in ways that mask pervasive and continuing inequalities” (2000, p. 10). In addition to increased stress, lowered self-image, and higher drop-out rates, McNeil describes another pervasive and equally disparaging cycle. She explains that in less privileged schools, which are known as “the schools of poor and minority children” (McNeil, 2000, p. 231), test preparation takes almost all instruction time. This puts more focus on content of lower difficulty and does little to develop any higher-order thinking skills. Students in these circumstances will learn less and become less engaged. As mentioned above, this lack of engagement tends to further lower expectations and leads to investing even more time in basic test preparation. These “already marginalized students” are then categorized and labeled based on their results, which are “taken to be objective evaluations” of their ability (Ellis, 2008, p. 1347), in a way justifying their lack of success and the systemic inequity of standardized education.

Standardized testing in Québec.

The findings in the background research above are based on research conducted in different education settings around the world. Whether or not and to what extent these findings are applicable to the system of standardized testing in Québec requires at least a terse analysis. For

successful completion of Québec secondary school, students are required to pass five uniform examinations (in English, French, History, Mathematics, and Science). These exams are sent to schools by the Ministry of Education, schools are required to administer them, and may not alter their content in any way. Usually, the exams are sent to schools very close to the examination date to limit teachers' opportunities to see the exams prior to students writing them. Three of these are written in secondary IV (grade 10) (History, Mathematics, and Science) and two in secondary V (grade 11) (English and French). All of the secondary IV exams are for content-based courses, while both of the secondary V exams are for process courses. All of the exams are required for secondary school completion, although the secondary V exams are written after students have applied to post-secondary educational institutions. This means that the secondary IV exams, in addition to being required for graduation, are also used to determine students' post-secondary acceptances/placements. The tests being studied in this thesis are for the mathematics courses in secondary IV. Further rationale explaining the choice of mathematics over other courses is provided in chapter 2.

Uniform mathematics examinations.

The Ministère de l'Éducation et de l'Enseignement Supérieur offers three different courses in secondary IV and V mathematics, the Cultural Social and Technical Mathematics Option, the Technical and Scientific Mathematics Option, and the Science Mathematics Option. Though the courses contain slightly different mathematical content and they focus on slightly different skills, all three of them have uniform examinations, and all three exams follow the same format.

These exams are end-of-year, cumulative, summative assessments. They are administered every year in June, and students have another opportunity to write an equivalent exam, though different in its content, in late June or early August (depending on the year), and in January the following academic year. Each exam contains six multiple-choice questions, four short-answer questions, and six long-answer or constructed-response questions which are typically contextual and might require the use of multiple skills/concepts from the course. The multiple-choice questions on the exam each have four response options, one of which is correct. For this section, students indicate their responses by shading the appropriate circle on a scan-tron sheet. The response sheet is sent to the government and is corrected there. A correct answer in this section is worth four marks (out of a total of 100) and an incorrect answer is worth zero. The short-answer section is one in which students do not show their work or process, but rather, write their final

answer on the backside of the multiple-choice scan-tron sheet. Classroom teachers grade these responses using a government-provided marking guide. They indicate the assigned grade for each question by shading the appropriate circle next to each student response on the backside of the scan-tron sheet. Most questions only allow for students to get zero marks or four marks. Exceptionally, questions which require specific notation, or those which have two parts to an answer (for example, a question whose answer is a set of coordinates with both an x-coordinate and a y-coordinate), allow for students to get two out of four marks. The long-answer questions are answered by students in the same booklet as where the questions are printed. There is room in the booklet for students to show their work and to explain their thought process. Each question is assessed out of 10 total marks, where four of the marks are attributed to the student's understanding of the problem and which concepts/processes are required to solve it, four marks are attributed to how the student carried out those processes mathematically, and two marks are attributed to the organization, clarity, justification, and correctness in the use of mathematical notation. These questions are corrected by teachers in their respective schools according to rubrics and marking guides provided by the government. The final numerical grades for each question are sent to the government along with multiple-choice and short-answer work, but the student work for this section stays in each respective school.

One important aspect of Québec uniform examinations that makes them standardized-test-like is that the government may adjust the results. This consists of two separate processes: moderation and transformation. Transformation is the process whereby the government will shift the entire province's results on an exam in a normalized fashion (the shift is greater for results closer to the mean than those farther away). Transformation only occurs on years and for exams whose overall results are considered to be too high or too low by the government. Moderation occurs every year and affects one of the two components of the final mathematics grade. At the end of the year, the mathematics grade that students receive is made up of two different competency results. One of them counts for 30% of the overall mathematics grade and comes from complex, multi-step, situational problems solved throughout the school year in class. This grade is unaffected by the uniform exam. The other grade counts for 70% of the overall mathematics grade. This grade is arrived at by taking an average of a student's uniform exam result and his/her "moderated" class mark from throughout the year. The "moderated" class mark is generated by the government using a statistical formula to ensure that every class throughout

the province has a “moderated” class mark whose mean and standard deviation is equal to the mean and standard deviation of the same class on the uniform exam. Essentially, the government adjusts the final class marks of students to more closely match their results on the end-of-year uniform exam. If a student’s result on the uniform exam is higher than his/her moderated class mark, the student received the uniform exam result as his/her result for the competency worth 70%.

The exams are “high-stakes” for students according to the definition provided by Abrams et al. in that their results are used to make “highly consequential decisions” for students’ continuing education (2003, p. 19). The results from secondary school standardized math tests in Québec are used by schools to determine student placement in secondary V (grade 11) and by CEGEPs and universities to determine admission (along with the results of other standardized test results and school results). School or school board funding is not affected by test results and teachers’ salaries are not affected by these results, and so, despite the pressure on teachers (from students, school administration, school boards, parents, and the government) to have students do well on standardized tests, the system in Québec is not considered to be “high-stakes” for teachers according to Abrams et al. (2003). However, through the moderation process, on a class-wise basis, if the difference between class marks throughout the year and standardized test results for a given class of students is sufficiently large, the teacher is sent an official letter by the government and is required to explain the gap. Similarly, the government may request to audit a teacher’s long-answer questions. Both possibilities likely increase the pressure felt by teachers.

In some ways, the mathematics uniform examination is not a true standardized test. Having long-answer, developed-response questions is not a characteristic of typical standardized tests. Furthermore, teachers correct a portion of the exam in-house, which introduces elements of subjectivity. Even though teachers are instructed to correct according to a uniform rubric and marking guide which is produced by the government, this is still atypical of standardized tests. However, there are many ways in which the Québec mathematics uniform examinations are like standardized tests, namely that the same examination is administered to all secondary IV students across the province, that some questions are scored completely objectively and a portion is corrected centrally by the government, that all results are collected by the government and then statistically transformed, that individual students’ results are used to determine whether or not they graduate, and that group data is used to rank schools. Overall, for the purposes of the current

study, the uniform mathematics exams in Québec have been considered to be standardized tests, although a discussion of potential limitations with using this definition has been included in chapter 4.

Viewing background research-based effects in the Québec context.

Concerns over the true “objectivity” of standardized tests are valid ones in Québec. Standardized mathematics tests are written by committees of teachers selected from various school boards across the province with direction from the government. They are approved by the ministry of education (Ministère de l’Éducation et de l’Enseignement Supérieur), and are then validated by another committee. Though many teachers are involved in the process, tests are written through a human process of negotiation and discussion, and thus, they are at least somewhat inherently subjective.

Since teachers’ results are monitored by the government, it is fair to say that standardized mathematics tests in Québec hold teachers at least somewhat accountable. However, the research findings questioning the validity of standardized tests as accountability tools are applicable in the Québec context. Similarly, though none of the referenced studies investigating the correlation between standardized test results and post-secondary academic success were conducted in Québec, the Québec context is not so different to suggest that these results would not be generalized here. Després et al. provide a reasonable explanation for why this result should be transferable, which is that a single test is unlikely to be more accurate and/or valid than the ongoing professional judgment of a trained teacher.

The idea that standardized tests may not actually measure student proficiency is also one which seems like it should generalize to the Québec context. Firstly, the studies referenced were conducted in diverse educational settings across the United States which suggests that the results are generalizable. Secondly, one of the mechanisms explaining this finding is that standardized tests often have certain topics and types of questions more frequently tested than others, making the test content and format at least somewhat predictable. This leads to a range of practices referred to as “specific test preparation”. In Québec, there are indeed several often-repeated types of test questions, and several predictable topics on standardized math tests.

Other practices of “teaching to the test” may be applicable in this context as well as might all of the secondary effects of standardized testing, although one thing which is at best only

partially applicable is the policy-linked aspect of instructional narrowing. By including constructed-response questions on the tests, the government seems not to have exclusively engaged in Ellis' idea of lowering the standards of what is considered "proficiency". However, it is worth noting that short-answer and multiple-choice questions still make up 40% of the standardized mathematics exam, and the specific skills required to answer the constructed-response questions are yet to be analyzed to determine whether they represent more or less challenging standards.

Background research on teacher research.

Given that the present study aims to collect data from teachers based on their beliefs, experiences, and general professional knowledge, what follows is a brief justification of the relevance of teacher opinions, as well as definitions of certain key terms.

Skott, Van Zoest, and Gellert explain that until the 1970s, mathematics education research had been primarily interested in investigating either mathematical curricular content or "clinical studies of student achievement", which focused exclusively on results devoid of context (2013, p. 501). It was only at this turning point in the 1970s that "teachers and teaching" became "pivotal concerns of mathematics education research" (Skott et al., 2013, p. 502). From that point forward, including most recently, they add, "research interest in teachers has certainly not diminished" p. 501. The authors explain this continued interest by pointing out that teachers are the link between instructed content and student achievement, the two previous areas of education research focus.

Skott, Van Zoest, and Gellert attempt to classify and categorize the body of research on teacher education. They claim that much of the research has focused on 1) teacher beliefs, 2) teacher knowledge, and 3) teacher identity, categories which they claim are important to understand the teacher's "role in instruction", "meta-mathematical understanding", and views of "themselves as professionals" (p. 502) . For the purpose of this study, I will briefly describe teacher beliefs and knowledge, since they are the categories targeted through teacher interviews.

Teacher "beliefs are considered mental constructions developed through processes of assimilation and accommodation as a response to new experiential realities" (Skott et al., 2013, p. 502). Teacher beliefs are "expected to determine teachers' perceptions and interpretations of unfolding classroom events as well as their own contributions to these events" (p. 502). It seems

pertinent to attempt to collect teacher belief information, since teacher beliefs about standardized testing would have developed as a *response* to the *experiential reality* of being mandated to assess students by this manner. Furthermore, if there are any effects of standardized testing on instruction in this study, then those effects would *unfold* in the *classrooms*, and *teachers* would likely *contribute to them*.

Teacher knowledge is a term which can be broken into several categories. Kennedy (2002) describes three in particular: “craft knowledge”, “systematic knowledge” and “prescriptive knowledge” (p. 356). She writes that “craft knowledge” is “largely acquired through experience” and it “tends to be a-theoretical and idiosyncratic”. “Systematic knowledge” is “acquired mainly through colleges and universities, research articles, journals and professional associations” and Kennedy explains that it “tends to be more theoretical, codified and abstract”. Lastly, she writes that “prescriptive knowledge” is typically “acquired through institutional policies” and she describes this knowledge as being responsible for teachers knowing what they are “supposed to” do (Kennedy, 2002, p. 356).

In the context of the proposed study, “craft knowledge” is the knowledge gained by teachers through teaching. This includes their observations and beliefs developed through their teaching experiences about how students generally learn best. It also includes the specific knowledge of how best to teach students the various elements of the required mathematics curriculum, how best to prepare students for standardized tests, how to balance the demands of a standardized test with best teaching practices. “Systematic knowledge” in the context of this proposal is basically its definition: knowledge having been acquired theoretically, through the studies associated with obtaining teaching certification, through ongoing professional development, or some other type of formal/informal training. This would be teachers’ theoretical knowledge of the course curriculum including why certain topics are covered rather than others, of instructional strategies, educational theory, etc. “Prescriptive knowledge” in the context of this proposal has to do with the knowledge that teachers gain from the standardized tests themselves, as well as the systems surrounding their creation, dissemination, use, and evolution.

Chapter 2

Study Design and Methodology

In this chapter, I describe the present study's design and explain why it was chosen to best meet the desired research goals. Also included in this chapter are the methodology with which the selected study design was carried out, a brief description of the study participants, and an explanation of how that data was analyzed. Efforts taken to enhance credibility and trustworthiness of the study results are briefly described.

Study Design

In general, a qualitative study design was selected for several reasons. Firstly, a qualitative study is better able to “document” the “process of teaching and learning” which can be seen more holistically in a “broadly understood social context” (Sfard, 2005, p. 398). This strength of qualitative research in education is part of the reason that Sfard's international teacher respondents overwhelmingly expressed a “preference for qualitative research” (2005, p. 398). Secondly, Kuzmic writes that by focusing on whole populations, and faceless quantitative data, “individuals have come to be viewed as powerless to resist” the “institutional forces” which shape education (1994, p. 16). Focusing on the beliefs, opinions, and experiences of a small number of individuals empowers the participants by valuing their experiences and by including them in potentially working towards a solution. It is the goal of the present study to have participating individuals feel empowered as they help develop an understanding of a process in the world of Québec education.

More specifically, the aim of the present study is to gain an in-depth understanding of one particular phenomenon, namely uniform examinations in Québec. To gain a multi-dimensional understanding of this phenomenon, it was important to seek insight into teachers' perceptions of the examinations in general; their format and content, the ways that they are administered and evaluated, and how well they align with teachers' educational philosophies. A complete picture of these examinations also required an exploration of how these exams might impact instruction and how students might be affected by their existence.

In this aim, a qualitative, phenomenological study design was selected. The primary “purposes of phenomenological research are to seek reality from individuals' narratives of their experiences and feelings, and to produce in-depth descriptions of the phenomenon” (Yuksel & Yildirim, 2015, p. 1). Yuksel and Yildirim go on to state that phenomenological research studies

use the “lived experience, perception, and feelings of participants” to “understand and describe a specific phenomenon in-depth” (2015, p. 3). Max van Manen further explains that a phenomenological study aims to understand the “essence of a phenomenon” which he describes as “a universal which can be described through a study of the structure that governs the instances or particular manifestation” of it (1990, p. 10). This highlights the importance of allowing participants to describe the structure, settings, and general context surrounding uniform examinations as they see them. In fact, van Manen states that the “universal or essence may *only* be intuited or grasped through a study of the particulars or instances as they are encountered in lived experiences” (1990, p. 10).

Interview questions were designed to facilitate conversation about teachers’ lived experiences with, perceptions of, and feelings about uniform examinations in Québec. The resulting data was analyzed in the hopes of approaching the “universal” or “essence” of the phenomenon in question. As mentioned in chapter 1, despite the fact that standardized tests are a well-studied phenomenon globally, the specifics of Québec uniform examinations are not as well explored. In order to initiate an exploration of the given educational context, the present study seeks to provide detailed, in-depth, open-ended responses from few participants rather than limited data from many participants. In pursuit of this goal, a case study was selected for the design of the present study.

The rationale for this case study, as Schram states, is the purpose of any case study: “to draw attention to what can be learned from a single case” (2006, p. 107). In this case, teachers in a single school and in a single department have been interviewed to begin to look at the perspectives and observations of teachers in Québec. Secondary school mathematics teachers are uniquely positioned to provide insight into the effectiveness of standardized testing in achieving its goals according to the QEP because of the role they play in the educational process. Teachers possess various professional and situational forms of knowledge which could help to explain potential direct and indirect instructional consequences of using standardized tests. These are the forms of knowledge which were briefly discussed in chapter 1. Examining the perspectives and observations of teachers in a single case study is a way to initiate an investigation. Further proposed research steps to be taken based on the present study are discussed in chapter 4.

Though it was believed, a priori, that many of the effects of mathematical uniform examination on mathematics instruction and learning would generalize to other subject areas with uniform examinations, only one subject area was chosen for this exploratory study. It was important that the chosen subject area was a “content-based” course (like mathematics, science, or history) whose end-of-year uniform examination is based heavily on the specific content knowledge acquired in the course. This is as opposed to a “process-based” course (like English or French) whose end-of-year uniform examination is more focused on a process (like analyzing a text, or writing a persuasive essay). A content-based course was desired for this study since, according to the background research presented, a content-based course is more likely to see certain instructional effects of standardized testing, or is at least more likely to see enhanced effects (alignment of curriculum to more narrowly focus on uniform examination content, affecting teacher pacing to ensure coverage of all examination topics, and various forms of “teaching to the test” are some examples).

For the purpose of an exploratory case study, if there are potential noteworthy effects of uniform examination on instruction and learning, I wanted a school subject that would be more likely to showcase as many of them as possible. Along the same lines of reasoning, a subject area whose uniform examination is conducted in secondary IV was selected, since the results of these examinations appear on students’ official transcripts *before* they apply to post-secondary educational institutions. This makes the secondary IV exams “higher-stakes” than those conducted in secondary V, where exam results are only returned after acceptances to post-secondary academic institutions have been given out. If effects of uniform examinations on students are to come up, they are more likely to do so in a higher-stakes case.

Finally, of the content-based courses whose uniform examinations are conducted in secondary IV, mathematics was selected over others (history and science) because it is the topic that I teach. This means that my knowledge and understanding of the subject area helped me to contextualize the interview responses, to quickly ask follow-up questions during interviews based on an understanding of participant responses, and to look for themes based on a subject-specific understanding of what responses were related despite different terminology used.

Setting and Context

The present qualitative, phenomenological case study was conducted in a single Montréal area secondary school. This is the school where I work as a secondary school mathematics teacher, which made the study site a convenient one. Other than being a sample of convenience, this school was an ideal one to choose for this study, since it is a school which has participated in various research studies in the past and a school whose administration and staff members believe in the benefits of research. Since it is the school where I work, a trusting relationship already existed between the researcher and the school administration. Furthermore, being a member of the school staff, a trusting relationship already existed between the researcher and the teacher participants. I believe that this helped the participating teachers feel that they were not being judged despite sharing their personal teaching experiences and opinions. Furthermore, I believe it allowed participants to feel immediately at ease, which is crucial for fluid and comfortable interviews (Bogdan & Biklen, 2007, p. 104) which, in turn, meant a greater richness of results.

Because the purpose of the present study is to begin an exploration of uniform mathematics examination in Québec and their impact on teachers, students, instruction, and learning, it is important to highlight some characteristics about the school where the study was conducted. The school is a private, English school in Montréal. These characteristics are salient, because, typically, there are differences across the province according to type of school, language of instruction, and school location. The most significant difference, numerically, is how private school students perform compared to public school students. Most years, this difference is in the range of 10% on the average exam result, and close to 15% on success rate (Ministère de l'Éducation et de l'Enseignement Supérieur, 2014). The difference between how English school students perform compared to French school students is typically much smaller. For example, in 2013, French school students did less than 2% better on uniform exams than English school students across the province (Ministère de l'Éducation et de l'Enseignement Supérieur, 2014). Results also tend to vary from region to region in Québec. Typically, Montréal and Québec City tend to have slightly higher results, and in particular, students in private school in Montréal and Québec City do significantly better than students outside of the major cities (Ministère de l'Éducation et de l'Enseignement Supérieur, 2014). These trends are the same for all three of the mathematics courses with uniform examinations in secondary IV, namely the Cultural Social and Technical Mathematics Option (abbreviated CST Option), the Technical and Scientific

Mathematics Option (abbreviated TS Option), and the Science Mathematics Option (abbreviated SN Option).

As an example of the difference between public and private institutions, figures 1 and 2 respectively illustrate average results and success rates on the June 2013 uniform mathematics examinations in Québec. Data from the June 2013 exam period was chosen to illustrate general trends since these are the most recent results published on the Ministère de l'Éducation et de l'Enseignement Supérieur website. In addition to comparing public and private schools, figure 1 shows the average exam result for students in the study setting school for each of the three mathematics courses. As seen in figure 1, the study setting school has results which are much more closely aligned with private schools, on average, than public schools. Figure 2 did not include success rates for the study setting school as this data is not published by the Ministère de l'Éducation et de l'Enseignement Supérieur for individual schools.

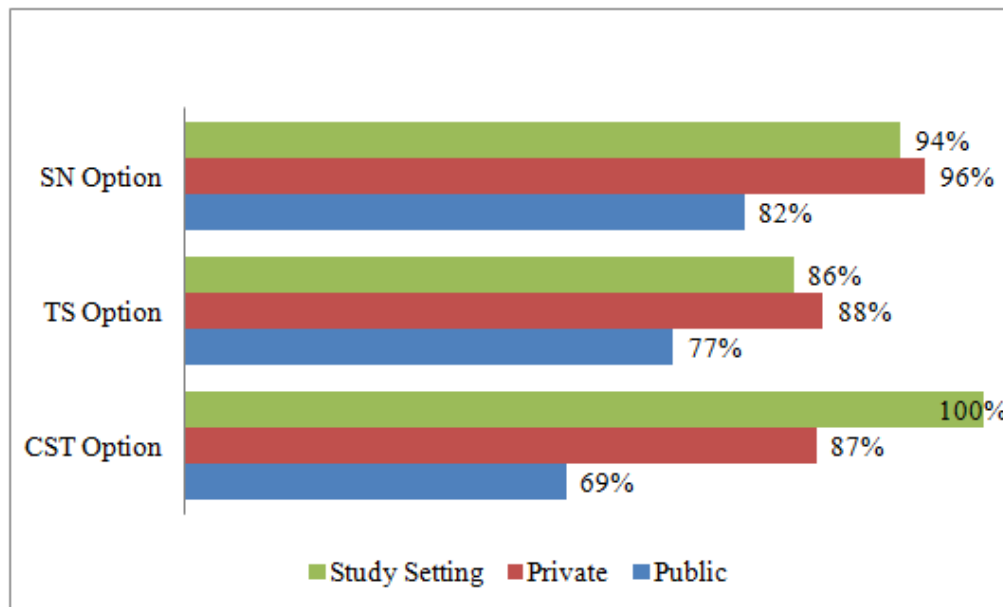


Figure 1. Average June 2013 uniform math exam results by academic institution.

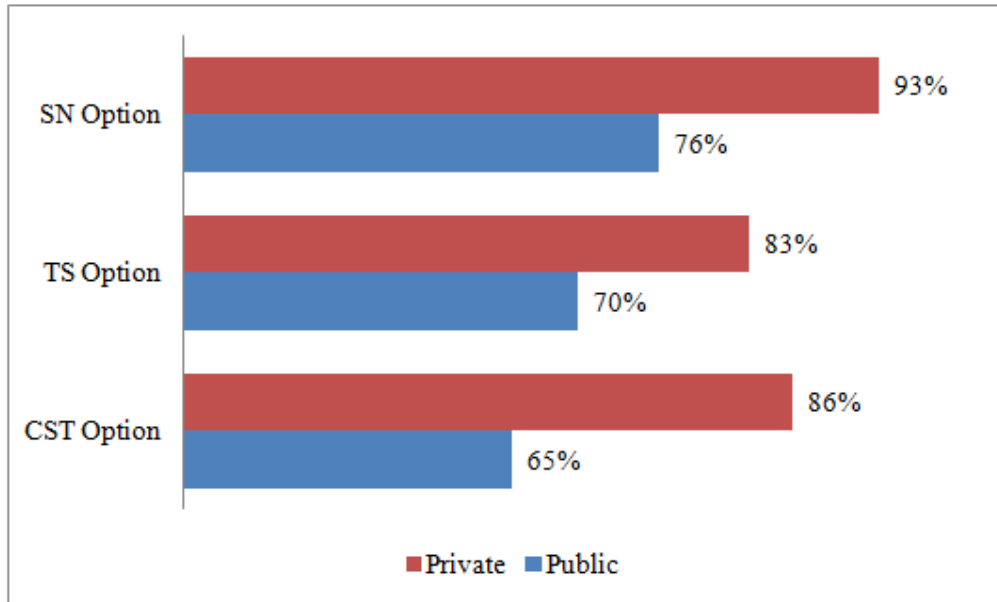


Figure 2. Success rates on the June 2013 uniform math exams by academic institution.

