THE NOVA SCOTIA ENGLISH 10 EXAMINATION: A CASE STUDY OF CONSEQUENCES

STEFANIE COX
Bachelor of Education, University of Toronto, 2009

A thesis submitted
in partial fulfillment of the requirements for the degree of

MASTER OF EDUCATION

in

GENERAL

Faculty of Education
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

THE NOVA SCOTIA ENGLISH 10 EXAMINATION: A CASE STUDY OF CONSEQUENCES

STEFANIE COX

Date of Defence: June 24, 2020

Dr. D. Slomp Thesis Supervisor	Associate Professor	Ph.D.
Dr. L. Fowler Thesis Examination Committee Member	Professor Emerita	Ph.D.
Dr. R. LeBlanc Thesis Examination Committee Member	Assistant Professor	Ph.D.
Dr. A. Murray-Orr External Examiner St. Francis Xavier University Antigonish, Nova Scotia	Professor	Ph.D.
Dr. D. Burleigh Chair, Thesis Examination Committee	Assistant Professor	Ph.D.

DEDICATION

To Wyatt and Everett, for your patience while I worked on this. I love you to the moon and back.

ABSTRACT

This study aimed to better understand the unintended consequences of the Nova Scotia Examination: English 10 for students and teachers. Two classrooms were involved in this multisite case study. Data was gathered through interviews with teachers and students, classroom observation, and document review. The study found that the Examination's limited scope and ambiguous questions threatened student self-efficacy going into the Examination. Document review revealed potential construct underrepresentation within the larger program due to a lack of compensation for constructs not represented on the Examination. Other unintended consequences included influence on students' perceptions of literacy and a reduction in teacher morale. Recommendations include the review of multiple-choice questions, increased communication about use of results, and the consideration of teacher-created summative assessment.

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge the support I have received from my husband throughout this process. Brian, I could not have asked for a more patient or helpful partner throughout the planning, execution, and writing of this study. You have provided company during so many late nights of writing, and your own work ethic and drive inspire me. Thank you.

I would like to acknowledge my wonderful colleagues, who have provided an ear, an opinion, and a helpful push when needed. Ann Marie, thank you for always having the answer to my intensely specific grammatical questions. Dr. Mike, thank you for the weekly check-ins and encouragement. Steven, thank you for lending your ear and your expertise in the field of local history.

I would like to thank my committee for their support throughout this long process. Leah, Robin, and Robert, I am blessed to have access to the wealth of knowledge you possess. Thank you for your constructive feedback and encouragement.

Finally, thank you to David, my thesis supervisor. Your own passion for assessment led me to this field of research. This thesis has been an epic adventure, and you have patiently waited for updates through cross-country moves, teacher strikes, the birth of my youngest son, and the COVID-19 pandemic. Thank you for answering questions any time of day, regardless of the time zone. Most importantly, thank you for your guidance and support.

TABLE OF CONTENTS

Dedication iii

Abstract iv

Acknowledgments v

Table of Contents vi

List of Tables ix

List of Figures x

List of Abbreviations xi

Chapter 1. Introduction 1

Autobiographical Beginnings 3

Situating the Context 9

Rationale and Objective 18

Research Questions 19

Chapter 2. Theoretical Framework 20

Multiliteracies 20

Validity Theory 21

Chapter 3. Curriculum and Assessment in Grade Ten 29

The Program of Learning Assessment for Nova Scotia 29

The NSE: English 10 30

Construct Representation in the ELA Curriculum and the Examination 37

Chapter 4. Review of Related Literature 45

Local Publication and Analysis of Data 45

Reporting on Racial Inequity in Nova Scotia 47

Consequential Validity Inquiries in Canada 50
Chapter 5. Methodology 54
Qualitative Inquiries and Constructivism 54
Multisite Case Study 55
Data Collection 56
Timeline of Research 63
Validity, Reliability, and Generalizability 64
Ethical Considerations 67
Chapter 6. Case Profiles 69
Case One: Eastern High 69
Case Two: Oceanview Collegiate 74
Chapter 7. Findings 82
Organization of Data 82
School Contexts 84
Classroom Procedures 94
Student Perceptions of Constructs 108
Perceptions of the Examination 111
Connecting the Themes 128
Chapter 8. Discussion of Findings 129
Conclusions 131
Chapter 9. Recommendations, Limitations, and Areas of Future Study 148
Recommendations 148
Limitations 159

Areas of Future Study 166

References 168

Appendices 177

A: Cognitive Levels of Selected Answer Questions 177

B: Teacher Interview Questions 178

C: Student Interview Questions 179

D: Sample Writing Rubric 180

LIST OF TABLES

Table 1: Overview of Provincial Assessments and Evaluations in Nova Scotia 30
Table 2: NSE: English 10 Reading Tasks 32
Table 3: NSE: English 10 Writing Tasks 32
Table 4: Curricular Strands and Focus Categories 39
Table 5: Speaking and Listening Sample Assessment Events 40
Table 6: Construct Representation within Strand 2 42
Table 7: Construct Representation within Strand 3 43
Table 8: Timeline of Data Collection 64
Table 9: Overlap of Validity Issues Within Conclusions 132

LIST OF FIGURES

Figure 1: Sample Multiple-Choice Question from *Lessons Learned* Document.....36
Figure 2: Codes and Child Codes Used for Initial Data Review 83
Figure 3: Sample Multiple-Choice Question from Exemplar.....112
Figure 4: Sample Multiple-Choice Question from Exemplar (2).....113
Figure 5: Significant Codes Distributed by School Name.....129

LIST OF ABBREVIATIONS

BLAC	Black Learners Advisory Committee
ELA	English Language Arts
EAL	English as an Additional Language
HRSB	Halifax Regional School Board (before April 2018)
HRCE	Halifax Regional Centre for Education (after April 2018)
IB	International Baccalaureate
IEP	Individual Education Plan (Ontario)
IPP	Individual Program Plan (Nova Scotia)
NSE	Nova Scotia Examination
NSTU	Nova Scotia Teachers Union
OSSLC	Ontario Secondary School Literacy Course
OSSLT	Ontario Secondary School Literacy Test
PLANS	The Program of Learning Assessment for Nova Scotia

CHAPTER 1: INTRODUCTION

Standardized assessment is a fixture of Canadian education. Currently, every province administers standardized assessments, most commonly in the form of English or Literacy assessments at the high school level (Quigley, 2018). The suggested benefits of these assessments include increased accountability and the monitoring of student outcomes. This is certainly the goal in Nova Scotia, where the Program of Learning Assessment for Nova Scotia claims provincial assessments "determine the effectiveness of curriculum delivery," "assist students to achieve outcomes," and "help teachers understand assessment principles and practices" (Government of Nova Scotia, 2019a). Low-stakes assessments are administered in grades three, six, and eight in Literacy and Mathematics. Two higher-stakes evaluations with an impact on student grades take place in grade ten: the first in Mathematics, and the second in English Language Arts. The grade ten English Language Arts Examination will be the focus of this study.

Recently in Nova Scotia, standardization has come under the spotlight for two reasons. The first is politics. During the 2016/17 school year, a stalemate during contract negotiations between the Nova Scotia Teachers Union (NSTU) and the government of Nova Scotia led to work-to-rule action. Following the work-to-rule, a Council to Improve Classroom Conditions, comprised of teachers, parents, and guidance counselors was created to find solutions to numerous problems, the first and second of which involved assessment and evaluation (Nova Scotia Teachers Union, 2017a). Their April 2017 Report lists 40 recommendations, among which was the recommendation to end provincial assessments and evaluations, since teachers "are in the best position to assess their students' progress, and additional provincial or board assessments take time away

from teaching and learning" (p. 4). The timing of the report, released ahead of a provincial election, created significant media attention.

Secondly, score discrepancies between different groups have stoked public concern. The recent introduction of self-identification surveys has allowed the Department of Education and school boards to attach racial information to results, thereby identifying discrepancies between the results of racial groups (Halifax Regional School Board, 2014; Chiasson, 2015). Public debate is frequently framed around allegations of declining achievement, weak literacy support, and unmet goals (Literacy help, 2015, p. A8; Willick, 2013; Willick, 2014). The progress of students of African ancestry is especially worrisome, with reports citing "alarming" literacy shortfalls in the province's two largest school boards (Jeffrey, 2014). Solutions have focused on assessment and accountability, but with opposite approaches; the optimal level of classroom oversight appears to vary with the stakeholder group posing the question.

Despite the availability of quantitative data, researchers have little insight into those most directly involved with and affected by the assessments themselves—students and teachers. This study aims to contribute qualitative information to the conversation. Through interviews, observation, and document review, the study examines students' and teachers' experiences with the Nova Scotia Examination: English 10 ("NSE: English 10" or "Examination") with the goal of better understanding the assessment's unintended consequences. A multisite case study approach allows differences between the two schools to add depth to the research. At minimum, these case study interviews lend a voice to the students and teachers who currently lack agency in the public discourse.

Autobiographical Beginnings

According to Merriam (2009), the researcher's position holds great significance to a study. The articulation of one's experiences, values, and perspectives may allow the reader to "better understand how the individual researcher might have arrived at the particular interpretation of the data" (p. 219). At the outset of this report, I am therefore including the following account of my experiences and subjectivities.

The OSSLT and OSSLC

I was first introduced to standardized assessment in high school. In tenth grade, I was one of tens of thousands of Ontarian students to write the second round of the Ontario Secondary School Literacy Test (OSSLT). The year prior had been the pilot year, and our older peers told fearsome stories about the length and demands of the test. Hoping to prepare us, teachers gave quizzes requiring scrutiny and attention to detail. We were warned that writing beyond the given space would constitute an incorrect response. Though English is my maternal language and I received high grades in English class, the high stakes of the OSSLT frightened me; the threat of being denied a high school diploma was too great to be ignored. When the day of the test arrived, I nervously opened my booklet to find familiar genres and formats: a news article, an essay, and multiple-choice questions, among others. I passed the test, as did my friends, and we wondered who possibly could have failed.

In 2010, I met those children. I had been hired at a Scarborough high school to teach two sections of the Ontario Secondary School Literacy Course (OSSLC), a remedial course for students who had failed the OSSLT. Most of the students in my classroom had recently immigrated to Canada and were in fact in grade twelve. As a

recent graduate of a Bachelor of Education program, I was eager to implement the latest in language teaching methods, to make my program authentic to my students, and to celebrate students' diversity. However, I quickly felt the limits of rigid curricular outcomes and the themes presented in the course textbook. I read stories about baseball legends with students who had never seen a game. I taught the components of news articles to students who, while unfamiliar with print news, were exceedingly familiar with electronic communication. Though most of my students achieved high marks in other courses, I watched them grow discouraged by the low marks they received on writing assignments. Furthermore, I worried about the high stakes of the course; like the OSSLT, failure of the OSSLC would result in a student being ineligible for graduation. Did my students' fate rest on my shoulders alone?

A guidance counselor soon quelled my anxiety: their unwritten policy was to alter the grades of failing students so they could graduate with their classmates. I felt conflicted. On the one hand, I no longer held the responsibility of deciding whether a student should graduate. On the other, I questioned the purpose of the OSSLC and OSSLT altogether. The goal of improving education "by helping to identify areas of learning that may need more attention" (Education Quality and Accountability Office, 2014, p. 3) sounded worthy, but I was unaware of any communication between test markers and classroom teachers. If the OSSLC was not identifying these areas of need, then why bother? The test and course temporarily segregated students from minority populations into a remedial course. The students placed in the OSSLC expressed a sense of discouragement at being labeled "illiterate," and I wondered if our efforts, though born of a desire to help all students succeed, could be doing more harm than good.

Assessment in the West

Shortly thereafter, I moved from the urban setting of Toronto to the northern reaches of rural Alberta. Many of the towns in which I worked had a smaller population than my previous school, and my diverse classrooms of different languages were replaced by unilingual classes of students who had never left the province. For all of my students, English was the first language, though a few also spoke some Cree, a language and culture about which I knew nothing. Indeed, when it came to the history of First Nations in Canada, I had a lot to learn.

After a few months of substitute teaching, I was hired to teach Social Studies at the local high school. The Alberta Social Studies curriculum places substantial emphasis on the legacy of colonialism, including residential schooling, relocation to reserves, and paternalistic policies (Alberta Education, 2005). In my classrooms, I saw first-hand the legacy of these policies in the attitudes of students of First Nations, Métis, and Inuit (FNMI) heritage. These students were more likely to have poor attendance, to exhibit defiant behavior, and to be disengaged from lessons. Navigating the cultural divide between students of FNMI heritage and their white peers was difficult while fielding racist comments from white students during class discussions. Managing defiant behavior was especially difficult after learning that some of my FNMI students had been raised in foster care, subjected to physical and sexual abuse, and had otherwise experienced significant trauma. Though I had arrived in the North with little knowledge of FNMI cultures and languages, my role demanded that I find ways to engage all my students by creating authentic and relevant lessons and tasks. For me, this included offering the option to present knowledge orally or musically, having one-on-one conversations with

students to assess understanding, offering choice when it came to research and reading topics, and working in conjunction with our school's FNMI support teacher and social worker to support students with a history of trauma.

Unfortunately, I was unable to make these same accommodations when it came to summative assessment. In grade twelve, Alberta's students write a number of "Diploma" exams. These exams are standardized across the province and, at the time I was teaching in Alberta, accounted for 50% of students' final grades in their core courses. In Social Studies, the exam includes multiple choice questions, many of which demand advanced reading comprehension. The exam also includes two written, essay-type pieces, one of which is a comparison of written or graphic texts. Though the students in my grade ten and eleven classes did not write the Diploma exam that year, our department modeled all unit tests and final exams on the Diploma exam in an effort to better prepare students for grade twelve. Many of the reading comprehension and written response pieces came from the course textbook and therefore borrowed standard themes and topics.

Most of my students, regardless of heritage, fared poorly on these summative assessments. The average grades for my locally designed classroom assessments were far higher than the grades on unit tests. I knew, from classroom conversations and formative assessments, that most of my students had a solid understanding of the course content. Something, then, about the format of the unit tests was limiting their ability to communicate that knowledge, and I suspected it was related to the reading of complex passages. I wondered whether the tests were placing undue emphasis on skills and knowledge that were external to the content of the course.

In an effort to help the students, I dedicated considerable class time to practicing the types of multiple-choice questions on the tests and modeling the writing of the written tasks. In doing so, I sacrificed class time that I could have spent on topics and genres that were more relevant to my particular group of students, and I lost the interest of many students of First Nations heritage. Drop-out rates for FNMI students in Alberta are staggeringly high (McClure, 2014) and retention of these students was a constant struggle in our school. Again, I wondered if the assessments were truly in the best interest of students, considering the effect they had on my classroom practices.

Systemic Advantage in the East

I now teach at a leading independent school in Nova Scotia, and I am once again in a context vastly different from the last. Most of my students have European heritage, speak English as a first language, and come from well-educated families. The school does not follow the Nova Scotia curriculum, nor do its teachers administer the province's standardized assessments. From Junior Primary (elsewhere called Junior Kindergarten) to grade eleven, when the International Baccalaureate program begins, teachers create their own curricula and assessments. For the first time in my career, I have the freedom to design a program that is relevant and authentic for my students. Teachers at my school teach outside the constraints of the provincial program. For instance, at the elementary level, we spend fewer hours teaching the core subjects of math and language arts in order to provide more recreational time outdoors.

My students are born into financial and educational privilege; the school itself is located in an affluent neighbourhood, and the annual tuition costs over $17,000. Through their parents' funding, these children are afforded another privilege—the privilege of

avoiding standardization—for the majority of their time in school. And yet, teachers at my school still teach traditional topics in a traditional way. We teach students who have been exposed to traditional literacies and genres by their parents—children who are expected to attend university and to obtain, at minimum, a Bachelor's degree. We teach this way because arguably, this *is* the program that is most relevant to the students' abilities, needs, and backgrounds.

While I teach my senior school French course, the International Baccalaureate assessments loom. Administration reminds me to teach themes that will be found in the later years of the IB program and to use IB-appropriate assessments. However, I am not bound by the IB program in the same way I was bound by the Alberta Diploma exam simply because of the change in context. My course median is 90%. I know my students will do well in grades 11 and 12. I feel no need to teach test-taking skills; I know my students will perform well on the exam because they have been sitting final exams since grade seven. I teach in a traditional way because it is what my students, their parents, and the school administrators expect. I teach in a traditional way because it is what my students respond to.

Without reflection, it would be easy to forget the challenges my Torontonian and Albertan students faced. However, reexamining my experiences side-by-side reveals a glaring contrast, and this contrast raises important questions. How do student experiences with assessments interact with classroom context? Do privilege and socioeconomic status affect a student's perception of standardization? Are the consequences of standardized assessments different for students from different socioeconomic backgrounds? This line

of inquiry, along with the context described in the following section, informs the research questions outlined at the end of this chapter.

Situating the Context

The political context surrounding Nova Scotia provincial assessments and evaluations is inextricably linked to the questions driving this study. The legacy of systemic discrimination, persistent score discrepancies, and discordant relations between stakeholders must frame the conversation. The following provides a brief overview of these topics.

A History of Systemic Discrimination

The most frequently cited example of systemic discrimination in Nova Scotia is that which persists against African Nova Scotians. Although the province takes pride in its historical involvement in the 19th Century Underground Railroad, it is more difficult to recount the insidious stories of racism and discrimination that exist in our provincial history. In fact, Nova Scotia is one of the few places in Canada with a history of slavery. Slaves began arriving before the outbreak of the civil war and continued during the war (Black Learners Advisory Committee [BLAC], 1994, p. 10). The approximate 3500 free Black Loyalists who arrived in Nova Scotia during this time were promised resources and equality, yet most of that land was ultimately given to their white counterparts (BLAC, 1994; Pachai, 2007).

After slavery ended, poverty and discrimination forced many African Nova Scotians into low-paying labour jobs in which they were exploited: “[b]ecause the Black Loyalists needed the work for basic survival, employers exploited them, paying them wages about one-quarter of the rate paid to white labourers” (BLAC, 1994, p. 11). In

turn, these black citizens were blamed for the lowering of wages by working-class white citizens (Pachai, 2007). The pattern of broken promises continued through the War of 1812, when thousands of Black refugees arrived in Nova Scotia and received little to no government assistance, and again in the 20th Century, when residents of Africville, an established African Nova Scotian community, were forcibly relocated (BLAC, 1994; Evans & Tynes, 1995).

Of particular importance to this study are the social injustices endemic to the provincial education system. From 1816 through the 1960s, separate schools operated for African Nova Scotian children (Pachai, 2007). The education provided at these schools could be best described as rudimentary. The topics covered were limited to those that would be useful for low-level labour jobs, thereby reinforcing the predominant power structures. Poverty played a significant role in denying access to education, as most African Nova Scotian parents could not afford the mandatory fees for school materials and heating the building (BLAC, 1994, p. 15). By 1880, four schools exclusively for African Nova Scotian children had been established in the City of Halifax, but their construction and maintenance suffered, and the instruction was "simple and elementary" (BLAC, 1994, p. 20). Though segregated schools disappeared in the 1960s, African Nova Scotians remained largely absent from the school system through the midcentury (BLAC, 1994, p. 27).

As a result of unequal access to education, among other factors, African Nova Scotians were more likely to be unemployed, and those who worked held lower-level jobs and earned less. Sadly, evidence of this social and economic disparity remains, and the historical attitudes of prejudice have survived. As recently as 1991, census results and

the Black Learners Advisory Committee (BLAC) survey reported that "the results for key socio-demographic variables – education, employment, occupation, and income – consistently indicate that compared with the average Nova Scotian, African Nova Scotians are severely disadvantaged in all areas," and across all age groups (BLAC, 1994, p. 35).

Since the 2015/16 school year, the province of Nova Scotia has undertaken the disaggregation of assessment results on a central level. Data sets demonstrate comparatively lower results for students of African heritage across all assessments and assessment components. The difference between average score ranges from 7 percentage points to 13, depending on the component in question (Government of Nova Scotia, 2019b). This trend is particularly pronounced in grade three reading comprehension tasks, which showed a 16% discrepancy between African Nova Scotians' average and the overall average in 2013 (Halifax Regional School Board, 2014).

The quantitative data available provides evidence that the legacy of systemic discrimination in Nova Scotia remains, not just for students of African ancestry, but also for students from the Mi'kmaq community. Fortunately, this particular problem has received attention in recent years. The Nova Scotia Provincial Literacy Strategy presents an urgent effort to reduce "achievement gaps," which commits to "raising the achievement of students of African descent and of Aboriginal heritage" (Government of Nova Scotia, 2017a, p. 4). On a qualitative level, other reports, such as the BLAC Report (1994), have described ongoing issues within classroom practice, social culture, and assessment and have made recommendations for growth within education. This particular report underscores the importance of a holistic approach, including expanding the

curriculum to include more culturally relevant material and providing professional development of teachers.

A New Plan for Action

In 2015, Nova Scotia's Ministry of Education published *Nova Scotia's Action Plan for Education.* In 2019, this Action Plan remains in place and continues to drive decision-making about classroom practices and assessment. Although education reviews are routine in every province, this particular review defines its focus almost immediately as math and literacy: "In the simplest terms, we want to ensure that our students do better, especially in math and literacy, and that they are better prepared to lead productive lives in our changing world" (Government of Nova Scotia, 2015a, p. 6). The former Minister cites accountability to the public and "real, tangible results" as a goal (p. 6).

The actions in the plan comprise four pillars, the first of which is "A modern education system" (Government of Nova Scotia, 2015a, p. 7). This pillar focuses mainly on the restructuring of the Ministry and school boards, including the addition of a Centre for Excellence to focus on innovations and success stories (p. 7). The second pillar, "An Innovative Curriculum," outlines a plan to provide more teaching time for math and literacy in the early years, access to modern technology, and a culturally relevant curriculum, among other items (p. 18). Within this pillar, assessments are given particular attention, with an outline to incorporate more standardized assessments in the early years (p. 19). Interestingly, the Action Plan claims that provincial math and literacy assessments in grades 3 and 8 will be replaced by teacher-designed assessments. The current PLANS website does not reflect this change. Regarding grade ten Examinations, little information has been made available, aside from the claim that more professional

development in the area of assessment will be offered, an assessment policy will be developed, and government educators will refocus on "assessment for learning" (p. 19). Pillar three, "Inclusive School Environments," outlines the new model for inclusion and behavioural tracking, as well as the implementation of an Achievement Gap Initiative (p. 30). Pillar four, "Excellence in Teaching and Leadership," includes measures to improve classroom working conditions through teacher education. For example, the Ministry promises to work with universities to create a focus on behavior management, literacy and math instruction, student assessment, and inclusive education (p. 33).

The Action Plan also outlines the importance of cultural awareness and equity. It acknowledges many of the barriers reported by the studies that will be discussed in Chapter 4, including a lack of curricular content related to marginalized groups (Government of Nova Scotia, 2015a, p. 29) and a lack of diversity among teachers (p. 30). One of the proposed measures would implement a province-wide Achievement Gap Initiative "to address persistent differences between groups of students in results in math and literacy" (p. 30). Along with this measure, the report promises the implementation of "an Inter-University Chair in Research for the Achievement Gap" and the creation of a team of "representatives from the department, school boards, and post-secondary institutions to support research into student achievement in math and literacy" (p. 30). To date, the proposed "Achievement Gap Initiative" has not be publicly discussed or reported on, to my knowledge. In September of 2019, however, The Inter-University Research Network accepted applications for funding "to support research on improving student achievement and well-being in Nova Scotia, including evidence-based ways to understand and address the achievement gaps" ("Inter-university Research," 2019).

Political Discord

Nova Scotia's Action Plan for Education was published only six months before the Nova Scotia Teachers Union contract with the province expired. By November of 2016, the government and NSTU leaders had twice reached a new labour agreement regarding a new contract, but both offers were voted down (Williams, 2016). During this time, news outlets began reporting escalating tension between the two stakeholders and in November of 2016, the NSTU voted to engage a work-to-rule, wherein teachers would "only do the minimum amount of work required by their contract" (Williams, 2016). The NSTU cited *Nova Scotia's Action Plan for Education* as a driving force behind the job action: "The flurry of objectives from the Minister's Action Plan without adequate time, professional development and support for implementation have added to the complexity and demand, and taken away time from teaching" (Nova Scotia Teachers Union [NSTU], 2017a). The adjustments proposed by the NSTU included "A fair salary package and negotiated benefits," "free and fair collective bargaining," "more quality teaching time," and improved learning conditions (NSTU, 2017a).

When the NSTU and the Ministry of Education reached an agreement in the spring of 2017, the government created a council of teachers, parents, and administrative personnel who would be discussing issues of classroom conditions. This "Council to Improve Classroom Conditions" was tasked with creating a list of recommendations. The 40 recommendations they released in April of 2017 included smaller class caps in the early years, more support in the form of teaching assistants for integrated classrooms, and doing away with, or pausing, provincial assessments (NSTU, 2017b). In response, the provincial government promised to adopt the following recommendations: (1)

"eliminating three provincially mandated assessments and exams," (2) "placing a five-year moratorium on new school board and provincially initiated assessments," (3) "maintaining suspension of Grade 10 exams," and (4) "maintaining suspension of one provincially mandated assessment for Grade 8 students" ("Province moves ahead," 2017) The purported intention was to determine which assessments were "the most crucial" and which could be let go due to potential overlap ("Province moves ahead," 2017). As of the 2019/20 school year, there appears to be no change to the Program of Learning Assessments. The grade three, six, and eight assessments to be administered as scheduled, as are the grade 10 Examinations (Government of Nova Scotia, 2019a).

Later in 2017, the province of Nova Scotia commissioned Dr. Avis Glaze—a former Ontario Education Commissioner and Senior Adviser to Ontario's Minister of Education—to provide recommendations about education administration in Nova Scotia. Her recommendations reflect the attitude that Nova Scotian education is in a state of disrepair: "the system is not working the way it should for students, parents, teachers, and principals. […] In many cases, Nova Scotia students are performing below average compared to the rest of the country and the lack of clarity and coherence in the system is contributing to that" (Education and Early Childhood Development, 2018). Published in January 2018, her recommendations include the establishment of three new independent bodies: a student progress assessment office, a college of educators, and an education ombudsperson. She also recommends the elimination of school boards and the implementation of one advisory council (Glaze, 2018).

With respect to assessment, Dr. Glaze acknowledges reduced scores among students of marginalized groups and recommends more disaggregation along lines of

"region, gender, and economic status, including traditionally marginalized communities" (Glaze, 2018, p. 36). Another recommendation is to create a standardized literacy assessment, or "a test of literacy skills, beginning in Grade 10. […] In grade 12, the province could provide a literacy course of similar standard for students who have not yet passed the test" (p. 36). This recommendation appears to mirror the OSSLT and OSSLC in Ontario.

In March of 2018, Zach Churchill, the Minister of Education, chose to adopt Dr. Glaze's recommendations writ large (Doucette, 2018). Immediately, the NSTU voiced disagreement with the report and frustration with a lack of consultation. The NSTU chose to withdraw its members from the Council to Improve Classroom Conditions (Quon, 2018). At the time of publishing, Mr. Churchill claims the recommendations are still "underway" (Ritchie, 2019).

Political Climate's Impact on the Research. I began my study at the height of this political tension, and its inflammation, stagnation, and eventual resolution influenced the timing of my research. Finding teachers willing to participate in a research study during a work-to-rule proved at first difficult, then impossible. The teachers with whom I spoke were supportive of the research but feared participation would constitute a violation of the terms of their union-led work action. Although the work action ended before June of 2017, I was unable to conduct my study in the spring because the provincial assessments were placed on hold for the duration of the school year. Schools had been given the option of administering the assessments, and the majority of schools chose instead to create their own assessment. As a consequence, results are absent from all reporting documents published by PLANS for the 2016/17 school year.

Perhaps more importantly, however, the political climate and history of racism described above impacted the study's design in significant ways. Initially, the goal of this study was to compare the experiences of students with African Nova Scotian ancestry with that of their peers when approaching the NSE: English 10. In this form, the University of Lethbridge ethics committee approved the study. However, the former HRSB took issue with the racial framing of the proposed research. The concern expressed by the HRSB involved the self-identification of students that would be required in order to gain access to students with African Nova Scotian ancestry. I was asked by the research panel to remove any mention of race from student interviews and classroom observation. Once this facet was entirely removed, my study was approved.

Absent this focus on race, my study continued to experience social barriers. Before gaining access to the first of two participating classrooms, I held a meeting with the English Language Arts department leader—herself a teacher, although not the teacher who participated in the study. She expressed significant concern about the aim of my research and, on multiple occasions, spoke about the school's history of racial tension. She explained that numerous researchers attempt each year to conduct studies on race at this particular school, and she expressed concern that such research could bring social or psychological harm to the students.

The net result of these interactions is a study that is quite different from the one initially proposed. Its rationale and objective remain relatively consistent, although the conclusions that can be drawn about the study's impact on students of African Nova Scotian ancestry are drastically limited. The collective discomfort around issues of race is evident throughout this study, and these accounts will add richness to the discussion of

findings in Chapter 9. These exchanges, including the limitations they present to research and my own journey to better understanding the local social climate, will be further discussed in Chapter 10.

Rationale and Objective

The historical and current political climate described in this chapter provides a rationale for this study. As the provincial government moves into a new era of education, it appears to have adopted a mindset shared by many other provinces—that the education system is something to be monitored, assessed, and held accountable to stakeholders. A limited focus on accountability, however, may obscure other, less beneficial, consequences of standardized assessment.

