The Formative Potential of Standards-Based Grades and Report Cards

LATHAM JENNER CAMERON

The Formative Potential of Standards-Based Grades and Report Cards

By Latham J. Cameron

Abstract

Following the initial standards reform movement in U.S. education to establish learning objectives for subject curricula, standards-based grades (SBGs) and standards-based report cards (SBRCs) are theorized to unlock the formative potential in these traditionally summative means. Research on this initiative is in its infancy and my thesis focuses on four areas which remain under-investigated: the degree to which the SBG and SBRC formative rationale holds true; SBG and SBRC impact on student motivation; student perceptions of SBGs and SBRCs; and how the SBRC form impacts student meaning making.

My mixed methods case study took place within the middle school of an American international school during the school's final years with traditional letter grades and its initial years of implementing SBG and SBRC reform. I investigated the meaning students constructed from grades and report cards with the two systems, and how the SBG and SBRC shift impacted student motivation. Results were interpreted using formative assessment theory, the work of Vygotsky and Wertsch, and Self-Determination Theory. Focus group findings revealed that students perceived letter grades uniformly as summative measures, but two contrasting student constructions of SBGs and SBRCs were found: an autonomy supporting formative function through connecting them to criteria rubrics, and a controlling extra step obscuring letter grade standing. It is concluded that schools hoping to maximize the formative potential of SBRCs should avoid a hybrid SBRC. Quantitatively, questionnaire results indicated the SBRC shift had minimal impact on the average motivation of the student population. Integrating focus group results with these quantitative findings, I argue that gains in autonomous motivation within some constructions of SBGs and SBRCs were negated by losses from other constructions.

These findings reveal the potential of SBGs and SBRCs to positively impact student learning and motivation, while also raising questions about their successful implementation for all students.

The Formative Potential of Standards-Based Grades and Report Cards

By

Latham J. Cameron

A thesis submitted for the degree of
Doctor of Philosophy

School of Education
Durham University
2020

Table of Contents

List of Tables

List of Figures

Glossary of Key Terms and Acronyms

Academic Self-Regulation Questionnaire (SRQ-A) – Self-Determination Theory questionnaire used within my study to measure students' motivation towards their academic work.

Assessment – The process of gathering, recording and making inferences about students' responses to educational activities.

Assessment for Learning (AfL) – Strand of FA with student-centred focus on learning gaps.

Assessment as Learning (AaL) – Strand of FA focusing on students taking evaluative power and generating their own feedback through self and peer assessment.

Common Core State Standards (CCSS) – Set of learning criteria for English and math, adopted across the majority of U.S. states and many American international schools.

Confidence Intervals (CI) – Inferential statistical measure used within my study to measure the preciseness of effect size estimations.

Effect Size (ES) – Statistical measure used to determine the size of effect of a given treatment; used within my study to measure size of effect of shift to SBRCs on student motivation.

Feedback – Information which allows students to understand their current standing and next steps in any given learning progression.

Formative Assessment (FA) – Inferences from assessment data which seek to identify a student's current standing in relation to a learning target and their next steps to close the gap.

Hybrid Standards-Based Report Card – An SBRC which includes an aggregated letter grade for each class in addition to the standards-based grades.

Letter Grades – The traditional grading letters (*A*, *B*, *C*, *D*, *F*) students are given to signify their performance on an assignment or for a term of study.

Mediational Means – The objects and constructs which mediate human learning.

Mediated Action – The action resulting from the irreducible tension between the unique identities of human agents and the mediational means of the environment.

Next Generation Science Standards (NGSS) - Set of learning criteria for science, adopted by many U.S. states and many American international schools.

Organismic Integration Theory (OIT) – Sub-theory of SDT which posits that the degree to which an individual's environment supports their basic psychological needs of autonomy, competence, and relatedness determines the individual's degree of internalized and autonomous motivation.