The average exam results of students in the study setting school are important, since relative student success on these uniform examinations may impact teacher beliefs, feelings, thoughts, and experiences. Given that students in this setting tend to do well on these exams overall, it seems less likely that negative issues would arise compared to a setting where students tend to do worse, and it seems much less likely that any negative issues that do arise will be exaggerated by the context of the study. The flipside of this, though, is that teachers in this context might have had experiences which are more likely to give a positive impression of these exams. In any case, some effects in participating teachers' experiences, perceptions, and beliefs might not generalize to accurately represent issues elsewhere in the province. A more complete discussion of this idea is presented in chapter 4.

Methodology

Permission to conduct the study with members of the school teaching staff was sought from the school administration. I, as the principal researcher, met with three school administrators to describe the purpose and methods of the study in person and with a written information letter (appendix A). Written permission was subsequently obtained. Once permission was secured, all seven secondary math teachers in the school were invited to participate with a personal invitation

and an information letter and consent form (appendix B). All seven participants read and signed the study consent form.

The voluntary teacher participants each took part in one-on-one, semi-structured interviews. Orafi and Borg describe semi-structured interviews as being “focused on some predefined topics but with flexibility to cover others which emerge” (2009, p. 246). Since the intention of the interviews was to give participants the opportunity to share their lived experiences, thoughts, and feelings about uniform examinations, having the flexibility to diverge from the original guiding questions was important. Similarly, since the present study is an exploratory one, it was crucial to give participants the latitude to voice issues important to them which may not have been in the scope of the original questions. Hancock and Algozzine recognize that because of this latitude and flexibility, “semi-structured interviews are particularly well-suited for case study research” (2011, p. 45).

The structured part of the semi-structured interviews consisted of prepared questions (appendix C) which were written before any of the interviews began. These questions were piloted in an interview with a secondary mathematics teacher from another school. This teacher was not a participant in the study, but was sought to provide feedback on the initial interview questions. To ensure that the piloting was effective, the volunteer teacher was asked to consider questions with a “critical state of mind” and was urged to “not just answer” questions, but also to “reflect critically on the usability” of those questions (Glesne, 2016, p. 106). The volunteer teacher was also tasked with verifying that the interview questions were well-aligned with the research questions, and that they were not overly leading. Sohn, Thomas, Greenberg, and Pollio note that, particularly in phenomenological studies, there is “often a difference between the research question and the interview question” which can lead to invalid results (2017, p. 129). They also warn against leading questions which direct “participants towards a specific aspect of their experience rather than letting them describe what aspects of their experience stand out to *them*” (p. 129). Based on feedback about the questions themselves, their “usability”, their alignment with the research questions, and their ability to let participants provide their own direction, the initial questions were amended. Attempts were also made after each interview to review questions based on how the participant interviews went. Questions were adjusted after the first three participants to ensure clarity of questions and to minimize confusion and ambiguity.

After the fourth, fifth, and sixth participants, it seemed that the remaining interview questions were clear, and thus did not undergo any further amendments.

The interviews took place on the school premises after school hours. Hancock and Algozzine recommend conducting interviews in a “natural setting” to “enhance realism”, and also to “seek a private, neutral, and distraction-free” location, if possible, to “increase the comfort of the interviewee and the likelihood of attaining high-quality information” (2011, p. 45). Using the school was convenient for both the participants and the principal investigator, making sure that interviews took place after school hours made the location private and distraction-free, and offering the participants a choice of using an office or a classroom as well as which specific room, gave them the opportunity to select the location which would make them the most comfortable.

At the start of each interview, participants were reminded of the purpose of the study and of the fact that results would remain confidential (Bogdan & Biklen, 2007, p. 103; Hancock & Algozzine, 2011, p. 46). Participants were informed that pseudonyms would be used, and they were asked to think about what pseudonym they would prefer. Participants were then asked to select their pseudonyms at a later date. All participants re-read and signed consent forms and all participants agreed to have their interviews be recorded. During interviews, I tried to follow some of the general interview advice of Bogdan and Biklen. For example, I tried to prime participants to “share their own ideas and observations” (2007, p. 107). I tried to keep in mind that importance of remaining neutral and avoiding being judgmental (2007, p. 108). In some cases, participants were worried that their responses might not be ‘what I was looking for’, and I assured them that whatever they wanted to say to answer the questions was absolutely the ‘right thing’. Finally, as per the rationale of conducting semi-structured interviews, I tried to “always be prepared to let go of the plan and jump on the opportunities the interview situation [presented]” (2007, p. 106). Though it was not always easy to adjust in real-time, if a topic came up which seemed interesting, relevant, of particular importance to the participant, or which the participant seemed to have more to say about, I tried to ask follow-up questions to get at as much of the important information as possible.

All participant interviews were recorded using the Microsoft Voice Recorder software for Windows 10 on the researcher’s computer. One interview was transcribed manually (listening to the recording on the same Voice Recorder software and typing the text), one interview was

transcribed using the voice-to-text software in Google docs, and five interviews were transcribed using the Temi transcription software. The interviews transcribed using the Temi software were all reviewed word for word manually after transcription, and software errors were amended manually. After transcription, participants were each given a typed transcription of their own interview and asked to approve without changes, approve with changes, or not approve the transcription. Participants were specifically informed that by “changes”, they could determine if: 1) there were errors in the transcription and something typed was not actually said, or that something said was not typed; 2) when reading, participants felt as if their responses did not accurately represent their beliefs, thoughts, or experiences; 3) there were any additional thoughts or experiences that participants wished to add. In any of these cases, participants were invited to offer their suggested changes.

Once all interviews were transcribed and approved, participant responses were organized and analyzed for emerging themes manually, according to the advice of Bogdan and Biklen (2007). They recommend that once the data is transcribed, the researcher “take long, undisturbed periods and carefully read your data at least twice” (p. 185). The purpose of carefully reading through the data is to “begin developing a preliminary list of possible coding categories” (p. 185). Dey (2003) adds that though a researcher reads to comprehend, “intelligibility is not the only, nor even [the] main goal”, but rather to “prepare the ground for analysis (p. 87). The emphasis in reading, Dey adds, is to explore “potential themes and topics” (2003, p. 88).

Bogdan and Biklen suggest that the next step in analyzing data is to develop codes and to assign codes to various units of the data. This process, called “coding”, allows to better “connect with the data to facilitate the comprehension of the emerging phenomena and to generate theory grounded in the data” (Basit, 2003, p. 152). The data was systematically read-over, and a series of words and phrases were established as a preliminary set of codes to represent repeating “topics and patterns” in the data (Bogdan & Biklen, 2007, p. 173). The codes were then assigned to units of data throughout. As a second wave of coding, I read through the original (un-coded) version of each transcript, and with the responses fresh in my mind, I went through my list of codes again to see if any significant responses seemed not to fit with any existing codes. At the same time, the list of codes was reviewed and adjusted to group, collapse, or separate data into categories. Finally, decisions were made about the final categorization of data with overlapping codes, and a final list of themes was determined.

Initially, at the proposal phase of the study, it had been my intention to look for differences between middle school and senior school teacher responses, but an insufficient number of middle school teachers made it difficult to establish a difference in thematic patterns. Some minor general observations about potential differences have been made in chapter 4.

Emergent Themes

Table 1 lists the themes having emerged from analyzing the interview data. Associated with each theme is a list of corresponding codes and in the final column of the table, the frequency, meaning how many participants discussed the given theme/sub-theme. Themes are listed alphabetically, and sub-themes are listed in descending order of frequency within each theme. In a few cases, participants made indirect reference to a given theme/sub-theme, and in other cases, participants made direct reference to a theme/sub-theme, but did not discuss it in very much detail. In these cases, a judgment call was made as to whether or not the data was significant enough to justify counting it towards the total frequency of the theme/sub-theme. In borderline cases, the data was not included, so as not to inflate the appearance of results. In these cases, a quote may appear in the results table for a given participant in chapter 3 even though the total frequency does not reflect it. Themes with frequency less than three were omitted from the table, but some of these are briefly addressed in chapter 3 and in the discussion of chapter 4, as they are of interest to the overall goal of the study, namely initiating an exploration of teacher beliefs, feelings, thoughts, and experiences surrounding uniform examinations and their impact on teachers, students, teaching, and learning.

Credibility and Trustworthiness

The credibility and trustworthiness of the present study are potentially affected by my personal beliefs and biases. Though precautions were taken to increase the objectivity and neutral tone of interview questions (by having a test subject review questions and by trying to remain neutral as an interviewer and researcher throughout the process), my beliefs and biases are subtly present in ways I might not have noticed throughout the process.

To ensure transparency in my interpretation of participant responses, raw interview data is provided in a separate table for each participant in chapter 3. The tables are formatted similarly to table 1, meaning that interview data is grouped according to the same themes and sub-themes

and in the same order. Furthermore, in the interest of transparency, my own views are disclosed in chapter 3. I filled in a results table with my own personal beliefs in a partial ‘self-interview’. In the discussion in chapter 4, results which closely aligned with my own personal beliefs are explored in the interest of determining whether these similarities are likely due to researcher bias, shared experiences with participants, or the legitimate existence of a particular phenomenon.

Table 1. Emergent themes and sub-themes, associated codes, and frequency of responses.

Themes and sub-themes		Codes	Frequency (out of seven participants)
Exam format	Multiple-choice questions	all or none, guessing, multiple-choice, no thought-process shown	7
	Long-answer questions	applying skills, conjecture questions, critical thinking, different areas, explain their thinking, give you so little but you can do so much, long-answer, paragraphs, rubric, situational problems, think abstractly	7
	Short-answer questions	final answer, marking guide, multi-step, process	5
Exam nature	Existence of a uniform exam	content knowledge, cumulative, curriculum, exam, grade 10, standardized tests, summative assessment, test, uniform exams	7
	Validity/reliability of the exam	Accurate, also going to perform, a reflection, indicative, certain questions show up, expect, if they haven't mastered, marks are better than understanding, material on test based on the most important things, predictable, represented, representation, skews the data, unpredictable, valid,	7
	The repetitive nature of questions	always on the test, every year, frequently, most years, past years, patterns emerge, repeat, repetition,	5

		repetitive, similarity, year after year	
	Moderation	aftermath, easy year, hard year, moderation, trained me, trust me	3
Instructional issues	Using old exams to prepare	help to study, end of year, exam prep, get used to, model, old copies, old exams, old exam questions, past exams, practice tests, review, teaching to the test	7
	Timing throughout the year	couldn't cover that material in enough time, deviate from the curriculum, forego other opportunities, get through it quickly, have to move on, incorporate more topics, left something out, pace, pacing, time, timeline, timing, slow down, spend more time, supplement	7
	Teaching specific test-preparatory skills	check the memory aid, draw connections, go through your notes, help prepare, in my experience, meta-skills, model after conditions, notice connections, practice tools, review package, shortcuts, skills that are applicable, study skills, teaching to the exam, timing,	4
	Teaching throughout the school year	could approach teaching a class differently, do a better job of teaching the skill, got something to go with, haven't tried, if it wasn't a standardized test, incorporate old exams, not what I should be doing	4

		with my class, spend more time, teaching to the test, throughout the year, we shouldn't be	
Learning issues	Student stress	anxious, care, CEGEP, confident, fear, future, gatekeeping, gateway, heavy marker, high stakes, impact, management, motivate, motivator, pressure, scary, success, stress	7
	Students not learning effectively	applying, don't feel like they taught for understanding, just going to focus on certain areas, looking for a shortcut, focus just on the exams, guidance of the teacher, trying to figure out what they're going to put on it, trying to figure out what they should do in order to have a good mark	6
	Developmental capacity of students	ages, development, developmental, handle, frontal cortex, maturity, not ready this year, ready for that thinking, years old, young for this/them	4
Systemic issues other than the exam		conjecture, develop well rounded students, did not give me what I was actually looking for, geometric proof, parental involvement, schools across our province are so different, the system, what is Alex good at?	5

Chapter 3 Presentation of Data

Plan for Data Presentation

Interview data were categorized according to the five emergent themes and the accompanying sub-themes. For each participant, a separate table presents original interview responses according to each theme/sub-theme. In the cases where a theme or sub-theme was not brought up by a participant, “N/A” was entered in the corresponding table cell. Each participant’s table of results is followed by a brief summary written by the researcher, wherein certain points are clarified or contextualized with additional information about secondary mathematics courses, the examinations themselves, or other issues relating to the education system. Before each table of results is a brief description of the participant to help the reader gain a more complete picture of each participant’s views, beliefs, and experiences. I have included my own results table after all of the participant results in the interest of transparency and trustworthiness in results.

Alice’s Interview Data Table and Summary

Alice is a female senior school mathematics teacher who also teaches photography. She has been teaching secondary school in Québec since 2011 and prior to this, she was an early childhood educator in Québec. She holds a master’s degree in mechanical engineering and a second one in teaching and learning. She has taught grades eight to 11 and has taught courses with uniform exams four times.

Table 2. Interview responses contributed by Alice

Theme/subtheme	Participant response
1. Exam format	
a) Multiple-choice questions	<i>The one that thing that I always feel bad about during those tests is that they won't know the approach that students will take to solve. It doesn't measure that at all. It doesn't give them the time to show that. The fact that it's time limited, I think that that's a big problem. If it wasn't time limited, then I think it would be a much more fair representation of what kids know... Well that's any multiple-choice question... I think that the</i>

	<i>real deeper understanding can only be seen in situational problems, I don't think that any C2 type test can really show the mastery that I would like to see... So, definitely, I also have a thing against multiple-choice questions.</i>
b) Long-answer questions	<i>...the fact that they have to explain their thinking more, it's one problem that is more complex, they really have to take time. And I like that part, they are forced to work through it slowly.</i>
c) Short-answer questions	N/A
2. Nature of the exam	<i>Time issue is a big problem for me. Like, really, when in life, ever, do you have to something in one hour? I don't think that's... that must change.</i>
a) Existence of a uniform exam	<i>I find it so degrading and humiliating and against everything that education should be... if you want to do the test, you want to see how well would you do on the... you know compared to other schools, you can. But I don't think everybody should. Like some students, like for example, our AP students, I think they enjoy that stress. I think they crave that. That's what gives them motivation. But half of SN and CST students, they don't need that, they don't want that, that's not useful for them.</i> <i>And</i> <i>You know the Waterloo contests that we're doing? I think that's more interesting, that's the result that I would look at more than our June exams. Like the questions are sort of unexpected, right? And they really require you to ... I'm curious about those results more than exams.</i>
b) Validity / reliability of the exam	<i>But I think that they overemphasize quadratic function and statistics. I find that strange, honestly, that statistics is given so much... and it's such a tiny topic in grade 10. That I find a little bit weird. Like, I would have more algebra in grade 10. I think that's not so well represented... working with polynomials, factoring, more... to make sure that they know this</i> <i>And</i> <i>If it wasn't time limited, then I think it would be a much more fair</i>

	<i>representation of what kids know.</i>
c) The repetitive nature of questions	<i>In grade 10, I honestly haven't been thinking like that. I wasn't looking to see if there was a pattern of questions that are repeating. I wouldn't say so.</i>
d) Moderation	<i>They should trust me... I mean, if they already trained me, right? I didn't decide to become a teacher and just came to school and nobody knows what I'm doing. They already trained me, and they should trust me. They gave me the right to teach, so if I'm saying that this student has finished the grade 9 topics, yes, he's done.</i>
3. Instructional issues	
a) Using old exams to prepare	<i>I am the coach, you are going to do this race and this is the terrain that you're going to be in, and if the weather is... yes, absolutely, test focused. I try to teach them throughout the year, and I feel honestly bad about it, because that's really not what teaching and learning should be about, And DE: how much time, would you say, at the end of the school year, do you devote to doing exam prep? ALICE: 2 weeks... more in the second half of the year, because I worry if I do that after every unit, that I will... that this exam stress is going to be present all of the time.</i>
b) Timing throughout the year	<i>ALICE: I hardly teach anything that I don't have to in grade 10. DE: Because it's just such a time crunch? ALICE: Yeah. But I will go over... I will try to teach them shortcuts more than in any other year. Like tricks how to remember...</i>
c) Teaching specific test-prep skills	<i>I will teach them shortcuts if I can. I will teach them to try to approach the exam in a way that's best for them. For example, to start with long-answer questions, or to cover the multiple-choice options and solve them. Those types of things you don't have to teach for situational problems... Or when they read a question to try and think of which unit it is, and to even write about the question, and then to check the memory aid and to</i>

	<p><i>compare a little bit to refresh their memories. If they're stuck... I will tell them to think about grade 10 topics: what's new? What haven't you seen ever before? What did you learn this year? So then, they will say, step function. So, is it to be expected on the exam? Yes. What about, I don't know, area and volume of a sphere, is it new, not really, so, I'm teaching them to look through the curriculum and try to see what's really important to... again... to help them prepare, but it is too much focus on the exam ... I will try to teach them shortcuts... like tricks to remember</i></p> <p><i>And</i></p> <p><i>I mean, I think that this year will, can, as you said, can be viewed as: "this is how you can study, if you have to do some exam",</i></p>
d) Teaching throughout the year	<p><i>... if there wasn't an exam at the end of the year, I could approach teaching a class differently.</i></p> <p><i>And</i></p> <p><i>I feel like that's not what I should be doing with my class. I should be really discussing with you what are similar solids all about, where do you see them, where do you use them, we shouldn't be even thinking about what is it that I'm going to test you on, and we are not in that situation. I'm not sure we would do that even if there was no test, because we are so used to testing them in this way.</i></p>
4. Learning issues	
a) Student stress	<p><i>It's more in the second half of the year, because I worry if I do that after every unit, that I will... that this exam stress is going to be present all of the time.</i></p> <p><i>And</i></p> <p><i>I will try to teach them shortcuts more than in any other year. Like tricks how to remember, even though they have a memory aid, maybe they don't need that, but because of the stress.</i></p>
b) Students not learning effectively	<p><i>I mean it's reality that, I think 80% of students are trying to figure out what they should do in order to have a good mark.</i></p> <p><i>as a student, how will you know that? Now, we are all good</i></p>

	<i>mathematicians, right? So you had, in high school, you had a feeling of yourself being good at math, so your stress was definitely lower than others. But imagine if you are not so confident about that topic, how will you make sure that... you don't suddenly have motivation to go through everything in two weeks, so... I would myself do that as well. I would look through the exams and try to figure out which units I have to know.</i>
c) Developmental capacity of students	<i>Maybe I'm not ready for this test this year. Why not do half of the test, and maybe next year, okay now we do it next year. I mean, we do it twice, but I still find it's way too, how can I say, I'm actually surprised that Québec, which is so progressive in so many things, in education, is so old-fashioned.</i>
5. Systemic issues other than the exam	<i>How many times are we discussing in school how many tests, put the tests on the calendar, the schedule for this, the schedule for that, we never discuss ok what is Alex good at? Where does he shine? In all of this, what is his strength in all of this? You know, that's forgotten And I guess under it all is my personal connection with it, because I was a good student, doing everything well, high school, university, managing everything and I felt cheated in a way that that did not give me what I was actually looking for. You know? I was always good in math, but what is the connection between the math and problem solving in class and the actual math that I would do as an engineer? That's not clear. So is that the profession for me really?</i>

Summary of Alice's interview.

Alice's wariness of uniform examinations, and really all tests, emerges in our interview. In the background and introduction questions that I asked her, she responds that the purpose of assessment in general is "to judge you". This idea surfaces again later when we delve into the specific format and content of the exams themselves.

Exams themselves (format, content & nature).

In terms of her thoughts about the content and format of the exam, Alice's answers speak to a more general philosophy about assessments. She responds that she is not a fan of the multiple-choice questions on the exam, but she moves beyond the exam when she explains why. She says that she dislikes those questions because that type of question does not allow students to demonstrate the "approach" that they "take to solve" the problems. Although she did not explicitly mention it, it seems that she feels the same way about short-answer questions, since there is no work shown, and no partial credit awarded for work which is partially correct (other than in exceptional cases).

Alice goes on to differentiate between the type of assessment with fewer, complex, contextual problems on which you "let me work on it, and you let me talk to friends, you let me investigate, you let me use books" and the type where "you give me ten questions and you tell me that you're coming back in one hour and a half". She raises the question of whether or not the uniform exams are valid at all, since she feels as if these two types of assessments measure "different qualities". She asks: "do you want to have somebody who is thinking, or do you want somebody who is able to do things fast?", because "for me, they are two different things".

Instruction.

Alice's approach to helping her students prepare for exams is related to her belief that limiting exam-writing time forces students to use a different skill-set than the one she would ideally want them to use. Consequently, she teaches her students shortcuts to save time, and strategies like starting with the long-answer questions to ensure that any questions left at the end of the exam are easier to complete, require less time, and are worth fewer marks. She also teaches her students to use meta-test skills, such as questioning which topics are new to the grade 10 course to determine which topics are more likely to be assessed on the exam. Another example is teaching her students to cover the multiple-choice options when solving the problem in order to solve it without the bias of seeing any provided options. Personally, she believes in a more relaxed form of studying, wherein a student discusses, thinks over, and contemplates the course content. She thinks that this form of studying is more closely related to what should be the purpose of learning and education, and she believes that is a more long-term useful practice than "cramming". It seems telling that this is the way she counsels her own sons to study, possibly the strongest endorsement there is.

In terms of exam preparation, Alice makes use of old exams for practice. She reserves the final two weeks of the school year for exam preparation using old exams as practice. Throughout the school year, she says that she also brings in examples of specific old-exam questions, but only in the “second half of the year”. She believes that bringing in exams earlier in the year does less to help students prepare than it does to ensure that “exam stress is going to be present all of the time”.

Student learning.

Alice’s main concerns with student learning have to do with the general stress felt around the exam and the specific feature of the time limit on the exam. Because of these, she feels that a teacher’s job becomes managing stress in addition to teaching math. She also believes that a history of failure in mathematics makes students more likely to be stressed and more likely to do poorly establishing a self-fulfilling prophecy and a cycle of failure. She talks about herself and the other math teachers in the department and points out that “we are all good mathematicians”. She goes on to say that students who fit that profile will tend to have a “feeling of [themselves] being good at math” and that this will help them to feel confident, able, and ultimately, to actually do well. On the other hand, she says, “if you are not so confident about that topic”, then your level of anxiety will probably be higher, and this might lead students to look through old exams and “try to figure out which units I have to know” rather than learning all course content.

Alice says that we do not have to than force students to feel anxious, unable, or like they have to narrow the focus of the course to get a good grade even if they do not understand the math. She suggests that, instead, we could and should offer more flexibility in when exam are written, or even if they are written at all. She suggests that a uniform exam could be optional for students. She says “if you want to do the test, you want to see how well would you do ... I don’t think everybody should”, and explains that for some students, the pressure of an examination like this is good, but that for others, the process itself is potentially harmful. She says that students could decide if they want to write the exam, and then if they choose to write it, there should be a process whereby students and their teacher determine when a student is ready to take the exam, and that it might not be on a specific, pre-determined day at the end of the year. She suggests: “why not do half of the test, and maybe next year, okay now we do it next year”. She goes to say that given that “we do it twice”, why not be more flexible in the options to redo the exam or simply do it for the first time at a different time. She makes reference to the fact that exams are

currently held in June, and then at two other times in an academic cycle, in late July or early August, and again in January. She proposes that if we have other time periods in which to administer exams anyways, that we should be counseling students to introspect and decide when they are actually ready to write the exam. Alice says that not every student will fit the mold of the ideal timeline and that we should encourage students to write the exam when they feel ready.

Other systemic issues.

Alice felt that the education system in general was failing to recognize students as individuals and thus, was not helping them to recognize their personal strengths, to reach their personal potentials, or to find the future professions that would lead to feel accomplished and happy. Although she did not directly mention uniform examination in this observation, she implies that they are mutually exclusive with her ideal educational world for two reasons. Firstly is the impersonal nature of exams. They are corrected by a machine in an institution where nobody knows any student rather than by teachers who get to know their students as individuals all year. Consequently, each student is viewed as a number in the system. Furthermore, everyone writes the same test rather than individualized assessments personalized for students. While this might not be a feasible solution, it adds to this prevailing impersonal atmosphere. Secondly, Alice does mention that extra pressure placed on this exam means that she spends less time teaching the way she wants to and the way she thinks she should. She effectively says that she spends less time helping students cultivate their individual areas of skill because she is so focused on the exam.