The objective of this comparative case study is therefore to better understand consequences of the NSE: English 10 at the classroom and student level. Using the lens of validity theory, the study examines unintended consequences for students and teachers. Methods of inquiry include observation, document review, and semi-structured interviews with teachers and students in two schools. The following outlines the questions driving the study:

Research Questions

What are the consequences of the NSE: English 10 for students and teachers?

i. How do students and teachers describe classroom practices in the two schools being studied?

ii. How do students and teachers describe the evaluation in the two schools being studied?

iii. Do any differences between these accounts pose a threat to the inferences drawn from test results?

CHAPTER 2: THEORETICAL FRAMEWORK

In recent years, schools across Canada have made efforts to make education relevant to all students, respectful of diversity, and less regimented. Indeed, many of the public documents reviewed in Chapter 1 promote these ideals. At the same time, rejection of "assembly-line" models of education has run coincident with a drive to increase accountability through standardization. As Slomp et al. (2014) observe, "[i]ncreases in diverse student populations in the Canadian educational system have been accompanied by a concomitant proliferation of large-scale testing" (p. 277). The push for accountability has brought an accompanying criticism of standardization, with validity theory and its branches at the core of current research. The following chapter outlines the various theoretical frameworks which frame this study's qualitative inquiry within the broader context of the existing literature.

Multiliteracies

The theory of multiliteracies, first posited by The New London Group, encompasses two observations: "the multiplicity of communications channels and media, and the increasing saliency of cultural and linguistic diversity" (Cadzen et al., 1996, p. 63). The first observation centers on the invention and use of new modes of meaning-making. According to the New London Group, "[w]hen technologies of meaning are changing so rapidly, there cannot be one set of standards or skills that constitute the ends of literacy learning, however taught" (p. 63). The second observation centers on the navigation of an increasingly interconnected world in which the negotiation of differences becomes vital to one's effectiveness, as "effective citizenship and productive work now require that we interact effectively using multiple languages, multiple

Englishes, and communication patterns that more frequently cross cultural, community, and national boundaries" (p. 64). Literacy pedagogy, therefore, cannot remain stagnant. While schooling was once used "as a way of standardizing national languages" (p. 68) and producing workers for traditional, hierarchical workplaces, the world now demands citizens that are flexible, innovative, and able to navigate diversity.

The New London Group argues the futility of standardization in the face of local diversity and global connectedness. They acknowledge anxieties about educational standards and the preference for a back-to-basics approach, and they argue these anxieties are fueled "in part by the sense that [...] there are still vast disparities in life chances—disparities that today seem to be widening still further" (Cazden et al., 1996, p. 61). Reducing these disparities and increasing access, however, requires the recruitment of different subjectivities and their "attendant languages, discourses, and registers" as resources for learning (p. 72). They argue that "old meanings of 'access' and 'mobility' are the basis of models of pedagogy that depart from the idea that cultures and languages other than those of the mainstream represent a deficit," and that the new role of pedagogy is "to develop an epistemology of pluralism that provides access without people having to erase or leave behind different subjectivities" (p. 72). Any standardization of language learning would therefore be counterproductive to the greater goals of education.

Validity Theory

The goal of validity theory is to validate (or invalidate) the use of an assessment tool for a given purpose or set of purposes by gathering evidence from a number of different sources. Historically, validity has been divided into three major branches—content, criterion, and construct validity. However, more recent literature on the topic

reorganizes the theory into one unified concept, (Messick, 1989; Kane, 2013; Cronbach, 1988). Construct validity, it is argued, subsumes both content and criterion validity "because such information about the range and limits of content coverage and about specific criterion behaviors predicted by the test scores clearly contributes to score interpretation" (Messick, 1989). In evaluating for construct validity, one must integrate "any evidence that bears on the interpretation or meaning of the test scores," including evidence relating to content and criteria (Messick, 1989).

Two major sources of error threaten construct validity. Firstly, if an assessment tool fails to accurately represent the entirety of the construct being measured, it can be said to suffer from construct underrepresentation. If the construct under question is literacy, for example, a reading and writing assessment may present construct underrepresentation by focusing on expository genres to the exclusion other genres. By contrast, construct irrelevant variance threatens the validity of an assessment when constructs external to the construct being measured affect a student's performance. In the aforementioned example of a literacy assessment, a reading task for which prior knowledge of the topic would support a student's ability to correctly respond to questions in a timely manner could be said to contribute irrelevant variance.

Construct validity evaluates the degree to which an assessment tool measures the construct it is meant to measure (Kane, 2006) and the appropriateness of the interpretation of results (Messick, 1989). The distinction between the assessment tool itself and its interpretation is integral: "what is to be validated is not the test or observation device as such but the inferences derived from test scores or other indicators—inferences about score meaning or interpretation and the implications for

action that the interpretation entails" (Messick, 1989, p. 5). In other words, the validation of the interpretation of a student's performance by teachers, administrators, policy-makers, and other stakeholders is of utmost significance, as is the validation of any actions taken in response to the results.

Consequential Validity

Within recent decades, social consequences have become more visible within the field of test validity (Messick, 1989, Cronbach, 1988). However, scholars express disagreement as to their importance to the validity argument. According to Messick, social consequences should only be considered if they are a result of test invalidity: "it is not that adverse social consequences of test use render the use invalid but, rather, that adverse social consequences should not be attributable to any source of test invalidity such as construct-irrelevant variance" (Messick, 1989, p. 11). In other words, negative social consequences deriving from an assessment tool should only render the tool invalid if they stem from some other source of construct invalidity. Cronbach (1998) is less accepting of negative consequences when he argues, "validators have an obligation to review whether a practice has appropriate consequences for individuals and institutions, and especially to guard against adverse consequences" (Cronbach, 1988, p. 6). Kane (2013) balances these two views with a cost-benefit analysis: "[i]f the perceived costs and/or any immediate negative consequences exceed the perceived benefits, the program will get a negative evaluation" (Kane, 2013, p. 49), regardless of the presence, or lack of, construct validity concerns.

Kane (2013) gives special mention to the positive and negative consequences of standardized assessment within the school system. Potential benefits, he argues, include

increased public confidence in the schools and an increase in student achievement within the constructs measured by the tests. Potential costs include the narrowing of the curriculum, the opportunity cost of time and resources taken away from other learning, and increased dropout rates (p. 54). According to Kane, the cost benefit calculus is not absolute. He asserts that while "[n]egative consequences [may] count against a decision rule, they can be offset by positive [ones]." Ultimately, an assessment tool with negative consequences can still be considered valid given enough evidence of positive outcomes (p. 54).

Race, Power, and Genre Ecologies

Others argue that this framework of consequential validity does not go far enough to address the power structures that are created and perpetuated by assessment. Inoue (2009), for example, argues for the recognition of racial validity as a separate line of inquiry into writing assessment practices. According to Inoue, writing assessment contributes to racial formations by reinforcing divisions. He argues that "…tests, like all assessment, tend to produce the very qualities, competencies, behaviors, and attributes they purport to measure" (p. 99) and by etching "social formations that otherwise would not be there" (p. 102). Power is inextricably tied to assessment, as it "defines students through their subjection to it" (p. 104) and is therefore particularly powerful with respect to race and disparities in assessment performance. He promotes seven questions for the validation of assessment technologies' decisions:

> How is race constructed or conceptualized in our writing assessments, and in our validity inquiries? What racial formations are produced, built into, or assumed in the power, parts, and purposes of our assessment

> technologies? In what ways do our assessment technology's power affect various racial groups differently? What racial biases do our technology's parts have? How do our purposes for assessment and the decisions we make from them affect different groups, or form particular racial formations? Are there conflicting purposes? What historical blocs, both in teachers and in students, do we assume function in our assessments, and what historical blocs can we actually find working in various racialized stakeholders? (Inoue, 2009, p. 110).

This focus on racial formations, he argues, is necessary if we are to address racism and unequal racial social arrangements.

Similarly, Mya Poe (2008) calls attention to the reorganization of power structures undertaken by assessment policy. Her work focuses on the American No Child Left Behind (NCLB) legislation. She argues that NCLB has normalized a withdrawal of authority from local school districts, students, and teachers, and a redistribution of said power to the federal level (p. 143). With respect to marginalized groups of test-takers, she agrees with Inoue's assertion that assessment can perpetuate social arrangements and that it often reflects social anxieties. She argues that "[a]ssessment is often a reflection of broader social tensions in a culture, ranging from immigration, national security, segregation and social stratification, affirmative action, and privatization of education" (p. 146). Her work also raises the concern that the disaggregation of data can contribute to the reproduction of educational stereotypes. While she acknowledges the potential benefits of reporting racial disparities in assessment performance, she asks whether such

reports could “reinforce public fears about the failings” of historically marginalized groups (p. 147).

A Validity Analysis Framework

Discrepancies and racial tensions are not unique to Nova Scotia. As the trend toward standardization and accountability popularizes in Canada, disparity in results is linked more frequently with group identity. As exemplified by Nova Scotian media (Jeffrey, 2014), these concerns may bolster prejudice and distract from substantive causes. They may also “reinforce public fears about the failings” of certain groups (Poe, 2008, p. 147) and draw attention from more relevant concerns, such as issues of validity. That said, diversity makes it difficult to compare circumstances between provinces and individual school boards. To this end, Slomp, Corrigan, and Sugimoto (2014) outline a common framework for consequential validity.

Based on models designed by Kane (2006, 2013), as well as White, Elliot, and Peckham’s (2015) interpretation of Kane’s model, the Slomp, Corrigan, and Sugimoto design provides “a systematic approach to collecting and integrating consequential validity evidence” (p. 279) by integrating consequential validity into each stage of the validation process for assessments. The authors acknowledge the burden placed on test developers and recommend the consideration of the assessment’s stakes and complexity before proceeding (p. 279).

In total, nine sources of validity evidence are offered. The first, construct definition, encourages caution on the part of test developers when attempting to assess constructs that are complex and difficult to define and may hold different definitions across groups of people, such as writing ability. Secondly, the question of construct

irrelevant variance draws attention to the influence of observation methods, testing contexts, and other constructs on the domain under investigation. Thirdly, the authors call attention to the design process. They recommend the explicit identification of an assessment tool's aims and purposes, as well as the inclusion of multiple stakeholders in the design process. Doing so, they contend, "is important to enhancing transparency and ultimately the validity of writing assessment tools" (p. 282). Next, scoring procedures intended to increase interrater reliability, such as design rubrics, may unintentionally limit construct representation, according to the study's authors. The fifth source of evidence involves the sampling plans put in place by test developers, since "[w]ithout an adequate sampling plan in place, it is difficult to determine the extent to which [a] difference in experience translates into different test scores for various populations of students" (p. 283). The next source of evidence, disaggregated performance, is necessarily linked to sampling plans. Disaggregating sampling plans by population can provide insight into the potential sources of irrelevant variance, such as culture, race, and socioeconomic status, which can influence student outcomes and call into question the inferences being made about scores. Construct remodelling includes the recommendation to examine student responses. Once disaggregated results have been identified across population groups, a study of student response processes can shed light onto the sources of the disparities. Intended and unintended implications are offered as the final two sources of validity evidence. They that the inclusion of consequential validity evidence in validity arguments strengthens them, despite the tendency (historically) to omit consequences from analysis. Within every step of framework, the authors offer first a question of construct validity, and secondly of consequential validity to ensure "that the

focus of data collection and interpretation is constantly grounded in the integrity with which the test measures the construct it was designed to measure" (p. 284).

CHAPTER 3: CURRICULUM AND ASSESSMENT IN GRADE 10

A narrow focus on the consequences of the NSE: English 10 first demands a review of the greater program, as a better understanding of Examination's context informs any discussion of its consequences. This chapter therefore provides an overview of all provincial assessments in Nova Scotia, including the NSE: English 10's administration, scoring procedures, evaluation criteria, and usage of results. This chapter also discusses construct representation within the Grade Ten English Language Arts Curriculum and the NSE: English 10. Implications of this construct representation will be discussed in Chapter 8.

The Program of Learning Assessment for Nova Scotia

The Department of Education and Early Childhood Development hosts a Student Assessment and Evaluation division that is responsible for the design of standardized assessments and examinations for the province. These assessments and examinations hold the umbrella title of "the Program of Learning Assessment for Nova Scotia" (PLANS). The division lists among its overall objectives, to (1) "develop and administer program assessments to determine the effectiveness of curriculum delivery," (2) "develop and administer student assessments to assist students to achieve outcomes," (3) "provide student achievement information to government for education decision making," (4) "help teachers understand assessment principles and practices," and (5) "publish accountability reports for all assessments and examinations, both for teachers and for the general public" (Government of Nova Scotia, 2019a).

Assessments are delivered in grades three, six, and eight. These assessments are formative in nature; that is, results are meant to support students and do not contribute to

summative grades. Conversely, the grade ten English and Math evaluations summarily evaluate a student's understanding of English 10 and Math 10 curriculum outcomes, and they contribute 20% of a student's final mark in those courses. The following table outlines the assessments, the constructs they claim to measure, the duration of each assessment, and the intended usage of results:

Table 1

Overview of Provincial Assessments and Evaluations in Nova Scotia

Grade	Constructs Measured	Duration (days)	Use of Results
Three	Literacy and Mathematics	4	Formative
Six	Reading, Writing, and Mathematics	4	Formative
Eight	Reading, Writing, and Mathematics	4	Formative
Ten	English	1	Formative & summative (20% of final grade)
Ten	Math	1	Formative & summative (20% of final grade)

The NSE: English 10

The NSE: English 10 is administered once during a student's grade ten year. All students registered in an English 10 course must take the Examination, except those who have an Individual Program Plan (IPP) in place for English Language Arts. An exemption

may also be granted by a principal, upon consultation with parents, in the case of a student who has been learning English for one year or less (Government of Nova Scotia, 2015b). A French version of the Examination, "L'ENE: Français 10" is administered within francophone schools.

Content of the Examination

The Examination consists of five sections: three reading tasks and two writing tasks. The three texts include prose, poetry, and visual/media text. During the three reading tasks, comprehension is evaluated solely through multiple-choice questions (referred to by PLANS as "selected answer" questions). In multiple documents, PLANS emphasizes the variety of "question types" used in these sections. Questions are designed "to assess students' performance at different cognitive levels," including literal comprehension, non-literal comprehension, and analysis (Government of Nova Scotia, 2015b). A more thorough description of these cognitive levels can be found in Appendix A. The number of questions offered within each cognitive level is roughly equal, although PLANS offers a small range to account for minor differences. The range for "literal comprehension" questions in the prose section, for example, is five to eight. The following table, adapted from a PLANS publication, (Government of Nova Scotia, 2015b) represents these tasks, the number of multiple-choice questions, the score value of each task, and the suggested time to be spent on each one:

Table 2

NSE: English 10 Reading Tasks

Task	Number of Questions	Points	Suggested Time
Prose	20	20	40-45mins.
Poetry	10	10	20-25mins.
Visual/Media Text	10	10	20-25mins.

There are two writing tasks included in the NSE: English 10, although one—the reading response task—overlaps with reading constructs. The first task, a persuasive essay, asks students to agree or disagree with a statement and to argue their point of view. For example, an earlier iteration of the Examination uses the statement "People of different ages can learn from each other" (Government of Nova Scotia, 2019c). The second task requires students, again, to agree or disagree with a statement. This time, they are asked to use examples from the earlier reading tasks to support their point of view. The following table lists the writing tasks, the score value of each task, and the criteria present in the rubrics:

Table 3

NSE: English 10 Writing Tasks

Task	Score	Criteria
Persuasive Writing	24	Ideas Organization Language Use Conventions
Reading Response	20	Ideas Organization Language Use Conventions

Designing the Examination

> With regard to design, the province has stated the following:
>
> The NSE English 10 is constructed according to precise specifications. Questions are written to match curriculum outcomes and then are field-tested with students. Field-test results are then analyzed and items that meet provincial standards are approved for inclusion on examinations (Government of Nova Scotia, 2017b, p. 10).

The NSE Advisory group, comprised of high school English teachers within Nova Scotia, "assists" in the development of examinations (Government of Nova Scotia, 2017b). The frequency and nature of consultations with this Advisory group are not discussed in any public documents. The design itself follows the development procedures outlined in the Nova Scotia Assessment Development Model (Government of Nova Scotia, 2017b)—namely: (1) Assessment Plan and Timeline, (2) Field Testing, (3) Refinement, (4) Administration, and (5) Reporting (Government of Nova Scotia, 2019d). Student Assessment and Evaluation of the Department of Education and Early Childhood Development facilitate all design efforts, reporting, and scoring.

Administration of the Examination

Security is ensured at every step of the examination process. Examination documents are carefully packaged and delivered to ensure security. All materials are numbered and tracked (Government of Nova Scotia, 2015b, p. 4), and teachers are not permitted to keep or reproduce items. Examination documents, including student work, must be returned to the Department of Education promptly following administration (p. 6). PLANS defines security of the Examination as validity issue: "Securing the NSE is

critical to ensuring that the evaluation of student achievement is valid and fair. Users of the Examinationination results draw conclusions about the achievement of students based on their scores" (p. 4).

The administration of the Examination takes place on one of two days per school year in January and June. Under usual circumstances, Examination writing takes place in students' regular classrooms under supervision of their ELA teacher. Students are permitted a print dictionary and a print thesaurus and are given three hours to complete the Examination with the option of an extra 15 minutes where deemed necessary. Students use a bubble sheet both to identify themselves and to answer multiple-choice questions during reading assessments. Students are not permitted to use scrap paper for note-taking. Students work alone, and teachers are not permitted to read items aloud or discuss the Examination with students (Government of Nova Scotia, 2015b).

Scoring Procedures

Each Examination is scored twice—first by the classroom teacher, and again during a summer scoring session by a team of English teachers representing the province's school boards. Classroom teachers are given a scoring guide, including rubrics, to aid in the scoring of the Examinations. Teacher-determined Examination scores represent 20 percent of students' final grades in the course. Conversely, the results of the summer scoring session are used to "provide information to policy makers on the implementation of the course curriculum and on standards of student achievement in relation to expected learning outcomes" (Government of Nova Scotia, 2015b, p. 1). No information is publicly available regarding scoring discrepancies. However, Mr. Clark, the the first teacher participant of this study, had participated in summer scoring sessions

and provided insight into this process: “I remember them telling us the reason they do the remarking is because teachers also are thinking about the student when they’re marking [whereas] they’re purely marking for data.”

Usage of results

According to provincial documents, assessment and evaluation results provide (1) “teachers with information to inform instructional planning,” (2) parents and guardians with information on their child’s performance,” and (3) “the Department of Education and Early Childhood Development and regional centres/school board with information to inform educational decisions” (Government of Nova Scotia, 2016). At some point following the central scoring of the NSE: English 10, schools receive an Item Description Report. This report “includes student achievement data at the school, regional, and provincial levels for all questions in reading on the NSE English 10” as well as “data on student achievement in writing” (Government of Nova Scotia, 2019c, p.1). Schools can use this data to identify “areas of strength and areas where changes in instruction and/or assessment can be made” (Government of Nova Scotia, 2019c, p. 1). The Individual Student Reports provided to parents are described as an overview of a student’s achievement within the three strands measured across “performance levels” (Government of Nova Scotia, 2016). These reports are provided to students and their parents following the provincial scoring of the Examination, which takes place during the summer.

As detailed earlier, results are disaggregated by self-identification and district, although it is unclear how the province uses these disaggregated results. These disaggregated results are available for public viewing on the government’s PLANS website (Government of Nova Scotia, 2019b). Disparities in student performance are

often cited within research and reports. The Action Plan for Education (2015), the Council to Improve Classroom Conditions (2017), the Raise the Bar report (2018), and the NSTU have leveraged the disaggregated data to demonstrate the need for change within the education system.

Internal review of the NSE: English 10

Within the last calendar year, a report on the 2013-2018 iteration of the Examination was published. This document, entitled *Lessons Learned*, is "intended to support classroom teachers and administrators at the school, regional and provincial levels by using information gained from the NSE: English 10 results to inform next steps for literacy instruction" (Government of Nova Scotia, 2019c, p. 1). It focuses specifically on "areas that students across the province found challenging based on evidence from examination results" (p. 1). The document examines the multiple-choice questions that students most often answered incorrectly and offers suggestions to better prepare students for similar questions. For instance, the following sample question is provided:

Figure 1

Sample Multiple-Choice Question from Lessons Learned *Document*

The old truism of "less being more" (lines 54-55) in the context of this article means that

a. buying less is better than buying more.
b. less packaging makes more sense.
c. the less you buy, the more you save.
d. using less packaging is becoming more popular.

To assist students in correctly answering similar questions, the document suggests that teachers build vocabulary by "flooding the room with vocabulary rich talk" and

"explicitly teach dictionary use," among other approaches (p. 10). To support student success in the written sections of the Examination, the document recommends posting samples of student responses from prior Examinations on the classroom walls, so students may "compare their writing to the samples on the wall and identify, either in a conference or in a journal, what conventions they should focus on to improve their writing" (p. 16, 19, 22, 26). Teachers are also encouraged to model reading and responding to these texts (p. 16, 19, 22, 26). The suggestion that teachers spend time explicitly discussing the Examination preceding the assessment itself is pervasive throughout the document, especially within the sections dealing with writing (p. 16-26).

Construct Representation in the ELA Curriculum and Examination

The NSE: English 10 aims to measure student achievement with respect to specific curricular outcomes. It does not claim to measure literacy as a broad construct, which distinguishes its objectives from the OSSLT in Ontario. Accordingly, this section first discusses the constructs represented within the ELA curriculum. It will then consider the Examination's representation of those constructs. A larger discussion of the implications of construct representation—and underrepresentation—will follow.

The Nova Scotia Grade Ten English Language Arts Curriculum

The Nova Scotia Grade 10 English Language Arts (ELA) curriculum is available in two formats: a 200+ page document published in 1997 outlining the philosophy behind the program (Nova Scotia Education and Culture, 1997), and a more recently published "Desk Blotter" (Government of Nova Scotia, 2019e) outlining course outcomes along with the key points of assessment and instruction in the course. Desk Blotters, published by the Department of Education, are provided only for English Language Arts courses

from grades four through ten. It is unclear why Desk Blotters have been created for ELA courses and no others, although they do provide an overview of the curriculum that is far more distilled and accessible than the original curriculum document. For the purposes of this document review, I will examine the Desk Blotter, as it represents a condensed version of the larger document, lists all course outcomes, and pays special attention to assessment opportunities.

The ELA Desk Blotter provides a portrait of a course program that is relatively balanced by today's standards despite the age of the program. As opposed to organizing the course by genres, such as short stories, novels, and theatre, the course is divided into constructs of communication. The three strands—Reading and Viewing (40-50%), Speaking and Listening (20-30%), and Writing and Representing (30-40%)—are given a percentage representative of the attention that should be paid to each one. Three to four focus categories divide each strand into varied components. Teachers must provide a minimum number of "Assessment of Learning Events" in each of the focus categories for the strands. The following table represents the three strands, the twelve focus categories, and, in parentheses, the number of expected assessments for each focus category:

Table 4

Curricular Strands and Focus Categories

Reading and Viewing	Speaking and Listening	Writing and Representing
Narrative (2)	Informal speaking (2)	Expressive Writing/ Representing (1)
Expository/ Informative/ Persuasive (2)	Formal speaking (1)	Poetic Writing/ Representing (2)
Poetry (2)	Performance (1)	Transactional Writing/ Representing (2)
Visual Multimedia (2)	Listening (2)	

To guide teachers in the creation of assessments, a substantial list of sample assessments is offered within each of the three strands. A large number of these assessments include opportunities to explore constructs outside what is assessed by the Examination. The sample assessments listed for the Speaking and Listening strand are exemplified in the table below:

Table 5

Speaking and Listening Sample Assessment Events

Focus	Sample Assessment Events	
Informal Speaking	• informal debate • informal presentation • small talk • retelling	• interview • conversation • conference • small-group discussion
Formal Speaking	• public speech • integrated multimedia presentation • seminar	• panel discussion • formal presentation • academic debate • persuasive speech
Performance	• dramatization • tableau • role-play • read-aloud • think-aloud	• choral speaking • improvisation • monologue • reenactment
Listening	• interview • completion of task according to oral instructions • critical response • debate	• personal response • notes • paraphrase • descriptive feedback • review of a presentation or performance

The other two strands offer an equally varied list of possibilities, including recipe writing, painting and sculpture, the creation of children's books, and guided visualizations (Government of Nova Scotia, 2019e).

The ELA curriculum provides a robust picture of language and literacy, especially when considering the sample learning activities and assessments in the document. The myriad constructs presented recognize the "multiplicity of communications channels and media, and the increasing saliency of cultural and linguistic diversity" (Cazden et al.,

1996, p. 63). In distinguishing between formal and informal speaking, as well as between expressive representing and transactional writing, the document, to some extent, acknowledges the "epistemology of pluralism" discussed by the New London Group.

However, the openness of the document leaves room for interpretation and adaptation. In theory, the three strands should represent a relative balance. Yet, in practice, a teacher may adjust the weightings to the following: Reading and Viewing (50%), Writing and Representing (30%), Speaking and Listening (20%). A teacher may also eschew the non-traditional literacies entirely within assessments, favouring essays, formal debates, and fiction—all of which are listed as sample assessments within the focus categories. The curriculum therefore contains the possibility for construct over- and underrepresentation. If a teacher's programming choices and assessment practices overlap with the Examination, those constructs and assessment practices will receive far more attention at the expense of others.

The NSE: English 10

To illustrate the relationship between Examination and curriculum, the following table maps the curriculum outcomes represented in the Examination onto grade ten ELA course outcomes. Relying upon the English 10 Information Guide's table of outcomes, I have struck through those that are absent from the Examination (Government of Nova Scotia, 2015b, p. 11-12). Green outcomes are those that are visibly underrepresented in the exemplars provided to students. For instance, outcome 8.3, "choose language that creates interesting and imaginative effects" is entirely absent from the writing opportunities and rubrics provided. Furthermore, some outcomes describe "a wide variety" of texts and strategies, or "various" techniques (4.1, 4.2, 4.3, and 9.1). Although

this language is open to interpretation, the Examination presents only three texts and two writing opportunities. It is difficult to accept that such a limited number of test items could reasonably be considered a variety. The Speaking and Listening strand has not been included, as it is unquestionably absent from the Examination.

Table 6

Construct Representation Within Strand 2

Reading and Viewing
GCO 4: select, read, and view with understanding a range of literature, information, media, and visual texts
• 4.1 read a wide variety of print texts • 4.2 view a wide variety of media and visual texts • 4.3 seek meaning in reading, using a variety of strategies • ~~4.4 use fix-up strategies to clear up confusing parts of a text and adjust reading and viewing rate according to purpose~~ • 4.5 demonstrate an understanding of the impact of literary devices and media techniques on the understanding of a text
GCO 5: interpret, select, and combine information using a variety of strategies, resources, and technologies

• 5.1 research information from a variety of sources
  ○ select appropriate information
  ○ analyze and evaluate the information
  ○ effectively integrate information in a way that meets the requirements of a learning task

GCO 6: respond personally to a range of texts
• 6.1 respond to texts regarding issues, themes, and situations, while citing appropriate evidence • 6.2 respond to texts by questioning, connecting, evaluating, and extending • 6.3 make thematic connections within print and media texts and public discourse • ~~6.4 demonstrate a wiliness to consider other interpretations of text~~
GCO 7: respond critically to a range of texts, applying their understanding of language, form, and genre
• 7.1 examine the different aspects of texts that contribute to meaning and effect • 7.2 make inferences, draw conclusions, and support responses to content, form, and structure • 7.3 explore the relationships of language, topic, genre, purpose, context, and audience • 7.4 recognize the use and impact of specific literary and media devices • ~~7.5 discuss the language, ideas, and other characteristics of texts and genres~~ • 7.6 respond critically to various texts • 7.7 demonstrate an awareness that texts reveal and produce ideologies, identities, and positions • 7.8 evaluate how gender, cultures, and socio-economic groups are portrayed in media

Table 7

Construct Representation Within Strand 3

Writing and Other Ways of Representing
GCO 8: use writing and other ways of representing to explore, clarify, and reflect on their thoughts, feelings, experiences, and learning; and to use their imagination

• 8.1 use writing and other ways of representing to
  ○ extend ideas and experiences
  ○ reflect on their feelings, values, and attitudes
  ○ describe and evaluate their learning processes and strategies

• ~~8.2 use note-making, illustrations, and other ways of representing to reconstruct knowledge~~
• 8.3 choose language that creates interesting and imaginative effects
GCO 9: create texts collaboratively and independently, using a variety of forms for a range of audiences and purposes
• 9.1 demonstrate skills in constructing texts for various audiences and purposes

• 9.2 create an organizing structure appropriate to the purpose, audience, and contexts of texts
  ○ use appropriate form, style, and content for specific audiences and purposes
  ○ use appropriate strategies to engage the reader/viewer

• ~~9.3 analyze and reflect on others' responses to their writing and multimedia projects and consider those responses in creating new pieces~~
GCO 10: use a range of strategies to develop effective writing and other ways of representing and to enhance their clarity, precision, and effectiveness
• ~~10.1 demonstrate an awareness of what writing/representation processes and presentation strategies work for them in relation to audience and purpose~~
• 10.2 consistently use the conventions of written language in final products
• ~~10.3 experiment with the use of technology in communication~~
• ~~10.4 demonstrate commitment to crafting pieces of writing and other representations~~
• ~~10.5 use a range of materials and ideas to clarify writing and other ways of representing for a specific audience~~

Viewing the tables above, construct underrepresentation becomes immediately visible. Notably absent from the constructs on the Examination is the entire Speaking and Listening strand, which is meant to comprise 20 to 30 percent of course assessments. Within Reading and Viewing, four of the nineteen outcomes are unrepresented. Within Writing and Other Ways of Representing, five of the eleven outcomes remain unaddressed. In other words, the Examination covers 20 of a possible 40 course outcomes, or 50 percent. Taking into account the outcomes PLANS claims to assess yet

are underrepresented within exemplars and rubrics—those listed in green above—only ten represented outcomes, or 25 percept of the curriculum, remain.