Prioritized Standards – Every set of standards has too many standards to meaningfully teach during a course of study. Teachers must identify and focus on the prioritized standards they deem most important.

Report Cards – End of term report used to communicate student learning over a term of study.

Reporting Standards – Including every prioritized standard for each class on the SBRC would make it too long and overwhelming. Instead, SBRCs communicate the reporting standards for each class: approximately four essential skills or areas of knowledge for that subject.

Self-Determination Theory (SDT) – Theory of motivation used within my study.

Sociocultural Forces – The cultural, historical and institutional forces which shape mediational means.

Standards – Learning criteria; what students should know and be able to do.

Standards-Based Grades (SBGs) – Grades which describe student progress within a progression of learning for any given standard (Extending, Meeting, Progressing, Beginning).

Standards-Based Grading – A set of grading practices leading to a standards-based grade; aimed at addressing issues of traditional letter grading systems and bringing focus to students' learning standing in relation to standards criteria.

Standards-Based Report Cards (SBRCs) – Report cards used to communicate student learning over a term of study. Achievement and behaviour are reported on separately and students receive multiple standards-based grades for each class's reporting standards.

Sutter International School (SIS) – A pseudonym for the American International school where I taught and conducted my study.

Summative Assessment (SA) – Inferences from assessment data which seek measure and summarize the amount of student learning during an elapsed time period.

Understanding by Design (UbD) – A framework from Wiggins and McTighe, used within schools to design curricular units with standards-aligned assessments and instructional activities.

Declaration

This thesis is the result of my own work and has not been previously offered in candidature for a degree at this or any other institution.

Statement of Copyright

Acknowledgments

Throughout the five years of my research, Jonathan Tummons and Peter Tymms guided me with deliberate and insightful feedback. Thank you both for sharing your immense expertise and knowledge while simultaneously creating a space for my own emerging ideas and understanding to take shape.

This journey began almost five years ago in Durham with a conversation I had with Per Kind. Per had an excitement for research that was contagious and a way of seeing you in the moment that gave me the courage to believe in this research project when it was in its infancy. I have missed him and I will always try to pay the encouragement he gave me forward to my own students.

I am thankful to all of my colleagues who helped me during this journey. In particular, I would like to thank Darnell Fine, Robyn Chapel, and Pete Lutkoski, all of whom spent generous hours with me brainstorming the logistical and conceptual frameworks to actualize this research.

I cannot pinpoint a definitive source of where my desire to pursue a PhD originated from, but it was surely in large part from my parents, Lynn Cameron and Chuck Sternberg. I feel lucky to have grown up with parents who modelled themselves as learners and encouraged me to be curious and to pursue my own sparks of interest.

Finally, to my wife, Amanda, thank you. I am afraid to tally the number of weekend hours we have spent together studying over the past five years in Southbank Centre overlooking the Thames. For this thesis you were my project manager, editor, and therapist, all while conducting your own studies. Thank you, a million times over for your patience, feedback and support. I love you.

Chapter One: Introduction

1.1 My Illusion of Perspective with Letter Grades

My teacher preparation courses in California during the early 2000's included very little education about the topics of assessment and grading. Disproportionate to the amount of education I had received on these topics, they were the focal point of a great deal of anxiety and stress - for both students and myself - upon beginning my job as a full-time teacher. During my first year of teaching, there was an immediate need to implement grading practices. In the absence of any taught strategies or deeper guiding theory I relied on the assessment and grading norms which I had internalized as a student, norms which were subsequently reinforced by my more experienced colleagues. I fumbled my way through forming a functional assessment and grading system which was somehow instantaneously enmeshed with my system of classroom behaviour management by using letter grades as rewards and punishments to guide student behaviour. Grading and the larger area of assessment were always on the periphery of my consciousness during these early years of teaching. They were part of the educational landscape, but just as one rarely questions the presence of the sky or the return of the morning sun, evaluating students on assignments and end of term report cards with the letter grades of *A*, *B*, *C*, *D*, and *F* had a sense of perpetual existence which seemed to dissuade a closer and more critical examination.