Bart’s Interview Data Table and Summary

Bart is a male senior school mathematics teacher. He also teaches wood working and is a member of the school’s outdoor education department. He has been teaching for 11 years in Québec and taught in Alberta before that. He holds a bachelor’s degree in education from McGill University with a physical education major and a minor in science. He has taught kindergarten to grade six in elementary school and also grades eight to 11 in secondary school. He has taught courses with uniform exams for five years.

Table 3. Interview responses contributed by Bart

Theme/subtheme	Participant response
1. Exam format	

a) Multiple-choice questions	<p><i>I think the multiple-choice section [does a particularly poor job].</i></p> <p><i>Actually, multiple-choice and short-answer sections, because there's no credit given for their work, and you don't get to see their work. So, you could randomly guess.</i></p> <p>And</p> <p><i>You're not with them when they're doing the problem. So you can't see the thought process. And especially the multiple-choice/short-answer sections: you get to see none of their thought-process, because none of their work is shown. How they approach a question... you don't really get to see that.</i></p>
b) Long-answer questions	<p><i>the long-answer section questions are usually pretty good to detect for deep understanding</i></p>
c) Short-answer questions	<p>See 1a) , plus:</p> <p><i>...just to refer back to the short-answer section, I just, I don't see its purpose, again, it's not a guess like multiple-choice, but because there is no work shown, what are we really assessing there? We're assessing if they can get an answer, we are not assessing whether they can actually perform a task.</i></p>
2. Nature of the exam	
a) Existence of a uniform exam	<p><i>So if it was just, I shouldn't say if it wasn't a standardized test, but if it was more blind, then they would have to prepare the whole year worth of material rather than just sections of it.</i></p>
b) Validity / reliability of the exam	<p><i>I think that, typically, your higher achieving students throughout the year are also going to perform well on the standardized tests, and your lower achieving students will perform lower, however, some of your lower achieving students at the end of the year will perform really well on that exam, because they spend a lot of time preparing, they go to tutorials and things like that, so all of a sudden, it's not a reflection of what they did throughout the year also. In some cases it could be a really high achieving student who performs poorly on the day of the exam, and all of</i></p>

	<i>a sudden, it's not a reflection of what they're actually capable of. Or it could be the kid who did nothing all year, but decided to prepare at the end of the school year and does well. So, does it really reflect a student or who they are as a person? I don't think so.</i>
c) The repetitive nature of questions	<i>Yeah. There's quite a few [questions that repeat]. Particularly the type of question, like most years, there's a conjecture type question, although this year there wasn't. But, yeah, there are multiple types of questions that repeat quite frequently.</i> <i>And</i> <i>There's so much repetition in the exams from year to year, the types of questions that you see from year to year, if they know that there's this type of question which is worth 10 marks, and they see it in 5 or 6 old exams, then that's what [they're] going to study ... I don't think it's good, because now they're forgetting, or they're not really applying the information they learned, all of the information they learned throughout the year, they're just going to focus on certain areas. So if it was just, I shouldn't say if it wasn't a standardized test, but if it was more blind, then they would have to prepare the whole year worth of material rather than just sections of it.</i>
d) Moderation	N/A
3. Instructional issues	
a) Using old exams to prepare	<i>I hate to say it, but because of the standardized tests, quite often at the end of the school year, you give them a bunch of standardized tests to write just so they get used to seeing the format and the types of questions that they're going to be assessed on, and then modelling on the board. How to approach a question. As I do the question, I talk out loud as I do it, so they can see my thought process was, so they can see how to attack a problem.</i> <i>And</i> <i>...because the way we write the questions in class, or the types of</i>

	<p><i>questions we give in class might be slightly different than the way they're worded on the exam, and the format also isn't always the same.</i></p> <p><i>I usually have them work in class and give them feedback as they're going, but I don't give them the solutions until after it's been corrected, because I want them to be able to attack the question without the crutch of looking at the solutions.</i></p> <p>And</p> <p><i>...they can see that there are patterns, and that will guide their memory aid production ... It's just a shame we have to do that to help them study, I guess.</i></p>
b) Timing throughout the year	<p><i>It affects the pace of my course for sure, because you know you need to get... just because there's so much content, that you have to get through it quickly enough that you can actually have time at the end of the year to address the standardized test. If it was not for the standardized test, I wouldn't push through the material as quickly, because I wouldn't need that time at the end of the year to prep the students for it.</i></p>
c) Teaching specific test-prep skills	<p><i>I teach the way I teach, and then gear it towards the exam at the end of the year.</i></p>
d) Teaching throughout the year	<p>See 3a)</p>
4. Learning issues	
a) Student stress	<p><i>...[the exam] has such an impact on them. Their future success in CEGEP, whether they can get into the program or not...</i></p> <p><i>I think, especially in grade 10, it's always in the back of their mind, because it's such a gateway course to CEGEPs, so I think that in the back of their minds... I don't know if it's a huge effect, but there is a stress that: grade 10 year, I have to perform well because there's a standardized test that's going to decide what I can do with my future.</i></p>
b) Students not learning effectively	<p><i>because there's so much repetition in the exams from year to year, the types of questions that you see from year to year, if they know that there's this type of question which is worth 10 marks, and they see it in 5 or 6 old</i></p>

	<p><i>exams, then that's what [they're] going to study. So I think they do gear their studying towards what they've seen in previous exams because there are patterns.</i></p> <p>And</p> <p><i>I don't think it's good, because now they're forgetting, or they're not really applying the information they learned, all of the information they learned throughout the year, they're just going to focus on certain areas</i></p>
c) Developmental capacity of students	N/A
5. Systemic issues other than the exam	<i>[the purpose of education should be] to develop well-rounded students, not just in one area: that means physical, mental, social, like, the whole package. Which is why I think things like outdoor ed. are so important.</i>

Summary of Bart's interview.

Bart's beliefs about the nature of instruction and assessment seem to be tied to his experiences teaching outdoor education and wood working, disciplines in which effective behaviours, skills, and practices have to be modeled rather than simply spoken about. He feels that his purpose as a teacher is to "make sure the kids get all the information they need in order to be successful" and that the purpose of assessment is to "see if the kids have actually ... acquired" and integrated it. He goes on to explain his belief that the way students learn to solve problems and apply concepts is to have them modeled. Even when Bart brings in old uniform exams as practice, he does not simply distribute them and leave the students on their own. He goes through the questions with his students, modeling the problem-solving techniques that he wants them to learn.

Exams themselves (format, content & nature).

Bart believes that the multiple-choice section of the exam is not helpful in assessing what students are able to do. He says this because students are able to guess if they do not have to show their work; and also, as a teacher, he cannot be sure what they were thinking when they did the problem. This is consistent with the emphasis that Bart places on modeling the processes and techniques that he wants to see his students carry out. If he cannot see the process, then he cannot properly assess. This belief about the importance of showing a student's process extends to his feelings about the short-answer section of the exam too. Without the work to go with students' final answers, he says that "we're assessing if they can get an answer, we are not assessing

whether they can actually perform a task”. The Québec education program refers to competency as a student’s ability to “mobilize knowledge” to solve problems (Ministère de l’Éducation, 2001, p. 185), a definition which highlights the active nature of problem solving. The importance that Bart places on seeing student process is in total agreement with this definition. Also consistent with this importance on being able to demonstrate process, Bart feels that the long-answer questions typically do a good job of “detecting [students’] deeper knowledge”.

For the most part, Bart believes that the uniform exams are valid in the sense that “typically, your higher achieving students throughout the year are also going to perform well on the standardized tests, and your lower achieving students will perform lower”. However, he points out that the exams could afford to be even more valid, since there are cases every year when students do better on the exam than throughout the year, or vice-versa, which he says is “not a reflection of what they’re actually capable of”.

Instruction.

Bart says that uniform exams do not have a big impact on the nature of his instruction other than affecting his pacing, since he saves time to review old exams at the end of the year. He says: “I teach the way I teach, and then gear it towards the exam at the end of the year”. The importance that Bart places on modeling problem solving skills throughout the year, applies to those practice exams too, as he goes through the problems with students in class and shows them how to solve the problems rather than have them rely on the crutch of looking through the solutions. Bart believes that students might still be successful on the exam without practicing old exams, but “not as successful”. He believes that this is because practicing old exams helps students to become familiar with the format, style, and wording of the exam questions.

Student learning.

Although Bart says that uniform exams do not heavily influence his teaching over the course of the year, he believes that the exams are ever-present in students’ minds, at least “in the back of their minds”. In discussing this stress, he explains why he thinks it’s there as well as what consequences he believes it has. He thinks that the stress that students feel surrounding the exams has to do with their high-stakes nature, specifically that they are “a gateway ... to CEGEPs”. As a consequence of this stress and the importance of the exams, he believes that students “gear their studying towards what they’ve seen in previous exams because there are patterns” and that

as a result, they end up “not really applying... all of the information they learned throughout the year”.

Other systemic issues.

Bart believes that the purpose of the education system should be to produce well-rounded students. His opinions of the uniform exam causing students to focus on a narrowed scope of the course and incentivizing students to learn how to answer the most frequently repeated question seem incongruous with his goal for the education system. Although Bart did not explicitly say that uniform exams contradict his fundamental beliefs about the education system and what it should aim to achieve, he brought up several issues for which his beliefs and uniform exams have mutually exclusive features.

Cleo’s Interview Data Table and Summary

Cleo is a female middle school mathematics teacher and she has been a teacher in Québec for five years. She holds a bachelor’s degree in education which she completed concurrently with a Bachelor of Arts degree in mathematics. She is currently enrolled in a master’s program in mathematics. She has taught grades seven and eight, and she has tutored students in grades seven to 12 and in CEGEP. She has never taught a class with a uniform exam, though she has tutored several individual students in courses with uniform exams.

Table 4. Interview responses contributed by Cleo

Theme/subtheme	Participant response
1. Exam format	
a) Multiple-choice questions	<i>I always think multiple-choice is a poor way of assessing student knowledge, personally. I love it personally as a student and I love it as a teacher. It's so easy to mark and sometimes it gives immediate feedback. I personally don't think it has a place in a formal assessment like that at the end of the year. It leaves a lot of room for kids to guess. Typically, I find the multiple-choice questions very sneaky from my experience. Especially like last year's was particularly sneaky. I don't like the way that they do that where a student might feel really confident and one answer and then see something else, and then second guess themselves. I</i>

	<p><i>prefer when students can show their thought process and work.</i></p> <p>And</p> <p><i>Sometimes I like the multiple-choice if it's done properly, like if it's done in a clever way, because sometimes it can guide students who remember their units, for example, or to remember, I don't know, a formula that's used there that they might be able to use later and that might trigger something which is useful. I think it can maybe build confidence in the beginning and then move forward.</i></p>
b) Long-answer questions	<p><i>because if they give you so little, but you can do so much with it. So there's almost always a question like that. It's always with the zeros because I am, as a teacher, as a tutor, we're so used to teaching like there's this method, this method, this method, if you've got the vertex and a point, if you've got this, if you've got that, whatever. And I always forget about using the zeros. Always. And so they always throw a question like that on there. And sometimes even as a tutor I sit there and I'm like, what are we missing? Until I realize that it's actually so easy, but it's not as commonly taught or that there's not as many of those examples. There's always at least one of those.</i></p>
c) Short-answer questions	<p><i>Short and long, that's fine. I mean really those are just like the depth of the question, right? The length of it. So I like those components. We give a lot of part marks in middle school in particular for process. Like, we actually very rarely care about the answers, right? So I tend to not like the multiple-choice very much.</i></p>
2. Nature of the exam	
a) Existence of a uniform exam	<p><i>I mean, in math in particular, I think we have to have some sort of testing at the end of the year. I think that's actually just appropriate to see what have you learned and how have you learned it and how can you apply it? I don't think our current standardized tests do that in a particularly fabulous way, but it's not horrible either.</i></p>
b) Validity / reliability of the	<p><i>there's an air of unfairness that's often felt, right? Like um, whether you</i></p>

exam	<p><i>do really poorly and you study really, really hard, and, and, you know what I mean? Like there's a lot of room for certain questions to show up and just because you studied three years ago 's practice tests instead of last year's, like that student that spent way less time preparing and doesn't understand as much might do way better than a student who does deeply understood, you know what I mean? Who just hasn't seen that particular kind of question before. Um, and so I definitely think that there, there would be that feeling of, well, I don't know what I'm going to get. And there's almost that feeling of why even put the time into this more, if it's a gamble.</i></p> <p>Also, see 4b)</p>
c) The repetitive nature of questions	<p><i>We've done tons of practice tests with lots of students that I work with. So I would say, typically, when I'm going over practice tests especially, I can say this is always on the test or there's always something like this, or they really like these kinds of questions.</i></p>
d) Moderation	N/A
3. Instructional issues	
a) Using old exams to prepare	<p>See 2c), and:</p> <p><i>I just say this is a really good question. They will likely give you something like this maybe with a different function or maybe with a different whatever</i></p>
b) Timing throughout the year	<p><i>Cleo: you can't always just focus on this one topic and then let them scrape by on the rest. So sometimes I have to move on even though I know they haven't 100% grasped that yet because I need like I've now got them to a point where they can start something and throw something down and make a few guesses</i></p> <p><i>DE: then movin' on...</i></p> <p><i>Cleo: Have to, right?</i></p> <p>And</p> <p><i>...Maybe if the teacher couldn't cover that material in enough time</i></p>

	<p><i>because they felt pressed then yeah, maybe the kids wouldn't feel motivated and the teacher wouldn't feel motivated to teach it and then that could drop off.</i></p>
<p>c) Teaching specific test-prep skills</p>	<p><i>Or like, oh, that was really intelligent, the way they asked that. Did you like, did you notice that? Um, they only gave you this, but really they actually gave you this, this and this. I like to draw more connections like that.</i></p> <p><i>And</i></p> <p><i>So I spend a lot of time with students about kind of seeing the similarities between "slope", or how the "a" value impacts what the function looks like or, um, I think it's so beautiful that you can always plot points. Always. Like, it never occurred to me when I was a student that if you're not really sure what the function is, just throw down some points and see if you can get a good idea. So I like giving the students practical tools like that. So even if they're so focused on memorizing the formula for whatever reason, I really like giving them these skills that are actually applicable to all functions or something that's very familiar and similar to all functions, so that they don't panic as much. That's one thing I know in particular that I focus on.</i></p>
<p>d) Teaching throughout the year</p>	<p><i>Because there are certain topics that I think are more valuable than others. Maybe it's a personal preference, maybe it's just what I think the students will more likely use, or whatever, that I think would be better to spend more time on, to make sure that they really understand. Whereas a lot of times, when I'm tutoring, especially for a standardized test, I'm really just trying to make sure they've got something to go with for all the topics. Um, you can't always just focus on this one topic and then let them scrape by on the rest. So sometimes I have to move on even though I know they haven't 100% grasped that yet because I need like I've now got them to a point where they can start something and throw something down and make a few guesses</i></p> <p><i>And</i></p>

	<i>But if it wasn't a standardized test, I think there would be more room to say, no, no, this is actually a really important life skill. We should focus on this. Don't worry about the rest of that stuff. This is what you need to focus on and the rest will come later or whatever. So...</i>
4. Learning issues	
a) Student stress	<i>For sure. Especially because, um, there is that fear of: my teacher isn't writing this, there's no control over what's on this. There's that fear, right? Like, what if my teacher has forgotten something? I mean there just is that fear.</i> <i>Also, see 3c)</i>
b) Students not learning effectively	<i>Like I got hundreds in math without really understanding it. Um, because I had to write those standardized tests, because the teachers had to teach for that, and they did a great job doing that. But did I do a great job of learning it? No. And I got away with it and I thought I was great in math and it was a huge shock moving forward.</i> <i>And</i> <i>Um, I really don't feel and I had wonderful teachers, don't get me wrong, but I really don't feel like they taught for understanding. They taught to cover material because they had to. Because there's so much to be covered and there are skills that you need if you're going to go further in math and you need, if you're going to go into psychology or whatever you want to go into. But I went through my entire math degree, not really understanding what this or that or this was and really stumbling through it, because I was super good at regurgitating. I was super good at teaching the surface... or understanding and, you know, displaying what I knew about the surface material, which wasn't good for much, right? Wasn't going to help me in my future career. Like I really struggled.</i> <i>Also, see 2b)</i>
c) Developmental capacity of students	<i>I think they're all kind of young for this kind of stuff and these techniques are all kind of superficial ... I actually think that it's a bit young for them to be... Not that they're incapable... I just, from my experience, those</i>

	<i>techniques have not proven to lead to deeper understanding one way or another.</i>
5. Systemic issues other than the exam	N/A

Summary of Cleo’s interview.

Cleo’s ideal perception of mathematics education involves preparing “students for their future[s] with the most realistic applications possible”. It seems that this is related to her own high school mathematics educational experiences as a student, in which she felt that despite receiving good grades, she never gained the “deeper understanding” that she felt would have helped her in her post-secondary studies. In her experience, having uniform exams in mathematics classes, even those with “wonderful teachers”, necessarily promotes teaching and learning in a way that leads to improved exam results over a deeper conceptual understanding. Cleo is a middle school teacher and, as such, she brings a slightly different perspective to the study. She does not teach a class with uniform examinations, and so has less interaction with administering the exam and the processes of moderation and transformation. Since she tutors students who take these courses, though, she has some different insight into student perspectives as well as experience with individualizing instruction for students in these courses.

Exams themselves (format, content & nature).

Cleo believes that exam multiple-choice questions are problematic because they leave room for students to guess and they do not provide any process evidence to help assess a student’s understanding. She explicitly states that the “process” is the part of a student’s answer that she cares about and that she “actually very rarely care[s] about the answers”. In addition to these concerns, Cleo finds many multiple-choice questions particularly tricky, or “sneaky” and makes reference to “last year’s exam” (the June 2017 exam) which was understood to be a particularly difficult exam. Despite these problems, Cleo points out that she does like certain potential characteristics of these types of questions, namely that a multiple-choice question can “guide students” to “remember their units... or to remember ... a formula that’s used there that they might be able to use later”, or to generally “trigger something which is useful” and “build confidence”. She does specify, however, that to generate these benefits, a multiple-choice

question would have to be “done in a clever way”, and she later says that these current exams do not contain particularly fabulous or clever multiple-choice questions.

In terms of the short and long-answer sections of the exam, Cleo explains that she likes the fact that students have to work through these problems. She describes content-specific examples of the types of questions that she likes best, including types of questions where very little information is given but where everything comes together, and types that combine different concepts together or which use different ways of understanding the same concept in a single question.

Cleo calls into question the validity of the uniform mathematics exams when she implies that success on these exams should predict future success in math studies, but says that, through her own experience, that they seem not to. She recalls that “I got hundreds in math without really understanding it” and that she then “went through my entire math degree, not really understanding what this or that or this was and really stumbling through”. She says that having a uniform exam promotes teaching to do well on the exam at the expense of teaching to understand more completely. She implies that a truly valid assessment would measure real mathematical understanding, and would necessitate the type of instruction and learning that align with that type of understanding.

Instruction.

When she tutors students in courses with uniform exams, Cleo makes use of old exams for practice. When she does, she tries to help students recognize the kinds of questions that are likely to be asked. She does not do this in the way that a teacher might have students memorize a group of questions, but rather she helps her tutees develop the meta-skills to recognize what qualities or characteristics make a “really good question”, or one in which “the way they asked that” was “really intelligent”. Cleo’s experience tutoring these courses is a different one than teaching a class of students. As a tutor, Cleo gets to know an individual student well and can tailor her instruction to his/her specific individual needs, she also feels time pressure though, because she still has a finite time to work with students, and sometimes she is called to do “last-minute tutoring”. Because of a combination of knowing an individual tutee well and being constrained by time, sometimes Cleo says she unfortunately gets to a point where a student does not understand a topic 100%, but she has to move on and get her students “to a point where they can start something and throw something down and make a few guesses”.

Student learning.

Cleo notes that in her experience, students are stressed about uniform math exams, more so than non-uniform end-of-year exams. She believes that a lot of this stress comes from students' fear that their classroom teacher is not the one writing or (fully) evaluating the exam. Because of that, she has observed that students worry about having “no control over what's on this” exam, and specifically worry about whether or not their “teacher has forgotten something”. Another part of student stress may come from Cleo’s opinion that students are simply not developmentally ready to properly handle an exam of this difficulty level. While she believes that students are capable of passing and even doing well on the exam, she thinks that they are “kind of young” to use the required techniques in a way that will lead to deeper understanding. In fact, Cleo believes that regardless of their developmental level, having a uniform exam leads students to prepare in a way that does not enhance their conceptual understanding.

Daphne’s Interview Data Table and Summary

Daphne is a female middle school mathematics teacher and she has been a teacher in Québec for 11 years. She holds a bachelor’s degree in honours mathematics from Queens University as well as a teaching certificate in mathematics and English from the University of Strathclyde in Glasgow. She is currently enrolled in a master’s program in mathematics. She has taught grades seven to 11. She once taught a course with a uniform math exam.

Table 5. Interview responses contributed by Daphne

Theme/subtheme	Participant response
1. Exam format	
a) Multiple-choice questions	<i>I really don't understand why there is that multiple-choice section, to be honest with you. I don't see why and I believe... they used to be four marks or zero marks. I don't know if that's still the case... but to me, I mean if, if we....going back to my philosophy of education, like I don't, I put zero emphasis on the answer. To me the answer is not the important thing and the multiple-choice is all about the answer. So I don't like if someone makes a small... and the way the multiple-choice is laid out is to kind of to trick you, because if you make the mistake they expect you to</i>

	<i>make then you'll pick one of the answers, right? So like there, then you get zero marks instead of four. So to me, that is taking away the entire process and making it about the answer, which I think is ridiculous.</i>
b) Long-answer questions	<i>... I do think from, from what I've seen from some of those sort of longer questions at the end, I have been impressed at the way that they make you tie together information from different units and... not all the time, but I think sometimes, definitely yes.</i> <i>And</i> <i>Some of the long-answer questions I think are great. Like I remember using, I've tutored for this level as well, and I know that looking back on old exams, those are questions that I've seen, is that: 'oh those are great, I'm going to use those in my class'... From what I've seen it. I think those are some really good questions that I think applied, I mean, they're not like a full on situational problem, but they do take from different areas and put it together. Right. Which I think I, I like that.</i>
c) Short-answer questions	<i>... But it's just the multiple-choice and even the short-answer, I just don't really. I don't really get the purpose of that.</i>
2. Nature of the exam	
a) Existence of a uniform exam	<i>... but I do think that it, it rewards you for being good at taking a test... So I think that it does put an emphasis on test-taking as opposed to really mastering the material.</i> <i>And</i> <i>I know that nothing shows you if you've mastered something better than teaching it. Right? But how do you evaluate someone's teaching of a topic? Um, I don't know. I think if I could think of a better way... because right now I give very traditional tests that are not standardized test. But they are very sort of typical old-school math tests and just because I can't think of a, a better way to do it, but I, I think that it's out there. I don't know what it is.</i>
b) Validity / reliability of the	<i>...my marks were better than probably my understanding of the material</i>

exam	<p><i>because I knew how to take these tests and I did a good job. I studied for the test and I excelled. It wasn't until I was teaching that I realized: 'Oh, I don't really understand this material'. Like I better go back and figure it out.</i></p> <p>And</p> <p><i>... do they pick the material on that test based on the most important things? Or is it just the ones that make the easiest and sort of test questions? I don't know. Not the easiest, but the easiest to make an interesting test question out of it. That lends itself to a test question. You know, I, yeah, I don't know. It's hard to say.</i></p>
c) The repetitive nature of questions	<p>See 3a), and</p> <p><i>... especially if you look at the past like three or four years, then you would really see. You could, you could have pretty good idea.</i></p>
d) Moderation	N/A
3. Instructional issues	
a) Using old exams to prepare	<p><i>From what I know of the sort of standardized tests, for me, the only way you would prepare was by doing old copies of standardized tests. Like for me, you would see patterns that emerged year after year and so I feel like the way to prepare for that test is, maybe more so than making sure you really understand things, is just trying to figure out what they're going to put on it based on what they've put on it before.</i></p>
b) Timing throughout the year	<p><i>I think if I was teaching a course where I knew there was going to be a standardized, I would feel more stressed because I would know that if I left something out and it was on the exam, that that's on me and I would think it would leave less, sort of freedom to deviate from the curriculum. If you want to do something that is of interest to your particular class, you can't, you can't do that.</i></p>
c) Teaching specific test-prep skills	N/A
d) Teaching throughout the year	<p><i>And I think I'm lucky that I, in that teaching grade seven, eight, I don't need to [teach content in a specific way based on how it is assessed on</i></p>

	<p><i>the exam] because I'm the one making the test. But if I were teaching grade 10, would I do that? Yes, I would.</i></p> <p>And</p> <p><i>...But it would, it would be hard not to test to the exam when you know, it... sorry to teach to the exam when you know the importance that everyone puts on that exam right now.</i></p> <p>Also, see 3b)</p>
4. Learning issues	
a) Student stress	<p><i>And I can tell you, my students ask me that like: 'Miss, are you making the test?' and for some reason it's so reassuring to them that I am, even though I don't know why, but the thought of some someone else making their exam, it's very scary for them.</i></p> <p>And</p> <p><i>... I think I was probably that type of student in that, high stakes made me want to rise to the occasion and perform even better. But for someone who was already a bit anxious, I think they would have absolutely the opposite effect.</i></p>
b) Students not learning effectively	See 3a)
c) Developmental capacity of students	N/A
5. Systemic issues other than the exam	N/A

Summary of Daphne's interview.