PLANS itself acknowledges an inability to capture many of the outcomes listed within the curriculum (Government of Nova Scotia, 2015b, p.11). Yet, they fail to provide justification for the exclusion of a large portion of course outcomes. Nor do the documents provide evidence that the Examination does, in fact, assess the outcomes it intends to assess. The absence of certain constructs on the Examination could have implications on instruction and assessment within the broader curriculum, and on how a student's final grade should be interpreted. Perhaps most importantly, using results to "draw conclusions about the achievement of students based on their scores" (p. 4) and to make policy decisions is called into question.

CHAPTER 4: REVIEW OF RELATED LITERATURE

A limited number of studies in Nova Scotia have examined literacy, reading, or writing assessments through the lens of multiliteracy and validity theory. Those conducted have tended to focus on quantitative rather than qualitative approaches. A review of the literature available within Nova Scotia will therefore be followed by a review of studies external to Nova Scotia that approach consequential validity through a qualitative methodology.

Local Publication and Analysis of Data

Although the province of Nova Scotia has in recent years released disaggregated results from provincially-administered standardized assessments, the new data does not appear to have accelerated further enquiry. In fact, I am aware of only one formal study (Thiessen, 2009) to consider the results and implications of provincial standardized assessment. Notwithstanding this dearth of research, the fact that more recent provincial test results were disaggregated based on racial self-identification has allowed some basis for comparison with Thiessen's 2009 review. Immediately stark are consistent differences between the results of student groups. Invariably, the assessments have produced results for African Nova Scotian and Mi'kmaq students which are lower than those of the general student population.

In 2009, Victor Thiessen conducted an analysis of standardized assessment administered by the Chignecto-Central Regional, Strait Regional, and Tri-County Regional School Boards. His study compared the achievement of African Nova Scotian students, Mi'kmaq students, students of Acadian heritage, and students of other European ancestry. On all assessments, "learners of a European heritage obtained the highest

average scores, were most likely to be in the top 25 percent of learners, and least likely to find themselves in the bottom 25 percent". African Nova Scotian students, by contrast, "typically fared worst" (Thiessen, 2009, p. 1). Interestingly, the differences between assessment scores tended to narrow over the long term for students of Mi'kmaq and Acadian ancestry. By contrast, the assessment continued to produce lower than average scores for students of African Nova Scotian ancestry. Furthermore, higher numbers of African Nova Scotian students were reported as "reading below grade level" and requiring test adaptations (Thiessen, 2009, p.1).

A review of recent disaggregated assessment data from across Nova Scotia results demonstrate variance between school districts, construct categories, and cultural groups, drawing parallels with Thiessen's earlier work. As mentioned, the province has released results from standardized assessments since 2012 (with a gap reflecting the work-to-rule in the 2016/17 school year). Beyond the raw scores made publicly accessible, the data is also disaggregated based on school district, construct categories and since 2016, self-identified cultural groups. Variance is apparent across all categories. For example, the Conseil Scolaire Acadien Provincial (CSAP) consistently achieves higher scores than the other districts across all categories (Government of Nova Scotia, 2019f, p. 2-4) and generally achieves scores five percent higher than those of the HRCE, despite many of their schools being located in Halifax. Likewise, scores in the Reading category are 10 to 13 percent higher than those in the Writing categories (p. 2). Across reading, writing, and mathematics, the Examination produces scores which are lower by at least five percent for students of African or Mi'kmaq ancestry (Government of Nova Scotia, 2019b).

Though assessments have adopted a different format since Thiessen's 2009 study, today's results reveal similar discrepancies within results.

Policy at the provincial level has been shaped, in part, by these results. The 2015 Action Plan for Education opens with a message of urgency regarding assessment results. The Minister of Education states, "[t]ime and again, test results show our students are falling behind in math and literacy, nationally and internationally" (Government of Nova Scotia, 2015a, p. 5). In response, the *Action Plan* proposes more teaching time for math and literacy, an increase in the number of standardized assessments during the early years (p. 19), and the implementation of an Achievement Gap Initiative (p. 30). The follow-up Nova Scotia Provincial Literacy Strategy (Government of Nova Scotia, 2017a) commits to "raising the achievement of students of African descent and of Aboriginal heritage," setting the specific goal of narrowing the gap "between their literacy achievement and that of their peers by 5 percent by 2020" (p. 4). Similarly, *Raise the Bar* (Glaze, 2019) cites "achievement gaps" and "underperforming" students (p. 36) as motive for further assessment. Dr. Glaze recommends an emphasis on centralized assessment and goes so far as to suggest the implementation of a Literacy test, similar to the assessment currently in place in Ontario (p. 36).

Reporting on Racial Inequity in Nova Scotia

Studies of racial inequity in Nova Scotian schools have become the backbone of racial equity policy in provincial education. In 1994, the Government tasked the Black Learners Advisory Committee (BLAC) with providing recommendations for the improvement of education for students of African heritage. This report, known as "The BLAC Report," reviews the province's history of racism, discrimination, and segregation

and argues for drastic change in the education system. The BLAC Report's recommendations have led to the implementation of a Racial Equity Policy (Government of Nova Scotia, 2002) and the creation of the African Canadian Services division at the Department of Education.

Among the report's recommendations is the establishment of "student assessment and testing practices for racial, cultural, and gender bias" (Government of Nova Scotia, 2003, p.12). In response to this recommendation, the Department agreed to continue reviewing provincial assessment practices for such biases. The Department also promised to "work directly with school boards and schools to discourage the use of tests that might discriminate based on race or culture" (p. 12). Currently, the Assessment Advisory Committee, along with item writers and reviewers, canvasses test items to ensure they are consistent with the standards of the *Standards for Psychological and Educational Assessments* and the *Standards for Fair Assessment in Canada.* However, no information can be found to verify the Department's claims. To my knowledge, there is no published evidence of validity inquiry aside from the claims cited above.

In 2009, Enidlee Consultants conducted a follow-up review of the progress made since the BLAC Report (Enidlee Consultants, 2009). A number of qualitative studies took place as part of this review, most focusing on the personal experiences and accounts of those directly involved in public schools. Participants in the studies included "parents, members of education committees, tutors, site coordinators of cultural and academic enrichment programs, tenants associations, African Nova Scotian organizations, students, teachers, regional educators and student support workers" (p. 24). Focus groups of African Nova Scotian students were organized with the help of student support workers

and Race Relations, Cross Cultural Understanding and Human Rights (RCH) Coordinators. The review involved six of the province's school boards and targeted community members of schools with a large population of students of African heritage, as well as schools with few to no students of African heritage. The study's topics were far-ranging, including twelve individual programs and organizations that have stemmed from or been enriched by the BLAC Report, including the African Nova Scotian Student Scholarship Program and the English 12: African Heritage Course, among others (p. 11).

Through interview and observation, the BLAC Report uncovered ongoing issues of racial inequity in the province's major school boards. For example, racism was far less likely to be acknowledged by administrators than by students and their parents (Enidlee Consultants, 2009, p. 45). Furthermore, the review of the Five Schools Project revealed limitations in studying assessment result discrepancies between students of African heritage and their peers. One of the project's mandates is to improve student achievement in literacy and mathematics. However, the data itself does not present a clear picture, as at least half of the African Nova Scotian students at the schools studies were exempted from writing assessments due to Individual Program Plans (Enidlee Consultants, 2009, p. 48). The review of the Five Schools Project also acknowledged the rigorous nature of curriculum heavily focused on literacy and numeracy and recommended teachers "receive support on embedding Africentric content" into their lessons (p. 50).

In the late 1990s, an external review of Cole Harbour District High School was commissioned by the Halifax Regional School Board "to make visible the complexities of issues found within the school" (Frank, 2002, p. 48), at a time when the school was under scrutiny due to racial conflict. The study was qualitative in nature; its participants

included “students, teachers, support staff, parents/guardians, school board personnel, members of various committees associated with the school, officials with the Department of Education and Culture, and a wide range of community members” (p. 4). The interviews were open-ended and took place over several months.

The existence of discrepancies with respect to literacy assessment was a theme that emerged through the interviews. According to the study, most teachers viewed the problem as “historical” in origin. They argued it was “[h]istorical in the sense that a history of marginalization by generations of exclusion from formal schooling, as well as historical in the sense of the inadequate preparation in the elementary schools [or] a recognition that elementary schools face many of the same problems” (Frank, 1997, p. 29). Teachers cited, among other factors, a lack of time, a lack of resources, and a lack of connection with the greater community as factors in low literacy achievement. Other themes that emerged included a need for support beyond what is currently allocated to all high schools, a more informed teaching staff, and a lack of communication between the school and its community.

Consequential Validity Inquiries in Canada

Numerous Canadian studies suggest that the standardization of assessment and scoring procedures can lead to construct underrepresentation, both within the assessment itself and in the broader curriculum. For example, some of the students interviewed by Kearns (2011) expressed confusion as to why they failed the OSSLT, considering their relatively strong performance in English class. Some specifically complained that the OSSLT did not afford them the chance to write about personal matters, as they would often do in English class. This confusion suggests a potential misalignment or narrowing

of reading and writing constructs represented within the OSSLT. Through interviews with teachers, Skerrett & Hargreaves (2008) found that the OSSLT similarly restricted the construct of literacy to more traditional concepts and reinforced out-dated teaching methods. Skerrett (2010) found the test was creating an overreliance on standardized scoring procedures and, once again, was working to narrow the curriculum to traditional constructs of literacy, despite the recognition on the part of teachers that the constructs were too restricted and harmed students' literacy learning. Ricci's (2004) case study of one secondary school suggested the OSSLT made significant impacts on the curriculum and professional learning of the school's teachers. For example, class time was dedicated to preparation for the test, thereby taking time away from other literacy learning opportunities. He observed unwillingness on the part of teachers to explore topics of interest with students when they were not related to the test itself. Furthermore, professional development time, which could be used for more relevant learning pursuits, was instead dedicated to the marking of practice tests.

Canadian studies that take an interview approach with teachers and students seem to best capture questions of construct irrelevant variance. For example, Fox and Cheng (2007), whose focus group-led study compared test-taker responses of students whose first language was English (L1) to English language learners (L2), were able to identify constructs external to the construct of writing that were being inadvertently measured by the OSSLT. Knowledge of test genre was one such construct, as a comparison of L1 and L2 focus group responses "revealed considerable difference in their knowledge of test genres" (Fox & Cheng, p. 18). Knowledge of narrative genre, with its "embedded social, cultural, and conventional meanings" (p. 13) further differentiated L1 and L2 test takers,

as did knowledge of vocabulary; many L2 students struggled with the words included in writing prompts and subsequently left entire written sections blank (p. 17). Similar focus group discussions with teachers revealed a focus on test-taking skills in Ricci's study (2004). Ricci's argument goes further, however, by suggesting the time and effort spent on test-taking skills in the classroom are threatening the validity of classroom literacy practices (Ricci, 2004, p. 354).

Many of these interview-based studies have found evidence of the further marginalization of already-marginalized groups of students at the hands of standardized assessment. Both L1 and L2 test-takers interviewed by Fox and Cheng (2007) described feelings of anxiety and pressure leading up to the test. However, L2 students were more likely to express feelings of frustration and disappointment both during and after the test. They conclude that the test may be "contributing to lowered self-esteem, a lack of confidence and a loss of efficacy for some of the students considered here—particularly within the L2 student group" (Fox & Cheng, 2007, p. 20). Similarly, student participants in Kearns's study (2011) expressed feelings of shame, embarrassment, and stress. Kearns argues such negative sentiments could impede students' ambitions: "[youth] saw their futures in jeopardy as a result of OSSLT failure" (Kearns, 2011, p. 120). Like the students in Fox and Cheng's study, Kearns's participants often brought different literacy practices and cultural norms to the test. She argues the OSSLT "privilege[s] some youth's cultural capital and devalues that of others" and thereby further marginalizes groups of students that already face barriers (p. 125). Though Ricci (2004) did not interview students, his interviews with teachers revealed what Ricci terms a "deskilling" of teachers and the shift of power from a local to a central level. This restructuring of authority may

be indirectly marginalizing groups of students. Anecdotally, it was discovered that one elite private school did not spend time preparing students for the test, in stark contrast with the hours dedicated to test preparation in the observed schools. Time was therefore used at the teacher's discretion. He argues such differences between public and private schools may, in fact, be contributing to class divide (Ricci, 2004, p. 354-355).

Each of the aforementioned has influenced the design of this study. Like those above, it employs interview methodology to collect data. This method emphasizes the first-hand accounts which are necessary for deep understanding of the contexts in which the assessments are written, as well as their unintended consequences. Additionally, student interviews are designed to explore course content and highlighted construct differences between the examination and course content, as they did for Kearns (2011), Skerrett and Heargreaves (2008), and Fox and Cheng (2007). Teacher interviews touch on power structures and explore teachers' opinions about test design and usage of results, similar to those in Ricci's study (2004). Furthermore, comparing the findings of this study with those of Kearns (2011), Skerrett and Heargreaves (2008), Fox and Cheng (2007), and Ricci (2004) provides an opportunity to support conclusions and adds depth to the discussion of findings in Chapter 8.

CHAPTER 5: METHODOLOGY

Methodological aspects of the studies described above provide a road map for this study. The following chapter provides a more detailed overview of the qualitative method, including methodology, data analysis, validity, and ethical considerations. A discussion of the researcher's subjectivity concludes the chapter.

Qualitative Inquiry and Constructivism

This study attempts to illuminate the ways in which evaluation and classroom practice interact and influence one another to create both intended and unintended outcomes. Put another way, it aims to "reveal how all the parts work together to form a whole" (Merriam, 1998, p. 6). This focus on understanding, as opposed to explanation, is characteristic of qualitative inquiry (Stake, 1995; Merriam, 1998). The research questions driving this study thus ask not *why* scores vary between groups of students, but *how* classroom practice and the students themselves may be affected by the evaluation.

The constructivist perspective that accompanies qualitative inquiry assumes that "meaning is embedded in people's experiences and that this meaning is mediated through the investigator's own perceptions" (Merriam, 1998, p. 6). An interview-based inquiry supports this perspective. Interview questions respect the idea that "reality is constructed by individuals interacting with their social worlds" (p. 6). Within this study, for example, interviews engaged participants in discussions that reveal their personal framework surrounding literacy and the ways in which those frameworks have been, and continue to be, socially constructed. The task of the researcher is not to discover knowledge, or to explain anything away. Rather, the researcher's task is to understand and narrate the

socially constructed realities of the participants, so he or she may provide readers "with good raw material for their own generalizing" (Stake, 1995, p. 102).

Multisite Case Study

A research study with such broad questions demands the narrow focus of case study methodology. According to Merriam (1998), "the case is a thing, a single entity, a unit around which there are boundaries" (Merriam, 1998, p. 27). For the purposes of this study, the classroom walls form the boundaries. It is precisely these boundaries that have allowed me to examine cases and participants closely. An in-depth study of two cases has provided the boundaries necessary to remain grounded—to not become overwhelmed by an overambitious study for which time or resources were lacking.

In fact, the objectively narrow focus of case study can allow for greater understanding of the entire system. As Cohen and Manion suggest (2003) "[t]he purpose of such observation is to probe deeply and to analyze intensively the multifarious phenomena that constitute the life cycle of the unit with a view to establishing generalizations about the wider population to which that unit belongs" (p. 106-107). A study of the entire province would be costly, time-intensive, and unrealistic, but a study of two well-chosen cases can enable the reader to transfer knowledge about the single case to other contexts. Indeed, "[t]he case report is, at its best, a 'portrayal' of a situation. It may read like a novel but it does so for the same reasons that novels read like novels—in order to make clear the complexities of the context and the ways these interact to form whatever it is that the case report portrays" (Lincoln & Guba, p. 214). A thorough examination of two individual cases allows for some degree of transferability to the general system.

The examination of two independent cases enhances the validity of the analysis. According to Merriam (2009), "[t]he more cases included in a study, and the greater the variation across the cases, the more compelling an interpretation is likely to be" (p. 49). The two cases described in this study vary greatly from one another—a method common to multisite case studies (Merriam, 2009, p. 81). As Stake suggests, the cases are chosen "because it is believed that understanding them will lead to better understanding, perhaps better theorizing, about a still larger collection of cases" (Denzin & Lincoln, 2000, p. 437). In fact, many of this study's findings gain significance when considering how the two cases are in conflict.

That said, limited time and resources naturally restrict the scope of research. As Hill (2004) explains, the intensity of understanding the individual story of each participant often prevents the researcher from examining a large number of cases (p. 9). To provide the kind of robust description that conveys to the reader "what experience itself would convey" (Stake, 1995, p. 43), I devoted significant time and effort to each case in the form of interview research, observational records, and intensive analysis after the research was conducted.

Data Collection

Data was collected through semi-structured open-ended individual interviews, classroom observation, and document review. The following section outlines each of these methods, provides rationale, and explains the logistics of their application within the study.

Selection of Cases

In designing this study, the goal was to select two cases that could "maximize what we can learn" while being mindful of the fact that generalization about the entirety of the school district would not be possible from such a small sample (Stake, 1995, p. 4). Instead of looking for schools that were themselves diverse enough to represent the entirety of the HRSB, I sought diversity between my two cases in the hopes that differences between schools would illuminate a range of the Examination's effects on classroom practice.

Ultimately, I achieved this diversity between cases, but political circumstances presented considerable obstacles. As described in Chapter 1, the 2016/17 school year was fraught with political tension. In early 2017, the teachers who had originally agreed to participate felt compelled to drop out of the study due to union restrictions. My study, like many others, was postponed for over a year. When the NSTU reached an agreement with the province and the work-to-rule was lifted, I contacted principals directly. Two principals connected me with teachers who taught grade ten English Language Arts. Happily, these teachers agreed to participate, and their classrooms were different in ways that proved significant to the study. An overview of each case is provided in Chapter 6.

Interviews

According to Steinar Kvale (2007), the qualitative interview is in line with constructivist epistemology:

> [it is] a key venue for exploring the ways in which subjects experience and understand their world. It provides a unique access to the lived world of the

> subjects, who in their own words describe their activities, experiences and opinions (p. 9).

Given the subjective, personal nature of my study, an interview format provided the most appropriate approach.

The most common criterion for categorizing interviews is structure. According to Sharan B. Merriam (2009), a highly structured interview presents a predetermined set of questions to participants in a rigorously guarded order and manner. In semistructured interviews, the questions may be a mix of more and less structured questions, or "all questions are more flexibly worded" (Merriam, 2009, p. 90). Unstructured interviews are more in line with a naturalist approach, insofar as they "are virtually impossible to design in any definitive way before the study is actually undertaken" (Lincoln & Guba, 1985, p. 214). For the purposes of this study, a highly structured interview format would not allow me to "access participants' perspectives and understandings of the world" (Merriam, 2009, p. 90). By contrast, an unstructured interview may not lead to worthwhile knowledge, instead "reproducing common opinions and prejudices" (Kvale, 2007, p. 8). I therefore chose a semi-structured individual interview format, in which participants' unique ways of viewing their lived worlds are respected. Questions were designed to be more open-ended and flexible.

Student and teacher interview guides have been included in Appendices B and C. These guides loosely follow the typology of Strauss, Schatzman, Bucher, and Sabshin (1988), in that they include some hypothetical, devil's advocate, ideal position, and interpretive questions. The questions also respect Merriam's (2009) assertion that "good interview questions are those that are open-ended and yield descriptive data, even stories

about the phenomenon" (p. 99) by avoiding yes-or-no and leading questions and providing time to ask probing questions. Student questions are written in a language that is accessible to young participants.

Teacher interviews were conducted following the first classroom observation in order to provide direction to interviews. Each teacher was interviewed once for a total of 90 minutes, and the interview took place in his or her classroom. During these interviews, teachers were asked to describe their background, their day-to-day activities, their approaches to English teaching, and their experiences with the provincial Examination. Interviews were recorded on a voice recorder, although some notes were taken by hand during the interview itself. On both occasions, teachers were asked brief follow-up questions to clarify and ensure the valid interpretation of certain statements. Profiles of each teacher will be provided in Chapter 6.

The selection of student interview participants was less intentional than originally planned. Initially, the intention was to use what Merriam (2009) calls "maximum variation sampling" in order to interview a sampling of the greatest variety of students, in terms of gender, socioeconomic background, and academic achievement. However, as explained in Chapter 1, the HRSB expressed concerns about student or teacher identification of cultural background and thus requested that I use an open sampling method, whereby any student who wished to participate in an interview—and who obtained consent from his or her parent—would participate.

The selection of student interview participants was therefore determined by consent form completion. In other words, all students who assented and whose parents had consented to their child's participation were interviewed. At Eastern High, this

amounted to a total of three students, although one ultimately declined. At Oceanview Collegiate, a total of five students were interviewed. Each of these students was interviewed once for a total of no more than 30 minutes. Student interviews took place in a quiet space apart from other students to ensure students were not influenced by their peers' presence. At Eastern High, the interview location was the teacher's empty classroom. At Oceanview Collegiate, we were provided with a book room.

During interviews, students were asked to describe their self-perceptions, their opinions of the Examination, and their daily lives. The "lived worlds" of the subjects formed an integral part of the research. Interviews provided insight into things unseen: "Interviewing is necessary when we cannot observe behavior, feelings, or how people interpret the world around them" (Merriam, 2009, p. 88). Again, these interviews were recorded on a voice recorder, and some notes were taken by hand. Student participants will be briefly described in Chapter 6.

Observation

Classroom observation formed an essential part of this study, both as a means of discovering interview questions and gathering information about potential issues of validity within the test and curriculum. Since I was attempting to form an accurate picture of classroom practices, I was careful not to intervene or interrupt normal classroom routines. This passive observation is common among case study researchers. According to Stake (1995), researchers carrying out case studies "try not to disturb the ordinary activity of the case […]. We try hard to understand how the actors, the people being studied, see things" (p. 12). However, passive research is arguably limited by practical realties. Denzin & Lincoln (2000) assert that all "observation involves the observer's

participation in the world being studied. There is no pure, objective, detached observation; the effects of the observer's presence can never be erased" (p. 634). While imperfect, the balance of the literature clearly endorses a passive approach.

Since qualitative research is interpretive by nature, the researcher's interpretation is at the core of observation: "we emphasize placing an interpreter in the field to observe the workings of the case, one who records objectively what is happening by simultaneously examining its meaning and redirects observation to refine or substantiate those meanings" (Stake, 1995, p. 8-9). This examination of meaning and constant renegotiating of research methods in response to observations is what keeps the study moving forward, and what keeps it grounded. That said, exercising caution when drawing hasty conclusions or interpretations is wise, as a "[g]ood case study is patient, reflective, willing to see another view of Θ" (p. 12). Accordingly, as a classroom observer, I aimed to be as invisible and non-interventive as possible while respecting the subjectivity of my interpretation of the observed classroom experiences. The research balanced Angrosino's (2005) "unobtrusive observation" and "participant observation," as I made myself and my research known to the students, yet refrained from participating in the classroom activities (p. 37-39).

Prior to the first observation, I visited each classroom and introduced myself. It was during this visit that I distributed consent and assent forms. The actual recording of observations took place in the classroom during normal class time. Each of the five observations spanned the length of the class from beginning to end—roughly 75 minutes. Observations did not take place during subsequent classes. Rather, I attended classes that fit the schedule of the teacher participant as well as my own work schedule. For each

observation, I chose a seat near the back of the classroom and, when possible, sat apart from other students.

According to Angrosino & Mays de Pérez (2000), "observations in natural settings can be rendered as descriptions either through open-ended narrative or through the use of published check-lists or field guides" (p. 674). For the purposes of this study, a combination of the two was used. Field notes in narrative form allowed for the fluid recoding of all classroom activities, and it "[let] the occasion tell its story" (Stake, 1995, p. 62). However, a check-list of literacy constructs presented in the provincial curriculum aided in verifying the constructs represented within classroom practice. I was careful to "keep focused on categories or key events" (p. 62) such as literacy constructs, test-taking strategies, and the presence of dialogue surrounding culture. Lastly, I re-read my field notes directly following each observation and recorded interpretive thoughts quickly thereafter.

Document Review

Since my perspective as researcher was limited to what I saw and heard during a limited number of observations and interviews, a document review became a valuable source of data about the greater program and system in place. As Stake (1995) asserts, "[q]uite often, documents serve as substitutes for records of activity that the researcher could not observe directly" (p. 68). The respective course outlines provide insight into curriculum and the constructs of literacy included in classroom practice, and the curriculum document for the course provides an important contrast to both course outlines and evaluation content. The Examination itself remained inaccessible to me,

despite a request for access. The student exemplars and Information Guide on the PLANS website were the next best sources of information about the Examination.

Timeline of Research

The NSE: English 10 is administered twice per year—once at the end of the first semester, and again at the end of the second. The first round of research, which was conducted at Eastern High, took place in January 2017 ahead of the first Examination. The second round of research, which was conducted at Oceanview Collegiate, took place in April, May, and June of 2018 before the June Examination.

This timeline was complicated by HRSB policies. According to the Board's *Guidelines for Researchers* (HRSB, 2016), research cannot take place during the month of June. This rule upset the intended symmetry between case studies at different high schools. At one of the participating schools, the June blackout period meant limited opportunity for observation preceding the June Examination, and prevented student interviews after practice assessments had been completed. Fortunately, an exception regarding classroom observation was obtained, classroom observation was permitted on one occasion in June. The timeline below represents the entirety of classroom visits, teacher interviews, and student interviews. Note that pseudonyms are being used for students, teachers, and the schools themselves to ensure anonymity:

Table 8

Timeline of Data Collection

Date	School (pseudonyms)	Agenda	Duration (minutes)	Names of Participants (pseudonyms)
January 9, 2018	Eastern High	Teacher Interview, Classroom Observation	90 75	Mr. Clark
January 23, 2018	Eastern High	Student Interviews, Classroom observation	30/student 75	Colton & Abi
January 26, 2018	Eastern High	Classroom Observation	75	
April 11, 2018	Oceanview Collegiate	Teacher Interview, Classroom Observation	90 75	Ms. Hatfield
May 2, 2018	Oceanview Collegiate	Student Interviews	30/student	Josie, Shelby, Ben, Jason, & Kate
June 13, 2018	Oceanview Collegiate	Classroom observation	75	

Validity, Reliability, and Generalizability

Qualitative research is drastically different from quantitative research. It is, by its nature, continuously shifting and subjective. As Merriam (2009) states, "reality is holistic, multidimensional, and ever-changing; it is not a single, fixed, objective phenomenon waiting to be discovered, observed, and measured as in quantitative research" (p. 213). Furthermore, every individual's interpretation and constructions of this reality will be unique, making any attempt at proving reliability virtually impossible.

With these difficulties in mind, multiple methods were chosen to enhance the validity and reliability of this study. Triangulation and peer review acknowledge negative or discrepant information and provide a description of investigator bias via reflexivity. Internal validity is enhanced through self-reflection regarding the potential for bias or subjective influence. External validity is supported by the detailed accounts provided of each individual case and the participants within them. These techniques enhance and compliment the "rich description" necessary for transferability (Cresswell, 2003; Merriam, 2009).

Though triangulation can take different forms (Merriam, 2009), this study relies upon numerous and independent sources of data to enhance "rich description". Multiple student and teacher participants allow for different perspectives and interpretations. Observation counterposed against classroom and assessment data allows for cross-checking. In this manner, themes become easier to identify and justify (Cresswell, 2003). The study also acknowledges "negative" or "discrepant information" that does not support its themes or argument. Focus on discrepant information can "add to the credibility of [the] account for the reader" (Cresswell, 2003, p. 196), and is therefore noted in both field notes and the study itself.

Peer review further enhances the validity and reliability of this study. My thesis advisor and thesis committee have made suggestions for reflection and revision throughout each stage of the writing process. They have provided an important outside perspective to ground the research by asking "questions about the qualitative study so that the account will resonate with people other than the researcher" (Cresswell, 2003, p. 196).

The examination of biases and perspectives further enhances this study's internal validity. Self-reflection helps explain the author's methodology in developing hypotheses (Merriam, 2009) and also "creates an open and honest narrative that will resonate well with readers" (Cresswell, 2003, p. 196). Ultimately, this technique exposes the subjective background that influences objective analysis. In this study, the personal history related in the first chapter provides relevant background information. Evidence of systemic discrimination in Nova Scotia has given me a critical lens through which to examine the NSE: English 10. My experience teaching high school in Ontario and Alberta, where classroom practices were heavily influenced by standardized assessment, leads me to expect similar influences here. These two factors colour the outlook of this study. The conclusion of this chapter will further describe the ways in which my personal experiences and identity may influence the research itself.