To my inexperienced eyes, my grading system was effective. Students arrived on time and behaved in class. The majority of students followed directions and appeared motivated to do their work. Enough students received *A's* in my class to positively reflect my ability to help them learn, but not so many as to signal a lack of rigour on my part. When I reflect back now, however, I vividly remember the ways students would gather around the class

grade sheets I posted throughout the semester, like moths drawn to a light: they couldn't resist it, yet as soon as they got close enough to read it they seemed to instinctively withdraw and move away…only to return again, simultaneously craving and fearing the letter symbols. I also remember the sense of extreme injustice some students expressed as they bitterly argued with me about the letter grade they had received. These conversations were often far removed from the assignment rubrics and information about the learning criteria, and I can only wonder now how many of those students who approached me were fighting back against a feeling of having their identity as a learner reduced to a single confining letter and the loaded meanings it contained about their ability and their future.

I wish that I had taken the time in those early years to examine the assumptions behind my grading practices and their impact on students. I became a teacher largely out of my passion for working with young people and wanting to empower them to think critically. How was it then that I had adopted assessment and grading practices which, in hindsight, largely seemed to inhibit the rationale behind my career choice? How had I unknowingly become a prisoner to the constraints of past educational methods which effectively reduced the humanity of students to objects to be controlled through a system of rewards and punishments?

Decades later, during the course of my PhD studies for this thesis, I came across the work of James Wertsch and found the language and theory to give name to this phenomenon. Human consciousness is mediated by the means of our environment. These mediational means not only afford us the very thoughts which make us human, they also inherently constrain us. Further, we are usually unaware of these constraints, a phenomenon Wertsch (1998) termed "the illusion of perspective" (p.41). My illusion of perspective with letter grades was not unique amongst educators. In recent years, however, many in U.S.

education have begun to recognize this illusion with the introduction of standards-based grades (SBGs) and standards-based report cards (SBRCs).

In the years that followed my assessment and grading baptism by fire, the standards movement in U.S. education, which began as an effort to establish clear learning criteria for schools across all 50 states, expanded to incorporate assessment and grading reform. I became aware of my illusion of perspective with letter grades and report cards as I began to learn about standards-based grades and report card reform. In collaboration with some of my colleagues, I began to reform my grading practices soon after arriving as a teacher at Sutter International School (SIS)[1] in 2012.

Starting in the 2015-2016 school year, senior leadership at SIS implemented a multi-year plan to transition the entire middle school from traditional letter grades and report cards to standards-based grades and report cards. This plan coincided with beginning my PhD studies and from the outset of my research I decided to investigate the impact of this reform. I hoped that combining my PhD research with the work I was doing as a teacher would provide valuable feedback to my school about their reform efforts. I felt strongly as a teacher that my PhD research should have a practical purpose and applicability.

1.2 Theoretical Context of the Study

Before discussing the problem of my study further, I will first examine the theoretical underpinnings of my thesis as they relate to grades and standards reform. I will then return to these topics in chapters two and three.

In education, assessment refers to "the process of gathering, interpreting, recording and using information about pupils' responses to educational tasks" (Lambert & Lines, 2000, p. 4). Assessment data used to quantify the amount of

[1] Sutter International School (SIS) is a pseudonym which I will use throughout this thesis to protect the anonymity of my study's participants.

learning which has taken place in any given course of study represents a summative use of the data, while assessment data used to guide future student learning represents a formative use (Earl, 2003). Formative assessment (FA) is theorized to support student learning by identifying learning criteria, providing feedback to students about their current standing, and identifying next steps students should take to close the gap between their current standing and the criteria (Black & Wiliam, 1998; Sadler, 1989). A key aspect of FA is that teachers can teach these evaluation practices to students so that students are able to self-assess and identify their own standing and next steps in learning (Dann, 2014; Sadler, 1998). Another important aspect of FA and the assessment process are criteria rubrics, which articulate levels within the learning progression for any given learning criteria and aid student self-assessment (Brookhart & Chen, 2015; Panadero & Romero, 2014).