When I asked Daphne about what she thinks is the purpose of assessment in education, she replied: “what do I think it *is* or what do I think it *should* be?”, because, she continued, “I do think they're different”. She explains that in an ideal world, assessment would be ascertaining whether or not students can take what they learned and “apply it in a different way” or “in a different scenario”. She says that in the real world, assessment often ends up measuring whether or not students can “repeat back” and “reproduce things” that they have seen. She says that the uniform mathematics exams are assessments from this real world rather than her ideal, and that

they promote the practice of students “trying to figure out what [the government is] going to put on it based on what they've put on it before” rather than “really understand[ing] things”.

Exams themselves (format, content & nature).

Daphne does not like multiple-choice questions mainly because of the importance she places on the process of solving a problem as opposed to the final answer. She says that in her opinion, “the answer is not the important thing and the multiple-choice is all about the answer”.

Furthermore, she thinks that “the way the multiple-choice is laid out is to kind of to trick you” which makes these questions even less appropriate. Though Daphne only briefly connects this idea to the short-answer section of the exam, it seems principally consistent that she would have a problem with that section too, since its focus on the answer over the process is similar to what happens in the multiple-choice section. Because of her belief that the process is more important than the answer, she takes issue with the all-or-none nature of these questions, and the idea that a student who makes a small error would get “zero marks instead of four” which she thinks “is ridiculous”.

Also consistent with the importance that she places on process over final answers, Daphne likes the long-answer section of the uniform exams. In fact, she goes so far to say “looking back on old exams”, the long-answer questions are the ones that make her think: “those are great, I'm going to use those in my class” which is a high level of praise from an experienced teacher. In particular she like that many of these questions combine topics from various units into one question, that students cannot guess, that they get to show their process, and that their work is evaluated.

Daphne questions the existence of a uniform exam because she believes that it is necessarily invalid. As mentioned above, she believes that having a uniform exam “rewards you for being good at taking a test” which she juxtaposes with “really mastering the material”. She brought up her own experience as a student, where she felt that she always tested well, but that she didn't gain a mastery of the course content by taking the course with a uniform exam.

Instruction.

Daphne believes that teaching a course with a uniform exam inherently affects the way the course is taught. In terms of using old exams to prepare, she remembers, from her experience, that “the only way you would prepare” for a uniform exam “was by doing old copies of standardized tests”. In terms of how this affects the day-to-day instruction of the course, Daphne

reminds me that she does not currently teach a course with a uniform exam, but that if she did, “it would be hard not to ... teach to the exam when you know the importance that everyone puts on that exam right now”. She implies that the importance of the exam for teachers, students, parents, and school administrators would put pressure on her, as a teacher, to make sure that students were specifically well-prepared to write the exam, something she believes happens to other teachers in the province.

Daphne also believes that uniform exams have an impact on a teacher’s timing and level of stress. She says that, teaching a course with a uniform exam, “I would feel more stressed because I would know that if I left something out and it was on the exam, that that's on me”. She understands that there is a lot of content in these courses, and that running out of time to cover one of them means that students will be disadvantaged on the exam. She expands on this idea to say that the time constraints “leave less, sort of freedom to deviate from the curriculum”, that even if “you want to do something that is of interest to your particular class, you can't”.

Student learning.

As mentioned above, Daphne believes that students can do well on a uniform exam without mastery of the course content. Part of this belief has to do with her experience seeing question-types frequently repeated on exams. Though she answers that some topics do not necessarily repeat every single year, she clarifies that “especially if you look at the past like three or four years, then you would really ... have pretty good idea” of what types of questions would make up the uniform exam. Consequently, she believes that the way students “prepare for that test” is about “trying to figure out what [the government is] going to put on it”.

Daphne believes that students are generally stressed about uniform exams, and she brings up the idea that this stress is not equal for all students. She remembers that as a student herself, having the stress of a uniform exam made her want to “rise to the occasion and perform even better”, but she acknowledges that “for someone who was already a bit anxious, I think [uniform exams] would have absolutely the opposite effect”. Daphne also brought up a specific aspect of uniform exams that she believes causes additional student stress: the fact that the exam is written and partially corrected by someone other than the classroom teacher. This is similar to an explanation given by Cleo about student stress, and it may have to do with the idea that over the school year, trust is built between students and a classroom teacher, student comfort and confidence grow, and that is threatened by an unknown quantity.

Edna's Interview Data Table and Summary

Edna is a female senior school mathematics teacher (she usually teaches one math class per year) in addition to being a full-time school administrator. She has been a teacher in Québec for 29 years. She holds a bachelor's degree in computer science with a minor in mathematics, a master's degree in educational technology, and she just completed a Doctor of Education (EdD) program. She has taught grades seven, eight, 10, and 11, and she has taught courses with uniform exams for 26-27 years.

Table 6. Interview responses contributed by Edna

Theme/subtheme	Participant response
1. Exam format	
a) Multiple-choice questions	<p><i>I'm not a fan of the multiple-choice questions, but do they assess the wrong skill-set? Not necessarily.</i></p> <p>And</p> <p><i>I know that they're tricky. They're tricky to do. Considering what they're trying to do with the standardized tests, I understand why they work.</i></p> <p><i>I think there had been some multiple-choice questions that are ambiguous.</i></p> <p>And</p> <p><i>But then that doesn't make up... that's only six questions. If the entire exam was that way, I'd object more to it. But the fact that... I don't think there's anything wrong with differentiating a student who's a solid 80-85 with a student who really can get a 96... And if it was an entirely multiple-choice exam, then yes, I would have a bigger problem with it.</i></p>
b) Long-answer questions	See 4c)
c) Short-answer questions	N/A
2. Nature of the exam	
a) Existence of a uniform exam	<p><i>I don't object to the tests themselves. I have no problem with the concept that a standardized test evens the playing field so that all students across</i></p>

	<p><i>the province have to meet the same standard. That's fine by me. I just don't like what happens after... the fact that teachers are judged based on the results... or schools are judged based on the results. I don't like what people do with the results.</i></p> <p><i>And</i></p> <p><i>And I don't object to content knowledge. I know a lot of people say, "well, there's too much content", or "we're teaching too much content" ... I kind of... I have no problem with that. I think that people do need to know things.</i></p> <p><i>And</i></p> <p><i>[the test is] well-constructed, the curriculum is solid... there's not much I would change in the course</i></p>
<p>b) Validity / reliability of the exam</p>	<p>See 2a)</p> <p>Also,</p> <p>DE: <i>Do you think it's possible if there are students who never gained a very good understanding of the course content, but they're good at writing this type of test?</i></p> <p>Edna: <i>Not this test... I really don't believe that if they haven't mastered the content... then I don't think so. That's my opinion.</i></p> <p>DE: <i>Do you think that there are students who have a legitimate mastery of the content who, for whatever reason, don't do well on a standardized test?</i></p> <p>Edna: <i>Um, I think it happens. I think we do see it sometimes. I think that function of stress.</i></p>
<p>c) The repetitive nature of questions</p>	<p>DE: <i>Do you know of any types of questions, you don't have to give me a specific, just a yes or no. Do you know of any types of questions that frequently recur on those types of standardized tests?</i></p> <p>Edna: <i>Yes.</i></p>
<p>d) Moderation</p>	<p><i>I don't object to standardized tests in general, what I object to is the aftermath and the fact that... what I object to is the moderation and the fact that we don't know how they're going to end up doing, the</i></p>

	<p><i>government takes that into their own hands. I object to the fact that there are the seemingly, although they're probably not, arbitrary cut-offs at CEGEPs for applications. I object to the fact, on the greater scale, that teachers and principals and schools have funding associated with the results of standardized tests. That's what I object to. I don't object to the tests themselves.</i></p>
3. Instructional issues	
a) Using old exams to prepare	<p><i>Edna: I mean, any other test, I always model after the conditions that they're going to have to do. I mean the course is designed specifically so they can do well on that test.</i></p> <p><i>DE: Do you think a teacher is more likely to spend extra time doing old tests, practice tests, doing test taking strategy?</i></p> <p><i>Edna: Yeah, absolutely. All the time.</i></p> <p><i>And</i></p> <p><i>I don't really see a problem with it. And I know people say you're teaching to the test and you're not expanding their mathematical ability, but if... there's not much I would change in the course.</i></p>
b) Timing throughout the year	<p><i>Edna: I mean I think, I think that if you're teaching one of these courses, you probably do forego other opportunities, other topics. Our talk of more focus on experiential education and stuff you could do with math... We probably forego a lot of that because of the time pressure, because we only have so many classes to prep for the exam. So yeah, I think it does impact how you teach.</i></p> <p><i>DE: And do you think that a teacher is likely to budget their time according to what things are more likely to be on the exam?</i></p> <p><i>Edna: Yeah absolutely, which again, is a reasonable thing to do.</i></p>
c) Teaching specific test-prep skills	<p><i>DE: do you, as a teacher, do anything specific to help students train for a standardized test versus any other tests?</i></p> <p><i>Edna: I don't know, because I've always taught boys the course with that test, so I don't know. I mean, any other test, I always model after the</i></p>

	<i>conditions that they're going to have to do. I mean the course is designed specifically to so they can do well on that test.</i>
d) Teaching throughout the year	See 3b) and 3c) And <i>I know people say you're teaching to the test and you're not expanding their mathematical ability, but if... there's not much I would change in the course.</i>
4. Learning issues	
a) Student stress	DE: <i>Do you think that students are generally stressed about standardized tests?</i> Edna: <i>Yeah.</i> DE: <i>Do you think that's, um, exclusively due to the phenomenon you were talking about earlier? How it's a gatekeeper thing? Are there other things as well?</i> Edna: <i>That's probably one of the biggest factors, yeah.</i> DE: <i>do you think that having a high stakes standardized test has the potential to be a motivator for some students to do well?</i> Edna: <i>For some, it could be, yeah. They might put more time in, they might take it more seriously. I think there's something, particularly in our setting with the private schools, with parental involvement... parents make a phone call to say, well, you know, it's out of our hands. We can't control it. Sometimes that's a good thing, you know, because parents can put a lot of pressure on to change the grade, you know if the exam wasn't easy. I'm sure parents do it everywhere, not just our setting.</i> DE: <i>Do you think that there are students who have a legitimate mastery of the content who, for whatever reason, don't do well on a standardized test?</i> Edna: <i>Um, I think it happens. I think we do see it sometimes. I think that function of stress.</i> And Edna: <i>Assessment has become gatekeeping for the next step in education</i>

b) Students not learning effectively	N/A
c) Developmental capacity of students	<p><i>I mean, content across the world tends to be varied developmentally. So you know, 15 year old are capable of understanding functions, you know? 14 to 15 year olds can handle functions. Sixteen to 17 year old can make it a bit higher. And 13-14 year olds are starting to think more abstractly, they can handle algebra. That's true across the world. So...</i></p> <p><i>I have a real problem with the conjecture questions. I don't believe that kids at that age, 15, 16 years old necessarily have that level of abstract thinking developed to answer conjecture questions... I don't think it should be in the course at all. And therefore, it shouldn't be on the exam... I just don't think they're ready for that thinking. They don't know how to answer it, they're not familiar with proofs. Yeah, and they generally do very poorly on them.</i></p>
5. Systemic issues other than the exam	See 2a) and 4a)

Summary of Edna's interview.

In our interview, Edna repeatedly situated uniform exams in a larger context than the classroom. Perhaps her administrative experience and recent graduate studies continue to shape her opinions and beliefs about uniform examinations and broaden the lens through which she sees them. She looks beyond the impact on students and teachers when she brings up the idea of “parental involvement”, referring to the fact that some parents call teachers or school administration trying to influence their children’s results. She also situates the idea of level-appropriate course content globally, when she brings up the idea of what skills a student should be able to master developmentally based on age. She thinks that the skills and knowledge required to be successful on the Québec mathematics uniform exams are comparable to those required in other provinces and countries around the world for students of the same age. In her big-picture conception of uniform exams in Québec, Edna sees them overall as being effective forms of assessment.

Exams themselves (format, content & nature).

Overall, Edna believes that the Québec uniform mathematics exams do a good job of assessing students' knowledge and skills. She believes that the content (other than conjecture questions) is level-appropriate and that the exam is valid, in the sense that it prevents false positives.

Specifically, she says that this test does not allow for a student who has not really mastered the content to do well. On the flip-side, though, she believes that there are exceptional cases of students who have mastered the content, who sometimes do poorly on the exam. This, she thinks, is an effect of student stress, however, rather than a general failing of the exam. She also implicitly makes reference to the reliability of uniform math exams, when she talks about how exam difficulty varies from year to year. Edna refers to years in which the “exam wasn't easy”, speaking to her experience with the variability in the difficulty level of the exams. This idea means that two students with the same level of understanding and skill would not necessarily obtain the same result in two different years.

In terms of the format of the exam, Edna does not, overall, like the multiple-choice section of the exam. She finds the questions purposefully “tricky to do” and notes that in her experience, there have been “some multiple-choice questions which are ambiguous”. However, despite her negative impression of this type of question, Edna looks at the big-picture and says that the imperfections of the multiple-choice questions are largely mitigated by the fact that there are only six of them. She explains: “I don't think there's anything wrong with differentiating a student who's a solid 80-85 with a student who really can get a 96”. She adds that “if it was an entirely multiple-choice exam ... I would have a bigger problem with it”.

Instruction.

Edna brings up an interesting philosophical point when she says that she doesn't know if she teaches any differently due to the uniform exam, because she has always taught the course with the exam. She does, however, point out that she believes that uniform exam courses are “designed specifically so [students] can do well on that test”. She also says that her primary role as a teacher is “to do whatever it takes to help kids be successful”. This is probably why Edna says that she always tries to “model” her assignments and assessments “after the conditions that they're going to have to do” on the exam.

In her roles as a teacher with many years of experience and as a school administrator, Edna believes that teachers in Québec spend extra time reviewing old exams compared to teachers teaching non-uniform exam courses. She says that some people might take issue with

the fact that teaching this course leads to a form of “teaching to the test”, but she also says that the because the test covers the course content well, and since the course content is level-appropriate, that in this situation, it is okay to teach to the test. As a final consequence of uniform exams on instruction, Edna believes that teachers “forego” other educational opportunities and other topics because of time constraints to ensure that students are well-prepared for the final exam.

Student learning.

Edna believes that there is a stress associated with uniform exams, particularly because of the high-stakes involved, specifically that results affect students’ entrance to post-secondary educational institutions. Her main issue with student learning, however, is not related to stress, but rather to the developmental level of students and what they are/are not able to do developmentally at the age when they write the uniform math exam. She believes that in terms of most of the course and exam content, the difficulty level is appropriate for students this age. However, she believes that conjecture questions are beyond students’ ability. A conjecture question is one in which students have to discover and justify a general pattern between two or more quantities. The government marking guide states that one allowed method of justifying a student’s work is for the student to generate three different numerical examples that each meet all of the criteria of the problem. Edna says that this “proof” type of question requires a “level of abstract thinking” that students at this age typically do not have and that this is why students generally do poorly on this type of question.

Other systemic issues.

As mentioned above, Edna discussed the phenomenon in which parents sometimes call to inquire, influence, or complain about a student’s results. She says that “parents can put a lot of pressure on to change the grade”. She acknowledges that this effect is one which is frequently seen in a private school setting, but also says “I’m sure parents do it everywhere, not just our setting”. She thought that the more objective nature of the exam and the fact that the grades were handled externally allowed her to tell parents: “well, you know, it’s out of our hands”, which she thought helped to limit the influence and pressure of parents.

Flora's Interview Data Table and Summary

Flora is a female senior school mathematics teacher who has been a teacher in Québec since the fall of 2009. She completed an undergraduate degree in biology with a minor in socio-cultural anthropology, a master's degree in non-clinical based psychiatry, and a master's degree in secondary mathematics education. She teaches grades nine to 11 and she has taught courses with uniform exams for two – four years.

Table 7. Interview responses contributed by Flora

Theme/subtheme	Participant response
1. Exam format	
a) Multiple-choice questions	<p><i>...the nature of it is an all or nothing, all right or wrong. The multiple-choice also lends itself to a student getting full credit for having no idea what to do, simply by guessing... 25 percent chance... there's four options ... sometimes the questions, maybe just the first question is simple enough, that it gives the student confidence. So sometimes that helps a little and get the student started. When I mark these exams, maybe the first one or first two are usually pretty strong. So at least it gives the students some confidence going into a difficult exam, which is a pro, so to speak. But then what ends up happening is that by question three, four, five, six, maybe there's too many answers that are similar. It's a little bit vague, it's not obvious. And then the student who probably knows what to do and probably if they were given an opportunity to explain their work, show their work, justify their work, you'd probably see that maybe they did understand 80 or 90 percent of the problem, but because of the nature of the scoring, they get it all wrong if they're just a bit off.</i></p> <p><i>And</i></p> <p><i>Usually the multiple-choice questions are not questions that require seven or eight steps. It is a one or two-step skill. So it does check whether the student understands the fundamentals of one or two skills. So what I do find is that the multiple-choice does cover topics throughout the year, but at a superficial level.</i></p>

<p>b) Long-answer questions</p>	<p><i>My experience with the long-answer questions is that for the most part, the response from the students is that they felt prepared. That the topics, the content, the style looks familiar, so it's predictable. They felt confident applying their skills. Uh, the rubric lends itself to some professional judgment as well as some specific detail of to have a mark it. So in terms of the long-answer, I think that, excuse me, that those topics are well chosen. They're predictable and in my experience, the students do well if not at least reflective of how they've been doing the rest of the year.</i></p> <p><i>And</i></p> <p><i>I think the rubric is in some ways very detailed and other areas very vague, because it happens so often when we are marking in the office, and a student does something and we don't see that work reflected anywhere in the marking guide and it's up to the teacher's professional judgment, the advice and feedback she gets from her colleagues about what they think is happening in this problem. So that might just be the nature of the job and that happens, but it would be nice maybe if the marking guide was a little bit more detailed or allowed for different iterations of how the problem could be solved. So I don't know if part of the testing for these exams could be actually some, focus groups or sample students completing these things and then looking at actual feedback of real kids solving these problems and maybe that dictating the marking guide a little bit.</i></p>
<p>c) Short-answer questions</p>	<p><i>I think an all or none question has to be a one step. Identify this ratio, identify this side. The questions about which pair are congruent or similar. Something that is really a one step, but anything that requires multiple steps, you're not getting a good sense. To say that the student gets zero when he can do three out of four steps is not really reflective of his ability. And what happens in the short-answer section is that especially in math, where the problems are notation heavy, but also conceptual understanding is required and reasoning and choosing the</i></p>

	<p><i>correct strategy, and then just carrying out the correct multiplication or sorry, calculation, there's so many different facets, so to say zero or two or four and the student can drop down to two out of four... 50 percent because a bracket was the wrong direction or there was just a miscalculation, or um, or the direction of an inequality was incorrect... So just minor things to drop you down... to say that now the student just understands 50 percent of that question is not actually reflective of his or her ability.</i></p> <p><i>... . I like the fact that they can't really guess, they have to show some understanding of what the question is asking and what strategies are required to solve it. It's the marking that I find very frustrating.</i></p>
<p>2. Nature of the exam</p>	
<p>a) Existence of a uniform exam</p>	<p><i>The content covered in the progression of learning appears in the standardized testing. Sometimes that content is not as overt or obvious to the student in the way that they've seen it before... So the mathematical topics and skills are covered, but sometimes they are hidden or maybe a little bit more obscure.</i></p> <p><i>And</i></p> <p><i>Anecdotally, I hear students say grade 10 is serious, grade is time to buckle down. I've heard them say things like grade 10 counts more. So, you could argue that this is putting more pressure on them, but you could argue that it's reminding them that it might be time to take your studies more seriously and understand that the things you're doing, do have effects and real consequences. And I don't think there's anything wrong with that. And as long as the students are in a community and among teachers and peers who support them and believe in their ability and help to instill their confidence, then I don't think that the added stress necessarily needs to be a problem.</i></p> <p><i>Also, see 3a)</i></p>
<p>b) Validity / reliability of the</p>	<p><i>And in my experience with scoring, I often see that the way that the</i></p>

exam	<p><i>students score in the multiple-choice section is not at all similar to the way in which they score in the long-answer section. They tend to do much, much better in the long-answer section where there's multiple steps, multiple opportunities to show reasoning and justify work. That, that work is much more reflective of their success over the year, and the multiple-choice section is hit or miss all or nothing.</i></p> <p>And</p> <p><i>... in terms of the long-answer, I think that, excuse me, that those topics are well chosen. They're predictable and in my experience, the students do well if not at least reflective of how they've been doing the rest of the year.</i></p> <p>And</p> <p><i>Unfortunately, my most recent experience is that the 40 percent of the exam, which is the short-answer and multiple-choice and that all or nothing, drastically skews the data to several percentage points below what I expect them to do and what they expect to see, which is unfortunate.</i></p>
c) The repetitive nature of questions	<p><i>Yes, I do see similarity year over year in terms of the type of questions being asked.</i></p>
d) Moderation	<p>N/A</p>
3. Instructional issues	
a) Using old exams to prepare	<p><i>Flora: Throughout the year, I try to incorporate old exam questions on check-in quizzes, on unit tests, throughout the year, and tell the students these are old exam questions, so they start to get a sense of the level of difficulty and the wording of the questions. And then as we get closer to the end of the year and closer to the exam, we look at these exams, um, in more detail and for longer periods of times. Exams as a whole, not just a question here, a question there ... Towards the end of the year, we practice timing more. So during the school year, I really want them to feel confident and really show me their level of mastery, and I don't put as</i></p>