This study also considers its external validity, or "the extent to which the findings of one study can be applied to other situations" (Merriam, 2009, p. 223). As a multisite case study, it aims to answer questions about the broader context, notwithstanding its narrow focus. As Merriam suggests, "[i]n qualitative research, a single case or small, nonrandom, purposeful sample is selected precisely because the researcher wishes to understand the particular in depth, not to find out what is generally true of the many" (Merriam, 2009, p. 224). Thus, by contouring the distinctions between its case subjects, this study encourages broader themes to emerge. Its application of those issues, however, is beyond its scope. Ultimately, the burden of transferability, according to Lincoln & Guba (1985), lies with the reader. The duty of the investigator is simply "to provide sufficient information about the context in which an inquiry is carried out so that anyone

else interested in transferability has a base of information appropriate to the judgment" (Lincoln & Guba, 1985, p. 124-125). In relating its observations and interviews, this study thus aims to provide as much detailed description as possible.

Ethical Considerations

Qualitative interviews are, by nature, morally complicated. As Mauthner notes, "[e]thical problems in interview research arise particularly because of the complexities of 'researching private lives and placing accounts in the public arena'" (Mauthner, cited in Kvale, 2007, p. 22). The interviews conducted for this study were particularly sensitive because of the youth and implied vulnerability of participants. Careful consideration was therefore given to the ethics of data collection and analysis. First, permission was sought from governing bodies. This study did not proceed until the research ethics board of the University of Lethbridge approved the proposed methods and permission had been granted by the HRSB.

Second, before conducting any observation or interviews, participants were asked to sign an informed consent form. The form acknowledged that the rights of participants would be protected through the data collection process; it provided information about the purpose and procedures of the study and explained their right to withdraw at any time, to ask questions, and to remain anonymous (Cresswell, 2003). In the case of teacher participants, the form served as an open acknowledgment that the investigation may create conflict, especially if a teacher were to disagree with instructions given by a superior. A written agreement was preferred to "…serve as a protection for both the interviewees and the researcher" (Kvale, 2007, p. 27).

Within this study, student participants' right to privacy was carefully guarded due to their young age and vulnerability. Within the discussion of findings, students, teachers, and schools are referred to with aliases in order to protect identities. A consent form at the outset of interviews explained to students who would have access to their responses, and that teachers, specifically, would not. (Cresswell, 2003).

Because student participants were asked to describe their own self-perceptions and academic experiences, steps were taken to moderate and mitigate the interview experience. Since discussions regarding test results are intensely private, this study acknowledges that respondents may have felt "embarrassed by certain questions, and they may tell things they had never intended to reveal" (Merriam, 2009, p. 231). In-depth interviewing has the potential to bring emotions to the surface or cause participants to acknowledge something they had not before. For these reasons, it was important to not become, in Patton's (2002) words, neither "a cold slab of granite," nor a therapist. During interviews, I was responsive to students' well-being and did not push if they seemed at all uncomfortable. Furthermore, I was prepared to make referrals to resources should a participant need them (Merriam, 2003).

CHAPTER 6: CASE PROFILES

The following chapter provides a narrative introduction to the two classrooms that participated in this study. This overview lends context to the detailed findings and analysis in Chapters 7 and 8, both of which are organized more specifically around the research questions. The case profiles employ an increasingly narrow focus—from the larger school context, to the classroom, to the teacher, to his or her lessons, to the students, and finally to the individual students who participated in interviews.

Case One: Eastern High

Eastern High is a large, urban high school with a population of roughly 800 students. The school is located in an area with historically lower socioeconomic households. It pulls its enrollment from a range of middle schools nearby. The school itself is housed in an older building reminiscent of 1960s architecture. Walking past classrooms, many seem to be empty. One is given the impression that the school's enrolment is lower than usual. In fact, during the teacher's interview, I learn that the construction of a new school in a neighbouring area will diminish their numbers further. Like most Nova Scotian high schools, Eastern High spans grades ten through twelve. As one walks down the hallways, the student body's diverse character makes itself evident; posters for African Heritage Month and Mi'kmaq cultural events decorate the walls. Students represent a number of visible minorities, although I only hear English spoken.

The school holds a history of racial division and tension. To a certain extent, a decades-old conflict still reverberates through the halls, at least in the minds of veteran teachers. During an initial meeting with the English department head, I am asked to explain my motivations and the purpose of my study. A steady barrage of researchers

hoping to study race within the school has made her wary of any research directly involving students. According to the department head, some of these researchers held dangerous views about race, and she is understandably protective of the students at her school.

The Physical Classroom

The classroom is used for English, History, and Family Studies courses, and it reflects these different subjects. Animatronic babies lie in boxes, waiting to be adopted by grade ten students—or, as Mr. Clark puts it during his interview, "waiting to be dropped on their heads." The walls are sparsely decorated – perhaps because the classroom is shared across subject areas. A poster of Che Guevera hangs next to a WWII-era poster urging the viewer to "Buy Victory Bonds." A number of artists' sketches are posted on the wall with no text, to be used as visual prompts for creative writing. The rest of the wall space is largely empty save the blank white boards at the front of the room. Instead of a traditional seating arrangement in rows, chairs and tables are set up in the shape of a "U" so the teacher may stand in the middle of students as he or she instructs.

Mr. Clark. The teacher involved in this case will be referred to as "Mr. Clark." Mr. Clark is a relatively young teacher with a dry sense of humour. At the time of this study, he was in his fourteenth year of teaching. Before being hired twelve years ago at Eastern High, he taught briefly at three other schools, including two high schools and one middle school. Mr. Clark's own educational background is largely in the field of History. He was attracted to Eastern High for its International Baccalaureate (IB) program. However, since the IB History courses alone at Eastern High could not constitute a full

teaching schedule, he was assigned English courses. He currently teaches grades 11 and 12 Standard- and Higher-Level IB History, along with English 10.

In front of the class, Mr. Clark exudes both playfulness and stern authority. Students often interrupt his lessons with inappropriate comments, but Mr. Clark plays along and rarely loses his patience. During one lesson, a student asks whether he can bring a turkey dinner to the Examination. Another asks whether he could bring a tent and sleeping bag. Mr. Clark responds with a simple "I've had enough of your ridiculous questions" and moves on with his lesson. Mr. Clark's students often tease their teacher—for being bald, for never smiling, or for never letting them have fun. During the final class of the semester, a student asks, "Can we have a roast of Mr. Clark today?" to which he responds, "Maybe at the end of class." Mr. Clark then suggests that the class should have spent time studying the characteristics of an effective "roast."

The Observed Lessons

In total, I am able to observe three lessons in Mr. Clark's classroom. The first lesson takes place in January, three weeks prior to the provincial Examination. During the lesson, Mr. Clark assigns an essay about *The Giver*—the novel they have just finished reading as a class. Students are asked to choose a topic from a list of options, and he offers time in class for students to begin planning their essays. My second and third observations take place during the week before the provincial Examination, while the teacher is preparing students for the provincial assessment. During the second observation, Mr. Clark walks students through two reading assessment exemplars provided by the Department of Education. One involves a narrative text; the other, non-fiction. Students are asked to read the texts and attempt to answer the multiple-choice

questions. During my third and final classroom observation, students plan a response to a sample writing prompt. Student reactions to these lessons, as well as student-teacher interactions during this Examination preparation, will be discussed in detail in the next chapter.

The Students

Despite Mr. Clark's playful approach to teaching, I quickly notice most of his students are disengaged during lessons and seatwork. At the start of the first observation, many students are using their cell phones, and their phones remain close at hand for the remainder of the period. Some students lay their heads on their desks. During one observation, two students wear earphones and make no contact with their teacher or each other. When Mr. Clark asks a question, very few students respond, and most avoid eye contact altogether. There are a vocal few who participate in discussions, yet I notice that the students who routinely participate are, for the most part, the same three to four students. On my third visit, Mr. Clark tries to engage a quiet student by asking which answer he had chosen. The student responds simply, "I don't want to talk about it."

Student Participants

Three students agree to participate in an interview. Sadly, one declines on the day of the interview due to the suicide of a classmate. The other two students will be referred to as Colton and Abi. Again, pseudonyms are being used for students and teachers to protect their identities.

Colton. During my two classroom observations, Colton is one of the most active, and most disruptive, participants in lessons. Colton sits with a group of other boys, and many of the questions he asks are clearly designed to make his friends laugh. When his

teacher explains the expectations for the provincial Examination, Colton asks whether he should bring a turkey dinner. Despite these behaviours, Colton expresses a great deal of respect for Mr. Clark during his interview. In his words, the entire year has been "awesome." Colton currently has a mark of 90% in the course. He explains that while he "usually" likes English class, he does not enjoy writing "a lot," as he believes his handwriting is poor, as well as his grammar and structure: "I've heard it from all my teachers." What Colton enjoys most about English class is creative writing. He especially enjoyed an open-ended writing assignment that incorporated student choice. Previously, Colton attended an independent school in the same city. Consequently, he has never written a provincial assessment before, and he expresses worry on multiple occasions during his interview. He believes his mark will be reduced by the Examination.

Abi. Like Colton, Abi is an active participant during my two classroom observations. She raises her hand on multiple occasions to answer questions when the class is discussing Examination exemplars. Abi has a high mark in the course; she tells me her grade is currently 96%. Abi has always found English "kind of easy," and she tells me during her interview that she has had a "passion for writing" since she was a child. She especially enjoys writing poetry and fiction. Abi remembers fondly the 20 minutes of silent reading offered by Mr. Clark at the beginning of many classes. Abi appreciates Mr. Clark's teaching style. She tells me he is "easy to talk to" and she enjoys how he "roasts" people. Abi grew up in the same small town as Colton, but she previously attended a public school and therefore has experience with provincial assessments. She expresses some confidence and a sense of preparedness during her

interview, although she expresses frustration during the second observation when taking up the multiple-choice questions on the exemplar.

Case Two: Oceanview Collegiate

The site of the second round of research is a large high school in a more affluent neighbourhood of Halifax. "Oceanview Collegiate," as it will be called in this study, is larger than Eastern High; the student population is roughly 1500 students. The building is much newer than Eastern High, and natural light spills from large windows into a centrally-located food court. Although located in a suburban area, the school draws from a variety of neighbourhoods. The student body reflects cultural and linguistic diversity; as I walk the halls, I hear students speaking at least three different languages. The classroom teacher I will later interview tells me their school community includes a number of refugee families from Syria.

During an initial meeting with the English department, I am greeted warmly by the department head—the teacher who would later become the teacher participant in my study. She describes her experience at the school, the students she teaches, and her course content. She is enthusiastically supportive of the study; for years, she has voiced concerns about the provincial Examination and feels ignored by the Department of Education. She complains that the texts on the Examination are "too generic" and "not responsive to the culture at this school." She excitedly invites me to observe her class at any time, and she offers me a number of documents that she feels may assist me, including a hard copy of the Nova Scotia English Language Arts curriculum.

The Physical Classroom

Upon entering the classroom, it is immediately evident that the room belongs to Ms. Hatfield alone. Every corner of the room is devoted to the subject she teaches. The walls are covered with English literature-related posters. Prominent are promotional materials for the film *Great Expectations*, Shakespearean plays, and motivational messages—the kind with an inspirational quotation below a scenic image. I also notice various checklists: one lists the steps for writing a persuasive essay—another, the steps for writing an expository piece. One poster defines the genres of literature as follows: play, non-fiction, fiction, poetry, and folk tale. Along the windows, a large rotating shelf contains various novels, all of which appear to be fiction. On the white board at the front of the class, Ms. Hatfield her personal email for students' reference. In the middle of the board is a brief outline for a persuasive essay; political cartoons decorate the other side.

At the front of the room, a placard designates a neatly organized box where students deposit cell phones as they enter the room. Every student seems to comply without protest before sitting down at a desk. Next to the cell phone box, along the front wall of the room, is a coffee and tea station—a desk with a kettle, coffee maker, various teas, cups, and additives. A few students take advantage as they settle into class. They have arrived early enough to make their beverages before the bell. The desks in the room are arranged traditionally, in rows. The class is relatively large; there are approximately 25 students present each day during my observations.

Ms. Hatfield

Ms. Hatfield is a seasoned teacher with 44 years of teaching experience. For the first half of her career, she taught overseas in the Bahamas, the United States, and

Singapore. For a time, she was principal at a school on a First Nations reserve in Northern Ontario. 21 years ago, she moved back to Canada and began substitute teaching with English language learners. Shortly thereafter, she began teaching English full-time at Oceanview Collegiate. She is a petite woman with a warm smile and a booming voice. As she speaks, she jumps quickly from topic to topic, explaining everything from metaphor to essay requirements with enthusiasm. When she walks down the hallway, students smile and wave. Many greet her with a simple "Hey, Ms. Hat!," using the shortened version of her name. In fact, every student in her class addresses her this way.

Like Mr. Clark, Ms. Hatfield has a relaxed, humourous approach to teaching and relating to her students. She teaches mostly from the front of the room, where she holds a gavel. When she is ready to begin, she strikes the gavel. When Ms. Hatfield gives the students a task during my first observation, she allows the students to work with music playing in the background. As the students discuss their assignment—an analysis of the poetic devices in Tracy Chapman's "Fast Car"—she wanders the room, discussing the cancer fundraiser the school had just completed, joking about how cartoonish she may have looked in front of the TV crew.

The Observed Lessons

I am able to complete two observations in Ms. Hatfield's classroom. The first takes place in April, two months before the provincial Examination. During this lesson, Ms. Hatfield begins a new poetry assignment, in which students will look to identify poetic devices within their favourite song lyrics. Ms. Hatfield has not yet explained to students why she asked them for copies of their favourite lyrics—the reveal will happen next week. For the duration of this lesson, students listen to and discuss the song "Fast

Car" by Tracy Chapman. The second and final classroom observation takes place in June, the week before the provincial Examination. A different teacher teaches this class, as Ms. Hatfield has gone on sick leave. For the purposes of this study, he will be referred to as "Mr. McGarrigle." During this lesson, students review the same reading exemplars as Mr. Clark's class. They also review the sample writing task. However, instead of asking students to write their own responses, Mr. McGarrigle asks students to read examples of student responses and grade them according to the standard rubric.

The Students

Although the school is diverse culturally, linguistically, and socioeconomically, Ms. Hatfield admits that her own grade ten course is relatively homogenous:

> This class has a mixture of kids but they're I think without exception born and raised here. That doesn't happen every day. And I think part of it could be because they're what we call 'band kids.' […] And for whatever reason music, in this neighbourhood anyway, seems to be more European based kids than the mixture of kids we have from Africa and the kids that we have from Syria and the kids that we have from Korea.

The students in this particular class, she explains, are mostly enrolling in the International Baccalaureate program. This means they are almost universally headed to university. I learn that most of these students have known each other for years, and that they are largely enrolled in the same courses. She considers them to be high-achievers who take school seriously.

Overwhelmingly, I am surprised by the frequency of student participation during observations. When Ms. Hatfield asks students, "Did you guys find any poetic devices?",

multiple hands are raised. One student shouts “It’s like the whole song!”. Students provide examples of imagery and point of view. Students make insightful comments about personal conflict. At one point, the class begins discussing gender and how the meaning of the lyrics may change with a differently gendered voice. A discussion of African American identity follows, and students make connections to other works they have read.

Although the class is animated, students easily settle into seatwork. The words “engaged,” and “tuned in” appear numerous times throughout my field notes. During the first classroom observation, while students are discussing the song, I listen carefully to the groups nearest my seat at the back of the room. All students are discussing the task at hand for the duration of the lesson. During the second visit, while students complete practice assessments, I record “one could literally hear a pin drop.”

Student Participants

When I distribute consent forms for student interviews, I expect an uptake mirroring that at Eastern High. However, the response is overwhelming. In total, twelve students consent to participate in an interview. On my second visit, several students are absent due to an event, and I ultimately interview five students.

Jason. The first student interviewed in Ms. Hatfield’s class is Jason. Jason speaks with confidence and clarity about his abilities and his future endeavors. He tells me that he plans to pursue either acting or computer programming. Since neither of these professions demand “a ton of essay writing or reading comprehension,” he is not concerned about his final grade. That said, Jason tells me his mark is currently 85%, and he predicts he will do well on the Examination. During Jason’s interview, we discuss the

independent reading projects in some detail. He tells me about a web comic he studied—once which was "significantly longer" than the novels chosen by his classmates. He appreciated the independence this project afforded him. Jason has come to Oceanview Collegiate from a smaller junior high school. His experience at Oceanview Collegiate has been "significantly better." With enthusiasm, he describes a number of clubs he has been able to join, including a programming club and a video games club.

Shelby. Shelby's interview runs slightly longer than the others, as she offers thorough answers to most of the questions. She is articulate and confident when speaking about her experiences. In Shelby's opinion, the grade ten English course has moved too slowly. She fondly remembers her grade nine English course as her favourite class, and she laments that this year, the class has focused on "very basic kind of English stuff" like "book assignments" and watching movies. Shelby tells me that she knows everyone in the class because of the band program. She enjoys the camaraderie, but she complains that they sometimes veer "off track." Shelby loves poetry and "any kind of writing." She also enjoys reading, but she tells me she prefers choosing her own novels. When our conversation turns to the Examination, she expresses great confidence in her prospects. She remembers the grade eight provincial assessment in detail, calling it her "favourite exam to do, because it was the easiest one." She praises the value of standardized assessment in general, citing the province's ability to gather "statistics on how kids are doing" and potential problems with teachers.

Josie. Josie finds this year's English course to be a dramatic change from last year. Previously, Josie was enrolled in French immersion, where her "English class was sort of a joke." She admits that she "honestly can't remember doing much in English

class [in grade nine].” Despite the change in workload and expectation, she tells me her grade is in the mid-90s. Josie enjoys reading poetry for pleasure, although she dislikes the critical analysis they undertake in class. She feels that she is a literal person and that the figurative language “goes over [her] head.” She enjoys writing essays, since she finds them formulaic and simple to produce. For next year, Josie has signed up for the Advanced English Language Arts course, although she fears it will not run due to low enrolment and scheduling issues.

Kate. Like Josie, Kate was previously enrolled in a French immersion junior high school, and her experiences sounds similar. She tells me that English class was “[our] free time to like, just relax and not really do a lot of work.” Kate is the first participant at Oceanview Collegiate to openly discuss academic struggles. On multiple occasions, she describes herself as a “slow reader.” What she worries about most when considering the provincial Examination is the time limit. Kate explains that throughout elementary school, she would often “mix up” her words, and that once she believed herself to be dyslexic. She describes reading generally as a “chore.” Her chief complaint about her English class this year is that it is “really loud.” Kate expresses reservation about sharing her own opinions, especially when they diverge from those of her classmates. Despite her self-professed lack of confidence, Kate tells me her grade in the course is currently an 89%. Again, she enjoys being at her new school and the camaraderie that comes with the band program, although she will be quitting the band next year in order to take French courses.

Ben. Ben is yet another student who was previously enrolled in French immersion. He is an active participant during classroom observations, and much of his

interview focuses on his love of debate. In regards to classroom conversations, Ben states "they can get off-track sometimes and, sometimes hilariously off-track. But it's always good because it kind of opens up more about other people and it just gets me more attuned to what's going on in the world." Ben enjoys English class generally, but he particularly enjoyed the novel studies. His grade is in the "high 80s and 90s." With respect to the provincial Examination, he is not worried. When asked whether he believes his grade on the Examination will affect his overall mark, he explains, "I don't think it's going to adversely affect it… adversely affect it like, significantly." When discussing the Examination, he demonstrates a nuanced understanding of assessment: "it can be hard to draw the line between where a student is just performing poorly and when the exam is actually constructed in a way that is not really a good test of a student's knowledge." He explains that he has had many conversations about this with his father, who is a university professor.

CHAPTER 7: FINDINGS

Research data from field observation includes field notes, course literature, and recorded interviews. Data was organized through multiple rounds of analysis with an increasingly narrow focus on the research questions. A discussion of the organization of data will be followed by an overview of findings within each code.

Organization of Data

The first round of data analysis was inspired by the study's research questions:

i. How do students and teachers describe classroom practices?

ii. How do students and teachers describe the evaluation?

iii. Do any differences between these accounts pose a threat to the inferences drawn from test results?

Given the volume of research data, research questions were reframed during the coding process. For example, to consider how students and teachers describe classroom practices, data revealing "Classroom Procedures" was emphasized. This included approaches to reading, writing, and alternative literacies, as well as teacher-led exam preparation. To consider how students and teachers describe the evaluation, data reflecting "Perceptions of Assessment" was coded. This included messages from peers and teachers, as well as positive and negative perceptions of the Examination voiced directly by student participants.

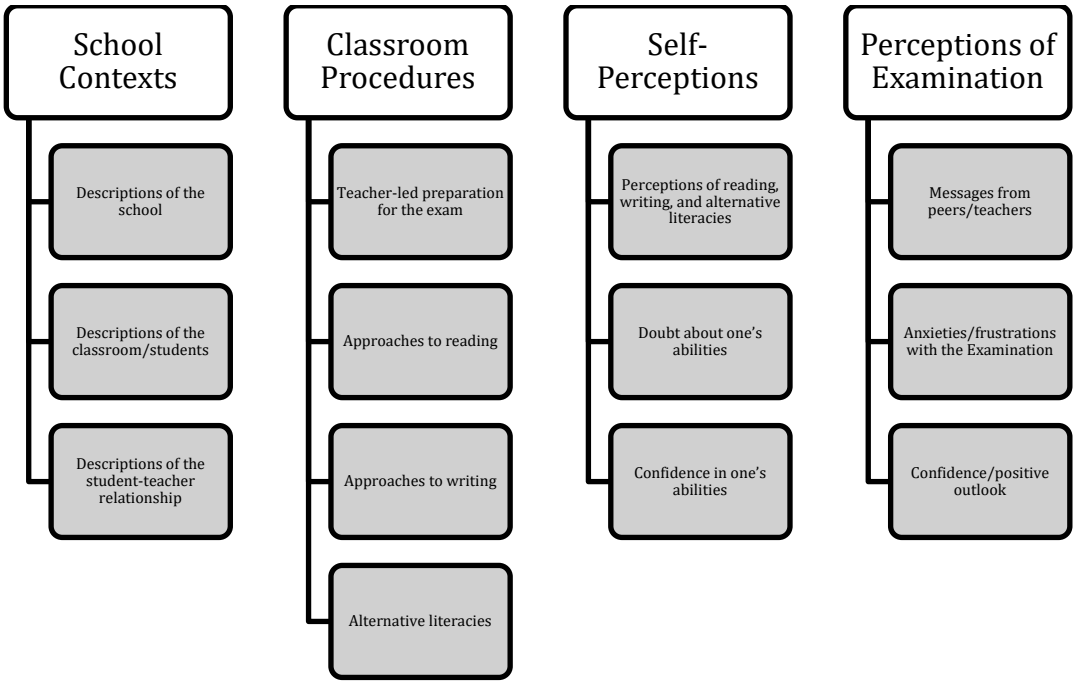
The coding process intentionally omitted the final research question (iii.), because any answer could only be inferred through analysis. However, contrast between schools, classroom procedures, participant perceptions and their interaction with the assessment was critical for post-research assessment. A code for "School Contexts" was generated to

include descriptions of the school, of the classroom and students, and of relationships within the classroom, including teacher investment in his or her students. A "Self-Perceptions" code was included to identify how students describe themselves. This code included self-doubt and confidence about one's abilities, and perceptions of writing, reading, and alternative literacies.

After completing the first round of data review using these codes, certain themes began to emerge. The following table presents a visual map of the codes used for data organization, and the sections thereafter explain emergent themes within each code subset.

Figure 2

Codes Used for Initial Data Review



School Contexts

This section outlines the information offered and observed about the larger school context and the classroom context of each case, including the ways teachers relate to their students. Themes of diversity, homogeneity, and teacher investment are explored.

Diversity

Participant descriptions of both schools almost universally emphasize the size and diversity of the student population. Both teachers describe their respective schools as "diverse" from a cultural and socioeconomic standpoint. When describing the demographics of his school, Mr. Clark explains, "there are about 800 students and they come from four really distinct communities that feed into this one school. […] So they're all very different communities that have similarities but also some pretty major differences that are all feeding into this kind of one building." According to Ms. Hatfield, the teachers at Oceanview Collegiate "like to call ourselves the largest and most diverse school of its kind east of Montreal." She cites the two full-time EAL teachers on staff and the recent arrival of students from Syria. Field notes reflect student diversity with respect to culture and language—at Oceanview Collegiate, multiple languages are spoken in the hallways, and at Eastern High, posters for cultural events and celebrations reflect the visible differences between the students.

The social benefit of diversity is a common topic within interviews at both sites. In response to the question, "How would you describe your school?", five out of seven student interviewees express enthusiasm for a more diverse social network. According to Josie, for example, "here, we met like, a lot of new people, a lot of new teachers. […] So, there were just a lot more people to choose from to like, be friends with". Ben is happy to

have found people with similar interests: “Because there’s so many different groups and, like, social circles and interests and stuff like that. And, so there’s… there’s no one that can’t find someone like them”. Mr. Clark shares this belief about finding like-minded friends:

> I think it’s amazing. It creates—I don’t know, students can always find like-minded friends in this building I find. You can always find somebody else that you can connect with, whereas I think in smaller schools or in schools where you have more of a homogeneous population, that’s not always possible I guess.

For the same reasons, Mr. Clark worries about the loss of diversity that may come when a new high school is built nearby:

> I think the new high school in [nearby town] that they’re building is going to affect us in a really negative way and I think it’s going to affect [nearby town] in a negative way. […] it’s going to just create this homogeneous – so if you grew up in [nearby town] […], you’re going to graduate with the same kids you started with and there’s going to be no differences.

Since the completion of this study, the construction of the new school has been completed. The student population of Eastern High has consequently been drastically reduced.

Classroom Diversity

Unlike those broader diversities highlighted above, classroom diversity is clearly distinguishable between case contexts. Mr. Clark’s classroom at Eastern High provided a general cross-sampling of the student population, as it was open to any student currently

enrolled in grade ten. There are differences between the students in his class with respect to confidence, participation levels, and attitudes toward the course. Mr. Clark speaks to these differences when he describes his students:

> I don't know, I worry about some of them in particular. I think some of them are going to be fine. […] Some of the quieter students I think are going to struggle with [the Examination]. Some of them even in the class require a lot of one-on-one help.

At least one student in his class will be exempted from the Examination due to an Individual Program Plan, and others are making use of classroom accommodations such as scribes.

Student disengagement is a theme that emerges within the "Descriptions of the Classroom/Students" code, especially at Eastern High. Colton, one of the two student interviewees from this classroom, recognizes his classmates' varying levels of engagement: "In this class, I think definitely half of the class is really into it. And then the other half is just not trying at all." When discussing her own performance on the Examination, Abi, another student interview participant, notes "but then a lot of us what say that [we don't need to worry] haven't really been I guess paying attention in class when [we] should be." notes from classroom observations provide what are perhaps the clearest indicators of student disengagement. The notes describe students who are "wearing headphones," "looking tuned out," and "not looking at the assignment." During the first classroom visit at Eastern High, while the students are choosing topics for an essay, I record that "some students have their heads on their desks and the teacher doesn't seem concerned." Colton and Abi contribute to teacher-led discussions during all three

visits, along with some of the students seated nearby. However, these few students are consistently the only participants, supporting Colton's observation that at least half the class is disengaged.

Conversely, Ms. Hatfield's classroom is more homogenous with respect to background and participation levels during classroom observation. During her interview, she explains that, because hers is the only English course offered on a full-year basis, its composition is mostly "band kids" to accommodate their unique schedule. Students are culturally homogeneous: "…for whatever reason music, in this neighbourhood anyway, seems to be more European-based kids than the mixture of kids we have from Africa and the kids that we have from Syria and the kids that we have from Korea." Indeed, the students in her class all appear to speak English as a first language, and few appear to belong to any visible minority. The question of cultural relevance with respect to course content will be examined in greater detail in "Classroom Procedures," but Ms. Hatfield's program appears to engage with more socially-critical topics than Mr. Clark's. She personally believes this focus on racism and social justice is less relevant to the students in this particular class: "Because none of this means anything to them for the most part because they're white." Her statement is somewhat inconsistent with student interviews and observations, however, as many of her students independently mention literature and topics that relate to issues of social justice.

Most of the students in this class, Ms. Hatfield explains, will move on to the International Baccalaureate program. During the three classroom observations, the students demonstrate significant engagement and motivation. Field notes record multiple observations that reflect this engagement, including "tuned-in", "multiple hands raised [to

answer a question],” “really insightful comments,” “students working quietly,” and “you could hear a pin drop.” They also describe relatively equal participation between boys and girls, and few students getting off-task. Ms. Hatfield’s cell phone drop-box may factor in her students’ focus, as they do not appear distracted by personal devices during observed lessons. Student interviews also reflect high levels of motivation. For example, one student laments the pace of the class and the “basic” nature of what they have covered. She comments that “[i]t felt like we got our work done a little slower than we usually could have. And we are starting some really exciting work and assignments and stuff, but there’s a lot of stuff that we did earlier, and it was very basic kind of English stuff.” In contrast with Mr. Clark’s students, those in Ms. Hatfield’s course face a much larger volume of writing and reading. This disparity will be further described in “Classroom Procedures.” Despite this larger volume, students do not complain about workload during interviews.