FA is a major rationale behind standards reform in U.S. education. Standards are the learning criteria students should know and be able to do for any given course. By 1998, every U.S. state had standards or was in the process of developing them (Marzano & Kendall, 1998). This initial phase of standards reform was then followed by corresponding initiatives to align grading and reporting practices with these learning criteria (Cox, 2011). SBG and SBRC reform sought to unlock a formative potential within these summative means, in part by directly responding to the constraints of the previous letter grade system (Guskey, 2015).

Letter grades have traditionally been constructed in ways which obscure meaningful feedback about students' current standing in their learning. Examples of these opaque construction techniques include: norm-referencing, which identifies student standing through comparison to other students, not fixed criteria; averaging, which distorts feedback by accounting for past performance and penalizes students for the developmental process of learning;

and assigning zeroes for late or incomplete assignments, a practice which distorts the aforementioned average (O'Connor, 2011). In particular, teachers have included behaviour measures in letter grades as a way to control student behaviour (Brookhart, 1994; Sun & Cheng, 2014). Along with these practices, letter grades have a long standing within U.S. society as markers of norm-referenced intelligence and gate keepers to future levels of schooling. These different factors have resulted in the strictly summative function of letter grades, something SBG and SBRC reform has attempted to address.

FA literature has called for reform efforts to account for how students learn (Assessment Reform Group, 2002). Inherent within FA theory is that unique learning gaps exist for each learner as a result of how they individually construct meaning according to their experiences and identity. FA literature has frequently used Vygotsky's theory of learning to articulate this two-way learning phenomenon (Baird, Andrich, Hopfenbeck, & Stobart, 2017; Black & Wiliam, 1998; Shepard, 2000). Vygotsky (1986) believed that learning is mediated by the means of our environment and that individuals interact uniquely with these means. Wertsch (1991, 1998) extended Vygotsky's work further to describe how mediational means are produced, and the ways in which they both afford and constrain the mediated action of humans. Within this study, grades and report cards are situated as mediational means, and the meaning that students construct from these means represents mediated action.

Assessment reform should also account for student motivation (Assessment Reform Group, 2002). Self-Determination Theory (SDT) posits that all humans are born intrinsically motivated to explore and understand the world around them. The degree to which an individual remains autonomously motivated depends on how one's environment supports their basic psychological needs for autonomy, competence and relatedness (Deci & Ryan, 1985; Ryan & Deci, 2000, 2017). Within FA literature, there are connections to a

formative use of assessment data supporting students' needs for competence and autonomy (Andrade, Du, & Mycek, 2010; Dann, 2014). As such, a formative use of assessment data should support more autonomous forms of motivation in students.

1.3 Statement of the Problem

From the early stages of my PhD research, I became aware that there had been very little investigation into standards-based grades and report cards. This was largely due to the fact that it was a relatively recent reform initiative. I identified four predominant gaps in the literature which I hoped my study would address. These problems framed the aim and scope of my study which I will discuss in the next section.

Firstly, grade and report card research has rarely investigated student perspectives. Letter grade research has focused disproportionately on teachers' perspectives and the measures teachers use in constructing grades (Brookhart, 2013a; McMillan, 2001), while the perspectives of students has been under-researched, with some notable exceptions (Thomas & Oldfather, 1997). Further, the initial research into standards-based grades and report cards has focused almost exclusively on parent, teacher and administrator perspectives, with investigations into student perceptions of this initiative almost entirely missing (Brookhart et al., 2016; Swan, Guskey, & Jung, 2014; Welsh, D'Agostino, & Kaniskan, 2013).

Secondly, standards-based grades and report cards were created with the intended purpose of unlocking a formative potential within these traditionally summative mediational means (Heflebower, Hoegh, & Warrick, 2014), but research on standards-based grades and report cards is still in its infancy and the degree to which they function formatively remains largely unknown (Brookhart et al., 2016).