	<p><i>much emphasis on timing, how fast can you do it. If a student wants to do a quiz in 10 minutes or 12 or 17 or 18, or redo a quiz... I'm not as focused on how long it takes them to master a topic, but towards the end of the year, we practice timing more. How long does it take you to do six multiple-choice? How long did it take you to do the six short-answer? So that kind of instruction is different at the end of the year.</i></p> <p><i>DE: Do you think a teacher who doesn't ever do any practice exam questions... do you think that students would be sufficiently prepared to write that uniform exam at the end of the year?</i></p> <p><i>Flora: I mean, you would hope that the answer would be yes. I don't have enough evidence to see that. I mean, I work with fantastic teachers, who employ some of the most current, best practices in pedagogy, and part of their practice is also to incorporate old exams. So I don't know... I don't have enough evidence to say that watching these teachers do incredible work, but not showing the students ever an old exam, whether they would do as well. I don't, I don't have enough evidence to say that. I would like to say yes, because it would be nice to know that people can do well on standardized tests without ever having seen an old one.</i></p>
<p>b) Timing throughout the year</p>	<p><i>Well, I guess I spend more time on topics that are more heavily represented content-wise. You kind of look at all the content and you're like, oh, these particular topics take up a large part of the year to teach, because there's more to it. So those, I guess just by nature, I end up spending more time on that...</i></p> <p><i>And</i></p> <p><i>I mean, it'd be always nice to supplement with other, with other topics. Definitely with more time, or more elective, or... it's always nice to incorporate more, uh, more topics. But, I usually spend time also... when it talks about like more time on which topics, topics that I know that the students will need for future courses. I'm more likely to spend more time on that, than more time because I know that there's gonna be a question on the exam.</i></p>

<p>c) Teaching specific test-prep skills</p>	<p><i>I will tell the student, oh, in my experience I see that there's often one multiple-choice like this and one long-answer is this type of question. I often just see a multiple-choice question...</i></p> <p>And</p> <p><i>Towards the end of the year, we practice timing more. So during the school year, I really want them to feel confident and really show me their level of mastery, and I don't put as much emphasis on timing, how fast can you do it. If a student wants to do a quiz in 10 minutes or 12 or 17 or 18, or redo a quiz... I'm not as focused on how long it takes them to master a topic, but towards the end of the year, we practice timing more. How long does it take you to do six multiple-choice? How long did it take you to do the six short-answer? So that kind of instruction is different at the end of the year.</i></p>
<p>d) Teaching throughout the year</p>	<p><i>So we might want to do a better job of teaching the [conjecture] skill in a way that's more interesting throughout the year and not just, "Oh, it's the last month before the exam, let's practice these conjecture type questions". But the way the students do it in their science class of gathering, putting together data and looking for relationships, we should do a better job of that in math class because that is actually a very strong skill.</i></p> <p>And</p> <p><i>Throughout the year, I try to incorporate old exam questions on check-in quizzes, on unit tests, throughout the year, and tell the students these are old exam questions, so they start to get a sense of the level of difficulty and the wording of the questions.</i></p> <p>And, see 3b)</p>
<p>4. Learning issues</p>	
<p>a) Student stress</p>	<p><i>I find that students stressed when they care a lot about the outcomes. So for students who care about the outcomes of these tests, it causes stress... ... I think our job as teachers is to calm them, remind them that they are prepared, remind them that they can be successful ... At our school, which</i></p>

	<p><i>is a high-performing, academic school where students do care about their academic success, the notion that there is a standardized test at the end of the year, definitely in my experience helps motivate this particular demographic because they care about it.</i></p> <p>Also, see 2a)</p>
<p>b) Students not learning effectively</p>	<p><i>I think that it depends on the guidance given by the teacher. So, if the teachers says to the student: “okay students, here's your study guide, here's your review, here are five old exams, this is our review class”, then yes, the students will do that. If the teacher guides them and said, “okay, let's start reviewing the topics, let's put the topics up on the board that we covered this year. What are the topics?” Get some of that from the student. Get them thinking about the topics. What did we learn about these topics? What skills did we need to know? Get them... teach them to study their notes first and their old tests first and their old homework first. Then let's look at these old exams and then let them relay what they know and what they've reviewed to applying it to the old exams ... Again, it totally depends on the guidance of the teacher. So if the teacher just has some stack of old exams, yes, they will learn, they will make sure they can solve that question. They will ask their friends, ask their teacher. But if, instead, the teacher puts up that question and says something like: “this type of diagram where there's lines and right angles and triangles and the final thing says, find this length”. Yes, let's walk. We can walk through all the steps and I can tell you how to find the length and then you guys could figure out how to find that length. But if, instead, the teacher prompts them, and says in our whole course, what skills do you know about finding lengths? And then get the kids to say Pythagorean Theorem or basic ratio trigonometry, or similar triangles. Then the kids are understanding all their tools about how do I find the length in this course. Oh, and then you have another student who says: “distance formula”. And then you talk about all those ways to find lengths and then you look at your diagram and say, well, do I have coordinate? No coordinates.</i></p>

	<p><i>Okay, so we're not going to do distance formula. Do we have any right angles? I see one over there. So then you keep the Pythagorean Theorem and the ratio trigonometry, you keep that on your short-list of skills. Do you see big triangles and little triangles? No we don't see any triangles. Okay, so maybe it's not a similar triangle problem. So it depends on the guidance of the teacher. Is the teacher saying, I don't know, we've never learned to solve this problem, what could we do? But if the teacher prompts them to say, what content do you know and what can be applied to this problem? ... Then they're better prepared for all the other iterations</i></p>
<p>c) Developmental capacity of students</p>	<p>N/A</p>
<p>5. Systemic issues other than the exam</p>	<p><i>Maybe some students might argue like, oh, these conjecture questions, conjecture type questions, which are questions where the student is given a particular set of constraints and they must come up with a hypothesis about the relationship between two quantities. So the question will give the students a scenario given this, given that, come up with a conjecture about the relationship between the height of this and the radius of that and the student has to put together some sort of hypothesis testing and then come up with a hypothesis. So a lot of the students would argue: "what does it have to do with the course?" This isn't any of the specific skills. But I actually think that that idea of students looking for patterns and coming up with ideas about relationships is the most powerful concept of mathematics. That is what mathematics is. That's what real mathematicians do, is that they look for patterns. That's what math is, it's patterns. So we might want to do a better job of teaching the skill in a way that's more interesting throughout the year and not just, "Oh, it's the last month before the exam, let's practice these conjecture type questions". But the way the students do it in their science class of gathering, putting together data and looking for relationships, we should do a better job of that in math class because that is actually a very strong</i></p>

	<p><i>skill.</i></p> <p>And</p> <p><i>... similar triangles and the idea of proving that things are similar, when am I ever going to need this in my life? But the idea of proving, of understanding scale models, and understanding that if you want to prove that something is true, you need to understand that sometimes there's just a minimum number of conditions you can prove to someone that are undeniable axiom, axiom concepts that are enough to prove that your claim is undeniable. Like those skills, I find very important. So even though at the surface level it might seem that these are very, you know, contrived ideas and not useful on the grander scheme, I find a lot of the skills are very applicable.</i></p>
--	--

Summary of Flora’s interview.

Several of Flora’s responses speak to a belief that teachers have a lot of potential influence/impact on students and that, consequently, they also have a lot of responsibility. She believes that uniform exams might lead students to prepare in a way that narrows their scope and promotes the wrong learning strategies, but that teachers have the responsibility of guiding students to prepare the right way. Similarly, she believes that a uniform exam course could be overly stressful for students if their teachers do not help them to feel calm, prepared, and confident. Also, when the idea of a specific exam topic (on which students do not typically do well) arose, Flora suggested that this issue could be solved by teachers integrating the topic more seamlessly during the school year.

Exams themselves (format, content & nature).

Flora repeatedly brings up the validity of uniform math exams throughout the interview. She compares sections of the exam and shares that, in her experience, the long-answer section results tend to be much more representative of how students perform throughout the year than either the multiple-choice or short-answer sections. She reflects on these less representative sections of the exam and says that though she finds some questions to be well-chosen as multiple-choice or short-answer (in the sense that they should be more straight-forward and require fewer steps to complete), not all of them are appropriately chosen. Additionally, she takes issue with the all-or-

none nature of both sections. She further differentiates between the two sections and says that she prefers the short-answer to the multiple-choice, because she likes “the fact that [students] can't really guess, they have to show some understanding of what the question is asking and what strategies are required to solve it”. She goes on to say, however, that she finds the “marking” of the short-answer questions to be “very frustrating”. She says that the marking guide is very prescriptive, and in her opinion, there are frequently cases where “minor things” like “a bracket [facing] the wrong direction”, “a miscalculation”, or “the direction of an inequality [being] incorrect”, lead a student to drop to 50% or 0% of a given question. She says that to give a student “zero when he can do three out of four steps is not really reflective of his ability”.

Instruction.

Flora believes that it is important to bring in old exams for students to practice. She wants students to be familiar with the style, “the level of difficulty and the wording of the questions”. Closer to the end of the year, she also tries to help train students to focus on timing to be successful when writing the timed uniform exam. When asked if she thinks that practicing old exams is necessary for students to be successful on uniform exams, Flora replies: “I would like to say yes, because it would be nice to know that people can do well on standardized tests without ever having seen an old one”. However, she continues to say: “I don't have enough evidence to say that”, and that, actually, in her experience, the teachers she considers to employ the most effective pedagogy bring in old exams themselves too.

Student learning.

This is the theme in which Flora's belief about classroom teacher influence really shines through. Flora says that it is possible for Québec math teachers to instruct in a way that leads students to learn a narrowed scope of the course content based on what is most likely to appear on the exam. She also acknowledges that students might be tempted to learn how to answer specific questions rather than learn the more general theory behind the question. She believes that “it totally depends on the guidance of the teacher” to ensure that students approach studying and learning in a way that leads to a deeper understanding of the course content.

Similarly, Flora believes that teachers can help students to deal with the added stress of uniform exams in a positive way. She acknowledges that students might be more stressed about the uniform exam than other assessments, although she believes that this added stress is not due to anything inherent in the exam itself, but rather that students are stressed about things they care

about or where they care to succeed. She says that it is the responsibility of teachers to “calm [students], remind them that they are prepared, remind them that they can be successful, [...] and that they’re going to do great”.

Other systemic issues.

Flora brings up some of the course topics that she believes students find less popular. It could be because students find them difficult or unfamiliar, because the process to solve them is legitimately difficult, or because students find the real-life usefulness difficult to find. These topics are conjecture-type proof questions and questions that involve triangle similarity, congruence, and general triangle proofs. Flora stands by these types of questions and provides extensive justification for why they should be a part of the course. This is consistent with her general stance that the exam/course content is appropriate, that the marking is the element of the uniform exams that causes the greatest problems for her, and that with the proper teacher stance and action, students can be successful.

Gidget’s Interview Data Table and Summary

Gidget is a female senior school mathematics teacher who is also the mathematics department head. She has been a teacher in Québec for 12 years and also taught in Scotland in British Columbia. She holds a bachelor’s degree in mathematics, linguistics, and English literature, a teaching degree obtained in Scotland, and a master’s degree in educational leadership. She has taught grades eight to 12 and she has taught courses with uniform exams for 13 years (12 in Québec and one in B.C.).

Table 8. Interview responses contributed by Gidget

Theme/subtheme	Participant response
1. Exam format	
a) Multiple-choice questions	<i>Hate it. Hate it. Brutal. A multiple-choice question for me is something that I find synonymous with maybe a history or biology, something where you’ve had to just memorize something, not necessarily a skill-testing anything. So what I’ve seen in the past is there are multi-step multiple-choice questions, which absolutely infuriate me because when a student circles their final answer, it has no indication to me what their thought</i>

process was, if they cheated, if they had a good guessing day. I mean maybe it could tell me as an exercise in class, if we talked through it, I could find out how he eliminated the other three options, which does show me part of his thought process, and I think is part of being a strong multiple-choice student, but I find in our tests, our exams, to have them be worth 4 marks or zero is absurd. And to have there be six of them... four of them? Six of them, so 24 percent of your exam, you know.

And

... I'm not sure a multiple-choice type question does the best job at evaluating the content covered in that grade 10 course.

I know that when I learned how to create multiple-choice questions, which was in Scotland, we were taught that the other three answers in the multiple-choice were called distracters, and you are supposed to predict the errors a student would make and put a little hat on it and make it look enticing and get them to circle it. And then when I was on the creation team for an exam, we certainly talked about what are likely mistakes students will make just to make it not obvious what the final answer is. I think if we had time to use multiple-choice properly, that would certainly be a technique.

And

... But the way that we do it in Québec, is that I don't mark them multiple-choice. I have to submit that to the government with a scantron, so I'm going to assume that that machine doesn't care why any kid picked 'b' over 'c', when 'c' was the correct answer. And whether or not that indicates that the students, the average student, has trouble remembering to flip the inequality when they divide by a negative, like I don't think that the government looks at that on that particular test. I think classroom teachers could when they give their assessments in class, but on that uniform test, I don't think that anyone looks at that.

And

... having been on a creation committee, the person leading the committee

	<p><i>for this provincial exam, so they wanted medium level questions, which I realized was subjective, not just easy level questions in the multiple-choice, so that must be coming from somewhere in the government. I think also what happens is after so many years of looking at those tests, you almost just get brainwashed that I've seen it 400 times, therefore matching the inequality with the shaded region is therefore in the easy category. And I think teachers, including myself, are guilty sometimes of failing to look at these through the lens of a student. Would I call it easy? Sure. Would they call it easy? I don't know.</i></p> <p><i>And</i></p> <p><i>I've seen it such that, um, yes, the first one or two are easier and they do get progressively harder and I've seen, essentially, that the multiple-choice questions, uh, could be used as clues to answer some questions later on in the test. If a student is calm enough while writing this highly prized uniform exam, to sort of have the wherewithal to be like, "Oh yeah, I've seen that".</i></p>
<p>b) Long-answer questions</p>	<p><i>I think at this level of mathematics, the grade 10... essentially intro to functions class going onto like a real pre-calculus, lots and lots of functions. A critical thinking skill is hugely important to be successful in that following year. So I think the long-answer questions certainly evaluate multi-step, critical thinking, combining concepts. So I often like those questions because I do think it is indicative of how a student will likely will do the following year, is he in fact prepared for the challenges that lie ahead? If these long-answer questions on this uniform exam have stumped him completely, if he was overwhelmed by the multi-step, he's never seen a question like that and can't apply his knowledge to a new situation... I think that's a good tool to help us throw up a flag. So generally, I like the long-answer questions. I like what they're trying to get at for the students and I'm rarely surprised with my students' results in that section because I get to sort of see, oh, he's still not squaring the binomial properly, you know, he made that error three times already this</i></p>

	<p><i>year. There it is again, you know, and I can see it. Whereas in the multiple-choice or the short-answer, sometimes I can't see that mistake, or a particular students is overwhelmed with decoding. So if there's two paragraphs that he's got to get through. I am not surprised when there's little work, reading a blank the paper because I know the student well. I think sometimes they do a little run around intro paragraph. Uh, which I think is fine. I think it's important for the kids to buy into the content a little bit knowing that there's a real world application and we can certainly do that at the grade 10 level, whereas a lot of the times when you're learning the fundamentals and the earlier grades it's, it's a bit more contrived when you try and make it a real life application. But given how many students are now coded with test anxiety and dyslexia and whatever, maybe we should be getting more to the point. I do appreciate when the questions, when the information in one of the paragraphs is already transcribed into the diagram. So for instance, if say the distance from Billy's house to Johnny's house is 56 kilometers in text that's already in the diagram. They don't need to make transcription errors. We're not evaluating that skill, presumably.</i></p>
<p>c) Short-answer questions</p>	<p><i>I think short-answer could be useful as a build up to the long-answer. Um, it may offer clues to some of the long-answer questions, it may help a teacher dissect exactly where the problem lies... the difficulty or the roadblocks for that particular student. But the way that these are marked, the provincial exam, is it can be zero or four marks... occasionally, the marking guide allows us to give two marks, but even then it's quite... quite strict. There were a few times where in our office, we discuss our marking and I think, you know, if I asked this question on a class test and marked it out of four, I would give the kid, given his mistakes and what I saw on the paper, three on four, even sometimes three and a half on four, but given the marking guide for the uniform exam, I'm not allowed to look at his work, I'm only allowed to look at his final answer and it's written in a particular line of a particular booklet. It ends up being zero on four. So</i></p>

	<p><i>does that zero really indicate his understanding of the material? So I think the short-answer can prove useful, but the marking guide makes it difficult.</i></p> <p><i>And</i></p> <p><i>I think it would be easy to come up with a marking guide similar to the long-answer questions. I know when we co-teach, so when like Edna and I worked together last year, we sat down and said, okay, before we came up with how much a question would be worth, we do our solutions first. If it was a four step question, we would put it out of 4 marks... one mark for each step, or if one of those steps was pretty hefty, it would be a five mark question. It was very clear where we would indicate our part marks, so perhaps not a formal rubric, but nevertheless we discussed exactly what a four on five would look like, or what a 3 on 5 would look like, and then there are some answers that that deviate from that and it's nice to have a colleague to bounce ideas off of. You can't do that with exam. You have to just follow the guideline and trust that your fellow teachers across the province are being just as rigid with following the guidelines.</i></p>
<p>2. Nature of the exam</p>	
<p>a) Existence of a uniform exam</p>	<p><i>I don't think there's anything wrong with summative assessment. I think when used properly, it can provide the same, the same feedback. It's certainly a succinct way of saying things. If you score 75 on 100 on a test that encompasses an entire topic, presumably the safe statement would be to say that there's still 25 percent of the material that you're shaky with, but that would mean... it would have to be a good assessment, right?</i></p> <p><i>And</i></p> <p><i>... I think math being a cumulative topic warrants a uniform assessment at the end of the year, perhaps more than certain other subjects. Because it is important that a student moving on to the next level of mathematics does have quick recall whether that's for a fact like, this is what an ellipse looks like, or for a skill: this is how you complete the square.</i></p>

	<p><i>But I think the one test is pretty harsh. I do think math, as I said before, being cumulative... It is important that on any given day, if you want to be successful in calculus, you have to recall ABC really quickly, and you know, that day, that's sort of synonymous with the day of your exam, that day, three weeks later when we start calculus all, you need to recall everything. If you couldn't do it on the exam day, is that an indicator that your path should be in this direction? I do like what we're doing now and allowing students to have the re-write opportunities. So the proverbial door is never really locked. Just the path to the door is a little bit different because of maturity ... Because it's math, I think there should be a cumulative exam, perhaps more opportunities... June, July, August before the next year.</i></p> <p><i>And</i></p> <p><i>You know? I think it's okay, and this is sort of, I guess the premise of the Québec reform from a few years back, that if you didn't learn it in September but you, you pick it up, something clicks for you by June, and you can produce this amazing exam. You know, I think that's, that happens and that's, that speaks to the kids' resiliency. So the provincial exam can have a positive effect, not just, not just negative.</i></p>
<p>b) Validity / reliability of the exam</p>	<p><i>Gidget: You know if... they do this with the advanced placement calculus exam, they'll sort of say all... they'll rank the kids with whoever got a five, four, three, two, and one score. They'll say for question one, everyone who got a five got it right. Everyone who got a one got it wrong, so arguably the easy question number one was easy for those that got a five, four, three, two, and those that got a one probably shouldn't be writing that exam. It's interesting to see that and then they'll even make a move, this is still college board, they'll even make a move that if the kids who scored a five, if the majority of the kids who scored a five got a question wrong, they may remove the question from the assessment.</i></p> <p><i>DE: they figure that it's not valid</i></p> <p><i>Gidget: Yeah, they'll figure is not valid or some answer was ambiguous or</i></p>

	<p><i>whatever. Trust the kids. But the way that we do it in Québec, is that I don't mark them multiple-choice. I have to submit that to the government with a scantron, so I'm going to assume that that machine doesn't care why any kid picked 'b' over 'c', when 'c' was the correct answer. And whether or not that indicates that the students, the average student, has trouble remembering to flip the inequality when they divide by a negative, like I don't think that the government looks at that on that particular test. I think classroom teachers could when they give their assessments in class, but on that uniform test, I don't think that anyone looks at that.</i></p> <p><i>And</i></p> <p><i>DE: overall, do students do well on this final exam if they do well during the year and vice versa?</i></p> <p><i>Gidget: I'd say it's unpredictable. So there's some years where, yeah, absolutely everybody does exactly as I predicted. There's some years where there are a few surprises. I think more dissecting needs to be done to figure out if that surprise is on account of the fact that the student didn't study, or parents just got divorced, or the assessment has something ambiguous on six questions, or maybe the government lined up the three uniform exams for the grade tens, one, two, three days in row. That makes it hard for them to study and keep focus. If I were to put a number on it, I'd probably say probably say 50 percent of the time. Not the majority. 50%</i></p> <p><i>Also, see 1b), 2d)</i></p>
<p>c) The repetitive nature of questions</p>	<p><i>Not necessarily every year, but you know, every two or three years... like oh yeah, there's another one of those.</i></p>
<p>d) Moderation</p>	<p><i>Write better exams so that they're more valid. Why are we moderating things as far as 15, 20 percent some years? It's insane. Is that a reflection of my teaching, the student's performance or the assessment itself?</i></p> <p><i>And</i></p> <p><i>... Well the fact that there's gossip, that there's an easy year, then the pendulum swings and there's a hard year. I mean, that's insane. How</i></p>

	<p><i>come you can't get it right? Why are you sending 'info sanctions' every two days leading up to the exam, because you've found typos or errors or this or that? Like how many eyes are on that exam? ... So I fail to see why it's so complicated.</i></p> <p>And</p> <p><i>I appreciate that the government does a percentage of teacher marks/exam mark. I think that is the way to go, but the fact that they moderate my mark based on their assessment, is a little weird. If you're asking for my mark, presumably it's because you trust and value me and then you reject it. So that just shows that they value the exam more. I don't know why we go through this song and dance. Just take the exam mark if that's what you want to do. But I do think it should be a combination of both. Because it's math, I think there should be a cumulative exam, perhaps more opportunities... June, July, August before the next year. So a combination between exam and the teacher's marks would be ideal...</i></p>
<p>3. Instructional issues</p>	
<p>a) Using old exams to prepare</p>	<p><i>Gidget: Um, I give a lot of a past exam questions so they're familiar with formatting and language.</i></p> <p><i>DE: ... do you think that would be sufficient for students to do well on the uniform exam?</i></p> <p><i>Gidget: I would love to say yes. But I don't think that's true. I think the reality of my view is our reality at this particular school, with the particular expectation that our clients have in terms of results, is that they need to have a past exam to feel confident going into the exam. When I give them a review package and they ask for a practice test and I cross the word review off and write "practice test", all of a sudden they're more willing to do that practice even though it's the same questions ... I think, I think content-wise, the students would be ready, but I think they've got this idea in their head that they need to do past exams and I know</i></p>

	<i>they get it from their parents, because starting in April. I get the emails from the parents: “when are you going to give a past exam as practice?”</i>
b) Timing throughout the year	<i>Yeah, I've, I'm absolutely 100 percent more stressed about pacing when there is a provincial exam. There are times where I would like to slow down and repeat, because I have all kinds of signs, whether it's glossy eyes or poor quiz marks or blank homework. The students have not mastered this particular skill or content, but I have to, I have to keep going, you know, rolling the dice that, okay, so that just, that just isn't their subject that isn't their piece, that maybe if I move onto the next piece, I can still keep that glimpse of confidence they haven't totally lost, start something new and get them to buy in again. Then there are other moments where a kid asks an unbelievable tangential question and I just want to run with it for two classes and learn something new myself, but that would destroys the timeline. Time doesn't allow for that.</i>
c) Teaching specific test-prep skills	<i>Even just like when we have tests in class, and I distributed review package, they'll often say, ‘so if I do this, I'm good?’ Right? Like, the idea of being thorough and making sure that... they're very dependent on every literal thing I say, so if I give them one past exam, and that's all I give, they will assume that that is enough. Whereas by distributing more, I guess what I'm trying to show is that there's a range of things, there's, you've got to go through your notes, you've got to go through these past exams, you've got to review all your tests. But there's certainly some that will focus just on the exams.</i> <i>And</i> <i>I think it is important to instruct [test-taking skills] at all levels starting even earlier than the grade 10 uniform exam, we should be teaching these particular study skills. And I'd like to think that I try to, but what I haven't been able to do is impress upon the students how important that is. So they still have blinders on, they think if they just do the exam versus pausing and noticing, they're like, “oh, on the inequality question across these three tests, I made the same mistake”. They have their blinders on</i>