Relationships

During the coding process, the theme of relationship building between student and teacher gained significance. The data coded for this category includes the way teachers relate to students and how students, in turn, respond to teachers. Of course, due to the limited time spent in the physical classroom, this data provides only a glimpse into such a complex phenomenon of relationship building. Nevertheless, these accounts and, in particular, the differences between classroom management and student-teacher relationships will add richness to the discussion of findings in Chapter 8. Both teachers relate warmly to their students during observations, but there are differences in the way

they transfer responsibility, in their general expectations of students, and in how students respond to their teachers.

During classroom observation, Ms. Hatfield relates to her students almost as though they are adults. Her "coffee club" is popular among her students, and the teacher who later replaces her keeps this coffee club running. In his words, "[i]t's so civilized. The kids are so civilized," and it "…makes [the students] feel like grown-ups." Many students take advantage, arriving with enough time to prepare a drink before class. During observed lessons, Ms. Hatfield engages students in meaningful conversations about complex topics like abusive relationships and gender identity. In turn, her students make insightful comments and thoughtful connections to their own lives and other texts they have read. During the "Fast Car" task, for example, students discuss the ways the meaning of the lyrics may change depending on the gender of the singer in response to one student's mistaken belief that the singer was a man.

During the observed lessons, Ms. Hatfield provides multiple opportunities for students to work in groups or independently. When Ms. Hatfield transfers responsibility to her students, the students use time well. During the first classroom observation, Ms. Hatfield allows the students to discuss song lyrics with music playing in the background. As the students discuss their assignment in groups—an analysis of the poetic devices in Tracy Chapman's "Fast Car"—she wanders the room, stopping to talk to groups on the way. Students work on this for over twenty minutes before discussing their ideas as a class, and they work without distraction; in fact, it is Ms. Hatfield who occasionally changes the topic. During later observations, her replacement teacher allows the same transfer of responsibility when he asks students to read and answer questions from

Examination exemplars. This time, the students elect another student to read a story aloud—an event that will be described in the "Explicit preparation" section. When students are answering comprehension questions or reading essays, they work in complete silence.

With respect to her teaching philosophy, Ms. Hatfield seems to value a relaxed, "fun" approach. During her interview, she states simply, "There's no reason it can't be fun." She goes on to explain,

> Yeah. I think if you're more relaxed with the kids and they will be with you then they're not so afraid to make mistakes. The kids that are most uptight, like that young man that was here, can be most uptight about getting it right.

Certainly, her students respond well to this approach during classroom observation. Field notes describe students who participate frequently with thoughtful ideas and opinions, and they work in groups in a way that is respectful of time and of their teacher's expectations.

Considering Mr. Clark's class in relation to Ms. Hatfield's, the student-teacher relationship stands out. During observation, students demonstrate disengaged behaviours such as use of personal devices and lack of eye contact with the teacher. Perhaps it is unsurprising, then, that Mr. Clark's approach to independent work involves greater oversight. Most of the observed lessons are teacher-directed, and when he does transfer responsibility to students, they are expected to work alone. For instance, during the first classroom observation, Mr. Clark asks students to quietly consider and write notes about their essay assignment. During this time, Mr. Clark circulates, speaking with students

one-on-one about their choices. When a student begins showing something on his phone to his classmate, Mr. Clark reminds him to "brainstorm on paper." During the second and third observations, Mr. Clark asks students to read Examination exemplars and attempt to answer comprehension questions. This is done independently at first, and responses are corrected aloud by the teacher. At no point during the three observations does Mr. Clark's class engage in meaningful discussions of complex themes like the students in Ms. Hatfield's class, although it is certainly possible that such discussions took place during other lessons.

During observation, the way Mr. Clark's students respond to their teacher is, again, distinct. Mr. Clark's students frequently shout inappropriate comments during lessons, often at the expense of their teacher. For example, students joke about Mr. Clark's lack of smiling and make comments about his receding hairline. He appears to take these comments in stride, smirking and continuing on with his lesson. In fact, his response to these comments is often to play along. When one of his students asks "Can we have a roast of Mr. [Clark] today?" he responds, "Maybe at the end of class." When interviewed, Abi describes Mr. Clark's playfiul sense of humour—specifically, she tells me she likes when Mr. Clark "roasts" people. During the second classroom observation, Mr. Clark explains what will be needed for the Examination: "All you need is what? One pencil. One pen." A male student blurts, "Can I bring a turkey dinner?" Another asks, "Should I bring my tent and sleeping bag?" When Mr. Clark rolls his eyes and responds, "I've reached the end of your ridiculous questions," the remarks stop. The ways in which exemplars are presented and received by students will be discussed in greater detail in "Examination Preparation."

Teacher Investment

The code of "Teacher Investment" identifies instances when teachers expressed worry or demonstrated explicit efforts to support students throughout the study. The teachers exhibit distinct concerns about their students during interviews and observations. At multiple times during Mr. Clark's interview, he worries that students will not engage with the material on the Examination. In particular, he worries about the quieter students:

> I worry about some of them in particular. I think some of them are going to be fine. Even like [Colton] I think is—he requires a lot of attention but he's actually quite smart. Somebody like that I'm not worried about at all. Some of the quieter students I think are going to struggle with it. Some of them even in class require a lot of one-on-one help and me sitting down with them and looking at things together. That isn't really an option when they're doing the provincials.

Classroom observation confirms this to some extent. During the observed lessons, Mr. Clark frequently "checks in" with students, although the students he interacts with are usually those who are most vocal, like Colton. After assigning a task, such as selecting an essay topic or completing a Provincial Examination exemplar, he moves around the room, visiting students who might need support. Many of these visits are simply reminders to stay on task. Mr. Clark tries to encourage his students when they express frustration. During Provincial Examination practice, Mr. Clark reminds his students, "You guys sell yourselves so short. You guys are very capable."

Conversely, Ms. Hatfield rarely, if ever, expresses concern about her students' work ethic, motivation, self-esteem, or ability to succeed on the Provincial Examination. Rather, she expresses a high degree of confidence that her students will generally

succeed. When Ms. Hatfield expresses concern, it is with respect to her other classes, particularly those with a higher ratio of EAL students. For example, Ms. Hatfield feels strongly that EAL students would benefit from full-year English classes, as opposed to semestered ones:

> …I've been saying ever since the semester system came in 97, 98 we're going to miss full year English. And we really miss it with our EAL kids because they could do their English 10 and then not have another English course until the second semester of the second year.

Her own grade ten class receives a full-year English course because of their commitments to the school band. Again, when discussing the Provincial Examination, her chief concern is that it will negatively impact students from her other classes—particularly English Language Learners:

> [t]he fact that we have such a diverse group of students who don't necessarily identify with the topics or genres that, oh, I'm going to be so rude, that the narrow minded thinking of the board or the DOE or whoever puts, probably DOE, putting this together doesn't see beyond the apple orchards of the Annapolis Valley. And if I put apple orchard on a number of tables with the group that I have— "What's an orchard Miss?"

Indeed, the only concern she expresses about her students is that they will "overthink" their answers. During our initial consultation, Ms. Hatfield describes how many of her students will think too hard about certain questions, expecting there to be some greater, deeper meaning, when the answers are more simple. During her interview, she explains how she prepares students to think like a test-writer: "I can see where this isn't quite the

full answer but you're going to have to ask yourself, 'But will they know the difference at the government level?' Give them the answer you think they want." When multiple students have selected the incorrect answer, she and the other teachers will often give them the mark anyway. She explains that "[t]he grade 10 teachers get in a group and say look, I'm on number five of this multiple choice and all my students have answered C but the answer's supposed to be B. But you know what C is the better answer so we gave it to them."

Classroom Procedures

This section outlines classroom procedures that were apparent during field observations and interviews, supplemented by documents such as course outlines and sample assessments. The section will detail and contrast approaches to writing, reading, alternative literacies, and Examination preparation.

Approaches to Writing

As in virtually any language classroom, writing plays a central role in the two classrooms studied. Both course outlines weigh "Writing and Other Ways of Representing" at 25% or more of the course mark. However, there are subtle differences to the way in which each teacher approaches writing and the genres he or she emphasizes and/or avoids.

Ms. Hatfield acknowledges her focus on persuasive essay writing; this genre is her starting point for writing with the class. She describes how she begins "with persuasive with my 10s obviously because that's where they're going with the Examination." Her students write a number of essays over the course of the year. Without clarification of the exact number, one student states there were "four or five big essays."

Each grade trials a different type, and during her interview, Ms. Hatfield describes the progression:

> As far as - because they have to write an essay, the 10s write a persuasive, the 11s write a comparative and the 12s write an analytical, that's all prescribed. Which is fine, you know, we divide up the responsibility and I think you build to analytical that way.

When asked what they believe teachers would say is the "most important thing to learn" in an English class, three out of five of the student participants in Ms. Hatfield's class name essay writing. According to one of her students, "I feel like, they would probably say that like, it's important to get essay writing skills down, because it's very useful for all sorts of different courses in university."

Among those forms of writing referenced in the data, journal writing and poetry are the only others present. During her interview, Ms. Hatfield explains that journaling prompts will often relate to current events: "[t]hey do free writing for their journals based on topics that - if we've had a - this has been a crazy week but perhaps, you know, the Humboldt hockey team, might have been a topic." Interestingly, she declines to emphasize creative writing. When asked, "[w]hat kinds of writing are your students producing?" she responds glibly, "[n]ot creative writing." She explains that "students do it all the time in junior high" and need no further practice.

Mr. Clark does not share Ms. Hatfield's view of creative writing. Rather, he and his students appear to embrace it as foundational. During Mr. Clark's interview, he is excited to tell me about a creative writing piece which relied on hand-drawn prompts posted on the wall. His enthusiasm is clear; he immediately presents me with the original

book and its accompanying images. When asked which parts of the class they enjoyed most, both student interview participants cite this creative writing task. In fact, for Colton, this is the only enjoyable assignment he mentions. In his words, "I did my creative writing piece about [the image] and I really enjoyed that because I just did what I wanted to do." Abi first mentions daily silent reading, then describes the creative writing task in glowing terms: "I loved the other writing part of it like creative writing. So that was pretty fun I guess." Earlier, she mentions her enthusiasm for creative writing, emphasizing the importance of personal choice and the ability to write "…fiction, just whatever I want to write about."

That said, Mr. Clark also places emphasis on essay writing. During the first classroom observation, students are asked to choose a topic in preparation for a five-paragraph essay. During his interview, Mr. Clark explains that students have been preparing for this persuasive essay through various informal in-class activities focused on essay writing and structure:

> This is a formal one but we do like – I do in-class essays where they just kind of do them in-class and writing. So there's that. I do sometimes, we'll just do like an introductory paragraph. So instead of writing out the entire essay, what you would do for an introduction or just the kind of bare bones but working towards just this kind of – this big formal essay that they have to do typed up.

There is no data to suggest that Mr. Clark's students complete the volume of essay writing that Ms. Hatfield's students complete. In fact, during Colton's interview, he references only the essay on *The Giver*: "We did one essay and I just wasn't really used to doing essays. So I didn't really know how to write an essay but I figured it out." In

comparison to Ms. Hatfield's students, it seems that Mr. Clark's students write at least three fewer essays over the course of the semester.

Approaches to Reading

Both teachers approach reading in a somewhat traditional way; students in the two classes engage in independent novel studies and study a novel together as a whole class activity. One assessment document provided to me by Ms. Hatfield is the outline for her independent novel assignments. The requirements include choosing "three novels to read independently" which students "have not previously read or viewed as a film." These novels are read over the course of the year with firm deadlines set for the completion of each one. She explains in her outline that "[t]ime will be provided during class time at least once a week for silent independent reading." Aside from the requirement that the novels must be at least 250 pages in length, Ms. Hatfield leaves student choice relatively open. Students may choose fiction, non-fiction, or autobiographical novels. During student interviews, one student references these novel studies as an example of a task he enjoyed: "I enjoy the studies on chosen novels that we're doing. Like, we get to choose a novel and do like, a reflection, or a project of some sort on those. And, I really enjoyed doing most of those."

Similarly, Mr. Clark offers an independent novel study assignment. However, Mr. Clark's students are expected to read only one self-selected novel. Like Ms. Hatfield, Mr. Clark offers time for students to read quietly at the beginning of class. Abi remembers this fondly during her interview: "So I mean like having just like – we had like 20 minutes of reading before we started class throughout the year. So that was really nice. It was kind of like relaxing too before we started our day." During Mr. Clark's interview,

he explains that this novel study is born of a desire to help kids discover a love of reading:

> I definitely try to foster I guess a love of reading. I love reading. I've learned so much just from reading and I find it's something that a lot of students don't realize that they like because maybe they haven't found the right novel or the right book. […] I try to really just foster a love of reading and try to get them to read, which is becoming more and more difficult each year.

Although he bemoans the lack of student interest in reading, by abandoning traditional expectations and rigid grammar lessons, Mr. Clark appreciates newfound flexibility. He observes that "in public schools now, we don't focus so much on grammar and stuff, so it's just a lot of—you can have fun with debate and getting people to I guess share their opinions and thoughts."

Mr. Clark recognizes that many of his students struggle to connect with the reading provided to them. "Cultural competence" is a term that he explores in-depth during his interview, and which will be detailed further in the "Perceptions of Assessment" section. With respect to his class novel, Mr. Clark feels somewhat restricted: "Well, you want it to apply to everyone. And even at that, it's still difficult. For instance, like a class novel, we're still kind of limited to what you have at your school. Like I can't go out and buy a class set of whatever." During his interview, Mr. Clark explains why he chose *The Giver* for his grade ten class:

> But I still try to like even for instance, I do *The Giver* in grade ten because it does talk about some of the issues that you can kind of connect to the whole cultural thing […] and connecting that to racism. So there's definitely—that's why I chose

> that one. But yeah, we're still kind of limited to the resources and depending on what high school you're in, your resources could be very difficult.

In other words, from a limited basket, he prefers and option that lends itself to the types of cultural and racial discussions that are most relevant and authentic to his students.

The students in Ms. Hatfield's class cover a much greater volume of teacher-selected reading material. Aside from the independent novel assignment, students also read "two to three" teacher-selected novels. Like Mr. Clark, Ms. Hatfield expresses a desire to find novels that are relevant to her students. She describes the process by which she chooses class novels:

> It's rather organic because you don't pick a book or a play until you see who's sitting in front of you. So you start with just—I have, I give them all a notebook first day of class, and I have a series of questions. I say 'I don't know you.' […] So they tell me about themselves. And my job that first night is to take all these notebooks home and read and it has to be at least a page and some kids write two pages. And that helps me.

However, Ms. Hatfield believes that the novels she selects are not personally relevant to the students in this particular grade ten class because, in her opinion, they cannot personally relate to issues of race and social justice. She believes her own personal connection to the materials is most important: "It has to be a topic I believe in and I'm engaged with. Because if I'm not how am I going to translate that to them?" She jokes that the students might ask "Can we stop talking about the African Americans and civil rights for heaven's sake miss?" By the time I completed my study, Ms. Hatfield's students had read Maya Angelou's *I Know Why the Caged Bird Sings* and Mebla Pattillo

Beals's *Warriors Don't Cry*. Again, data gathered through observation and interview did not support Ms. Hatfield's assertion that her students did not enjoy or connect with these works and topics.

Alternative and Critical Literacies

Although the two classroom teachers largely focus on traditional novel study and writing forms, there are some examples of alternative literacies within their courses. These literacies fall outside traditional the literacy constructs of novels, poetry, essays, and short stories. One way that Ms. Hatfield creates personal relevance for her students is by allowing them to self-select poetry in the form of song lyrics. During my first classroom visit, students are beginning this task by analyzing the lyrics of Tracy Chapman's "Fast Car." Although students are analyzing music, they are analyzing the lyrics in a traditional way, identifying poetic devices within the songs. When asked about their favourite tasks during interviews, students do not reference this assignment, nor do they reference any other classroom materials that would be considered alternative to traditional language learning approaches. However, it seems some students do select non-traditional texts for their independent novel assignments. For example, Jason was excited to tell me about a web comic he studied: "It's similar to a comic book, except on the Internet in a significantly longer format. It's like a novel. It's like a very long novel with pictures, the one that I did." The options for presenting knowledge within these novel assignments are themselves quite open-ended and represent several different literacies. Included among the choices are a novel scrapbook, a movie advertisement, a sculpture, a song, and a life-sized model.

In Mr. Clark's class, students have at least two opportunities to engage with alternative literacies. One document he provides for review is the "Someone Knows Something" assignment, for which students are required to listen to two episodes of the popular CBC podcast and respond to questions about it. The questions themselves are not simply content-based—rather, they encourage students to think about the elements of oral storytelling. Students are asked to make predictions about the content of the podcast and to reflect back on their predictions. Prompts ask: "Were either of the predictions you make at the beginning of class close to the general content of the podcast? Why is *Someone Knows Something* an appropriate title for this podcast?" Mr. Clark asks simple content-related questions, as well as higher-order questions that demand critical thinking from the students. He poses, "[h]ow would you describe the tone of David Ridgen's voice in the podcast?" and "[w]hat effect did it have on you as a listener?" His *Black Mirror* assignment also strays from the "traditional." For this task, students watch an episode of the popular Netflix show and write personal responses. During her interview, Abi remembers this task as "pretty cool." During one classroom observation, students ask the teacher if they could "please watch that show again?"

The question of critical literacies—whether students are engaged in examining texts and viewpoints with a critical lens—is more difficult to address within the limits of this study. Using only data gathered through observation, interviews, and document review, the degree to which students critically engage with texts appears to be minimal in both classrooms, but particularly so in Mr. Clark's course, perhaps due to a lack of resources. The course outline provides an overview of genres only—specific texts are not listed. However, comments made by Mr. Clark during his interview reflect an attitude of

defeatism regarding cultural relevance. He laments his inability to purchase his own resources. When asked whether he tries to provide relevant resources for his students, he responds "...it's still difficult. For instance, like a class novel, we're still kind of limited to whatever you have at your school." He goes on to explain that the main novel for the course, *The Giver*, has undertones that can be connected to racism and other social issues. However, the essay task distributed during the first classroom observation does not include a topic in line with such nuanced social issues, save one option about euthanasia. The extent to which students critically engage with exemplars will be explored in the next section.

Assessment Preparation

To prepare students for the NSE: English 10, both teachers spend time purposefully reviewing its content in class. However, data also suggests that both teachers prepare students throughout the year, albeit less directly. In discussing teacher-led preparation for the NSE: English 10, it will be therefore be useful to distinguish between explicit and implicit preparation.

Explicit Preparation. The two classroom teachers prepare students explicitly for the NSE: English 10 with the same resources. In both cases, teachers make use of the provincially-designed exemplars for narrative, informative, and multimedia texts. Students in both classes also examine exemplars and rubrics for written responses. Although the two teachers use identical documents to prepare their students for the Examination, there are notable differences to framing, to emphasis placed on certain skills, and to student reception.

The second classroom observation at Eastern High marks the first day of explicit Examination preparation in Mr. Clark's class. Mr. Clark distributes copies of the narrative exemplar and the nine accompanying multiple-choice questions. For the first 30 minutes of class, students are expected to read the text quietly and attempt to answer the multiple-choice questions. Immediately, Mr. Clark offers a tip to support students in answering multiple-choice questions, and it sounds as though it has been said before: "Remember, you're looking for the best fit, the best answer there." Multiple times, students respond with frustration to the options given for each question, and Mr. Clark reminds them to choose the "most correct option" and to use "deductive reasoning." After students have attempted to answer the questions themselves, Mr. Clark corrects them with the class.

Mr. Clark's interview suggests this focus on multiple-choice questions is deliberate. When asked how he prepares his students for the NSE: English 10, he references the exemplars, and then immediately brings up multiple-choice questions:

> I basically have - and you can get them off the department of education website. They're basically just exemplars. And that's really what I use to – so starting I would say the week before the exam starts, I'll talk to them about – one thing that they're not really used to is the whole multiple choice thing. So yeah, I do the exemplars with them, I talk a lot about multiple-choice. That's something they're really not all that familiar with.

When asked how he teaches his students to answer multiple-choice questions, he emphasizes the skill of deductive reasoning: "I mean deduction is usually a safe way kind of getting rid of those answers that you absolutely know are not the correct answer and

then trying to kind of work with what you have." Interestingly, he also feels the need to explicitly teach the filling out of bubble sheets:

> I mean getting them to fill out that bubble sheet takes a good 20 minutes at the beginning because I've never really like – when you think about it, it's something they've never really done and it's their first time. And yeah, it takes a little like – you got to go through it step-by-step, you've got to colour in each letter for your name.

Test-taking skills therefore consume a somewhat significant amount of time in this class, at least in the final week of the semester.

During classroom observation, a focus on test-taking skills seems to take priority over personal connections when working with Examination documents. Field notes reflect "little to no reflection, discussion, or connection" with respect to exemplar texts. The emphasis during each task is on finding the most correct answer among a number of distractors, especially when students become frustrated with their lack of success. Following the first exemplar—a narrative piece—students are visible and audibly upset. Of the students who participate, none appears to have chosen every correct answer. When the teacher asks students to move on to the next story, an expository piece about Search and Rescue Technicians, the mood in the classroom improves. The subject matter involves high-stakes rescues and helicopter stunts, whereas the previous story had recounted the experience of a middle-aged stepmother. The piece seems to appeal more to students immediately, as a few of them are overheard discussing it. Mr. Clark attempts to start a conversation, asking the class "What did you think? Fairly interesting? Anyone

want to do this job?", but few students offer ideas. Instead, the conversation quickly returns to the correct answering of the multiple-choice questions.

Student responses to the explicit Examination preparation are similar to the remarks described in earlier sections. Again, the students who participate in these lessons are the same few who regularly participate, two of which were interviewed. Students generally appear disengaged—field notes describe a number of students looking at their cell phones after Mr. Clark asks them to read quietly. When Mr. Clark tries to engage a student who has not raised his hand, the student responds "I don't want to talk about it." Interestingly, Mr. Clark does little else to engage students who are not participating. I write "More than half the students never say or ask anything" and "Very few kids are actually doing the work," and "the teacher doesn't seem to mind." One student sitting near me listens to his headphones for most of the lesson. I later learn through Mr. Clark that this student has an IPP and will therefore not be writing the Examination. There are also complaints—"What? They're gonna make us read all *this*?"—and questions about the grade value of the Examination: "If you completely fail this exam, how far down will your grade go?."

During the third and final classroom visit at Eastern High, I observe Mr. Clark preparing students for the persuasive essay section of the Examination. Mr. Clark reads the writing prompt to students—"It has been said that written texts often reflect human strengths and weaknesses"—and asks them to start noting some of their own ideas. One student interjects, "We're writing an essay?" Mr. Clark explains the expectation that students relate the prior readings to this writing prompt, and then he walks the class through the potential connections found in each one. He reviews the marking rubric for

this piece before moving on to the next writing piece. During the lesson, Mr. Clark engages with the students one-on-one, offering support where needed. He suggests to one student that he look for another point and assures another that he can use personal pronouns. Interestingly, when he talks about the Examination, he often uses the pronouns "they," as opposed to "we." For example, "They have a blank page that they call a planning page" and "This is how they're going to mark you."

The explicit Examination preparation observed in Ms. Hatfield's classroom is not carried out by Ms. Hatfield. Unfortunately, she was on sick leave during this portion of the study, and a substitute teacher replaced her for the last month of school. The week before the NSE: English 10 was administered, Mr. McGarrigle uses the same exemplars as Mr. Clark, but with a different approach. He begins with one of the writing assessments. Instead of asking students to write their own responses, Mr. McGarrigle distributes samples of student writing and attempts to score them using the standard rubric. Field notes reflect student engagement during this activity: "Students are working quietly. You could literally hear a pin drop. No discussion. No headphones." When Mr. McGarrigle takes up the activity, the class learns they have been overly critical of the student samples; in all but one case, students grade the samples lower than the score they actually received. The highest scoring student sample received a perfect grade of 16. A student responds, "So this is the one you want us to sound like?," and another remarks "But the language used is basic." Students seem surprised by the simplicity of the sample and relieved that the expectations are lower than they had predicted.

Like Mr. Clark's class, Ms. Hatfield's class begins their preparation for the reading comprehension section of the Examination with the narrative exemplar.

Immediately, a male student offers to read the piece aloud to the class. His classmates respond with enthusiasm to the idea, shouting "Yes!" He reads articulately and with expression, emphasizing such words as "primordial" and "undulates." When he reads the phrase "legs and arms akimbo," another student stands up to act out the motion. The students listen with attention, laughing when appropriate and clapping at the end.

Students are asked to complete the nine multiple-choice questions. Again, the class completes this activity with focus. However, when the teacher begins taking up the answers, there is some commotion as students realize they have answered incorrectly. During this portion of the class, the mood changes dramatically. Students become somewhat disruptive, voicing their frustration for the class to hear. I overhear such comments as, "There should be a choice for 'all of the above!'" and "Do I just get it wrong even though it's so close?" Another student remarks, "There's only one right answer on the Examination even if there are three right answers," to which Mr. McGarrigle responds "I didn't design the exam, I'm sorry!". There is some combativeness between students and teacher when students disagree with a given answer. One asks, "How are we supposed to get that?" The teacher responds, "Read it." The student, visibly frustrated, responds "I did read it though, and I got another answer." During both visits, this is the only open frustration students demonstrate.

Implicit Preparation. Both classrooms underwent intensive, explicit preparation during the week leading up to the NSE: English 10. However, both teachers prepare their students in subtle yet significant ways prior to the final week of class. As discussed earlier, Ms. Hatfield has students write multiple persuasive essays. Her focus on this essay genre is deliberate, and she is open about her goal: "…I start with persuasive with

my 10s obviously because that's where they're going with the exam." Her students appear to infer this connection; three remark during interviews that essay-writing skills are what teachers would consider the most important outcome in high school English courses. Kate articulates this idea: "I think, [our teachers] would say how to structure like, an essay or a paper. Because that's something that you're going to use like, throughout your life. Like, and how to make like, I don't know how to say it…how to put words together to make them sound professional." Similarly, students in Mr. Clark's class stress the importance of essay structure. Colton observes, "And my writing's – you have to have structure and all that stuff. And I think that's what they always – they mainly tell me." Abi alludes to Mr. Clark's encouragement throughout the year: "All week we've been doing our studying but also, throughout the year like Mr. [Clark]'s always like bring up different points. He's like, 'See this would be helpful for the exam coming up.' Or like, 'These different tools, you guys should pay attention to this because it will be helpful.'" In this way, the Examination—particularly the essay section—is demonstrably present in the minds of students and teachers before explicit Examination preparation takes place.

Student Perceptions of Constructs

On review, the data shows strikingly divergent student perceptions of literacy constructs between schools. Students' perceptions of their own abilities also stand in contrast. The following summarizes observations and conversations with students specifically on the topic of constructs within the English Language Arts program. In other words, this section explores student opinions and self-reported abilities in relation to constructs, but not in relation to the NSE: English 10.

Among student interview participants, perceptions of reading are largely positive or neutral in both classes. For these students, reading is an enjoyable pastime. When asked whether he has historically enjoyed English class, Jason, Ms. Hatfield's student, responds, "[f]or the most part. Like, I never really enjoyed writing but that was…but, I've always been very into reading and stuff like that." For Abi, a student in Mr. Clark's class, the 20 minutes of silent reading provided during class was a positive experience: "I love to read. So I mean like having just like – we had like 20 minutes of reading before we started class throughout the year. So that was really nice. It was kind of like relaxing too before we started our day."

Where attitudes turn somewhat negative is with respect to analysis and teacher-selected texts. This is particularly true for Ms. Hatfield's students. Ben, a student of Ms. Hatfield, enjoys reading until he is asked to analyze it: "I've been kind of on the fence about English because, some things I really like, like, discussions and some reading can actually be really nice. But analyzing poetry and analyzing other written works and things of that nature is not something I particularly enjoy." Josie, his classmate, echoes this sentiment: "I mean poetry I like, by myself and like, by itself. But, I don't like, how we analyze it in English, because I'm a very literal person. So, like, all that figurative language and stuff just like, goes right over my head." Two students in this class express reduced interest in reading when the text is chosen for them. According to Shelby, "I think my least favourite would be class book or a classroom book that we're all reading because it's really hard to get into a book if you don't like it." This opinion sits in contrast with the same students' otherwise positive view of reading.

Perceptions of writing in both classes diverge, particularly with respect to the genres students prefer. The students interviewed in Mr. Clark's class express clear enthusiasm for creative writing. Both referenced Mr. Clark's creative writing assignment as an example of a task they enjoyed. For Colton, the exercise was positive because he was afforded a choice: "…I did my creative writing piece about [an image] and I really enjoyed that because I just did what I wanted to do." Abi describes her love for creative writing outside of school: "I've always had a passion for writing since I was a kid." The only other genre referenced in student interviews is essay writing. It seems that, for Colton, the volume of writing involved in essay writing is the issue: "I don't like essays but I mean I enjoy pretty much everything. I don't like writing a lot. I mean I don't like writing lots and lots of writing."

Conversely, in Ms. Hatfield's class, the students interviewed embrace essay writing. According to Josie, the prescribed structure of essays makes them easier to grasp: "I do like the essays. I mean, I don't like putting the effort into them. But, I like, that they have like, their format and there's not really…like, if you know what you're doing then you really can't mess up an essay." During classroom observation, many expressed a sense of confidence and comfort with the persuasive essay section of the NSE: English 10. These reactions are explored in greater detail in the next section. Another student expresses enthusiasm for all forms of writing: "…in general I love doing poetry and any kind of writing. I especially love free writing, writing your own story or if she gave me a topic it just - I really love fictional writing but I just enjoy nonfictional writing like essays and all that."