A third issue within standards-based reform is that there is no set form which SBRCs take, apart from reporting separately on achievement and behaviour, and reporting on key learning standards for each class (Guskey & Bailey, 2001; Guskey, Jung, & Swan, 2011). From the early inception of SBRCs, some schools have created hybrid SBRCs which include the standards-based grades and aggregated letter grades for each course as a way to find a middle ground between increasing formative feedback and the traditional expectations of parents and external gatekeepers, such as universities (Marzano, 1998, 2006). It is unclear if some SBRC forms impact students differently than others and what meaning students construct of hybrid SBRCs, as SBRC research has yet to fully investigate this topic.

Finally, FA reform has called for initiatives to consider student motivation (Assessment Reform Group, 2002). Standards-based grades and report cards are theorized to function formatively, a function which should lead to more autonomous forms of motivation in students, but initial research has lacked investigation into their impact on student motivation (Brookhart et al., 2016).

1.4 Aim, Scope and Research Questions

My study took place in the middle school of an American international school where I taught, SIS. SIS middle school consisted of four grade levels (Grades 5-8) with 470 students aged 10-14. During the time of my study, the school transitioned from the final years of traditional letter grades and report cards to the initial years of standards-based grades and report cards. The focus of my study was initially solely on report cards, but during the early days of data collection it became clear that students perceived their end of term report card grades and the grades they received on assignments throughout the term in fluid and often inextricable ways. As such, I broadened the scope of my study to include both report cards and grades in general.

Even before I had identified the lack of student perspectives in grade and report card literature, I had decided to prioritize investigating student perspectives of grades and report cards because of my own beliefs of the need to empower students in schools. Further, my personal experience has been that the adults in schools all too frequently overlook or minimize students' ideas and perspectives. Because my study took place before and after SBG and SBRC reform had been implemented, I sought to understand students' perceptions of the old and new forms of grades and report cards. As my research and theoretical understanding developed, I wanted to know if the meanings students constructed from standards-based grades and report cards revealed a formative function as theorized in literature.

This aim of my study is expressed in my first two interconnected research questions:

> **Research Question 1**: *What meanings do students construct from traditional letter grades and report cards?*
>
> **Research Question 2**: *What meanings do students construct from standards-based grades and report cards?*

A key aspect of these two questions was to see how students' constructed meanings from grades and report cards changed as a consequence of the standards-based shift. Another key aspect of these questions is my use of the word "construct." It signifies the Vygotsky and Wertsch underpinnings of my study to account for students' socially mediated, two-way meaning making process, which I will discuss in the next chapter .

In conjunction with understanding the meanings that students constructed using these mediational means, my study also sought to measure how the SBG and SBRC shift impacted student motivation. This aim was intended to address calls within literature for FA to account for student motivation and the lack of research in this area with standards-based reform to date. This aim of my study is expressed in my third research question:

Research Question 3: *How does the standards-based grade and report card shift impact student motivation?*

I also aimed to investigate if Research Question 2 or Research Question 3 were impacted by the form of the SBRC. As I will discuss in chapter five, SIS chose to use a hybrid SBRC in the majority of the middle school grade levels, and a different SBRC form for others. This presented the opportunity to compare the responses of these different grade levels in the data collected for both of these research questions to investigate if the form of the SBRC impacted results. I hoped that this information could provide a valuable addition to SBRC literature.

My data collection of student perspectives through questionnaires and focus groups lasted over the course of two school years at SIS: the final year with traditional letter grades and report cards, and the first year that standards-based grades and report cards were implemented. In later chapters I will reflect on this and question if my design allowed enough time for systemic assessment reform to take hold at SIS.