	<p><i>no matter what I say, and perhaps that's in my instruction... my instruction for how to study is maybe more in passing rather than a designated class activity, so that might be why they're dismissing it. I think it may be a maturity thing. You know? I have some kids when I... I've told them for multiple-choice, it's not, it's not important enough to just select 'B' as your answer, you have to be able to rule out why 'A', 'C', and 'D' are wrong. And I now have a few students that will actually, on their tests, circle 'B' and then beside 'A', 'C', and 'D' put a little descriptor of why those are wrong. So those kids heard me. I mean, maybe they would have done it themselves and they just needed the tiniest of queuing, but other kids don't unless I make it an activity. Unless there's sort of four marks for this question. One for the right answer, three for telling me why the others are wrong. I haven't gotten them to do it.</i></p> <p><i>And</i></p> <p><i>I mean the schools across our province are so different, right? Look up North, if a kid shows up more than half of the days of school then I think they get really excited. So how are you supposed to teach in an exploratory way knowing that you also have the uniform exams? I think they just do what they can up there. And I think of like the friend I mentioned who's a Phys ed teacher who's teaching mathematics. If he's not in a department that's offering all kinds of resources and his math skills aren't much sharper than the students he's teaching, his sole resource is those past exams. So he ends up, arguably inadvertently, teaching just to the exam and then caters his instruction to that.</i></p> <p><i>Unfortunately, yes.</i></p>
d) Teaching throughout the year	See 3a, 3b, and 3c
4. Learning issues	
a) Student stress	<p><i>Gidget: I don't know if that's a reflection of the culture that we've created that CEGEP has this arbitrary benchmark to get into science, that the parent wants to be able to say you can get a car if you get a 90, or like...</i></p>

	<p><i>you know, we just put so much stress and importance on that one exam. I find it a valuable tool as an educator to show me what the student has accomplished over the year and how well they're prepared for the following year given that it's cumulative subject mathematics. But to have this be such a heavy marker on what happens to them.</i></p> <p><i>DE: do you think that students are stressed about standardized tests?</i></p> <p><i>Gidget: Yes.</i></p> <p><i>DE: Um, do you think that for some students it can be a positive stress that it's a motivator?</i></p> <p><i>Gidget: Yeah, I'm sure there are studies that prove that stress is a motivator for adults.</i></p>
<p>b) Students not learning effectively</p>	<p><i>I'm not sure if it's <u>because</u> of the uniform exam, or if it's because they're looking for a shortcut. Even just like when we have tests in class, and I distributed review package, they'll often say, "so if I do this, I'm good?" Right? Like, the idea of being thorough and making sure that... they're very dependent on every literal thing I say, so if I give them one past exam, and that's all I give, they will assume that that is enough. Whereas by distributing more, I guess what I'm trying to show is that there's a range of thing, there's, you've got to go through your notes, you've got to go through these past exams, you've got to review all your tests. But there's certainly some that will focus just on the exams.</i></p>
<p>c) Developmental capacity of students</p>	<p><i>I don't know why we pick grade 10 as the year. It goes back to the government. I don't know why they do a lot of what they do. Um, I think maturity and development, and meta-skills, and just the development of the frontal cortex, everything for this particular age group is so varied. You know, at 12 or at 16, you can see the same maturity. Or you could see 14 and 14 you can see two totally different characters and how they approach studying. So I do think it's arbitrary that we've picked grade 10.</i></p>
<p>5. Systemic issues other than the exam</p>	<p><i>I don't know why the government does what they do half the time. Sometimes I think it's to save money, sometimes I think it's to facilitate... the scantron... easy to run through a machine... you don't have manpower</i></p>

	<p><i>to pay. Sometimes I think it's because they have a system, you know, being Québec high school curriculum than the CEGEP entry level and all that. There's the system that is too daunting to amend. Yet, all of the things I just said are within the government's power to adjust.</i></p> <p>And</p> <p><i>I don't think the government has done what it can to ensure that there is the professional integrity of teachers is where it should be. You know, I think we care too much sometimes and we would write a reference letter or send marks that are maybe on the generous side, because we just want the best for that student, and that student really wants to be an engineer. So we send what is... we make that happen. And then there are the people that aren't math teachers, so I don't know how your marks can really have that much weight. Again, I think this is a problem that the government can resolve. You think of other professions where dentists and doctors and psychiatrists take oaths before they begin.</i></p> <p>And</p> <p><i>...the education world has a major hill to climb, uh, in that we have created a culture where marks are the most important thing and that goes for post-secondary applications, using that uniform exam, to just basic questioning in class where teachers often pose a question to the class and all they're looking for is the one correct answer. The students have been conditioned to just make sure they get everything right and get 100.</i></p> <p>And see 3c)</p>
--	---

Summary of Gidget’s interview.

In her interview, Gidget jokes that “I’m not here to solve problems; I’m just here to list them”. It seems that this is the type of joke which is funny because of how obviously un-true it is, since throughout her interview, far from simply listing problems, she provides thoughtful insight into specific aspects of uniform math exams that she thinks are problematic, and then she does, in fact, generate many constructive suggestions to solve those problems. Gidget draws on her extensive experience as a teacher, her experience as a department head, and her experience

having worked on exam creation committees to recommend solutions to problems ranging from the minutia of exam format and content up to the nature of the education system in Québec in its entirety.

Exams themselves (format, content & nature).

Gidget believes that there should be a cumulative end-of-year mathematics exam since, she says, concepts build on each other in math, making it a cumulative subject. She also believes that it is justified for this exam to be generated by the government and for the same exam to be distributed across the province. She recognizes that the training and experience of math teachers across the province is not equal and that this could disadvantage students in certain regions/schools if their teachers are unable to create fair, valid, end-of-year assessments. Gidget believes that in the ideal circumstances, this exam could also be helpful in providing information about student knowledge and ability, but for this to be justified, “it would have to be a good assessment”. She goes on to point out various reasons why the current uniform exam is not necessarily a particularly “good assessment”.

Firstly, in terms of exam format, Gidget believes that the multiple-choice section of the exam is poorly designed. The fact that there are ever multi-step problems in this section, she says, “absolutely infuriate[s] me”. This is because, she says, she is unable to assess what students are thinking when they go through the multiple steps of the problem. Gidget believes that multiple-choice questions are purposefully written to have incorrect answers which are plausible and meant to distract students, and that the potential positives of building confidence and cuing students to important information are only helpful in the case that the exam is not a stressful situation which, she says, it is.

For the other sections of the exam, Gidget says that the long-answer section questions are generally well-selected in terms of content, well-constructed in the sense that they are usually contextualized and have an interesting lead-in paragraph, and representative of students’ ability. In terms of the short-answer section, Gidget thinks that the questions are usually well-selected in that they are a good lead-up to the long-answer questions, and well-constructed in that they do not require too many steps, but that the marking guide is problematic. She says:

There were a few times where in our office, we discuss our marking and I think, you know, if I asked this question on a class test and marked it out of four, I would give the kid, given his mistakes and what I saw on the paper, three on four, even sometimes three and a half on four, but given the marking guide for the uniform

exam, I'm not allowed to look at his work, I'm only allowed to look at his final answer and it's written in a particular line of a particular booklet. It ends up being zero on four. So does that zero really indicate his understanding of the material? So I think the short-answer can prove useful, but the marking guide makes it difficult.

Secondly, Gidget brought up problems with uniform exams in terms of their validity and reliability. As for validity, she says that in her opinion and experience, it is unpredictable whether a student who does well during the year will do well on the exam. In fact, in a scathing estimate, Gidget says that, in her experience, only 50% of the time will a student's exam results be reflective of year-long results. In terms of reliability, Gidget says that it is "insane" to have such a large difference between year results and exam results, and with her further explanation about the "pendulum", suggests that this large difference is a failure of the exam. The "pendulum" refers to the idea that teachers across the province generally agree, after having seen the exam, that a uniform exam in a given year was either "easy" or "hard". Historically, years with particularly "difficult" exams tend to be followed by years with exams which are considerably easier. Gidget suggests that if it's possible to control the difficulty of the exam, it should just consistently be at the correct level.

Instruction.

Included in the many different teaching techniques she uses throughout the school year, Gidget regularly brings in old exam questions for students to become familiar with "formatting and language". When asked whether or not she thought that students would be able to succeed on the exam without practicing old exams, she responds "I would love to say yes, but I don't think that's true". This emphasizes how important she thinks old exams are in preparing students for the end-of-year uniform exam.

In terms of the pacing of the course, Gidget brings up two, almost opposite, scenarios and explains why both are problematic. The first refers to "times where I would like to slow down and repeat" but that she can't afford to because she has to keep up with the pace. The second is when she says that, sometimes, a "kid asks an unbelievable tangential question" and, educationally, she feels like the right thing to do is to take extra time to enrich student learning with a topic brought up out of student interest, but that time simply "doesn't allow for that".

Finally, in terms of specific test prep, Gidget believes in teaching students how to handle the format of the exam by teaching them meta-test skills. For instance, she tries to teach students

to answer multiple-choice questions by simultaneously finding the correct answer and doing the work to prove that all other options are incorrect. Similarly, in addition to course content, she wants students to learn how to study for tests, and to learn how to learn effectively. She believes that “we should be teaching these particular study skills” in every math class, even those before uniform exam years to provide students with tools to learn effectively early on.

Student learning.

On the note of teaching students how to learn, Gidget finds that a major impediment is the fact that some students are “looking for a shortcut”. These students do not have the practice of “being thorough” or making sure that they use all possible resources to ensure that they’re prepared. Gidget says that she thinks this is a phenomenon that occurs with or without the existence of a uniform exam, but that either way, she attempts to help students realize that “there’s a range of things” to do to be prepared for any assessment, including going through class notes, looking over class tests, and also going over old exams. In uniform exam classes, she acknowledges, that there are some students “that will focus just on the exams”.

Other systemic issues.

Gidget situates the uniform exam in all of Québec, and brings up several systemic issues that occur province-wide. The first issue has to do with teacher training. As mentioned above, she believes that schools in some contexts have no choice but to employ a teacher with no math experience or educational background. She also believes that some math teachers will end up with colleagues who might not communicate effectively, share material, or confer when correcting. She says that if these realities are problematic to the government, then the government should implement changes. Secondly, she thinks that teachers should have more flexibility to use their professional judgment when correcting exam questions. If the use of professional judgment is problematic to the government, then she says the government should take steps to fix that issue too, by having teachers take professional oaths, by changing official teacher training in Québec, or by some other method. Finally, she brings up an issue which she believes occurs in Québec and around the world: students being overly-focused on numerical results rather than legitimate learning. She says that we encourage this improper focus by using numerical results as entrance guidelines for schools, as cut-offs for “honor roll” or other distinctions, and for generally praising students for their grades.

Researcher's Interview Data Table

Table 9. Researcher self-interview results.

Theme/subtheme	Participant response
1. Exam format	
a) Multiple-choice questions	<p><i>I believe that the multiple-choice questions are probably the most repeated and predictable in their format and content. Most years, I can pretty confidently predict what five out of the six questions will be. In terms of their usefulness in assessing student ability, I think that multiple-choice questions do the worst job. They provide no difference in feedback between a student who made a minor error, a student who made a major error, or a student who had no idea what to do. Similarly, a student who did not know what to do but guessed correctly and a student who knew exactly what to do and did it perfectly would get the same result. A question like that seems like a poor assessment tool.</i></p>
b) Long-answer questions	<p><i>I think that most of the long-answer questions are grade level appropriate, content appropriate, and difficulty level appropriate. That being said, I think that some exams have one or two questions which are unclear, ambiguous, or too difficult. I think that questions like these being worth 10% of the uniform exam is far too high a weighting.</i></p>
c) Short-answer questions	<p><i>I think that the short-answer questions, while they're predictable in their format and content, do their job in forcing students to carry out some work to obtain a final answer (rather than having four options provided). I think that these types of questions have to be carefully selected to: 1) not require too many steps to solve; 2) allow for certain part marks to be given for certain specifically selected pieces of evidence that partial work was done correctly; 3) allow for a final answer (with no work shown) to give a teacher some idea of what a student did to obtain it (assuming that a reasonable approach was taken). I don't think that questions are always chosen according to these criteria, but when they are, I like these questions.</i></p>

2. Nature of the exam	
a) Existence of a uniform exam	<i>I think that a final uniform exam provided by the government is not necessary in this course. I think that a “complementary” exam would be just as good if not better. This is an exam produced by the government, but one on which teachers may change wording, numerical values, or entire questions altogether.</i>
b) Validity / reliability of the exam	<p><i>In my experience, uniform exams are somewhat valid, in the sense that strong students typically do well, and that students who do not understand the content will likely not do well. However, in every exam cycle, there are some “surprises”. These call into question the overall validity of the exam. Would students writing a differently-written, formatted, and styled exam on similar content do as well? I think probably not. I think that some of the success of students has to do with their specific preparation for this exam. That also calls into question the validity of the exam.</i></p> <p><i>In terms of reliability, I think that there have been easier and more difficult exams, which means that there is not perfect consistency. I think that it’s unreasonable to ask for zero variability in difficulty level, but I think that there could and should be less than what there is.</i></p>
c) The repetitive nature of questions	<i>I think that a student who practices old exams will see several types of questions repeat.</i>
d) Moderation	<p><i>I think that the moderation process puts way too high a weighting on the exam. In the case of a year with a particularly difficult exam, the government will bring an entire class’ year-long marks down to better align with exam results, and then, assuming that the exam results were low across the province, then raise grades across the board, with the added amounts being greater near the mean than at the extremities.</i></p> <p><i>Shifting grades down to then raise them up seems like a waste of time and resources. Furthermore, it renders the process far less transparent, which I think causes stakeholders to lose faith in the system.</i></p>

3. Instructional issues	
a) Using old exams to prepare	<i>I use old exam to help prepare students when I teach the course. I try to teach in way to help students gain a legitimate understanding of the course topics, and my goal is to have students be able to generalize this knowledge to new situations. Theoretically, my goal is to provide students with the skills and knowledge to be successful on the exam even without looking at old exams. The reality is, though, that students who, in my experience, have not completed old exams for practice do not do as well on the exam. Maybe this is a failure of my teaching, maybe it's a failure of the exam, or something in between.</i>
b) Timing throughout the year	<i>Having an exam affects the way I time my teaching throughout the year. I refrain for bringing in what I think are interesting mathematical topics if they don't align with what will be on the exam. Maybe this is partially the purpose of a uniform exam, but I think that it makes my class less interesting and gives me less leeway to teach to student interest and real-world usefulness.</i>
c) Teaching specific test-prep skills	<i>I probably wouldn't spend time teaching "conjecture" type questions if they weren't on the final exam. In any class that I teach (not just those with uniform exams), I try to teach the students skills to approach a test with a backward-design lens. This includes looking for clues in other questions, purposefully and explicitly eliminating incorrect multiple-choice options, showing all work you CAN do even if you cannot do all of it, making up a value to be able to continue if you're stuck.</i>
d) Teaching throughout the year	See point about conjectures in 3c) And <i>I teach students certain tricks about how to answer questions specifically based on how certain types of questions are typically written or presented.</i>
4. Learning issues	

a) Student stress	<i>I definitely think that students are stressed about uniform exams. The effect that this has is different for different students, but there seems to be something different about a government exam than an in-house final exam in students' eyes.</i>
b) Students not learning effectively	<i>I think that, for the most part, students prepare themselves to do well on the test as opposed to gaining a deep, meaningful, long-lasting understanding of course content. I think that if you re-tested students with a similar assessment tool a week or two after the uniform exam, I suspect that the average result would be lower. I think that a lot of the "learning" that happens is short-term and superficial.</i>
c) Developmental capacity of students	<i>I think that some students are more mature than others in any given year, but I think that the course content is level-appropriate.</i>
5. Systemic issues other than the exam	N/A

Chapter 4

Discussion and Conclusion

Discussion

Major findings are discussed and contextualized under the themes and sub-themes presented in tables 1-9. Format and nature of the exams have been collapsed into one, though all of the sub-themes are still present. In order to justify the interpretation of major findings and to facilitate readability, some ideas and quotes from individual participant summaries are repeated in the discussion. Efforts have been made to limit these redundancies where possible.

Exams themselves (format, content & nature).

Multiple-choice questions.

The multiple-choice section of the exam consists of the first six exam questions, each with four response options, one of which is correct. Students indicate their final answer (with no work shown) on a scan-tron sheet which is sent to the government to be corrected. All seven participants brought up the multiple-choice section of the exam, and all seven participants had at least some problem with this section. Edna brought up the idea that her problems with multiple-choice questions were mitigated because these questions only account for 24% of the exam, and Flora and Cleo found potential benefits, namely that if straight-forward, these types of questions could build student confidence and potentially cue them to important details. Both Flora and Cleo did acknowledge, however, that the questions are not always simple or straight-forward.

Overall, there was overwhelming consensus that the multiple-choice questions did a poor job of assessing student understanding or ability. One common problem was that participants viewed these questions as being unintentionally ambiguous, or worse, purposefully tricky, and that having a student select an incorrect multiple-choice option which was designed to be appealing was not an accurate measure of a student's proficiency. Furthermore, some teachers felt that some multiple-choice questions require students to complete multi-step mathematical procedures, and that these were inappropriate for this section. Several teachers also explained that multiple-choice questions are not particularly helpful in diagnosing what type of error the student made, because they are assessed on the final answer only and no work is shown. Furthermore, because no credit is given for work and no work has to be shown, multiple-choice

questions allow for students to guess an answer. There is no penalty for guessing, and in fact, guessing the correct answer provides a student with as many points as if they knew exactly how to solve the problem. Finally, and related to the fact that no work is shown, participants did not like that these questions are assigned a full four marks for a correct answer and zero for any error. This, the participants felt, devalues the process involved in solving a problem versus having the final answer, and leads to misrepresentative results for students who are able to do most of the work required but obtain none of the marks to show for it.

Short-answer questions.

The all-or-none nature of short-answer questions made them overall problematic for participants too, even though some participants felt that they were a good build-up to the long-answer section and others simply liked that there were no response options given. The short-answer section of the exam has four questions for which students give their final answer without any work or process shown. The majority of questions only allow for students to get zero marks or four marks, and only exceptionally (for questions which require specific notation or which have two distinct parts to an answer) can students get two out of four marks. Most teachers reported that throughout the school year, in order to provide rich feedback and help guide students, they would provide partial marks on assessment questions for work which was partially correct. Flora summarizes the issue when she says that the student's result who "gets zero when he can do three out of four steps is not really reflective of his ability". Specific frustration arose over the idea that the exam marking guide mandates that teachers provide either zero or two marks out of four for what teachers considered minor errors. Several teachers believe that the marking guide could be amended to allow for partial marks which would be better representative of a student's understanding or ability.

Long-answer questions.

The long-answer section of the exam consists of six constructed-response questions for which students show their work/process in a booklet and which are evaluated by classroom teachers out of ten total marks (four for understanding, four for mathematical processes, and two for organization). Whereas most participants felt that the multiple-choice section of the exam did a poor job of assessing students' ability and knowledge because they do not get to showcase their thought or problem-solving processes, most felt that, overall, the long-answer questions did a good job of allowing students to demonstrate these competencies. Overall, participating teachers

thought that these constructed-response questions were often situated in real-life contexts, and that they often combined concepts from different course chapters, both of which were well-liked features of the questions. Although the general consensus on long-answer questions is that they are an effective assessment tool for measuring student skills and knowledge, some complaints arose. Gidget thinks that the contextual, situational introduction to long-answer questions can be overly text-heavy, which can mean that a student's inability to answer a question could be a measure of his/her reading inability rather than a lack of his/her mathematical skills or understanding. Flora also comments on the marking guide for long-answer questions, saying that she finds it "in some ways very detailed and other areas very vague".

Existence of a uniform exam.

All of the participants brought up the idea of having a uniform exam at all. Most of them felt that there should be some type of formal summative assessment at the end of the year, since as Gidget says: "math being a cumulative topic warrants" it. The idea that a summative exam was appropriate, though, was not unanimous. Alice thought that a uniform exam should be optional for students. She says "if you want to do the test, you want to see how well would you do ... I don't think everybody should", and explains that for some students, the pressure of an examination like this is good, but that for others, the process itself is potentially harmful. Although the other six participants thought that a cumulative final exam should take place for math courses, they did not agree on what form that exam should take. Several teachers felt that a similar exam without a multiple-choice section would be better. Some added that being able to provide part marks for the short-answer section would even more effective at assessing student ability. Edna thought that "most of" the current exam as it is does an effective job of assessing skills and knowledge, and Bart brought up a different idea altogether, specifically, that the test should "more blind" to encourage students "to prepare the whole year worth of material rather than just sections of it".

Validity /reliability of the exam.

All seven participants brought up the idea of either the validity of the exam, its reliability, or both. Several participants think that the long-answer section of the exam is valid and generally does a good job of assessing student knowledge and skill, and as mentioned above, Edna thinks that "most of the exam" does a good job. Overall, though, the participants questioned the validity of the exam both explicitly and implicitly. Explicitly, most of the validity issues are related to

multiple-choice and short-answer questions, although some teachers generally feel that the format, nature, and year-to-year predictability of the exam mean that it is at least partially invalid anyways. Much of the justification for this was brought up implicitly. For example, many participants point out that students do not always achieve the same level of success on the exam as during the year, which seems to be a question of validity. There was disagreement about how widespread this mismatch is, ranging from rarely and primarily due to stress, up to frequently (one teacher estimated it at 50%). Flora reports that student exam results are frequently “several percentage points below what I expect them to do and what they expect to see, which is unfortunate”. Another example is that some participants discussed the exam promoting success at “taking a test” rather than “really mastering the material”. Some teachers justify this belief by saying that, in their experience, student exam success does not seem to correlate with post-secondary success in math, which seems to imply that the exam is less valid.

In terms of the reliability of the uniform exams, although fewer teachers brought up this idea, those who did were unanimous in their belief that the difficulty level of the exam is not constant, and that this variability affects students differently depending on their level of mastery of the course content, their individual levels of anxiety, and possibly other factors too. I think that specific exam questions changing from year to year means that some students will inevitably find one year’s exam less/more difficult than another based on which course topics they understand better/worse. However, the sense from the teachers who brought this up was that the variability was sometimes quite high and that this should not be the case.

The repetitive nature of questions.

Five out of the seven participants discussed the fact that the exams are repetitive in nature, and one additional participant simply stated her belief that they are repetitive. Participants agreed that some question-types repeat every year, and others every few years. Daphne points out that to get a good idea of what will be on a uniform exam, you could “look at the past [...] three or four years”. Most participants thought that this repetition could lead some students to learn a sub-set of the course content and/or to learn to solve specific question-types rather than learn more generally theory. Bart says “if [students] know that there’s this type of question which is worth 10 marks, and they see it in 5 or 6 old exams, then that’s what [they’re] going to study” and adds that as a result, “they’re not really applying [...] all of the information they learned throughout the year”.

While most participants recognize the repetitive nature of exam questions, very few feel that they teach course content in a way that teaches students to solve specific exam-type questions rather than learn the underlying theory, and several teachers report that they specifically take action to avoid this practice. Edna brings up an interesting philosophical point when she says that teaching in an exam-specific and focused way might be something that teachers would do, but admits that because she has always taught courses with uniform exams, she doesn't know whether or not she teaches the content differently because of the exam. Even though other teachers did not bring up this idea, it is interesting to consider how that might apply to any teacher having taught a particular course which has always had a uniform exam. Finally, some teachers thought that, despite trying to avoid teaching in an exam-focused way, some students study in a way that is particularly exam-focused, but this is discussed further under the heading: *students not learning effectively*.

Moderation.