Perceptions of the Examination

The "Perceptions of Examination" code is the largest contributor of data, with 89 unique quotes and observations. In fact, the three child codes—positive/neutral statements, negative statements, and messages from peer/teachers—are themselves so large, and so much of the respective data overlaps, that numerous themes exist within and between them. What follows is therefore an overview of the emergent themes, which does not neatly follow the child codes, but rather gives an overview of the themes discovered within them.

Doubt and Negativity

By far, there are more negative comments made about the NSE: English 10 than positive. This is the case throughout teacher and student interviews, as well as classroom observations. Generally, student participants in Mr. Clark's class are more pessimistic about the Examination than students in Ms. Hatfield's class. Colton, for example, gives a snapshot of student perceptions in his class:

> Everybody's stressed out. Everybody's stressed out about them because a lot of people aren't doing good. Not a lot of people do well. I was surprised like I have friends and they're really smart. Now I'm doing better than them. And they're like, 'I'm worried about failing.'

Some of this stress, it seems, may be coming from a lack of feeling prepared. Both Colton and Abi in Mr. Clark's class expressed concern that they had not prepared enough in class. After being asked whether he felt prepared for the Examination, Colton responds, "[b]etween you and I, no. Because I don't think we did anything except for today. We didn't really do anything that would get us ready for the Examination. So I

don't know. I did some research last night and I'm trying to figure out what I could do." Colton's concern about a lack of preparedness was expressed after the first day of in-class Examination preparation.

Multiple-Choice Questions. In both classrooms, the multiple-choice questions become a source of stress during and following the completion of sample assessments. Students' frustration with these sections of the Examination is first made visible during classroom observation. At Oceanview Collegiate, students are vocal and argumentative, debating many of the answers. Negative reactions are loudest when the teacher reads the correct answer for the following question (C):

Figure 3

Sample Multiple-Choice Question from Exemplar

3. The phrase "discordant harmony" describes

a. both the sounds and the movements of the crowd.

b. peoples' screams as well as the music playing on the loudspeaker.

c. the noises made by the crowd as well as people's emotional state.

d. the similarities between the noise of the crowd and the music playing.

In answering this question, roughly half the students had chosen an incorrect answer. One shouts "Why?!". Question 9 elicits similar responses (the correct answer is D):

Figure 4

Sample Multiple-Choice Question from Exemplar (2)

3. Which word best captures the meaning of the title "Raging Waters" in the context of this story?

a. adventure

b. anger

c. danger

d. excitement

One student, having chosen A, asks, "[d]o I just get it wrong even though it's so close?" Another echoes this sentiment, asking whether they will have the opportunity to explain their answers in writing: "[t]here will be a section with short answer, right? Oh, no? So it's just multiple-choice and essay? That's stupid."

Aside from classroom observations, it is difficult to assess how students feel after practice assessments, since their interviews took place earlier in the semester. However, Ms. Hatfield's interview reveals negative perceptions about the multiple-choice questions in particular. She believes that many of the answers are too similar:

> [b]ut writing a multiple-choice test question takes skill. With my sociology backgrounds trust me when we design questionnaires and things of this nature, you have to have an obvious distractor but you can't have it so close that you could argue it's almost the same, you know.

She also believes there are too many multiple-choice questions, and that students should be afforded the opportunity to answer in writing: "[t]oo many multiple-choice questions. I much prefer my students to tell me what they think is going on."

The reaction to the multiple-choice sample assessments in Mr. Clark's classroom is similarly negative. Again, questions three and nine caused the most confusion and disagreement. However, Mr. Clark's students react differently than those in Ms. Hatfield's class. Rather than arguing, I record a number of dismissive comments such as "[t]hat's why I hate this kind of thing" and "I'm gonna do really bad on this exam (laughing)." Again, the students contributing to the discussion are generally the two who participated in student interviews. When the teacher asks a quieter student—a student who has not been participating at all—what he thought about a question, he responds with, "I don't want to talk about it," refusing to share his answer.

Students' discouragement is evident during interviews. When asked generally how she felt about the in-class practice assessments, Abi complains that the multiple-choice questions were too difficult: "I find [the multiple-choice questions are] opinionated in a way. So you could misinterpret it almost because the answers are pretty similar. That's the thing I don't like about it." Colton agrees that the multiple-choice questions are unclear: "You get all those different answers and some of them are like—I don't know. They're the same." Like Ms. Hatfield's students, Abi would prefer the option to explain her answer in writing, as in a short answer response: "Like I find it's more better if you'd be able to write down your opinion instead of choosing what the options they give you," and she feels that the multiple-choice questions are a less accurate representation of her abilities: "Yeah like because I do good at English but when we're doing these practice things like I didn't get a lot of them right."

Mr. Clark would prefer the opportunity to grade written responses. He admits that he talks "a lot" about multiple-choice questions, since his students are largely unfamiliar with the format. He believes the answers provided are unreasonably difficult:

> I've been going over some of the exemplars in myself questioning whether I'm choosing the correct answer. […] It's not all of them but there are certainly some examples where I just don't know it's the wording and how they're worded. I don't think they achieve anything.

When his students express defeatism during the in-class Examination preparation, Mr. Clark demonstrates sympathy. While students struggle and question themselves, he is warm and understanding, remarking "I can see why you thought that [was the correct answer]" and continually offering reassurance and advice about how to choose the most correct answer.

Examination Content. Even more unreasonable to Mr. Clark, however, are the texts chosen for the Examination. He feels strongly that students lack a connection to Examination content due to a lack of relevance. If he were to design his own exam, he explains, he would choose different texts that were connected to students' interests:

> I definitely would choose different text. I don't know. I don't like the notion of a standardized test. I feel like it goes against that—I mean everything that we're taught. I mean right now for the past really two or three years, cultural competency has been just a huge part of our professional development and I love the notion of it and I get it. But then you see these standardized tests, which to me kind of fly completely in the face of what we're being told we should be doing

> kind of trying to tailor stories and find stories that work for it depending on who you have in your class that semester.

Mr. Clark explains that many of the school's staff meetings and professional development days have emphasized the importance of cultural competence and making course content relevant to students, and he believes the Examination's standardized content contradicts this message.

Mr. Clark feels strongly about cultural relevance and student connection not only because of professional development days, but also because he believes it is linked to student engagement. The year prior to this study was unique in that provincial exams were cancelled during the work-to-rule, and Mr. Clark was able to administer his own Examination. The opportunity to create an exam using content that was familiar and relevant to the students made visible the lack of engagement students were previously experiencing with the Examination. He explains:

> Again, I don't know if it's because they knew that content but they still required I guess a little extra help during the actual exam but not as much because I think it's just – they were being tested on things that they were already kind of familiar with. They weren't being surprised where with the provincial like the entire thing is a surprise.

He sees a clear connection between student disengagement and a lack of connection to, or familiarity with, Examination content and format. Multiple times, he describes the Examination as "cold" and "removed" from the students: "whereas with the provincial, it's like this cold document that just is shipped in at the last minute. We give it and then

we ship it back out and it has nothing to do with really […] anything we've done throughout the semester."

The obvious drawback to a lack of student engagement is reduced Examination scores and, consequently, reduced grades. Mr. Clark claims that many of his students, historically, have chosen to put little effort into the Examination. He attributes students' lack of investment in Examination scores to a lack of connection to the Examination itself:

> To be totally honest, I know a lot of students and I don't know how if it's passed down from grade to grade but a lot of students will calculate what they need on the provincial to maintain a decent mark. So when I said I've had students fail, I had some of my most – some of my smartest students fail and they fail because they don't care about it.

Mr. Clark believes that some students feel less invested in the outcome when they know they cannot fail the course. This idea is somewhat confirmed by student comments during classroom observations. One student asks, "[i]f I completely fail this exam, how far down will my grade go?". During the next class, Mr. Clark explains that students with a grade above a certain threshold will not fail the course, regardless of their score on the Examination. A student responds, "[o]h, I'm golden. I'm good. I don't need to do this." A student in Ms. Hatfield's class echoes this sentiment, despite the fact that her interview took place prior to in-class Examination preparation. According to Josie, "some kids just don't care. Like, even kids who are smart and could do great on the exam don't care enough to do it. They just want to pass."

Recycling the Examination. At the time of research, the Ministry of Education had administered the same version of the Examination since its inception. Both teachers admonish this practice. Ms. Hatfield expresses frustration that her suggestions for content improvement have gone unanswered. Mr. Clark agrees with her: "…I've always said like just how hard can it be to just change the story? There are so many stories out there, just find another story and they don't. That's probably, yeah, I would say two or three years, it's been the exact same one." Although Mr. Clark reports that new content was provided for the 2018/19 school year, there is no data to suggest that Examination recycling will not continue.

Teacher Design vs. Provincial Design. Given the choice, most students express preference for an Examination prepared by their own teacher over the NSE: English 10. However, this preference is weighted heavily toward Mr. Clark's class, where negative perception of the Examination is most apparent. Mr. Clark's beliefs about weakened student engagement are reinforced by his students' stated preference for a locally-designed exam. Both of Mr. Clark's students believe an exam designed by Mr. Clark would be more familiar. Abi explains:

> I'd probably maybe feel more comfortable because like knowing like what his I guess – knowing how he teaches and knowing what he'd probably put on the exam I guess. They give you what the plan but they just kind of like throw at you like different ideas I guess like I'm more comfortable with somebody I know writing – doing the exam instead of like just complete strangers, in that kind of way, yeah.

Colton takes it a step further. He believes Mr. Clark's exam would be better suited for their class because Mr. Clark knows the students and is invested in their success: "I think the exam questions are all just so general like I think Mr. [Clark] would like know what our class is at and give an exam that would actually be good for us. So yeah, definitely, I'd rather Mr. [Clark] write the exam."

By contrast, roughly half of students interviewed in Ms. Hatfield's class express preference for the provincial Examination over one created by their own teacher. More precisely, two expressed preference, and one would be content with either. Unlike those at Eastern High, some students in Ms. Hatfield's class appeared to prefer to provincial Examination for its familiar format. Shelby explains, "I think I'd be fine with either but if I had to pick I'd probably pick the provincial exam because that's what we've been taking for years and that's what I've been expecting." Interestingly, two also cite confidence in the quality of the provincial Examinations as a reason they would be happy to take it. Ben states, "I think the provincial exam, to its credit, could be good generally, because it's what the Department of Education expects their high school students to know." Shelby feels "safer" taking the provincial Examination because she expects it has been thoroughly vetted:

> I think I'd feel safer with the provincial exam because I kind of have an idea of what it's going to be and I know that it's been checked many times. I know that's something they've been doing for years and so it's not something they're going to just put and stuff. It's going to be checked by many people and not just one person.

Her words contrast the opinions of both teachers regarding assessment quality. Perceptions of usage and quality will be further explored in "Fairness, quality, and use of results."

Confidence and Positive Outlook

Despite the strongly negative perceptions held by teachers and some students, others hold positive outlooks on the NSE: English 10—particularly in Ms. Hatfield's class. There, most students interviewed are not worried about their performance. They express confidence in their own abilities and preparedness. All of them hold high marks—the lowest mark cited during interviews is an 87. Ben explains, "…I'm not particularly worried about it because my mark is so high. So, it… even if it brings it down, it shouldn't bring it down to a very disappointing level for my standards." Jason is confident he has mastered the required skills: "…from what I heard, it's going to be just… it's going to be like, nothing based on specific things we've done in class and just general English skills, which, I think I'm fairly adept at the English language most of the time. So, I think I'm ready. Yeah." Even Colton in Mr. Clark's class, who had previously complained about pressure, expresses confidence that he will do at least "okay."

External Messages. Messages from peers and teachers regarding the NSE: English 10 are mixed. Largely, the students interviewed in Ms. Hatfield's class are being told that the Examination is simple and covers skills they have already mastered. Shelby's comment reflects the general attitude, "[w]e haven't really talked about it. The only thing I really know is you can't really study for it. There's nothing to study for it. Yeah, that's about it." Josie confirms this again, and she feels particularly confident about essay writing: "[f]rom what I've heard, it's just like, you have to write two essays and

answer some questions. And, all my friends that took it last semester told me that it's not a big deal. So, yeah, I feel prepared. Just cause like, essays are something that I can do."

Messages from teachers are minimal and generally reassuring. Students in both classes remark that little has been said about the provincial Examination. Colton states, "…they're trying to get us ready for the exam, which is nice but nothing really. To be honest, they haven't been talking about it as much as I'd though they'd be." Abi expands, "I haven't heard too much talk from teachers to be honest, really just [Mr. Clark] but he said if you just like try your best and your hardest that you should do fine with it." When asked what they would say to a student who was nervous about the Examination, both teachers emphasized the belief that the Examination is simple and that their students are generally capable of meeting the standard. Ms. Hatfield reminds students about their abilities and their experience with the genres, "[y]ou can read. You can write. We can have conversations as we've had in class. We've done a lot of practice writing through our essays. […] You're smart. You've done well. You have absolutely nothing to worry about." Mr. Clark would tell students "that it's pretty easy really. It's not that difficult, the multiple choice again can be difficult depending."

Prior Examination Experience

At least two students indicate their prior experience with standardized assessments could help prepare or enhance their confidence. Shelby details the provincial language assessments she took in grade eight. She recalls how "…it was my favourite exam to do. It was the easiest one." She remembers the format of the assessments and expects the format of the grade ten assessment to be similar:

> [y]es, the grade eight ones, the ones we had in elementary school. As long as they've been following the same kind of format but then if it's changed then maybe I won't be prepared but I feel very confident. […] They have usually a bit of writing and they ask you a few questions, it's usually multiple-choice. There's not a lot of short answers but there is a few later on. But it all—they give you a little splurge of writing and then you answer questions and at the very end it gives you a topic and an essay it wants you to write. And they give you about two pages, three pages or so.

Shelby's expectations reflect the format of the NSE: English 10. In contrast, Colton in Mr. Clark's class did not take the grade eight assessments, since he had previously attended an independent school. Colton feels unprepared due to his lack of experience: "I just don't know what to expect. I mean because I wasn't in the public school system last year and the year before that." Following one classroom observation, Mr. Clark expresses this same concern. He explains that some of the "feeder schools" for Eastern High choose not to administer the grade eight assessments, and that the students from these schools could be disadvantaged as a result.

Fairness, Quality, and Use of Results

The issues of fairness, assessment quality, and usage of results arose organically during most student and teacher interviews. Preconceptions about how results were being used and opinions about the quality of the Examination were a natural topic when discussing Examination content and student results. However, there were some differences in the presumptions shared by the different groups of students.

Generally, students in Ms. Hatfield's class express confidence that the provincial Examination is an accurate measure of what students need to know. They feel that the Examination represents a baseline of ability in language arts. During her interview, Shelby expresses the belief that the provincial Examination is useful in its ability to evaluate teachers:

> But I think the provincial exam in general is an important thing because it's just a good way for them to get statistics on how kids are doing. And if they see that tons of kids in this high school they didn't do well with this exam, then they know that something was going on with the teachers likely.

Shelby believes that the Examination can provide the Department of Education with valuable data about teaching quality. As discussed earlier, Ben feels the provincial Examination reflects "what the Department of Education expects their high school students to know."

Despite the validity some credit to the Examination, many students in both classes question whether one standalone assessment could accurately represent the entirety of their knowledge and skills. Colton and Abi from Mr. Clark's class express this concern, as do Ben, Kate, and Josie from Ms. Hatfield's. Stress and pressure, Colton explains, could negatively affect results: "[b]ecause sometimes just that amount of work and just the amount of time we have is just stressful and it just overloads you and you just can't do it." Kate worries that the province would only receive one snapshot of who she is as a person:

> [b]ecause, like, you could just have an off day. Like, everybody has off days. Like, maybe something happened the day before. And, you're like, not like, in the

> best frame of mind to write the exam. And, then, that's what goes on like, your Provincial exam. That that's what the Province knows of you.

Abi echoes Kate's sentiment in her own interview when she questions the province's ability to draw accurate conclusions about student ability:

> ...I don't get nervous for exams but say if I was really nervous… They could think that a person like if they didn't do well on the exam that they didn't do well throughout the year.

Josie believes it would be dangerous to use the Examination to evaluate teachers, since it fails to capture the student's work throughout the term. She observes that "…other kids who learned a lot, it may not show on the exam, because, it's not like, a periodic assessment, it's just a final assessment. So, I don't think it's a great way to judge like, teachers across the province and how the students are learning." Josie also raises the question of personal relevance of results. She agrees that the results hold some significance, in that they are included in a student's final grade. However, she recalls that "in junior high and elementary […] they mailed us our results like, months later. So, nobody cared anymore. Nobody even remembered what they did."

Teachers express similar concern about the use of results to evaluate educators. Mr. Clark is unsure whether this is happening, but he would "…hate to think that my effectiveness as a teacher is being judged on that exam." He criticizes Examination content in particular, suggesting that he "would need it to be much – a very different exam if you're going to judge me on it…" That said, Mr. Clark is unsure how or if results are being used at all. He suspects his department head may receive them, but has never sought them out:

> I have no idea. I know I think the school might get something back. I never get anything back. Once those exams leave my room, I've never seen or heard anything about them again mark wise or anything. I think my department head might get some of that information and I'm sure I guess if I had the time, I could go investigate and see but then it becomes I just don't have time.

The fact that Mr. Clark has never sought out the results indicates he does not have use for them personally. It is unknown whether his department head receives them.

Ms. Hatfield believes the results are, in fact, being used to evaluate teachers, and she believes results are an unfair measure of teaching quality:

> [They think] teachers aren't doing their job if the results are poor as opposed to thinking that the test was a poor instrument, you know? There's a real disconnect and I say even a disrespect of teachers at the DOE/the board, which is now nameless and I don't know why because they were all teachers most of them.

The topic of assessment fairness leads to a discussion of oversight generally, and Ms. Hatfield complains, "[t]hey don't come in. They don't talk to us. We're only addressed when there's some kind of problem perhaps with a student and the parent has called the board […]. But they're not seeing us and what's going on."

Ms. Hatfield also questions assessment validity. She believes it is dangerous to make general assumptions about the entire province based on one assessment:

> Because if you're going to be - if people are seeing this as a way of taking all the results of the test, all the grade 10 results across the province and see how many passed and what sections the kids passed and failed from the text to the poem to

> the essay, that might tell them about the students. We like to think it tells them about the students but look at the different school boards they're coming from, look at the different schools they're coming from. They're so different and yet you're giving a standardized test that doesn't allow for diversity, that's a big problem.

She believes there are significant differences between schools and districts and that these differences would affect test results. Simply put, in Ms. Hatfield's view, the content of the Examination is not appropriately relevant for students across all jurisdictions.

Locally-Designed Assessments in 2017/2018

In the 2017/18 school year, teachers enacted a work-to-rule and provincial assessments were deemed optional by the Ministry of Education. Given the negative perceptions discussed above, it is perhaps unsurprising that both teachers in this study chose not to administer the NSE: English 10. The format and content of the assessments administered in their place provide a unique window into the structures and genres valued by each teacher. The theoretical question, "How would you design your own exam?" is made literal, as these teachers were able to do exactly that.

For Mr. Clark, final assessments in 2017/18 were designed to reflect the values he describes above. Although his final assessment mirrored aspects of the provincial—insofar as students were expected to read certain texts and examine a visual—the most dramatic difference lies in the familiarity of texts. Students had either previously discussed the texts in class, or they were connected thematically to something they had examined together. Mr. Clark believes that students were more engaged as a result:

> "…structurally, [my exam was] kind of similar to the provincial where they had to read certain texts or look at a visual but again, it was the connection. So instead of writing about some text that they're seeing for the first time, they were maybe writing something about something that we've read together, we discussed together that they—you could just tell in their answers that they cared because it's something that we had done. It's something that we had already kind of discussed. And it's still the exam is always getting them to look at it in a new way, I guess in a way that we haven't done in class, which is how I guess you evaluate but they just care, whereas with the provincial, there's certainly a notion of "I just don't care."

Mr. Clark was also able to incorporate a listening component to his final assessment. Part of the exam involved viewing a short YouTube video and answering questions about what they had seen and heard. He was therefore able to integrate, at least partially, the "Speaking and Listening" strand of the curriculum. In contrasting his own approach, Mr. Clark criticizes the provincial Examination for its singular focus: "with the provincial I think that entire strand that's one of the main strands that we're supposed to be—[Interviewer: Giving equal weight?] Yeah, that just kind of gets left to the wayside."

Ms. Hatfield admits that she had little to do with the design of the assessment in 2017/2018. Since she only teaches one grade ten course, the other grade ten teachers created an assessment that was shared amongst them. Again, the format of their locally-designed assessment was similar in format to the provincial Examination insofar as it included reading comprehension and persuasive writing. But unlike Mr. Clark's exam, Ms. Hatfield's included texts that were unfamiliar to students in content and theme: "But

I think because the exam we designed is to the same extent to the provincial exam it's text based, skill based, it's not based on what we've taught and thought about and wrote about during the semester." However, she believes the texts and writing prompts chosen were more connected thematically: "[w]here you try - what we do is kind of a persuasive essay that is in some way linked, because there's often a thread in what the text we've picked, that there's a link that now we're going to ask them to make a judgment based on what they've just read." She impugns the NSE: English 10 content for lacking a common theme, and for its reliance on texts that seem "trivial."

Connecting the Themes

The data presented above represents a small sampling of the interactions and literacy work happening in each classroom. Two to three classroom observations, single interviews, and a few select documents can hardly paint a thorough portrait. However, the goal of a case report is to become a "portrayal of a situation" (Lincoln & Guba, p. 214). Rich description provides an opportunity to examine the nuances of each case, and the comparison of coded data within the two classrooms studied raises important questions. Namely, questions emerge about how material is presented, the ways students respond to teachers' expectations, and about how students perceive the Examination, both in terms of its importance and the usefulness of the data produced. These questions act as guide for the conclusions that will be discussed in the next chapter.

CHAPTER 8: DISCUSSION OF FINDINGS

The close examination of two specific cases provides a clearer understanding of how these particular students and teachers experience the English 10 course and provincial examination. The findings in Chapter 7 have provided answers to the first two research questions:

i. How do students and teachers describe classroom practices?

ii. How do students and teachers describe the evaluation?

Some degree of transferability can be assumed from this small sampling, particularly when considering the range of perspectives between the two classrooms. The classroom cases share a common grade level and school district, but they present notable differences with respect to teaching style, course content, student participation levels, and attitudes toward the Examination. Some of the more salient differences are made visible below:

Figure 5

Significant Codes Distributed by School Name

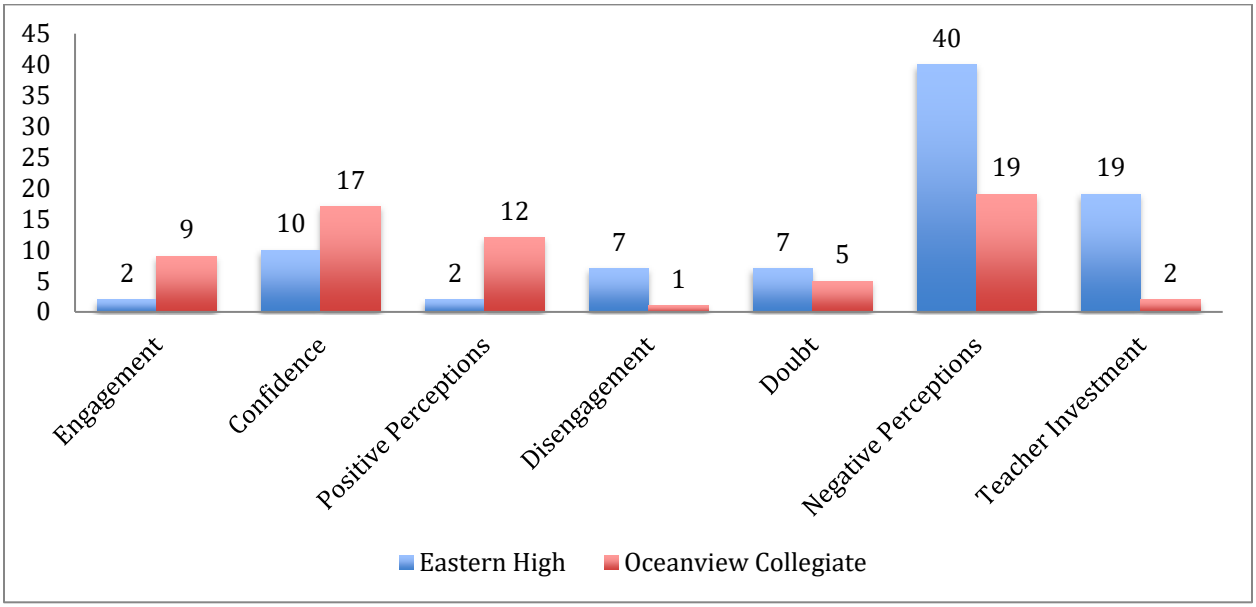


Figure 5 reflects the distribution or usage of listed codes by school name. In other words, the chart visualizes the number of times each code applies within the sources of data from each school. These numbers should not be considered a complete picture of attitudes within the class, as they are not gathered from a survey of all students and teachers. Rather, they constitute a snapshot of the attitudes within each of the two classrooms.

The first three codes connect with the next three codes on a spectrum from positive to negative: The inverse of engagement is disengagement, the inverse of confidence is doubt, and the inverse of positive perceptions is negative perceptions. Generally, Eastern High presents a greater percentage of "negative" codes. Analysis reveals lower observed evidence of engagement and students that are more likely to express self-doubt than self-confidence with respect to their abilities. Negative statements about the provincial Examination are noted at both schools—significantly more frequently than positive statements—but they are verbalized twice as frequently at Eastern High. By contrast, data sets from Oceanview Collegiate are decidedly more positive overall. Observed student engagement is predominant, and student interview participants are more likely to express self-confidence with respect to their own abilities. Although some student interview participants hold negative views of the Examination, students at Oceanview Collegiate are more likely to express positive or neutral attitudes than their peers at Eastern High. Finally, coding reveals higher levels of "investment," or concern, from the teacher at Eastern High. This captures when a teacher worries about his or her students, or when he or she make explicit and observed efforts to generate student

engagement and enthusiasm. In other words, teacher investment seems to inversely correlate with student engagement within this study.

Immediately, the range between cases becomes the most pronounced feature in the data. Other impressions are more obvious; it seems only rational that disengaged students would be more likely to express self-doubt, or that their teacher may strive harder to capture their interest. Likewise, it is logical that a teacher of highly engaged students is less concerned about poor outcomes. However, the most striking contrast between these two classrooms is the variance of the case subjects themselves—how the participants experience the assessment differently, and how this informs the validity of the Examination. The focus, then, moves away from the cases as singularities, and toward the distinctions between students' and teachers' experiences with the Examination. The ways in which these differences interact with the first two research questions inform our answer to the third: Do any differences between these accounts pose a threat to the inferences drawn from test results?

Conclusions

The open-ended, exploratory nature of this study yielded insight into multiple facets of validity. The conclusions that follow offer evidence of three problems—namely, (1) threats to construct validity, (2) unintended (adverse) consequences, and (3) construct irrelevant variance. Many of these conclusions present more than one threat to assessment validity. The following chart visualizes the overlap. The chart also demonstrates the significant potential of the Examination to contribute adverse consequences, as this category is the most strongly represented among the conclusions.

Table 9

Overlap of Validity Issues Within Conclusions

	Threats to Construct Validity	Unintended Consequences	Construct Irrelevant Variance
Impact on Program Assessment	✓		
Limiting of Critical Engagement	✓	✓	
Influence on Student Perceptions		✓	
Standardized Scoring Rubrics	✓	✓	
Influence on Motivation	✓	✓	✓
Prior Assessment Experience			✓
Restriction of Test Design	✓	✓	
Lack of Transparency		✓	
Lack of Validity Inquiry		✓	

Impact on Program Assessment

The Examination may contribute to construct underrepresentation within the larger program. The Program of Learning and Assessment for Nova Scotia (PLANS) admits its own limitations. PLANS does not claim to evaluate the entirety of the English 10 curriculum, let alone the larger construct of literacy. According to PLANS, the NSE: English 10 evaluates "*selected* curriculum outcomes" (emphasis added) (Government of Nova Scotia, 2017b). This is further qualified on the PLANS website, which explains that

only “elements of the [listed] curriculum outcomes that can be measured on large-scale assessments will be included” (Government of Nova Scotia, 2019a). Of the 40 listed curricular outcomes, the NSE: English 10 evaluates elements of ten (See Tables 7 and 8 for a visual representation). In other words, 20% of a student’s grade depends on elements of 50% of the curriculum. As explored in the previous chapter, many of the outcomes the Examination intends to measure are, at most, incompletely represented, narrowing the scope of the Examination even further. Such narrowing of outcomes creates two potential issues.

The first implication of a narrow evaluation is the potential for imbalanced assessment. When the Examination is restricted to 50% of curricular outcomes, a gap is left. To fill this gap, teachers would need to place greater emphasis within their own assessments on the constructs left absent on the Examination. The most obvious shortfall is the Speaking and Listening strand, which should account for 20 to 30% of a student’s final grade. Other absent constructs include the writing process, the consideration of alternative interpretations of text, illustration and other ways of presenting knowledge, and the use of technology. The case studies present little evidence of these types of assessments. Course outlines give some attention to speaking and listening, but few related assessments are either listed in course documents or referenced during interviews. By and large, the assessments referenced include essays, novel studies, and poetry analysis—such assessments were identified and coded far more frequently within data sources. In other words, what receive emphasis in the program are the same constructs measured by the Examination.