Moderation and transformation are the two methods by which the government alters exam results. Transformation is the process through which the entire province's results on an exam are shifted and moderation is the process through which the government alters each class group's marks from during the year to more closely match the same class' exam results. Because more participants brought up moderation, this was the label of the sub-theme, but all three participants who brought up ideas under this category discussed the government altering student results in one way or another, and all of them seemed to feel insulted and powerless by the alterations. Insulted that their professional judgment, in the eyes of the government, is insufficient to determine how well a student mastered the content of a mathematics course with a uniform examination, and powerless, in the sense that after spending a year teaching, guiding, and coaching students, that someone else should make the final decision about how they did. Edna says that she doesn't like what happens once the exams leave the school, Alice says that the government should trust her judgment, and Gidget says that it is "insane" to have to moderate "as far as 15, 20 percent". She goes on to discuss the "pendulum", a reference to the idea that teachers across the province can generally agree on which year's uniform exam was either "easy" or "hard". Historically, years with particularly "difficult" exams tend to be followed by years with exams which are considerably easier. Gidget suggests that if it is possible to control the difficulty of the exam, it should just consistently be at the correct level.

Instructional issues.

Using old exams to prepare for the exam.

All seven participants reported using old exam to help students prepare for the final exam. Some of them modeled all of their assessments from the start of the year after the uniform exams and others only brought in old exams in the final, review portion of the course close to the end of the year. Alice, for example, thinks that bringing in old exams too soon would increase student stress. Participants have different styles and rationales for using the final exams. Some, generally, think, like Daphne, that “the only way” to prepare students for exams is “by doing old copies of standardized tests”. Others were more specific in their rationale. Bart, Flora, and Gidget said that they wanted their students to become accustomed to the “format”, “types of questions”, “level of difficulty”, and/or “language” or “wording” of the old uniform exam questions, and that using the actual old exams was the best way to achieve this.

Most teachers agree that having students practice with old exams is necessary for students to be successful at the end of the year. Bart, for instance, thinks that students might be able to succeed at the end of the year without practicing old exams, but that they wouldn't do as well as with the practice. Flora says she had no evidence to suggest that students could be successful without practicing old exams, and Gidget says that she would “love to say” that students could be successful without old exam practice, but that she just does not “think that's true”. Edna makes explicit reference to the practice of “teaching to the test” explained in chapter 1, when she says: “I know people say you're teaching to the test and you're not expanding their mathematical ability, but if... there's not much I would change in the course”. Edna goes on to say that she believes the content of the course is already quite narrow and that the old exams cover the course content well. She implies that, based on these beliefs, using old exams as a teaching tool or guide is no different than using any other teaching strategy.

Timing throughout the year.

All of the participants brought up the impact that uniform exams have on timing/pacing throughout the year. Gidget sums up the general feeling of the participating teachers when she says: “I'm absolutely 100 percent more stressed about pacing when there is a provincial exam”. A reason that emerged to explain this stress is the large amount of content in the grade 10 math courses. Bart says that “because there's so much content”, a teacher is forced to “get through it

quickly”, and teachers are forced to move even more quickly than in other years to be able to leave time “at the end of the year to prep” for the exam. There is also a fear, reported several teachers, of not moving quickly enough, not having enough time to cover a topic, and then having that topic appear on the exam, which would disadvantage students.

Direct consequences on teaching and learning also emerged. One impact mentioned by several participants was the idea of spending more time on topics which are represented more heavily on the exam. Whether this is positive or negative, though, is unclear, because the relative weight of topics on the exam may be the government’s way of communicating the relative importance of certain course topics. Another impact, which teachers view more clearly as negative, is the inability to pace instruction according to student needs. Teachers feel that there is not sufficient time to slow down and re-explain a topic or concept even if students need it. On the other end of the spectrum, while participants feel, like Flora, that it is “always nice to supplement” the existing course content with other topics which might be of interest to students and/or teachers, timing does not allow for that either. Edna sums this up when she says that it is common for teachers to “forego other opportunities”, including topics and activities that would be more hands-on and experiential for students.

Teaching specific test-preparatory skills.

As discussed in chapter 1, teaching “specific test-preparatory skills” refers to the practice of taking course time to frame the content being taught in the context of a test, and also spending time learning about the format and nature of the test itself with the objective of improving student test results. Since all of the participants take the time to review old exams, they all spend time looking at course content through the context of uniform exams. In terms of learning about the format, several participants help students by pointing out questions which frequently repeat, that question-types are likely to appear in a specific section of the exam, or are generally “good exam questions” to review. In terms of learning about the nature of the exam with the purpose of improving student test results, there are a number of things the participants brought up. Flora helps students work on their timing for the exam, and Alice teaches her students to start with the long-answer questions to answer the highest value questions while there is time and students are still fresh. Gidget and Alice teach their students specific strategies to answer multiple-choice questions and Cleo teaches her tutees general skills for dealing with questions dealing with functions (which make up a large part of the course).

Finally, some participants report that they help students prepare for the exam by providing them with general study skills, like systematically going through course topics and skills when preparing, learning to use multiple sources to be maximally prepared, and learning that last-minute studying is not effective. Gidget says that these skills should be taught even earlier than grade 10, and Alice says that, in some ways, the entire grade 10 math course can be a lesson in “how you can study if you have to do [any] exam”.

Teaching throughout the year.

Although the timing of the course and the use of old exams for practice are necessarily a part of this category, this sub-theme groups together responses having to do with other ways that having a uniform exam affects teaching throughout the year (not just during end-of-year review). There is still some overlap with other sub-themes, in the sense that some teachers reported that they used old exam questions throughout the year, not just during review time, but only other effects not yet discussed are dealt with here.

Alice says, generally, “if there wasn’t an exam at the end of the year, I could approach teaching a class differently”. She explains that, for instance, she should be spending more time helping students to understand the meaning of complex topics by looking at real-world examples, showing students where they might have already seen examples in their lives, helping them to understand how and why the topic is useful, but that because of the exam, she doesn’t do that. Cleo says that she would want the flexibility to really focus on the topic(s) that students will need later in their studies or later in life, even if those topics are less important for the exam, and although Daphne doesn’t teach a course with uniform exams, she says that if she did, she would surely teach the content differently.

Learning issues.

Student stress.

All seven participants made reference to the idea that students are stressed about uniform exams. Most teachers felt that this was a negative impact of uniform exams and that student results would be affected by stress. Some teachers, though, were not convinced that student stress was all negative. Flora, for example, says that student stress is neither positive nor is it negative, but rather an indication that students “care a lot about the outcomes”. She says that it becomes the job of the classroom teacher to “calm them, remind them that they are prepared, remind them

that they can be successful”. Other teachers think that the stress might be positive for some students and help them to prepare for the exam in a more serious way. Some participants make a distinction between students who are generally high-achievers and those who are not. They say that for higher achievers, stress is likely to have less of an impact, or that it might help those students “rise to the occasion and perform even better”, as Daphne says. Conversely, stress is likely to have much more of a negative impact on students who are already likely to be lower achievers in math.

Two explanations emerged as to why uniform exam stress is different than stress on other assessments. The first has to do with the high-stakes nature of the exams for students. Bart says that the exam is “always in the back of [a student’s] mind” because the uniform exam course is “such a gateway” to post-secondary education. This sentiment about the importance of the exam and the high-stakes for students was echoed by other participants too. A second emergent explanation has to do with the uncertainty of what will be on the exam, because, as opposed to regular assessments during the year, it is someone other than the classroom teacher creating the assessment. Some teachers suggest that this feeling of powerlessness might come from the fear that a classroom teacher might have missed a topic that could end up on the exam, or might have taught something in a way that could make the exam question unfamiliar to students. I believe that over the course of a school year, students build a trusting, comfortable relationship with their teacher and they grow to believe that their teacher will not assess them in a way that they might see as being “unfair”. Similarly, I think that students learn, over the year, how a classroom teacher assesses, so there is less uncertainty about an end-of-year exam made by the teacher. Having an unknown assessment means less comfort in these respects and higher stress.

Students not learning effectively.

Six out of the seven participants brought up the idea that students who take the uniform exam mathematics courses do not learn the content as effectively as they could. Several teachers believe that, from a student point of view, success in a course like this is equivalent to obtaining a good grade on the exam. Alice says “I think 80% of students are trying to figure out what they should do in order to have a good mark”. Cleo says that in order to get a good mark, students learn to “regurgitate” information, or display their understanding of the “surface material”. Other teachers think that in order to get a good mark, students “gear their studying towards what they’ve seen in previous exams because there are patterns”, as Bart says. Several teachers believe

that this leads students to learn a sub-set of the course content, prioritizing the topics and types of questions that are most frequently repeated on old exams. Bart believes that this happens at the expense of students learning how to apply “all of the information they learned throughout the year”. Flora brings up an interesting point about this, which relates to her thoughts about teachers teaching to the test. She says that the way students learn the course content “depends on the guidance given by the teacher”. She continues to explain that a teacher who guides students to really think about the course topics and what skills were learned over the year, who directs students to look at other review material than old exams, will help students to learn effectively.

Developmental capacity of students.

Four participants brought up the idea of students either being able or not being able to complete or comprehend some of the exam content based on their developmental level at the age of exam completion, and all of them expressed different opinions on the issue. One participant, Cleo, believes that students are “too young” for the “techniques” required by the course which, ultimately, leads to superficial learning. One other participant, Edna, believes the opposite, that the course content (with one exception) is designed so that students of this age “can handle” it. Two other participants believe that being developmentally ready for the course depends on the individual student, and that there is no prevailing trend. Gidget captures this idea when she says “I think maturity and development, and meta-skills, and just the development of the frontal cortex, everything for this particular age group is so varied”.

Systemic issues other than the exam.

This theme became akin to an “other” category, which, on one hand, is inconsistent with the methodology of systematically searching for emerging themes. On the other hand, though, this theme allows for a discussion of interesting and potentially relevant topics which were brought up by fewer than three participants (the criteria for inclusion in the themes and sub-themes of table 1). The scope of the issues brought up under this theme range from specific, having to do with the exam itself, up to the general, dealing with the overall purpose of the education system.

Issue #1: existential questions about the exam.

Gidget asks hard-hitting questions about the exam, its nature, its format, and even its existence. She wonders why we have this exam, and suggests a multitude of motivations other than being the best way to evaluate students. She suggests that uniform exams might “save money” by

“limiting manpower” because the multiple-choice questions are corrected by a machine, or that Québec needs a specific system to facilitate CEGEP applications and admissions, and that often, it’s easier to keep the system that you have and know than to fix it. She suggests that maybe the system in Québec is “too daunting to amend”.

Issue #2: other uniform exam benefits.

Edna brings up a benefit of uniform exams that other participants did not mention, which is that it provides objectivity, or at least the appearance thereof. She says that this is important given the pressure that parents can put on a teacher or school to review and/or increase a student’s marks. She says that the uniform exam allows teachers and school administrators to say “it’s out of our hands”.

Gidget brings up another benefit of uniform exams that was not mentioned by other participants. She discusses the variability of teacher experience and training across the province. She says that while an experienced teacher who has a background and training in math could create a valid and fair end-of-year assessment, there are many teachers in Québec who would not be able to do so. Similarly, whereas a teacher with training and experience has the professional judgment to correct a uniform exam without a prescriptive marking guide, others do not. She says that the uniform exam levels the playing field, and offers a tool to teachers who are not able to create a better one. She does go on to say, though, that in terms of professional judgment and integrity, that the government has the ability to change the circumstances and qualifications of math teachers if this is the only reason to have uniform exams. She brings up other professions and suggests that more could be done to ensure that the “professional integrity of teachers is where it should be”.

Issue #3: the education system as a whole.

Bart and Alice look beyond all of the specifics of the exam to the nature of education itself. Bart says that our job should be to “develop well-rounded students” and he implies that the nature of instruction and learning with a uniform exam course do not align with this mission. Alice says that the purpose of education should be to discover what individual students are good at and to help them develop their strengths. She shares her own story to explain her belief. She remembers being “a good student, doing everything well, high school, university, managing everything” and then graduating university and entering into the workforce as an engineer, a job using lots of math. In her job, she discovered that “the connection between the math and problem solving in

class and the actual math that I would do as an engineer” was unclear. In fact, she says, she felt as if she became an engineer because she enjoyed math as a student and then realized that the reality of her engineering job was nothing like the math she loved as a student. She believes that the education system should help students to find their strengths and help guide them to jobs and future lives that allow them to shine, to do what they love, and ultimately, to be happy. She says that having a uniform exam forces teachers to teach in a way that does not allow them to help students find their true directions.

Contextualizing Findings in Literature

Many of the emergent themes and sub-themes from the present study are consistent with findings in the literature review contained in chapter 1, and some seem to be unique to the Québec context. As opposed to the thematic organization in the results and discussion, this section will be organized according to categories in the literature review section of chapter 1.

Fundamental issues with standardized testing.

Some participants in the present study explicitly report the belief that success on Québec uniform math exams is not correlated with success in post-secondary mathematics studies, and several participants believe that their own assessment of students is more accurate, valid, and reliable than a student’s result from the uniform exam. Though not empirically verified, these beliefs are consistent with findings in studies looking at predictors of post-secondary academic success (Geiser & Santelices, 2007; Geiser & Studley, 2002; Hoffman & Lowitzki, 2005; Rothstein, 2004).

Furthermore, most participants believe that multiple-choice and short-answer questions, specifically, do not accurately measure student ability or proficiency for a multitude of reasons. Some participants in the present study also believe that Québec uniform math exams promote “test-learning”, timed mathematics performance, and/or “regurgitation” which participants feel are all different than deeper conceptual understanding that we should be trying to assess. These beliefs are consistent with findings that standardized tests do not accurately measure student proficiency (Amrein & Berliner, 2002; Graue & Johnson, 2011). The flipside of this is that many findings in the literature suggest that standardized tests do not promote higher-order thinking because they do not allow students to solve complex problems. Findings here suggest that

teachers feel as if Québec uniform exams do not have this same weakness because of the long-answer section of the exam.

How standardized testing affects teaching and learning.

Findings from the current study suggest consistency with McNeil's findings that standardized tests have a general impact on teaching and learning (2000). Few ideas related to the policy side of existing literature emerged in this study, but many ideas emerged related to the practical side. Firstly, several participants report that they themselves, or that they believe other teachers in Québec, teach differently because of uniform math exams. Some of those participants specifically say that the way they teach exam courses contradicts their own educational beliefs and understanding of best-practices. This is consistent with findings by Abrams et al. (2003) and Anderson (2012).

Secondly, almost all participants in the present study believe that uniform exams affect their course pacing and that this leaves less time to review material not yet mastered or to bring in non-exam material that would be of particular interest to students and/or teachers. These findings are consistent with previous findings (McNeil, 2000).

Next, almost all of the participants in the present study report that Québec uniform exams are repetitive in nature. Most participants report taking measures not to "teach to the test", although many think that other teachers in the province might not be so judicious. Even with these measures taken though, all participants use old exams in one form or another to help students prepare. This means that students have access and exposure to old uniform exams. This means that at least some students are likely to recognize the repetitive nature of exam questions. Most participants believe that at least some students are led to study a sub-set of the course curriculum or will learn how to solve specific exam questions at the expense of learning more general theory. These findings are consistent with other findings in the literature (Després et al., 2013; Jennings & Bearak, 2014).

Finally, several participants in the present study report taking course time to frame the content being taught in the context of uniform exams, spending time teaching about the format and/or nature of uniform exams, and/or teaching explicit study skills to help students be more successful on uniform exams. These findings are also consistent with prior findings in the literature (Cohen, 2015; McNeil, 2000).

Secondary effects of standardized testing.

Various secondary effects emerged which are consistent with the findings in the literature review, and one which is inconsistent. All participants in the present study report that students writing Québec uniform math exams are more stressed about uniform exams than other assessments, although beliefs about the primary root of this stress are varied. This belief is consistent with findings by Abrams et al. (2003). Several participants believe that exam stress has a larger impact on lower-achieving than higher-achieving students, which would be consistent with findings by William (2010). Socioeconomic inequities and student engagement appeared in the literature review but did emerge in the present study.

Findings from the literature review suggest that standardized tests can lead to student disengagement because these tests often encourage teaching minimum standards and place lowered expectations on students (Abrams et al., 2003; McNeil, 2000). Contrary to these findings, some participants from the current study report that some students appear to be more motivated by uniform exams, and no participants brought up issues of student disengagement. It is possible that these differences have to do with the Québec context, although, I believe that it is more likely due to the specific setting of the study. Because the current study was conducted in a school where, typically, students perform well on uniform math exams, issues of disengagement are less likely to emerge.

Possible issues unique to the Québec context.

Certain emergent ideas were not related to findings in the literature review presented in chapter 1, and may be unique to the Québec context. Firstly, several participants discussed the marking guide for short-answer questions and, specifically, the belief that several marks are deducted for minor errors or inaccuracies. Québec uniform exams are only partially centrally corrected and the fact that these questions are marked by classroom teachers but according to prescriptive government guidelines is not standard everywhere. Participants in the current study believe that these questions yield results which are not accurate representations of student ability.

Secondly, the idea of moderation, though it was not brought up by all participants, was the source of very polarized opinions. The process through which the Québec government alters student class results from throughout the school year on a class-by-class basis is also not common widespread practice. Participants felt insulted, devalued, and powerless as a result of this practice.

Thirdly, though there was disagreement between participants about the conclusion, the idea of the fit between the difficulty level of the course curriculum and students' developmental abilities was called into question. This might be something that requires further investigation.

Additionally, issues having emerged under the theme of *other systemic issues* include the possibility that uniform mathematics exams may occur for the wrong reasons, which calls into question the justification for their existence. Other benefits of uniform exams emerged including limits on parental involvement and helping to eliminate inequalities in teacher experience and training across Québec. It might merit investigation as to whether parental involvement and discrepancies in teacher training and experience are particularly relevant issues in Québec versus other contexts.

Big-picture questions arose about whether or not Québec uniform exams are in alignment with the goals of our education system. This may require further investigation, but preliminary results from the present study suggest that uniform mathematics exams in Québec are not perfectly aligned with the philosophy of the QEP according to teachers. In chapter 1, it was reported that uniform exams were implemented in Québec for several reasons, including to “establish a summary of progress” and to “report to the population and provide information on the state of education”. It seems that based on participant responses, these goals are at least partially met by uniform exams. Some results suggest that the “summary of progress” offered by uniform exams may be incomplete or somewhat misleading if uniform exams do not, in fact, measure mathematical proficiency perfectly. Similarly, some participants feel as if the information that would be reported to the population might be misleading because of the extent to which exam results are moderated and/or transformed, which is sometimes very high. Another goal, namely to “support progress in student learning”, seems to have been severely called into question by participant responses. Several participating teachers feel as if uniform exams do a poor job of supporting student learning and many think that the existence, nature, and format of exams actually lead students to learn less effectively. In terms of the some of the benefits the Québec government hoped to achieve: low cost and ease to administer to large numbers of students were only briefly discussed (and not questioned) by one participant, although these were not included in the scope of the original interview questions. Another supposed benefit, however, is the reliability of standardized tests, a claim that was questioned by most participants. In Inchauspé's original report, it was highlighted that closed-response questions like multiple-

choice and short-answer are problematic and do less to promote and/or evaluate higher order thinking skills. In this sense, the participants in the present study agreed unanimously. They report that the multiple-choice and short-answer questions on uniform exams are the ones that they feel do the worst jobs of evaluating student proficiency whereas the long-answer questions, they overall feel, do a much better job.

Strengths, Limitations, and Methodological Observations

The participants in the present study have an average of over 10 years of teaching experience, and all of the participants rated themselves as ‘familiar’ with the secondary IV uniform mathematics examinations in Québec in terms of the content and the format. When asked to rate their familiarity with Québec uniform exams on a 10-scale, all participants rated themselves seven or higher, and the only seven was reported by a teacher who does not currently teach a grade level with a uniform exam. Additionally, all of the participants have studied education, many have studied mathematics in post-secondary settings, and several have completed or are in the midst of graduate degrees. All of this suggests that the data obtained in the present study is based on relevant and extensive professional experience and training. Furthermore, the specific professional experiences and educational backgrounds of the participants are varied which, I believe, increases the likelihood that commonalities in data are legitimate rather than coincidences or bias due to participants working in the same school.

Various steps were taken before interviews took place to eliminate leading questions and to limit confusion/ambiguity. Choices were made concerning timing and location of interviews in order to maximize participant comfort and, I believe, the relationship between participants and the principal investigator added to interviewee comfort as well. By taking steps to increase participant comfort, it is more likely that the data obtained is accurate and that the maximum number of ideas emerged. Finally, in an effort to increase trustworthiness and credibility, my own opinions were included in chapter 3 of this thesis. I believe that some of my responses are quite different from those of the participants, and those that are similar are no more similar to participant responses than some participant responses are to each other. Consequently, I think it is fair to say that interview responses have been analyzed and interpreted with limited researcher bias.

Because of the exploratory, case-study nature of the present thesis, findings are based on a small sample of participants. While I believe that the data helps to paint a robust picture of uniform math exams in the opinions and experiences of teachers, it does so at the expense of generalizability. The lack of generalizability is somewhat mitigated by the fact that the participants brought up different ideas and expressed differing opinions, in many cases, on similar topics. However, having all participants come from the same school introduces a source of selection bias which is impossible to disentangle. Given that there was consensus, or near-consensus on several themes, it is impossible to tell if and/or to what extent this occurred because of the participants teaching in the same school. Being from the same school means that the participants are likely to share common experiences which may be different from teachers in other schools in Québec. Specifically, because the setting of the present study is a Montréal-area, private school, where students typically do well on uniform math exams compared to the rest of the province, certain types of experiences are more likely to have occurred than others. For example, no issues relating to student disengagement arose in interviews though they may still be relevant to other Québec teachers. Similarly, no issues relating to socioeconomic status and its correlation with exam results arose. The fact that the present study was conducted in a private school, however, implies that there is more heterogeneity in student socioeconomic status, which makes this issue less likely to have emerged in this setting. It may, however, still be relevant in other Québec settings. These are simply two examples which appeared in a review of the relevant literature and which did not emerge in the current study. It is certainly possible that other relevant issues may exist in Québec outside of the setting of the current study.

Aside from potentially relevant issues and relationships not emerging because of selection bias, a case study also has the potential to “understate or overstate” relationships that have emerged (George & Bennett, 2005, p. 24). Again, because the setting of the present study is a school where students typically do well on uniform math exams, emergent issues related to poor student results, student difficulties comprehending or mastering course content, and/or student anxiety because of previous failure, are likely to be understated compared to other parts of Québec. Conversely, emergent issues related to student success, ability to comprehend, and/or drive and desire to succeed, are likely to be overstated comparatively.

The exams investigated in the present study have been described in chapter 1, and although there are many ways in which these exams are similar to standardized tests, they are

not, strictly speaking, standardized tests themselves. This did not affect data collection or interpretation of results, though it is possible that some connections that have been suggested between current results and existing literature are not as strongly aligned as they may seem. One example is that uniform math exams in Québec are high-stakes for students, but not for teachers. This means that teachers in Québec are less likely to experience as strong a pressure as teachers elsewhere to teach students content-wise and/or style-wise in a way that promotes exam success over theoretical understanding. Given that participants in the current study believe some of these practices to occur in Québec, though, it seems that this harm is not significant. Another example is that uniform math exams in Québec are not completely objective assessments. The long-answer questions and the short-answer questions, to a certain extent, allow for some interpretation which means that there is some leeway in allowable answers. This suggests that students are less likely to prepare for them in a single, standardized way. Again, since the idea of students learning either a reduced sub-set of course content or how to solve specific problem types did emerge in the present study, it seems that this difference may not have been significant either. Other potential difference might exist, though, which should be kept in mind.

Some methodological choices may also have had an impact on results. Firstly, though it was intended to improve the quality of interview questions, I amended questions after the first few interviews. That means that later interviews likely had fewer sources of question confusion which may have affected the overall quality of earlier results compared to later ones. Some of the changes made were also based on my feeling that some questions were too leading. This means that the results from the first few interviews may contain additional bias. Secondly, I conducted the first two interviews close together in time, and then waited several months before conducting the other five due to personal constraints. The inconsistency in time between interviews affected my comfort as an interviewer, which may also have affected interviewee comfort and possibly quality of results. Thirdly, the average duration of my first three interviews was significantly shorter than the average duration of my final three interviews. It is possible that part of this difference is due to the fact that two out of the first three participants were middle school teachers, and possibly had less to say about uniform exams (which are conducted in senior school), but based on my own sense of comfort, I think that I also improved as an interviewer as the process went on. Based on my observation, but not supported by any statistics, it seems as if the ratio of participant text to interviewer text is higher in transcripts of later interviews than

earlier ones. As I continued to conduct interviews, I believe that I got better at waiting for participant responses and asking for further clarification rather than moving forward to the next question. In retrospect, I believe that it would have improved the quality of data had I returned to each participant for a second round of interviews after having completed the first round.