Given the constraints of the Examination and the data gathered from case studies, it is difficult to accept that all teachers will sufficiently fill the gap left by the Examination—or indeed, that they should be required to. In this case, a student's final grade would not represent his or her success with respect to the entire curriculum. The Examination itself may measure the constructs it sets out to measure. However, Messick (1989) contends that validity lies with interpretation, not the assessment itself. The validity argument here lies with the interpretation of the student's grade itself. In other words, how accurately does a student's grade reflect his or her mastery of all course outcomes? If, as we see in these two case studies, a student's grade is largely gathered from 50% of the curriculum, a student's final grade may not be interpreted to be an accurate measure of all curricular outcomes.

The second implication of a narrow final assessment involves the narrowing of course content. The restriction of curriculum prior to a standardized assessment is a phenomenon observed in other Canadian studies. Skerrett and Hargreaves (2008) found that the OSSLT narrowed the concept of literacy within course content and worked to reinforce outdated teaching methods. In a later study, Skerrett (2010) again found the OSSLT to be narrowing curriculum, despite an acknowledgment on the part of teachers that such a narrow view of literacy was potentially harmful to students. The two cases in this study demonstrate a similar restriction of course content. Evidence gleaned from data sources include Ms. Hatfield's emphasis on persuasive essay writing, Mr. Clark's explicit teaching of answering multiple-choice questions, their shared focus on traditional narrative texts, and a deficiency of tasks or assessments involving alternative forms of

communication. A focus on traditional genres is particularly visible in Ms. Hatfield's classroom.

The larger program's constructs are further narrowed in the *Lessons Learned* document (Government of Nova Scotia, 2019c). Published in response to test items posing the most difficulty, the document offers instruction to support students in the development of skills assessed by the Examination. Arguably, the document asks teachers to devote more class time to discussing the Examination. At many points, the document suggests posting sample written responses from a previous iteration of the NSE: English 10 so students may "compare their writing to the samples on the wall and identify, either in a conference or in a journal, what conventions they should focus on to improve their writing" (p. 16, 19, 22, 26). Teachers are also encouraged to model reading and responding to these texts (p. 16, 19, 22, 26). These recommendations, coupled with evidence gathered through observation and document review, suggest the Examination itself narrows the scope of literacy practices in the classroom. In emphasizing the same constructs evaluated by the Examination time and again, the English 10 program presents a narrow view of literacy and of the Nova Scotia English 10 curriculum.

Limiting of Critical Engagement

Taking this line of inquiry a step further, the Examination may limit critical literacies, both within the frame of the assessment and within the larger program. Data from this study suggests students are engaging with the texts during Examination preparation in a superficial way. Field notes observe a focus on multiple-choice question and a general lack of student engagement in both classes. Student engagement with texts seems especially poor within the data from Eastern High. These observations run

contrary to what the province suggests in guiding documents. The *Foundation for the Atlantic Canada English Language Arts Curriculum* explains that students "need opportunities to critically examine different experiences and perspectives within social and cultural contexts" (Nova Scotia Education and Culture, 1997, p. 42). *Teaching in Action* asserts that "to be successful and in the larger world, students require multiple literacies. These include, but are not limited to, critical literacy, media literacy…" (Government of Nova Scotia, 2018, p. 54).

To a certain extent, teachers may be limited, as suggested by Mr. Clark, by the texts available to them. Whether the Examination itself can be held responsible for the lack of engagement seen elsewhere in the program cannot be answered. As studies have demonstrated in other provinces, however, standardized assessments can influence classroom practices in fundamental ways (Kearns, 2011; Skerrett & Hargreaves, 2008; Skerrett, 2010). It is also difficult to accept that a relatively high-stakes final assessment would not have some influence on a student's overall perception of literacy and on his or her relationship with texts, as will be explored in the next section.

Influence on Student Perceptions

Construct underrepresentation may influence students' perceptions of literacy. Although the Nova Scotia English Language Arts 10 curriculum and its guiding documents present a reasonably comprehensive view of literacy, the genres that receive the most emphasis within lessons and assessments likely shape the way students view literacy. Universally, students involved in the study express a belief that the genres represented on the Examination are "most important." When asked which concepts were most emphasized by their English teachers, most students cite the very genres

emphasized by the Examination—namely, essay writing, grammar, and structure. Many express a preference for standardized assessment in general, citing its familiar structure and simple writing expectations.

Such comments suggest that frequent standardized assessments may shape student perceptions of literacy. The New London Group (Cazden et al., 1996) warn that standardizing language education could similarly restrict the construct of literacy. They underscore the increasing saliency of linguistic diversity, as well as the importance of new "technologies of meaning." We may wish to ask whether the limitations of standardized assessment may contribute to an overall deficiency in language teaching. It is reasonable to accept that students would expect a final exam to reflect the most important learning outcomes of course. This is especially true if said exam is being used to make policy decisions, as is the case with the NSE: English 10. If students are being led to believe that the five-paragraph essay is the foundation for all types of writing, or that one's understanding of a text can be captured by multiple choice questions, we may be limiting students' potential with respect to other means of communication.

Standardized Scoring Rubrics

Scoring rubrics may contribute to a narrowing of writing constructs. The NSE: English 10 uses two scoring rubrics, which are identical save the addition of "Quality of Text Comprehension" for the Reading Response task. This rubric has been included in Appendix D. Slomp et al. (2014) discuss the effort on the part of test designers to create rubrics that promote inter-rater reliability and consistency: "…matters of organization, correctness, and choice are often emphasized because the criteria can be scored more consistently" (p. 283). This appears to be the case with the NSE: English 10's scoring

rubrics, which evaluate only the development of a main idea, organization, transitions, language use, and conventions (Government of Nova Scotia, 2019h). This emphasis on easily graded elements of writing, however, comes at the expense of others: "More complex aspects of the construct, such as creativity, critical thinking, or metacognition, however, are underrepresented by many rubrics" (Slomp et al., 2014, p. 283). Indeed, provincial Examination rubrics give no mention to creativity, critical thinking, or other complex constructs. Student interviews suggest a connection between scoring rubrics and students' perceptions of writing; many cited a preference for the type of essay writing on the Examination due to its formulaic and "simple" nature. Considering these comments, along with official recommendations that students and teachers pay to Examination scoring rubrics prior to the assessment (Government of Nova Scotia, 2019c), it is easy to see how scoring rubrics could work to narrow perceptions of literacy or, at minimum, of writing constructs.

Influence on Motivation

The format and content of the Examination may negatively affect motivation. The question of why some students expressed a strong desire to succeed when others adopted a more apathetic attitude toward the Examination raises the issue of motivation. Much of the current research around motivation agrees that student motivation is influenced by multiple factors, including boredom, difficulty, perceived relevance, consequences, and self-efficacy (Finn, 2015; Linnenbrink & Pintrick, 2002; Wigfield et al., 2002). Indeed, the data sets from this study reflect a number of these problematic factors.

Data sources from this study suggest some students do not care about their Examination results and, in some cases, are choosing to fail. Students at Eastern High

asked about the minimum score they would require, and one commented that he was "golden" whatever the outcome, since his grade was high enough to allow him to pass the course. During his interview, Mr. Clark recalls previous students who had deliberately failed the Examination. A student from Oceanview Collegiate claims to know students who have chosen to fail the Examination in the past. The most obvious source for this apathy would be a lack of consequences, as the students mentioned above believe they will ultimately pass the course. However, data suggests other factors may be at play—namely, relevance, perceived difficulty, and self-efficacy.

Firstly, a lack of cultural relevance or linguistic congruence may be one reason students are less engaged with the Examination. During interviews, both teachers express a desire for the selection of more culturally responsive texts on the provincial Examination. Teachers find the present offering to be largely reflective of the culture and age of the Examination's designers, rather than the culture and age of the test takers. Both teachers speak to the importance of cultural relevance within their program, but Mr. Clark expresses greater frustration. He expresses the concern that teachers are expected to make these efforts, yet provincial assessments are not held to account. The document "Nova Scotia Examinations: English 10 – Lessons Learned" explicitly underscores the importance of cultural relevance within assessment. The document states,

> [i]nstruction and assessment practices should be culturally responsive. Culturally responsive pedagogy is teaching that connects a student's social, cultural, family, or language background to what the student is learning; it nurtures cultural uniqueness and responds by creating conditions in which the student's

> learning is enhanced. It is critical that learning opportunities are relevant and meaningful to students, so they are responsive to students' learning needs. (Government of Nova Scotia, 2019c, p. 78).

The document goes on to encourage teachers to "adapt" the contents of the document "in order to respond to students' various cultural and life contexts" (p. 78). PLANS expects assessment and learning to be culturally relevant to all students, yet fails to acknowledge the contradiction inherent in this expectation. In so doing, PLANS impugns the validity of its own assessment. A lack of cultural relevance for students of diverse backgrounds may pose a threat to the consequential validity argument if it can be shown to reduce test scores among these groups.

Furthermore, students in Mr. Clark's class react with frustration to a number of words and expressions on the practice assessments. "[L]ike finding a golden egg," for example, is widely misunderstood by students, as are the words "undulating" and "akimbo." The strongest reaction in both classes, however, is to a question about the meaning of "discordant harmony." According to Ryan & Whitman (2013), "sociolinguistic patterns can impact positively or negatively on students' motivation and performance depending on the degree of congruence or compatibility reflected in assessment practices" (p. 174). A number of factors could be at play with respect to sociolinguistic differences—culture, background, or maternal languages, for example. Although I cannot point directly to one of these factors within this study due to a lack of data, the fact that students are unfamiliar with so much of the language in the assessment suggests construct irrelevant variance.

Secondly, the degree of difficulty presented within practice assessments may challenge students' self-efficacy. Motivation theory generally accepts that assessments encourage intrinsic motivation when they are moderate—not to easy or too difficult (Finn, 2015, p. 7). Students and teachers in both classrooms express strong frustration with the level of difficulty and ambiguity presented in the multiple-choice questions. In fact, both teachers admit difficulty in selecting the correct response. However, Mr. Clark's students generally express a defeatist attitude in response to the questions, while Ms. Hatfield's students become argumentative. This discrepancy suggests that students with higher levels of self-efficacy entering an assessment may be less discouraged by overly difficult questions. The students in Mr. Clark's class, including those who had previously expressed apathy about their score and negativity about their chances of success, exhibited a loss of confidence during practice assessments. Considering the negative effect low self-efficacy can have on motivation, the difficulty and ambiguity of the Examination threatens the validity of the scores produced by the Examination.

If students are choosing to fail or investing less effort, results will not represent an accurate portrait of student ability. The Department of Education could be drawing the wrong conclusions about certain schools or districts; even a small percentage of students choosing to fail the Examination would dramatically affect overall results. This possibility directly challenges the Examination's construct validity as defined by both Messick (1989) and Kane (2006), and it further challenges the appropriateness of using results to make any decisions about the education system or student performance.

The factors that may be causing a discrepancy in motivation—relevance, difficulty, and self-efficacy—need to be explored further for unintended consequences

for two reasons. Firstly, a reduction in motivation for certain groups of students will almost certainly lead to lower scores within these groups. These lower scores will inevitably, and inaccurately, suggest so-called "achievement gaps," and the reporting of these discrepancies may lead to negative assumptions about groups of students within the larger public. Secondly, reduced motivation may be a symptom of a larger issue. Students in both classes demonstrated their discouragement during practice assessments. When discussing the Examination, many of Mr. Clark's students made jokes and comments about failing. It is possible that some of the comments and jokes—perhaps even the choice to fail—are a form of discourse. What appears to be ambivalence on the part of students may be, in fact, a reaction to a poorly designed and delivered test. Test designers should consider the possibility that students are choosing to fail because they simply do not connect with the Examination for the reasons presented above.

Prior Assessment Experience

Prior experience with Nova Scotia's provincial assessments may influence outcomes. The contrasting accounts of Shelby and Colton raise the possibility that prior assessment experience may benefit students writing the provincial Examination. Shelby, in Ms. Hatfield's classroom, feel confident that the Examination will resemble provincial assessments she has taken in the past, and she accurately describes the format of the assessment from memory. She explains that she was comfortable with this format and, consequently, not worried. By contrast, Colton, in Mr. Clark's classroom, has never taken a provincial assessment before and expresses a high level of concern on multiple occasions. Mr. Clark himself cites the concern that some of his students have come from schools where provincial assessments are not administered and are therefore less

prepared. In their study of L1 and L2 test takers' experience with the OSSLT, Fox & Cheng (2007) found that students with test-taking knowledge held an advantage over their peers without. Although all students interviewed for this study spoke English as a first language, prior test-taking experience may play a similar role in levels of confidence and, potentially, in student outcomes.

Restriction of Test Design

The exclusion of teachers from test design may produce unintended consequences. The way teachers approach and discuss the Examination may contribute to student perceptions. As Ms. Hatfield surmised, "I often wonder if it's the attitude the teachers have about the provincial Examination, it kind of sets them up, you know, in a way. They have an attitude about it." During observations, both classroom teachers used polarized terms like "us" and "them" when discussing the provincial Examination. Both teachers also apologized—albeit in a humourous way—for the perceived unfairness of the Examination and reminded students that they did not write it themselves. Mr. Clark laments the "coldness" and unfamiliarity of the provincial Examination. The language used by teachers may be reinforcing some of the disengagement their students feel.

Again, this phenomenon could have negative implications for student outcomes. If students are less engaged with an Examination because of the way their teacher describes it, we can consider the Examination to possess an unintended consequence. Taking the argument further, however, one can ask why teachers feel so disconnected from the Examination in the first place. The primary reason offered by Ms. Hatfield and Mr. Clark is a lack of teacher input. Both teachers felt unheard by the Department of Education. Although some teachers were consulted during the initial design process, the

stagnation of Examination content proves little to no meaningful conversation with teachers in the years since. Slomp, Corrigan, & Sugimoto's (2014) framework includes the examination of test design as a factor that may impact consequential validity. According to the authors, "a more inclusive design process should lead to increased support from a range of stakeholder groups, reduce the potential for bias, and generate more appropriate assessment decisions" (p. 282). In this study, the lack of teacher input in assessment design may contribute a negative influence on teacher perceptions of the Examination. Arguably, these negative perceptions are being passed along in some measure to students.

Lack of Transparency

Confusion about the dissemination and use of results could be contributing to misinformation, mistrust, and a loss of respect for teachers. All student and teacher participants express a lack of awareness about the province's use of results. Teachers suspect results are used to evaluate educators or entire schools. Students hold various ideas, including teacher evaluation and the comparison of schools. Generally, however, a lack of certainty and a sense of resignation pervade; students and teachers are simply unsure as to why the exam takes place at all.

The lack of communication regarding usage of results directly challenges the objectives listed by PLANS:

- develop and administer program assessments to determine the effectiveness of curriculum delivery
- develop and administer student assessments to assist students to achieve outcomes

- provide student achievement information to government for education decision making
- help teachers understand assessment principles and practices
- publish accountability reports for all assessments and examinations, both for teachers and for the general public. (Government of Nova Scotia, 2019a)

If PLANS is not using results to support teachers, many of these objectives are immediately discredited. The government of Nova Scotia cannot claim to "assist students to achieve outcomes," or to "help teachers understand assessment principles and practices" without effort and communication to this end. Moreover, if teachers themselves are unaware of where and why results are published, it only follows that the general public would be left equally blind. Accountability must suffer as a result.

Perhaps more importantly, a lack of communication about the assessment tool's purpose or benefits can contribute to a public perception that schools and teachers require accountability measures. The provincial Examination represents a transfer of autonomy from teachers to the provincial government, similar to what Mya Poe (2008) warns against in her own writing on NCLB legislation. This transfer communicates a message of distrust—a message that the government can design a better assessment. Student interviews reveal some evidence of support for "monitoring" or "measuring" schools and teachers. The unintended consequences of the Examination may therefore include public (including student) mistrust of the system and teachers specifically.

Lack of Validity Inquiry

A lack of validity inquiry may reinforce power structures. Inoue (2009) asks, "How is race constructed or conceptualized in our writing assessments, and in our

validity inquiries?" and "How do our purposes for assessment and the decisions we make from them affect different groups, or form particular racial formations?" (p. 110). In the case of the NSE: English 10, a lack of critical engagement with disaggregated results prevents us from answering these questions. Document review revealed little inquiry into the sources of discrepancies. Rather, official publications have accepted these discrepancies as evidence for policy change. The *Action Plan for Education* (Government of Nova Scotia, 2015a) calls for the implementation of an Achievement Gap Initiative, and funds have since been allocated for research into this area (Inter-university Research Research Network Application for Funding, 2019). The *Raise the Bar* report (Glaze, 2018) cites "achievement gaps" as justification for further disaggregation and the implementation of a new standardized literacy assessment similar to Ontario's OSSLT. The assumption that the problem lies with students, and not with the assessments themselves, is implicit within these decisions.

The lack of discourse surrounding the validity of the assessment itself is surprising considering the data produced within this study. Student reactions to the questions within reading tasks are strongly negative; many express anger, frustration, and discouragement. Teacher reactions are similarly negative, with one teacher apologizing and reminding students that he did not write the questions himself. Ms. Hatfield's claim that she, along with other English teachers at her school, often accept "wrong" answers calls into question the validity of the answers. These observations directly challenge the suggestion that the assessment itself is valid.

It is, of course, possible that some form of validity inquiry beyond field-testing is happening. However, very little information is publicly accessible on the topic. Nor is

there literature available that speaks to any decision-making that comes from the Examination, besides the claim that "[r]esults from the provincial scoring session are used to provide information to policy makers on the implementation of the course curriculum and on standards of student achievement in relation to expected learning outcomes" (Government of Nova Scotia, 2015b). Without more information, an analysis of the Examination's possible influence on racial formations or power structures cannot take place.

CHAPTER 9: RECOMMENDATIONS, LIMITATIONS, AND AREAS OF FUTURE STUDY

This final chapter presents a number of recommendations arising from the Conclusions in Chapter 8. These recommendations are made in response to the personal experiences of the students and teachers within the two cases being studied. The conclusions in Chapter 8 cannot, given the small scope of the study, be considered conclusive for the entirety of the province. As such, areas of further research are also suggested to enlarge the scope of this study and to better understand the complex system in which these assessments take place.

Recommendations

Field test Multiple-Choice Questions Offered on Practice Assessments

The PLANS website claims questions are field-tested before inclusion on the Examination. These questions are then analyzed, and items that "match provincial standards are approved for inclusion on examinations" (Government of Nova Scotia, 2017b). PLANS does not explain whether the questions on practice assessments are field-tested in the same way, if at all. Data gathered within this study suggests many of these questions are confusing, ambiguous, and needlessly frustrating for students. Both teachers admit to being unable to select correct responses themselves. More importantly, ambiguous and confusing questions may be reducing students' self-efficacy and thereby threatening motivation. Field-testing these questions ahead of publication may reveal such problems, allowing questions to be eliminated or clarified ahead of student use.

Review the Inclusion of Multiple-Choice Questions

Of course, doing away with multiple-choice questions entirely and allowing students to answer in writing would arguably be a simpler and more effective strategy than field-testing. Multiple-choice questions comprise a significant portion of the Examination, and they were the largest contributor of frustration and otherwise negative comments during the study. Teachers expressed nearly equal frustration with grading these questions, with one teacher going so far as to grade "incorrect" answers as correct when they are particularly ambiguous. Asking students to answer questions in writing may not only prevent student frustration, but also provide evaluators with a more accurate picture of student comprehension.

In the *Lessons Learned* document (Government of Nova Scotia, 2019c), the issue of student confusion with respect to these "selected answer" questions is portrayed as a lack of skills on the part of students. In response, teachers are encouraged to flood the room "with vocabulary rich talk" and "explicitly teach dictionary use" (p. 10). The message embedded within this recommendation is that the problem lies with the students, and not with the assessment tool. The possibility that students may have been able to correctly respond to these difficult questions had they answered in writing is not acknowledged, although both teacher participants expressed this opinion in their interviews. The outcomes allegedly measured by these reading comprehension questions—4.3 seek meaning in reading, using a variety of strategies, for example—may be clouded by one's ability to deduce the most correct answer. In other words, the Examination may measure skills external to the constructs it claims to measure. The multiple-choice questions therefore pose threats to all three sources of validity, since (1)

the assessment tool may not accurately measure the proposed construct, (2) the assessment tool may measure test-taking skills external to the constructs listed, and (3) the assessment tool may have the unintended consequence of challenging students' self-efficacy and reducing motivation. Undoubtedly, there are benefits to assessing reading comprehension in this way, such as efficiency in grading and eliminating subjectivity in the evaluator. However, these benefits must be weighed in comparison to the numerous potential drawbacks (Kane 2013).

Review the Relevance of Texts According to Provincial Standards

Both teachers expressed concern that the texts chosen for the Examination were disconnected from their program, irrelevant to students' personal lives, and otherwise poorly chosen. In particular, Mr. Clark criticized the selection of texts on the Examination, as he believed their lack of cultural relevance contributed to student disengagement and, consequently, to poor results. The possibility that a lack of cultural relevance may be contributing to poor results for students from marginalized communities is harshly out of line with provincial policy around cultural responsiveness.

The *Lessons Learned* document (Government of Nova Scotia, 2019c) document emphasizes the importance of cultural responsiveness. The first point made in the "Key messages in instruction and assessment" section states, "[i]nstruction and assessment practices should be culturally responsive." It goes on to suggest, "[t]he content of this document may be adapted, as needed, in order to respond to students' various cultural and life contexts" (p. 76). The *Nova Scotia Action Plan for Education* itself emphasizes cultural relevance within teaching and assessment in one of its pillars (Government of Nova Scotia, 2015a, p. 18). The generic nature of the assessment conflicts with these

messages. Using the provincial standards outlining curriculum delivery, a review of chosen texts may be useful in identifying cultural representation gaps. Providing opportunities for student and teacher feedback could also highlight issues of irrelevance, as outlined in the next recommendation.

Collect and Review Teacher Feedback

The teachers involved in this study feel unheard and powerless regarding multiple issues. Their lack of agency with respect to the Examination leads to greater frustration, which could be passed onto their students. In both classrooms, teachers employed a language of "us" versus "them" when discussing the Examination with students. During observations, teachers spoke in apologetic terms about the Examination, occasionally reminding students that they took no part in its design. This reinforcement of the disconnect students already feel from the Examination may discourage and disengage students even further. Providing a platform from which to gather teacher feedback may reduce some of the disconnect leading to this approach. If teachers feel their voices have been heard, and that their concerns have helped reshape the Examination, they may be more likely to "buy in" to the assessment.

Furthermore, teachers will likely contribute useful and valid ideas. Both teachers involved in this study raised issues with respect to content validity, cultural relevance, and use of results. They expressed frustration with the Examination being "recycled" with a lack of cultural relevance among the texts chosen, and with the ambiguity of many multiple-choice questions. Their concerns were well-informed by years of experience and first-hand knowledge of the context in which the Examination is delivered. Collecting and analyzing teacher feedback on an annual basis may help PLANS to identify issues in

a more efficient and timely manner. Among the nine sources of validity offered by Slomp, Corrigan, & Sugimoto (2014) is design process. The inclusion of teachers in the design process via feedback may be of use to both the content and consequential validity of this particular assessment. Another positive consequence of such an approach is the rebalancing of power from central back to local, at least on a small scale, to counter the negative consequences of power reorganization as outlined by Poe (2008).

Increase Transparency about Purpose and Use of Results

The teachers involved in this study are confused about how Examination results are used. One is entirely unsure, and the other expresses the misguided belief that the province may be using results to evaluate teachers or schools. Students are equally unaware; although some believe results are being used to guide program delivery and policy, others assume results are used to measure teacher effectiveness. As expressed in Chapter 7, the assumption that the Examination is in place to monitor teachers or schools may contribute to feelings of mistrust on the part of students and the larger public, as well as resentment on the part of teachers.

The belief that the Examination is evaluating teaching is unfounded. That this belief may contribute unintended negative consequences is a sad fact, considering how easily it could be avoided. Increasing transparency about how exactly the results are used at the provincial level can allay concerns on the part of teachers and students. Currently, the only document that speaks to the use of results is the general “About PLANS” site. The six bullet points listed in Chapter 7 represent the only information given about the use of results, and they are intended to apply to all standardized assessments and evaluations in Nova Scotia. Creating a separate page to give real examples of how results

guide policy creation, inform education decision-making, or "assist students to achieve outcomes" (Government of Nova Scotia, 2019a) could relieve and correct some of the concerns and misconceptions described above.

In the unlikely event that there are no examples of positive consequences, PLANS may consider undertaking a cost-benefit analysis regarding the consequences of the Examination similar to the process suggested by Kane (2013): "[i]f the perceived costs and/or immediate negative consequences exceed the perceived benefits, the program will get a negative evaluation" (p. 49). At present, I have been unable to find evidence of specific benefits or usage of results within public documents. Until evidence of benefits is produced, this analysis cannot take place.

Increase Transparency About the Use of Disaggregated Data

Many of these recommendations rest on the assertion that students who are more engaged, who "buy in," or who trust in an assessment's purpose will perform better. Within this study, the students demonstrating the highest levels of participation and engagement are those from backgrounds of privilege—nearly all of the students in Ms. Hatfield's class speak English as a first language and are white. Many grew up in the same neighbourhood. By contrast, Mr. Clark's class represents a greater cross-section of the school as a whole. His students are from varied neighbourhoods and "feeder schools," and some are members of visible minorities. Both teachers independently raise the unique challenges facing students from marginalized groups. But when Ms. Hatfield discusses issues of consequences or fairness, it is always regarding her other classes, in which she teaches student refugees and students who speak English as an additional language.

Currently, results for the NSE: English 10 and the NSE: Mathematics 10 have been disaggregated according to student self-identification (Government of Nova Scotia, 2019b) and regional centre (Government of Nova Scotia, 2019f). The data displays relatively minor differences in results between non-identifying students and those who identify as having African, Mi'kmaq or other Indigenous ancestry. Discrepancies are more pronounced in the results from the NSE: Mathematics 10. Data also reveals differences between school districts; the Cape Breton district achieves scores over 10% lower, on average, than the Acadian francophone school board, for example. Certainly, data could be further disaggregated to reveal differences between urban and rural schools, or between native English speakers and English language learners, for example. It is arguable whether score discrepancies between groups are driving tangible change in curriculum, professional development, or assessment design, or whether results are informing discussions about assessment validity. Unfortunately, without more information, it is impossible to determine whether the disaggregation itself is doing more harm than good by "reinforc[ing] public fears about the failings" of historically marginalized groups (Poe, 2008) or "form[ing] particular racial formations" (Inoue, 2009).

Include Speaking and Listening Components

Although the Examination represents a culminating evaluation of the grade ten English Language Arts program, the Speaking and Listening strand is entirely absent from the assessment. Considering that the Examination comprises 20% of a student's final grade, the attention required to fulfill Speaking and Listening within other course assessments is unrealistic. This becomes especially true when considering the extra time

and attention given to constructs within the Examination itself. The inclusion of a listening component, such as a short podcast or recording, could help address this imbalance. Mr. Clark designed such a component when he created his own final exam in 2017. Alternatively, teachers could be asked to create and administer this portion of the Examination themselves, with the final weight of the provincial Examination reduced accordingly.

Admittedly, including a listening or speaking component on the Examination would have its challenges. Among other possibilities, it may require accommodations for children with hearing difficulties and open up the likely possibility of technical difficulties if technology was used. However, the exclusion of an entire curricular strand, when only three are represented within the course, is too large a gap to justify. Avoiding underrepresentation within the larger program should overcome any need for efficiency in test design.

Assess Other Forms of Writing

Currently, the Examination requires students to write two essays: one persuasive essay, and one reading response. The essays are framed in somewhat different ways, as described in Chapter 2. However, their rubrics are identical save the addition of "Quality of Text Comprehension" for the reading response. Ultimately then, the Examination asks students to demonstrate writing skills in the same way twice. This uniformity in writing stands in stark contrast with the varied nature of the curriculum and the sample assessments suggested in the English Language Arts Desk Blotter. If only for the sake of validating its content, the opportunities for writing on the Examination should be more varied.

Of special significance to this study are the potential consequences of this uniformity in writing. The students in Ms. Hatfield's class were engaged with essay writing, citing it as a preference over other forms. The teacher herself focuses heavily on persuasive writing with her students, asking them to write upwards of four essays over the course of the year. By contrast, Mr. Clark focuses on individual essay components, taking a paragraph-by-paragraph approach. His students are less engaged with essay writing and spoke instead of a preference for creative writing. The inclusion of a creative or narrative piece in the Examination could increase levels of engagement, especially for students who are at risk of being less engaged, and therefore at risk of doing poorly.

Acknowledge the Implications of Construct Underrepresentation for Teachers

One consequence that arose during the study was the likelihood that the larger program offered to students would become imbalanced when factoring in the underrepresentation of numerous constructs on the Examination. When the final exam represents 25 to 50% of course content, yet comprises 20% of a student's final grade, it is difficult to accept that missing constructs would be given appropriate attention during instruction and assessment before the examination. This becomes especially true considered against the backdrop of extra attention given to genres like the persuasive essay in preparation for the Examination.