Answers to Research Questions

The two research questions of the present study are restated below followed by what the data suggests to be partial answers to those questions.

- 1) How do secondary school mathematics teachers in Québec view provincial uniform leaving exams in mathematics?

There was disagreement among participants about whether or not there should be uniform mathematics exams in Québec. Some participants think that there should be, others think that there shouldn't be, and some believe that there should be summative exams, but not necessarily produced and evaluated by the government. Even the participants who agree with having uniform math exams, however, do not think that the format, content, or nature of the existing exam is perfect. Many problematic exam issues arose, particularly surrounding the multiple-choice section of the exam, and the marking guidelines for the short-answer section. Generally, there was consensus among participants that the long-answer section of the exam is effective at assessing student proficiency. Another problematic issue that arose has to do with the validity and reliability of uniform exams, both of which were called into questions for various reasons. Finally, though few participants brought up the idea of government evaluation, and moderation in particular, those who did discuss it were in strong agreement that it is detrimental to teachers feeling trusted and empowered by the government.

- 2) What impact, if any, does the existence of uniform examinations have on Québec secondary mathematics instruction, according to secondary school math teachers?

Participant data strongly suggest that the existence of uniform examinations does, indeed, have an impact on instruction, according to secondary school math teachers. As for the form and specific details of this impact, several ideas emerged. Firstly, it was nearly unanimous in the present sample that there were effects on course pacing, and most participants had experienced the consequential effects of moving too quickly to guarantee deeper student learning and/or

having to limit opportunities to explore non-exam topics of student or teacher interest. Secondly, all participants brought up the idea of student stress as an impact of uniform math exams. Participants believe that the high-stakes nature of the exam for students, the personal importance for some, and/or the idea that someone unknown decides the content and partial evaluation of the exam, make students more stressed about uniform exams than other forms of assessment. As brought up by several participants, this stress is likely to affect lower-achieving students more significantly than higher-achieving students which further calls into question the validity of uniform exams.

Finally, at the intersection of different themes and sub-themes, is the impact of uniform exams on teaching, learning, and specific preparation for the final exam. There was consensus in this case study that teachers use old uniform exams as practice with their classes. There was also agreement among participants that the exams are repetitive in nature, and that similar questions appear, especially over a cycle of two or three years. Teachers in the present study generally take action to avoid teaching in a way that promotes students learning a sub-set of the course content or having students learn to solve specific questions as the expense of general theory. However, most of the participants teach students about the nature of uniform exams and provide guidance about how to prepare for, effectively write, and/or pace oneself during the exam. In addition to this, most participants believe that “teaching to the test” is something that other Québec teachers do or might do, and that “learning to the test” is something that some students do anyways. This seems reasonable, given that teachers in the present study find uniform exams repetitive, that all students have access to the exams, and that teachers believe that because of the importance of the exam, students will do what they can to do well on the exam. In light of these beliefs and experiences, it seems likely that the old exams will have an influence on how students prepare themselves for the final exam. Teachers have undergone training, they have experience, professional judgment, and pride in their professionalism, so it makes sense that teachers would take action to ensure that they are not teaching in a way that promotes exam success over true understanding. However, students do not have the same training, experience, or big-picture thinking with respect to the uniform exam or the course content as a whole. Students only have one year to experience this course and its exam, and I think that this makes them more likely to make use of the repetitive nature of the exam.

Pedagogical Implications & Recommendations for Future Research

Firstly, it is important to reiterate that the present study is an exploratory case study. Because of its nature, the present study can “make only tentative conclusions” on how the emergent relationships and issues “generally contribute” to other contexts, even within Québec (George & Bennett, 2005, p. 25). Because of this, the pedagogical implications are largely localized and the recommendations are exclusively intended for future research and should not be over-extended to suggest any immediate widespread changes in practice or policy.

Based on the findings of the present study, there are certain pedagogical implications for my own personal teaching practice. Firstly, I have discovered various study skills and exam preparation techniques used by my colleagues. Through study interviews, I have come to learn how these various techniques are used and about their effectiveness. I will begin to use some of these techniques with my classes. Similarly, ideas for different approaches to conjecture questions than the ones I currently use emerged in participant responses, and I will try to use them with my students. Additionally, given the strong emergence of ideas relating to student stress, and given that uniform exams are not likely to be eliminated any time soon, I believe that it is worthwhile to investigate and consider bringing techniques to my own classes to help reduce student stress.

There are also some findings from the present study that I believe warrant further action on the local level. Given the strong participant agreement about issues with multiple-choice questions in general and short-answer question marking in particular, I believe that the next step is to seek the opinions of teachers in the association of schools to which the school in the present study belongs (this association is like a school board but for private schools). If similar findings apply in the larger group, it might be reasonable to bring the issues to the attention of government officials with the backing of a larger group. Given the fact that some participants felt very strongly about exam moderation, I think that the same strategy of seeking the opinions and experiences of a larger group of teachers is warranted.

One more major recommendation can be made without hesitation or reservations, which is that a wider-reaching, more formalized investigation of the impact of uniform mathematics exams on instruction and learning in Québec should be conducted. Educational reform, even if it only targets the exam of a single subject area in a single academic grade, is a process that requires significant justification through extensive documentation. Both this documentation

process and the reform process itself are lengthy in time and financially costly. Therefore, an exploratory study, particularly one conducted in a single school, is not sufficient to justify changes to exams administered across the entire province.

The ultimate goal of this future research is to inform potential policy change. Consequently, it is recommended that this future study be official in the sense that the government initiates and recognizes it. It is also recommended that it be administered through a partnering academic institution rather than directly by the government itself. This is recommended to increase objectivity and to encourage participants to be forthcoming and comfortable, given the possibility that teachers might feel themselves in an uncomfortable power-relationship with government officials.

Based on the rationale for and findings from the present study, it is recommended that a future official investigation into the impact of uniform mathematics exams on instruction and learning in Québec be conducted in two phases. It is suggested that the first phase be qualitative in nature like the present study. Perhaps semi-structured interviews could be conducted with teachers from various school boards (including some private schools) from across the province. In this sense, the first phase of the recommended study would be a continuation/extension of the present study to include a wider, more diverse group of participants and, consequently, experiences. I also think that it would be beneficial to include at least one new teacher (in his/her first three-five years teaching) and at least one experienced teacher (with over 10 years of teaching experience) from each school board. To facilitate data collection, these interviews could be conducted remotely, as there is no reason to believe that this would affect the quality of data collected. The themes that emerged from the present study could be used to inform the study design and potential interview questions of this initial qualitative study. The goal of this larger study would be to confirm what issues and/or relationships exist surrounding/between uniform math exams and teaching and learning.

Based on the themes that emerge from this first phase, a second phase of official investigation should aim to collect a more comprehensive set of data in the aim of determining to what extent the issues and/or relationships from phase one exist. At this point, I think that quantitative data would be appropriate. After the initial phase, a larger sample size would be ideal to determine the extent or strength of relationships. It would be more feasible to collect and interpret quantitative data from this larger sample, and quantitative data is more appropriate to

ascertain effect size/strength. Furthermore, quantitative data would allow for a more granular analysis of various issues and could be helpful in prioritizing actionable steps.

Finally, it had been the original intention in the current study to examine potential differences between middle school and senior school teachers in their opinions and experiences. Having only had two middle school participants, however, it was not possible with any amount of confidence to establish commonalities between the two that were different than trends among the other five participants. It is recommended that future research into this topic examine potential differences between teachers based on whether or not they teach the grade level with uniform exams. Similarly, as there was only one male participant in the present study, it was not possible to examine differences between genders or commonalities within. It would be interesting for future research into the topic to examine the potential impact of gender as well.

Conclusion

In conclusion, the present study helps to enrich the current understanding of uniform math exams in terms of teacher opinions and it provides preliminary insight into teacher and learner practices surrounding uniform math exams in Québec. The results from the current study suggest that uniform math exams in Québec have an impact on secondary school mathematics instruction, and should be used to inform further research into uniform mathematics exams in Québec with the ultimate goal of improving educational practice and policy.

References

- Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions on statewide testing programs. *Theory into Practice, 42*(1), 18-29.
- Amrein, A.L. & Berliner, D.C. (2002, March 28). High-stakes testing, uncertainty, and student learning Education Policy Analysis Archives, 10(18). Retrieved August 4, 2017 from <http://epaa.asu.edu/epaa/v10n18/>.
- Anderson, K. J. B. (2012). Science education and test-based accountability: Reviewing the relationship and exploring implications for future policy. *Science Education, 96*(1), 104-129.
- Basit, T. N. (2003). Manual or electronic? The role of coding in qualitative data analysis. *Educational Research, 45*(2), 143-154.
- Bogdan, R. C., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theories and methods* (5th ed). Boston, MA: Allyn & Bacon.
- Cohen, J. (2015). Challenges in identifying high-leverage practices. *Teachers College Record, 117*(7), 1-41.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record, 106*(6), 1047-1085.
- Després, S., Kuhn, S., Ngirumpatse, P., & Parent, M. J. (2013). *Real accountability or the illusion of success? A call to review standardized testing in Ontario*. Action Committee Task Force Report. Ottawa: Action Canada.
- Dey, I. (2003). *Qualitative data analysis: A user-friendly guide for social scientists*. London: Routledge.
- Ellis, M. (2008). Leaving no child behind yet allowing none too far ahead: Ensuring (in)equity in mathematics education through the science of measurement and instruction. *Teachers College Record, 110*(6), 1330-1356.
- Geiser, S., & Santelices, M. V. (2007). Validity of high school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes. University of California, Berkeley Center for Studies in Higher Education Research & Occasional Paper Series: CSHE.6.07.

- Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8(1), 1-26.
- George, A. L., & Bennett, A. (2005). Case studies and theory development in the social sciences. Cambridge, MA: MIT Press.
- Glesne, C. (2016). *Becoming qualitative researchers: An introduction* (5th ed.). Toronto, ON: Pearson.
- Graue, E., Johnson, E. (2011). Reclaiming assessment through accountability that is “just right”. *Teachers College Record*, 113(8), 1827-1862.
- Hancock, D. R., & Algozzine, B. (2011). *Doing case study research: A practical guide for beginning researchers*. New York, NY: Teachers College Press.
- Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *The Review of Higher Education*, 28(4), 455-474.
- Inchauspé, P. (1997). *Reaffirming the mission of our schools: Report of the task force on curriculum reform*. Québec, QC: Ministry of Education of Québec.
- Jennings, J. L., & Bearak, J. M. (2014). “Teaching to the test” in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher*, 43(8), 381-389.
- Kennedy, M. M. (2002). Knowledge and teaching. *Teachers and Teaching*, 8(3/4), 355–370.
- Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.
- Koretz, D. (2003). Using multiple measures to address perverse incentives and score inflation. *Educational Measurement: Issues and Practice*, 22(2), 18–26.
- Kuzmic, J. (1994). A beginning teacher’s search for meaning: Teacher socialization, organizational literacy, and empowerment. *Teaching & Teacher Education*, 10(1), 15-27.
- McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge

- Ministère de l'Éducation et de l'Enseignement Supérieur. (2014). *Results on the June Ministerial Examinations*. Retrieved from:
<http://www.education.gouv.qc.ca/en/references/publications/results/detail/article/results-on-the-june-uniform-ministerial-examinations/>
- Ministère de l'Éducation du Québec. (2001). *Québec Education Program*. Québec: Gouvernement du Québec.
- Orafi, S.M., & Borg, S. (2009) Intentions and realities in implementing communicative curriculum reform. *System*, 37(2), 243–253.
- Phelps, R. P. (2000). Trends in large-scale testing outside the United States. *Educational Measurement: Issues and Practice*, 19(1), 11-21.
- Rothstein, J. (2004). College performance predictions and the SAT. *Journal of Econometrics*, 121(1-2), 297-317.
- Schram, T. H. (2006). *Conceptualizing and proposing qualitative research*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Scogin, S. C., Kruger, C. J., Jekkals, R. E., & Steinfeldt, D. (2017). Learning by experience in a standardized testing culture: Investigation of a middle school experiential learning program. *Journal of Experiential Education*, 40(1), 39-57.
- Sfard, A. (2005). What could be more practical than good research? On mutual relations between research and practice of mathematics education. *Educational Studies in Mathematics*, 58(3), 390-411.
- Skott, J., Van Zoest, L., & Gellert, U. (2013). Theoretical frameworks in research on and with mathematics teachers. *ZDM The International Journal on Mathematics Education*, 45(4), 501-505.
- Sohn, B. K., Thomas, S. P., Greenberg, K. H., & Pollio, H. R. (2017). Hearing the voices of students and teachers: A phenomenological approach to educational research. *Qualitative Research in Education*, 6(2), 121–148.
- van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. New York, NY: SUNY Press.

William, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 30(2), 107-122.

Yuksel, P., & Yildirim, S. (2015). Theoretical frameworks, methods, and procedures for conducting phenomenological studies in educational settings. *Turkish Online Journal of Qualitative Inquiry*, 6(1), 1–20. Retrieved from <http://dergipark.ulakbim.gov.tr/tojqi/article/view/5000093521/5000087061>

Appendix A
Proposed Invitation Letter & Written Consent Form – School Administration



INFORMATION

Study Title: *Standardized Testing in Quebec Mathematics Education: An Exploratory Study of the Perspectives of Secondary Mathematics Teachers in a Private School in Montréal*

Researcher: Dan Elbling

Researcher's Contact Information: elblingd@selwyn.ca

Faculty Supervisor: Arpi Hamalian

Faculty Supervisor's Contact Information: arpi.hamalian@sympatico.ca (tel. 514 848 2424 ext. 2014)

Dear school administration,

I am doing research towards writing my MA thesis in Educational Studies at Concordia University and I would be delighted if you would agree to participate in the project. The current project seeks to add to our understanding of standardized testing and its impacts on mathematics education in the Quebec education system. Specifically, the study aims to investigate the perspectives of secondary mathematics teachers with respect to standardized testing.

The teachers being sought to participate are senior school and middle school math teachers. Participation in this study would mean taking part in a 45-60 minute one-on-one interview after school hours for the teachers in question. With your permission, the interviews will take place on the school premises or at a safe location of your choice. Interviews will be recorded with participant permission, transcribed, and analyzed for themes. The transcript of each interview will be shared with the participant for vetting with a deadline of two weeks after receipt of the interview transcript. The ultimate goal of the research project is to publish results, though the names of the school and participating teachers will not be published.

The successful completion of my MA thesis depends on this research project. Please note that Concordia University will give permission to conduct this study for my MA thesis based on the clearance from the school administration to allow me to conduct this study as planned. I am now

seeking your permission to send my invitation letter to eight colleagues who teach math at the middle and senior school levels.

If you have any questions or concerns, please do not hesitate to email me. Thank you for your time and consideration.

Dan Elbling
elblingd@selwyn.ca

Appendix B
Proposed Invitation Letter & Written Consent Form – Participant

Dear math teacher,

I am doing research towards writing my MA thesis in Educational Studies at Concordia University and I would be delighted if you would agree to participate in the project. The current project seeks to add to our understanding of standardized testing and its impacts on mathematics education in the Quebec education system. Specifically, the study aims to investigate the perspectives of secondary mathematics teachers with respect to standardized testing.

Senior and middle school math teachers are being sought to participate. Participation in this study would mean taking part in a one-on-one interview after school hours, which will last approximately 1 hour. Interviews will be recorded with participant permission, transcribed, and analyzed for themes. The transcript of each interview will be shared with the participant for vetting with a deadline of two weeks after receipt of the interview transcript. You will have two weeks to edit, revise, add information or delete some parts of the interview if you so wish. The ultimate goal of the research project is to publish results, though the names of the school and participating teachers will not be published.

If you would consent to participating in this study, please fill out the form below. Please note that the school administration has given permission to conduct this study. The interview based narrative and information you provide are totally confidential. I as the researcher and my thesis supervisor are the only ones who will have access to the information provided. Your name will never be used. You will be assigned a pseudonym. You should also note that you can decide during the interview that you no longer want to participate in the research project. You can also decide to have your data excluded from the research project after the interview has been completed with the last withdrawal being two weeks after you receive the transcript of your interview for vetting. You can withdraw your participation by simply advising me, Dan Elbling (elblingd@selwyn.ca) and/or my supervisor (arpi.hamalian@concordia.ca)

Once you return the vetted transcript, data analysis will begin, and data can no longer be withdrawn. If you have any questions or concerns, please do not hesitate to email me.

Completed consent forms can be returned to me in person at school, or scanned via email to elblingd@selwyn.ca

Thank you for your time and consideration.

Dan Elbling
elblingd@selwyn.ca



INFORMATION AND CONSENT FORM

Study Title: *Standardized Testing in Quebec Mathematics Education: An Exploratory Study of the Perspectives of Secondary Mathematics Teachers in a Private School in Montréal*

Researcher: Dan Elbling

Researcher's Contact Information: elblingd@selwyn.ca

Faculty Supervisor: Arpi Hamalian

Faculty Supervisor's Contact Information: arpi.hamalian@concordia.ca

Source of funding for the study: No funding

You are being invited to participate in the research study mentioned above. This form provides information about what participating would mean. Please read it carefully before deciding if you want to participate or not. If there is anything you do not understand, or if you want more information, please ask the researcher Dan Elbling (elblingd@selwyn.ca)

A. PURPOSE

The purpose of the research is to gain insight into teacher perspectives and practices regarding standardized testing in Quebec and its potential impact on instruction.

B. PROCEDURES

If you agree to participate, you will be asked to participate in a one-on-one interview. Dan Elbling, the researcher, will conduct the interview in an un-used classroom at the school after school hours. The interview will be recorded to facilitate transcription.

Questions will focus on the participant's experience teaching and opinions about uniform exams in the Quebec mathematics education program.

In total, participating in this study will take approximately one hour (plus whatever time it takes to read and approve a draft of the interview transcript – maybe another hour at most).

C. RISKS AND BENEFITS

No special risks are involved in your participation of this research project. The interview results and narratives you provide are strictly confidential.

Potential benefits include: being able to reflect on your professional practices with the possible contribution to a better understanding of your own practices as a math teacher.

D. CONFIDENTIALITY

We will gather the following information as part of this research: your insights, perspectives, and practices concerning standardized testing. Specifically, what (if any) impacts standardized tests have on students, on teachers, on day-to-day assessment, or other facets of mathematics instruction.

We will not allow anyone to access the information, except myself (the researcher) and my faculty supervisor. We will only use the information for the purposes of the research described in this form.

The information gathered will be coded by use of pseudonym. That means that the information will be identified by a code. The researcher will have a list that links the pseudonym to your name. The faculty supervisor will review data in which names have already been replaced by pseudonyms, but will also have access to consent forms with participants' real names.

We will protect the data by encrypting it. We intend to publish the results of the research. However, it will not be possible to identify you in the published results. We will destroy the data five years after the date of the successful defence of the MA Thesis.

F. CONDITIONS OF PARTICIPATION

You do not have to participate in this research. It is purely your decision. If you do participate, you can stop at any time. You can also ask that the information you provided not be used, and your choice will be respected. If you decide that you don't want us to use your information, you must inform the researcher at the latest, two weeks after you receive the transcript of your interview. There are no negative consequences for not participating, stopping in the middle, or asking us not to use your information. In order to withdraw your participation until the last deadline mentioned above you should inform the researcher, Dan Elbling (elblingd@selwyn.ca) and/or my thesis supervisor, Arpi Hamalian (arpi.hamalian@concordia.ca).

G. PARTICIPANT'S DECLARATION

I have read and understood this form. I have had the chance to ask questions and any questions have been answered. I agree to participate in this research under the conditions described.

NAME (please print) _____

SIGNATURE _____

DATE _____

If you have questions about the scientific or scholarly aspects of this research, please contact the researcher. Their contact information is on page I. You may also contact their faculty supervisor (arpi.hamalian@concordia.ca).

If you have concerns about ethical issues in this research, please contact the Manager, Research Ethics, Concordia University, 514.848.2424 ex. 7481 or oor.ethics@concordia.ca.

Appendix C

Sample Interview Questions

The questions below are intended to initiate a fluid conversation. Consequently, the actual interview may not follow this list of questions as a script, but as questions to refocus the interview, or to get respondents to begin sharing their opinions, beliefs, and knowledge. Respondents who provide more information and/or who lead the conversation will be encouraged to do so.

Background questions:

1. How long have you been a teacher in Quebec?
2. Have you taught mathematics the whole time?
3. What different grade levels have you taught?
4. How many years have you taught courses with standardized tests?
5. How would you outline your philosophy as a teacher?
6. What do you consider to be your primary goal as a mathematics teacher?

Questions about standardized tests:

1. How familiar would you say you are with standardized mathematics tests in Quebec?
(rate yourself on a ten-scale)
2. Which level(s) of math do you most frequently teach?
3. Have you found any questions or types of questions which frequently recur on standardized tests?
4. Are you aware of any course topic which is seldom/never tested on standardized tests?
 - a. If so, which one(s)?
5. Are you aware of any specific skill within a curricular topic which is seldom/never tested on standardized tests?
 - a. If so, which ones(s)?
6. Do you feel that there are any topics which are important for students to demonstrate mastery of the course curriculum which are not frequently tested on standardized tests?
 - a. If so, which ones(s)?

7. Do you feel that there are any specific skills which are important for students to demonstrate mastery of the course curriculum which are not frequently tested on standardized tests?
 - a. If so, which ones(s)?
8. Do you feel that there are questions on previous standardized tests which did not accurately assess one or more topics in the course curriculum?
 - a. If so, which ones(s)?
9. Do you feel that there are questions on previous standardized tests which did not accurately assess one or more skills in the course curriculum?
 - a. If so, which one(s)?
10. What are some instructional methods you use frequently?
 - a. Do you feel that these methods are effective in allowing students to learn basic concepts?
 - b. Do you feel that these methods are effective in allowing students to learn and practice applying required skills?
 - c. Do you feel that these methods are effective in helping students to be successful on standardized tests?
11. Do you do anything specific to help students to be/feel prepared for the standardized test?
 - a. If so, what specifically?
12. Do you devote any class time to showing students old standardized tests?
 - a. If so, how much?
13. Do you devote any class time to practicing old standardized test questions?
 - a. If so, how much?
14. Do you teach any topics which you feel are unimportant, but which you know are likely to appear on the standardized test?
 - a. If so, which one(s)?
 - b. How much time would you estimate you spend on this/these topic(s)?
15. Do you feel that students are stressed about standardized tests?
16. Do you feel that standardized tests help student to properly study the course content?
17. Do you feel that previous standardized tests help guide students to study appropriately?

18. Do you feel that studying previous standardized tests causes students to focus on a narrowed scope of the course content?

Overall questions:

1. Do you believe that standardized tests are effective at evaluating students' basic knowledge of the topics/concepts in the Quebec mathematics curriculum?
2. Do you believe that standardized tests are effective at evaluating students' mastery of the prescribed Quebec mathematics curriculum?
3. Do you believe that standardized tests accurately evaluate students' basic knowledge of the prescribed Quebec mathematics curriculum?
4. Do you believe that standardized tests accurately evaluate students' mastery of the prescribed Quebec mathematics curriculum?
5. Do you believe that standardized tests have any direct effects on your instruction?
 - a. If so, what effect(s)?
6. Do you believe that standardized tests have any indirect effects on your instruction?
 - a. If so, what effect(s)?
7. Do you believe that standardized tests have any direct effects on students in terms of their day to day school experiences?
 - a. If so, what effect(s)?
8. Do you believe that standardized tests have any indirect effects on students in terms of their day to day school experiences?
 - a. If so, what effect(s)?