At minimum, acknowledging construct underrepresentation means taking proactive steps to make teachers aware of the curricular imbalance caused by the Examination. Teachers should not be expected to discover this imbalance themselves and adjust their own assessments accordingly, as this process would be time consuming and

prone to error. Rather, professional development time should be devoted to the topic of construct representation within the course.

Allow Teachers to Create their Own Final Exam

Of course, many, if not all, of the recommendations suggested above could be disregarded with the elimination of the Examination. If teachers can create their own final exams, they are more able to design an assessment that is responsive to the students in their classroom, represents a greater proportion of the curriculum, and promotes student engagement. Certainly, there is no guarantee that every teacher would produce an exam that accomplishes these things. However, interviews with the teachers involved in this study illustrate the drive and motivation to provide an assessment that is responsive to the students in each classroom. Ricci (2004) found that the transfer of power inherent in standardized assessment can contribute to what he termed "deskilling" of teachers. Arguably, the province is losing out on the skills of its own teachers by not allowing them to use the knowledge of their own unique students in creating more relevant final exams.

The NSE: English 10 is intended to provide an accurate and objective portrait of student achievement in order to evaluate curriculum delivery, guide education decision-making, and increase accountability (Government of Nova Scotia, 2019a). Ironically, the efforts put in place to ensure inter-rater reliability and objectivity may be working to make the Examination more subjective. These efforts, such as the use of simple scoring rubrics, limiting constructs to those which can be assessed on a standardized test, and the re-grading of tests by external teachers, lead to important differences in the way students perceive and engage with the Examination. Many potential threats to student engagement

and motivation are eliminated or mitigated when teachers are invited to create their own assessments.

Employing Slomp et al.'s approach to unintended consequences provides some structure to this conversation: "Taken collectively, does the evidence gathered so far indicate that the assessment has achieved the purpose or goals for which it was designed? […] If not, how does this failure contribute to (a) particularly adverse impacts, (b) impacts on populations, and (c) impacts on relevant educational systems?" (p. 284). The evidence gathered within this study suggests that the answer to the first question is a firm "No." Answers to the subsequent questions have been described in earlier chapters. The Examination seems to negatively influence student motivation for some students and therefore has the power to reduce scores in a way that is not reflective of true student knowledge and skills. The impact of reduced scores can be seen throughout political discourse in Nova Scotia, particularly when there are apparent discrepancies between the scores of certain groups of students. The impact on educational systems goes beyond the shaping of curriculum. The *Raise the Bar* report's recommendations for increased standardization, testing, and disaggregation demonstrates the power of publicized test scores. Eliminating the Examination and inviting teachers to create their own exams largely mitigates these adverse impacts on students, the general public, and educational policy.

If the Examination were traded for of locally-designed final assessments, the province would lose the disaggregation of data and the ability to monitor student outcomes. However, it is impossible to know how results are guiding decision-making until PLANS reports more explicitly on its use of results. Greater understanding of how

the Examination benefits teachers and students can assist us in determining the relevance and fairness and sustainability of the Examination itself.

Limitations

This comparative case study was deliberately narrow in its focus. In restricting observation and interviews to two classrooms, the observer gains a clearer, richer picture of each context and the experiences of students and teachers within them. Inherent to this narrow focus, however, are certain limitations. These limitations, as well as others that arose unexpectedly during the research process, must be acknowledged for what they are. This section will outline the limitations that may have impacted the scope and reach of conclusions.

A Narrow Sampling of Student Participants

Initially, this study was intended to examine the experiences of students with African Nova Scotian ancestry. The proposed sampling method involved teachers in the selection process. Because of the sensitive nature of sampling within and among different races and cultures, the HRSB requested that I remove all constraints on sampling and open interviews to any students who wanted to participate. The students who expressed interest in participating were, by and large, from similar backgrounds. All student participants were members of the visible majority, spoke English as a first language, and typically received high marks in English classes. My relatively homogenous sampling may be related to the sampling process, although it may also have been influenced by my own identity as a white, English-speaking woman.

Many of the conclusions drawn from this study implicate the Examination's consequences for students marginalized on the basis of race, socioeconomic status, or

language. However, I was unable to personally interview any students about this very issue. Without speaking to these students directly, it is difficult to validate any statements about consequences as they concern marginalized students. It would be extremely useful to gain access to these voices in any future study of unintended consequences, especially if the province has any intention to further disaggregate results, as suggested by Dr. Avis Glaze (2018). A more intentional sampling would be necessary for this to take place.

A Limited Timeframe

To reduce the scope of the study to a manageable size, classroom visits were limited to two to three observations per class. Observations were intentionally scheduled during the week before the provincial Examination in order to better understand teacher and student responses to practice assessments. A narrow focus allowed me to more closely examine the relationship between the Examination and classroom practice, but it left me with less information about the larger program. I relied upon student and teacher interviews, as well as classroom documents, to inform my discussion on classroom literacy practices. Observing more classes throughout the semester or year would help to broaden the perspective of classroom practices, as would gaining access to all assessments used within the course.

Furthermore, student interviews were limited to one session per student. Because of the timing of the Examination at the end of the semester or year and the HRSB's (now HRCE) research policy, interviewing students after the Examination was not possible. Limiting student interviews to one session before the Examination limits the extent to which I can understand their experience with the Examination. Similar to the study completed by Fox & Cheng (2007), conducting interviews during or, at minimum, after

the Examination, could shed light on the consequences following the assessment and confirm or refute ideas held about unintended consequences. For example, levels of student engagement with the assessment could be verified in this way.

A Lack of Access to the Provincial Examination

Electronic documents, including outlines, grading rubrics, and sample assessments, provide insight into the format, evaluation, and types of tasks present on the Examination. However, access to a version of the Examination itself was not granted for the purposes of this study. Access to the Examination, in past or current iteration, would expand the degree to which analysis regarding content and construct representation could be undertaken. One important factor in student perceptions of the Examination seems to involve reactions to multiple-choice questions. These questions were a source of stress and frustration in both classrooms, but had a larger impact on students in Mr. Clark's classroom in terms of classroom practices and student self-esteem. Many of the questions on practice assessments were too ambiguous or confusing for the teachers to answer correctly. Whether field-tested questions on the Examination are any clearer remains unanswered. Exit interviews with students post-Examination could shed light on this question, but gaining access to the document itself could inform a larger discussion of validity.

The Position of Participants and Researcher

Although the researcher's identity is not a limitation in and of itself, it can complicate questions of access. Just as there are accommodations we can make to help participants feel comfortable, there are "characteristics that we can't do much about: our gender, relative age, perceived racial or ethic category, for example. The best we can do

with these factors is to gauge how they might be interpreted by the people being studied…" (Angrosino, 2005, p. 7). With respect to student participation in interviews, my own identity as a woman of the visible majority problematizes the question of access. Students who belong to other visible groups, especially those who have experienced marginalization, may be less willing to participate. Compounding this lack of access is the possibility that students who are less engaged in class are less likely to participate in an interview. My own teaching experience has taught me that students who struggle, who feel disconnected from the teacher, or who lack confidence are less likely to volunteer for any initiative and therefore less likely to participate in an interview. Therefore, despite an attempt to gain maximum variation sampling through open access to interviews, it is unlikely that the findings from this study's interviews accurately reflect the range of attitudes and identities in each class.

The Influence of Preconceptions

As researcher, I must acknowledge the evolution of my own ideas about local history, racial tension, and the language surrounding assessment throughout the duration of this study. These ideas have undoubtedly shaped the design of the study, the focus of interviews, and the interpretation of data, whether intended or not. Furthermore, I will discuss the insight I have gained into local preconceptions, as these insights have contributed to interpretation of the data and present possible paths to further research.

Political Tensions. When I arrived with my family in Nova Scotia over six years ago, I was unfamiliar with its historical context. Consequently, the frames through which I viewed the topic of standardized assessment were borrowed from other provinces in which I had lived and worked. The legacy of racism and the extent to which it continued

to thrive, first introduced through my husband's experience with different population groups as a criminal prosecutor, came as a surprise. I immediately drew parallels between the social contexts for African Nova Scotians and other marginalized groups in Ontario and Alberta. I allowed this study to be influenced by the design of studies in Ontario and Alberta because I expected the contexts to be similar, and this gave me a sense of confidence. Ultimately, this confidence was challenged.

When the HRSB took issue with the original design of this study, I struggled to understand their reasoning. Having heavily researched African Nova Scotian history prior to proposing the study, I felt confident that the research could offer a sensitive, yet objective, exploration of assessment in Nova Scotia. Their concerns lay, however, with the selection of student interview participants based on culture or ancestry. The HRSB prohibited any self-identification or teacher-identification of student cultural background. Previously, I had not considered the ethical complications of this facet of the design. My conversations with the HRSB therefore enriched my understanding of ethics and methodology in general. Making the changes necessary to gain HRSB approval was difficult, but I gained insight into the complexities and sensitivities of human research when working with vulnerable groups like adolescents from marginalized communities.

As discussed in Chapter 6, my study was later questioned and challenged by the leader of the English Language Arts department at Eastern High. The concerns she expressed regarding issues of race—despite the absence of any such focus in the study—echoed those of the HRSB. Her reaction was difficult to hear, not only because I felt that my study was ethically sound, but also because I had expected a positive reception; prior to this meeting, her counterpart at Oceanview Collegiate had expressed enthusiastic

support for the study. Through this difficult conversation, however, I began to understand the importance of sensitivity when working within politically tenuous contexts. Rather than pushing back, I asked questions about her experiences. She related stories of numerous other researchers proposing studies with a focus on race. She described the school's history, as well as the evolution of standardized assessment in the Nova Scotian education system that she had witnessed and participated in. Indeed, it was through this conversation that I gained significant insight into both the history of local assessments and the extent of the differences between the two participating schools. This insight, in turn, influenced the interpretation of my findings.

The exchanges described above illustrate the importance of being receptive to feedback and open to ideas that challenge one's own preconceptions. However, these exchanges also point to a collective discomfort to confront or discuss issues involving race and, to some extent, to discuss problems with provincial assessments. The HRSB asked specifically that I remove all mention of race from my study. To paraphrase one phone conversation, I was told that the HRSB cannot be seen to be supporting research which might make their assessments "look bad." Eastern High's ELA department leader took issue not only with the potential for my study to involve issues of race, but also with my study's focus on the Examination itself. She expressed the opinion that the current iteration of the provincial Examination is much improved in comparison older versions and questioned my motivation to study it. These conversations cannot be taken to represent the entirety of opinions within the region. However, they present two accounts of research being challenged and altered. Given the relative lack of qualitative research in

the field of standardized assessment in Nova Scotia, these two accounts provide some insight into why more research may not be happening.

Language. My own preconceptions about assessment rhetoric initially shaped the language of the study and, consequently, shaped the interpretation of some of the data. After a number of years working in this province, I became somewhat accustomed to the typical rhetoric surrounding race and education. The first versions of this document borrowed the term "achievement gaps" liberally and discussed the measures being adopted to "fix" or "close" these gaps, as this language is pervasive through official documents such as *Nova Scotia's Action Plan for Education*. It was only through conversation with my thesis supervisor that I began to critically interrogate the language being used to discuss assessment in Nova Scotia. These conversations brought new insight to the analysis of documents. When we reject the premise that there are gaps in achievement, we can then begin to critically examine the assessment, as well as the documents like *Lessons Learned*, which itself is based on the premise that lack of student success comes from a problem with either teaching or the students themselves, rather than a problem with the assessment itself.

Reflecting on my own acceptance of this rhetoric sheds light on the public's concern reflected by the media, as well as many of the comments made by students during interviews regarding the validity of the assessment. Certainly, it is understandable that many would accept that "achievement gaps" represent real gaps in the knowledge and abilities of our children. In fact, many who promote this representation do so with the best of intentions, calling for educational reform to support the students who fail as echoed in *Nova Scotia's Action Plan for Education.* Accepting or promoting this

representation, however, inherently accepts the validity of the Examination. It is only though critiquing this representation that we can begin to examine the appropriateness or validity of the Examination. Indeed, there is more work to be done.

Areas of Future Study

Beyond the limitations outlined in the previous section, there remain numerous directions for an inquiry into standardization and its consequences in Nova Scotia. Compared to larger provinces like Ontario and Alberta, relatively little discussion has occurred here. The topic itself is contentious; a removal of standardized assessments has been a talking point in focus groups and politics in recent years, and yet the assessments generally remain unchanged. In this section, I outline two possible areas for future research that arose during this study.

The Consequences of Standardized Assessments in Younger Grades

Perhaps the most obvious next step for research would an investigation of the provincial assessment in the younger grades. In particular, the Nova Scotia Assessment: Literacy and Mathematics in Grade 3 could shed light on unintended consequences. Since its inception, this assessment has been the cause of scrutiny and media alarmism, as disaggregation has revealed significant score discrepancies between certain racial groups (Jeffrey, 2014). Throughout this study, I frequently encountered the message that the NSE: English 10 is relatively innocuous in terms of its weight and implications for classroom practice. By contrast, the public feedback available for the grade three assessment is concerning. Parents with whom I have spoken have described anxiety, stress, and resentment over the administration of a four-day-long test at such a young age, and the removal of children from school while it is administered. This study revealed the

potential for a standardized assessment to restrict literacy perspectives, to limit classroom practices, to disengage students, and to create teacher resentment. It is difficult to accept that this would not follow, perhaps to an even greater extent, in the grade three context.

The Relationship Between Motivation and Standardized Assessment

The question of student motivation rose unexpectedly during my study. I did not anticipate that interviews and observations would produce so much data about this topic. Between the two participating classrooms, there was a marked difference in participation, transfer of responsibility, and student engagement with practice assessments. Research into motivation and self-efficacy theory revealed the potential for multiple explanations. Unfortunately, the scope of this particular study was too small to determine with confidence which of these factors may play a role.

Further qualitative research including interviews with students before and after the Examination could shed light on these differences in motivation. It would be useful to compare two or more classrooms to gain a broader perspective on such a large topic. A better understanding of what influences student motivation can, at minimum, contribute to greater validity in test design. However, it can also help test designers, at the provincial and local level, to better engage all students. Finally, exploring student motivation would contribute important information to the greater discussion of standardization and validity, and could shed light in particular on discrepancies revealed within disaggregated data.

REFERENCES

Alberta Education. (2005). *Social studies: Kindergarten to grade twelve*. Retrieved from http://education.alberta.ca

Angrosino, M. V., & Mays, K. A. (2000). Rethinking observation: From method to context. In Denzin, N. K., & Lincoln, Y. S. (Eds.), *Handbook of qualitative research* (pp. 673-702). Thousand Oaks, CA: SAGE Publications.

Angrosino, M. V. (2005). *Projects in ethnographic research*. Long Grove, IL: Waveland Press, Inc.

Black Learners Advisory Committee [BLAC]. (1994). *BLAC report on education: Redressing inequity – empowering black learners*. Retrieved Nov. 25, 2014, from http://ednet.ns.ca/acs/publications.

Cazden, C., Cope, B., Fairclough, N., Gee, J., et al. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60.

Chiasson, M. (2015, April 8). Students not measuring up to school board's literacy goals. *Truro Daily News*. Retrieved October 22, 2015, from http://trurodaily.com.

Cohen, L. & Manion, L. (2003) Research methods in education. (4th edition). New York, NY: Routledge.

Creswell, J. W. (2003). Research design: Qualitative, quantitative, and mixed methods approaches (2nd edition). Thousand Oaks, CA: SAGE Publications.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer and H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.

Denzin, N. K., & Lincoln, Y. S. (Ed.). (2005). *Handbook of qualitative research*. (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Doucette K. (2018, January 24). N.S. government to move ahead with recommendation to eliminate school boards. *Global News*. Retrieved November 20, 2019, from http://globalnews.ca.

Education and Early Childhood Development. (2018). *Glaze Report provides recommendations to improve Nova Scotia's education administrative system*. Retrieved from http://novascotia.ca/news/release.

Enidlee Consultants Inc. (2009). *Reality check: A review of key program areas in the BLAC Report for their effectiveness in enhancing the educational opportunities and achievement of African Nova Scotian learners.* Retrieved from https://ednet.ns.ca/acs/publications.

Evans, D. & Tynes, G. (1995). Telling the truth, reflections: Segregated schools in Nova Scotia. Hansport, NS: Lancelot Press Ltd.

Finn, B. (2015), Measuring Motivation in Low-Stakes Assessments. ETS Research Report Series, 2015: 1-17. doi:10.1002/ets2.12067.

Fox, J., & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test takers. *Assessment in Education, 14*(1), 9-26.

Frank, B. (2002). "Miracles could happen here": Report on an external review of Cole Harbour District High School. *Education Canada, 42*(3).

Glaze, Avis. (2019). *Raise the bar: A Coherent and responsive education administrative system for Nova Scotia.* Retrieved from https://www.ednet.ns.ca/sites/default/files/docs/raisethebar-en.pdf

Government of Nova Scotia. (2002). *Racial equity policy*. Retrieved from https://studentservices.ednet.ns.ca/sites/default/files/RacialEquityPolicy_Webversion.pdf.

Government of Nova Scotia (2003). *Report of the BLAC implementation review committee*. Retrieved from https://ednet.ns.ca/acs/publications.

Government of Nova Scotia. (2015a). *The 3 Rs: Renew refocus rebuild—Nova Scotia's action plan for education.* Retrieved from http://www.ednet.ns.ca/education-action plan.

Government of Nova Scotia (2015b). *Nova Scotia examinations: Information guide*. Retrieved from http://www.ednet.ns.ca/english10/documents.

Government of Nova Scotia (2016). *Interpreting the Nova Scotia Examination: English 10 Individual Student Report.* Retrieved from http://ednet.ns.ca/english10/documents.

Government of Nova Scotia (2017a). *Nova Scotia Provincial Literacy Strategy.* Retrieved from http://ednet.ns.ca/nova-scotia-provincial-literacy-strategy-0.

Government of Nova Scotia (2017b). *2015-2016 Nova Scotia Examination: English 10*. Retrieved from http://plans.ednet.ca/english10.

Government of Nova Scotia (2018). *Teaching in action, grades 10-12: A teaching resource.* Nova Scotia Curriculum Documents. https://www.ednet.ns.ca/files/curriculum/TIA10-12.pdf.

Government of Nova Scotia. (2019a). *About PLANS*. Retrieved from http://www.ednet.ns.ca/about-plans.

Government of Nova Scotia (2019b). *Nova Scotia Assessments and Examinations Results for Students with Mi'kmaq or other Indigenous Acestry and Students of African Descent.* Retrieved from http://plans.ednet.ns.ca/disaggregated-results.

Government of Nova Scotia. (2019c). *Nova Scotia Examination: English 10 Lessons Learned.* Retrieved from http://www.ednet.ns.ca/english10.

Government of Nova Scotia (2019d). *Nova Scotia Assessment Development Model*. Retrieved from http://www.ednet.ns.ca.

Government of Nova Scotia (2019e). *English Language Arts Desk Blotter*. Nova Scotia Curriculum Documents. https://www.ednet.ns.ca/files/curriculum/.

Government of Nova Scotia. (2019f). *Nova Scotia Examination: English 10.* Results. https://plans.ednet.ns.ca/results.

Government of Nova Scotia (2019g). *Information about Individual Student Reports for students, parents, and guardians.* English 10 Documents. Retrieved from http://plans.ednet.ns.ca/english10/documents.

Government of Nova Scotia (2019h). *Analytic rubric for scoring persuasive writing*. English 10 Documents. Retrieved from http://plans.ednet.ns.ca/english10/documents.

Grills, S. (1998). *Doing ethnographic research: Fieldwork settings.* Thousand Oaks, CA: Sage Publications, Inc.

Halifax Regional School Board. (2014). *Focus on learning: Grade 3 and 6 provincial literacy assessment report.* Retrieved from http://hrsb.ca.

Halifax Regional School Board [HRSB] (2016). *Guidelines for Researchers*. Retrieved from http://hrsb.ca/about-hrsb/program/research.

Hill, C. E. (2012). Consensual qualitative research: A practical resource for investigating social science phenomena (1st edition). Washington, DC: American Psychological Association.

Inoue, A. B. (2009). The technology of writing assessment and racial validity. In C. Schreiner (Ed.), *Handbook of research on assessment technologies, methods, and applications in higher education*. Hershey, PA: Information Science Reference.

Inter-university Research Network Application for Funding (2019). Retrieved from ednet.ns.ca/ERP/iurnfunding.

Jeffrey, D. (2014, May 29). 'Alarming' literacy shortfalls among African-Nova Scotian students. *The Chronicle Herald*. Retrieved November 25, 2014, from http://thechronicleherald.ca.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.

Kearns, L. (2011). High-stakes standardized testing and marginalized youth: An examination of the impact on those who fail. *Canadian Journal of Education, 34*(2), 112-130.

Kvale, S. (2007). Doing interviews. London, UK: Sage.

Kvale, S. & Brinkmann, S. (2009). Doing interviews: Learning the craft of qualitative research interviewing. Thousand Oaks, CA: Sage.

Lincoln, S. L. & Guba, E. G. (1985). Naturalistic inquiry. Beverly Hills, CA: Sage.

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review, 31*(3), 313-327.

Literacy help falling short (2015, March 5). *The Chronicle Herald*, p. A8. Retrieved February 19, 2015 from http://thechronicleherald.ca.

Merriam, S. B. (1998). Qualitative research and case study applications in education (2nd ed.). San Francisco, CA: Jossey-Bass.

Merriam, S. B. (2009). Qualitative research: A guide to design and implementation (3rd Ed.). San Francisco, CA: Jossey-Bass.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Nova Scotia Teachers Union [NSTU] (2017a). *Report of the Council to Improve Classroom Conditions.* Retrieved from http://ednet.ns.ca/docs/classroomcouncilfinalreport2017.pdf.

Nova Scotia Teachers Union [NSTU] (2017b). *Better Education for a Better Nova Scotia: Teachers are Ready to Act to Improve Learning Conditions.* Retrieved from http://nstu.ca.

Pachai, B. (2007). The Nova Scotia black experience through the centuries. Halifax, NS: Nimbus Publishing Ltd.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Poe, M. (2008). Genre, testing, and the constructed realities of student achievement. *College Composition and Communication, 60*(1), 141-152.

Province moves ahead with eight recommendations to improve classroom conditions. (2017, March 29). *CBC News*. Retrieved 30 September, 2019, from http://cbc.ca.

Nova Scotia Education and Culture, 1997. *Atlantic Canada English Language Arts Curriculum* [Program of Studies]. Nova Scotia Curriculum Documents. https://www.ednet.ns.ca/files/curriculum/ela10_12web.pdf.

Quigley, B. (2018). *Youthful bookworms: Students' experiences with critical literacies in the context of the Ontario Secondary School Literacy Test* (Doctoral dissertation). Retrieved from http://ruor.uottawa.ca.

Quon, A. (2018, February 2). N.S. teachers' union pulls out of provincial council on improving classroom conditions. *Global News.* Retrieved 30 September, 2019, from http://globalnews.ca.

Ricci, C. (2004). The case against standardized testing and the call for a revitalization of democracy. *The Review of Education, Pedagogy, and Cultural Studies, 26*, 339-361.

Ritchie, Sarah. (2019, January 23). Chaos versus order: One year later, how both sides of education debate view the Glaze Report. *Global News*. Retrieved 30 September, 2019, from http://globalnews.ca.

Ryan, T., & Whitman, A. (2013). The Inequity and Effect of Standardized Literacy Testing for First Nations Students: An Ontario (Candian) outlook. *The Journal of Educational Thought (JET)/Revue de la Pensée Éducative, 46*(2). 161-181.

Skerrett, A. (2010). "There's going to be community. There's going to be knowledge": Designs for learning in a standardised age. *Teaching and Teacher Education, 26*, 648-655.

Skerrett, A. & Hargreaves, A. (2008). Student diversity and secondary school change in a context of increasingly standardized reform. *American Education Research Journal, 45*(4), 913-945.

Slomp, D., Corrigan, J., & Sugimoto, T. (2014). A framework for using consequential validity evidence in evaluating large-scale writing assessments: A Canadian study. *Research in the Teaching of English, 48*(3), 276-302.

Stake, R. E. (1995). The art of case study research. Thousand Oaks, California: SAGE Publications.

Thiessen, V. (2009). Identity, equity, and performance: mathematics and reading literacy in Nova Scotia public schools. Retrieved March 3, 2015, from http://www.tcrsb.ca.

White, E., Elliot, N., & Peckham, I. (2015). Very like a whale: The assessment of writing programs. Logan, UT: Utah State University Press.

Willick, F. (2015, January 22). Reading, writing, 'rithmetic skills drop. *The Chronicle Herald*. Retrieved February 19, 2015, from http://the chronicleherald.ca.

Willick, F. (2013, January 22). Report: Literacy goals still a challenge. *The Chronicle Herald*. Retrieved February 19, 2015, from http://thechronicleherald.ca.

Willick, F. (2014, April 19). Extra reading help not working for many Nova Scotian students. *The Chronicle Herald*. Retrieved February 19, 2015, from http://thechronicleherald.ca.

Williams, Cassie. (2016, November 28). Teachers in Nova Scotia to start work-to-rule job action Dec. 5. *CBC News*. Retrieved 2 August, 2019, from http://cbc.ca.

Wigfield, A., Eccles, J. S., Roeser, R. W., & Schiefele, U. (2007). Development of achievement motivation. In W. Damon & R. M. Lerner (Eds.), *Child and adolescent development: An advanced course.* (pp. 406-436). John Wiley & Sons, Inc.

APPENDIX A

Cognitive Levels of Selected Answer Questions on the NSE: English 10

Adapted from *Nova Scotia Examinations Information Guide*

(Government of Nova Scotia, 2015c)

Literal Comprehension questions are designed to elicit responses that indicate the student has comprehended explicit information in the text.

Non-literal Comprehension questions are designed to elicit responses that indicate the student has comprehended implicit information in the text such as inferences, connotative meanings, idioms, and figurative language (e.g., simile and metaphor).

Analysis questions are designed to elicit responses that indicate the student has thought critically about texts by analyzing, synthesizing, and evaluating the explicit and/or implicit information in the text.

APPENDIX B

Teacher Interview Question Guide

- Could you describe your teaching and classroom practices, generally?
- What are your goals, as an English teacher?
    - Where do these goals come from? What factors motivate them?
- What are you doing in your course to prepare students for the test?
- Suppose I was your student, and I was nervous about the test. What advice would you give me?
- Some would say the evaluation is necessary to judge the effectiveness of teachers. What would you tell them?
- In your opinion, can a teacher influence a student's outcome on the test? How?
- If you could create your own test, what would it look like?
- If the test didn't exist, how would your program change, if at all?
- Is it important for students to pass? Why or why not?
- What inferences can you make about students from their test results?
    - What inferences do you think administration makes?
    - What inferences do you think the government makes?
    - Are these inferences justifiable? Why or why not?
- How do you think your students performed/will perform?
- Have any of your students ever failed the test? Why do you think these students failed?
- Suppose I was your student, and I failed or did very poorly on the Examination. What would happen?
- How would you describe the social climate of your school?

APPENDIX C

Student Interview Question Guide

- How's it going in English class this year?
- Do you usually like/do well in English class?
- What kinds of activities do you like/not like doing in English class?
    - Can you give me an example of something you did in class this year that you enjoyed? Something you read?
- If I asked the English teachers at this school, "What's the most important thing for kids to learn in English class?", what do you think they would say?
- Do you feel prepared for the provincial exam?
    - What makes you say that?
- How do you think you will do on the provincial exam? How did the practice questions go?
    - What makes you think that?
- What sorts of things have you heard about the test from your teachers? Other students?
- Do you think this examination matters? Why or why not?
- Do you think your mark on the test will be better or worse than your mark in the course?
- What English class do you plan on taking next year?
    - How will your mark on this test effect your decision?
- How would you describe your school? What's it like to go here?

APPENDIX D

Sample Writing Scoring Rubric

Analytic Rubric for Scoring Reading Response (ENG 10)

NOVA SCOTIA

Ideas (please note: Ideas includes Ideas 1 and Ideas 2)	
Ideas 1 – Quality of Text Comprehension	
4	Demonstrates a thorough understanding of the text(s) offering an insightful response with strong support from the text(s).
3	Demonstrates a clear understanding of the text(s) offering a thoughtful response with relevant support from the text(s).
2	Demonstrates a limited understanding of the text(s) offering a simplistic response with vague and/or minimal support from the text(s).
1	Demonstrates a misreading or significant misunderstanding of the text(s) offering an irrelevant response with minimal or no support from the text(s).
Ideas 2 – Quality of Writing	
4	A main idea is distinct and is developed through vivid and relevant details.
3	A main idea is clear and is developed through relevant details.
2	A main idea is evident and is somewhat developed through details, some of which may be irrelevant.
1	A main idea is not present, or a main idea is not developed with details, or writing does not address the prompt.
Organization	
4	The writing is skillfully organized with skillful use of varied transitions.
3	The writing is organized with effective use of varied transitions.
2	The writing is somewhat organized with vague or mechanical transitions.
1	The writing is lacking organization; there is little or no evidence of transitions.
Language Use	
4	Language use contributes to vivid and skillful writing.
3	Language use contributes to clear and fluid writing.
2	Language use contributes to vague or mechanical writing.
1	Language use contributes to confusing writing, or there is little evidence of language use.
Conventions	
4	A variety of consistently correct conventions contribute to enhanced communication.
3	A variety of generally correct conventions contribute to effective communication.
2	Errors in conventions are noticeable, but communication is coherent.
1	Errors in conventions contribute to confusing communication, or there is little evidence of conventions.