I close this section on the aim and scope of my study by recognizing that I only investigated one American international school, and further, this was a school in which I was both a full-time teacher and researcher. SIS charged hefty tuition rates and correspondingly, the students who attended SIS were typically from affluent and high socio-economic backgrounds. As an international school, these students also represented a unique blend of nationalities and cultures. The case within this study is not intended to represent a sample to be generalized to a larger population of schools. Instead, I aim to provide a rich description of the school site within this study and to situate my findings within broader theory and research to form analytic generalizations (Yin, 2018), a stance I will discuss in much greater depth in chapter four. Finally, my role as an insider at this institution provided many affordances as a researcher, but it also came with inherent constraints as I collected data from students and senior leadership who

knew me first and foremost as a teacher. The data and findings from this study are constructed and situated in this context and different than the data which would have been collected from a researcher who was an outsider to the institution.

1.5 Significance of Study

In answering Research Questions 1 & 2, my thesis explores the extent to which the formative rationale behind standards-based grades and report cards holds true from the student perspective.

My thesis constitutes a new application of Wertsch's work on mediational means and mediated action. To date, previous studies using Wertsch have not situated grades and report cards as mediational means to better understand the affordances and constraints they place on the mediated learning of students.

In answering Research Question 3 and situating the findings within broader SDT theory and research, my thesis addresses calls to account for the motivational impact of formative assessment.

A key under-researched aspect of standards-based report cards is considering the form that they should take. As I have discussed, many schools use hybrid SBRCs which include an aggregated letter grade for each course. It is unclear how the inclusion of the letter grade symbols within hybrid SBRCs impacts student motivation and the meaning students construct of these mediational means. The findings of my study serve a practical purpose in investigating if one form of SBRCs is more ideal than others in supporting student learning and motivation.

Finally, my study is significant because of its focus on students' perspectives of grades and report cards. As I have discussed in the previous sections of this chapter, this perspective is drastically underrepresented in the field and it is essential to understanding the impact of these mediational means on students.

1.6 Structure of this Thesis

In the next chapter I will provide an in-depth account of the theoretical underpinnings of my study. I first examine assessment theory to consider the role that assessment plays within student learning and specifically what it means for an assessment to function formatively. After exploring the gap metaphor within FA, I identify the specific FA process I theorize to take place within students' use of SBGs and SBRCs: self-assessment with criteria rubrics. Next, I examine the concept of mixing the uses of assessment data. This examination will show that beyond a formative and summative dichotomy, it is possible for assessment mediational means – such as grades and report cards – to function both formatively and summatively.

As I mentioned earlier in this chapter, FA theory is predicated on the belief that student learning is a socially mediated, two-way process. To understand this learning dynamic I will next discuss the work of Vygotsky and Wertsch. This discussion will focus on two key concepts: the unique way each individual constructs understanding and the socially mediated nature of learning. I then use Wertsch's extension of Vygotsky's work on mediational means to consider how mediational means are produced and consumed, focusing on the affordances and constraints these means place on the mediated action of human agents.

Finally, in the second chapter I will use SDT to account for the energy and forms of regulation students bring to the classroom. I begin with an overview of the SDT basics and then more closely examine its sub-theory with particular importance to this thesis: Organismic Integration Theory (OIT). OIT posits that environmental support of an individual's basic psychological needs determines the degree to which that person experiences autonomous motivation, with meaningful differences within extrinsic states of motivation. I then examine research on the educational benefits for students who experience more

autonomous forms of motivation, before closing with an examination of the specific SDT implications of assessment feedback and grades.

In chapter three I will examine the empirical element of this study, grades and report cards, and situate them within the theoretical framework established in chapter two. I begin this chapter by reviewing the history of letter grades using Wertsch's framework. This examination will reveal that letter grades are infused with historical, cultural and institutional forces which precipitate a summative function of these means. Next, I review research which has revealed that letter grades are constructed in opaque ways and used within schools to control student behaviour, something which should lead to less autonomous forms of motivation and further enforce their summative nature. After reviewing letter grades, I will examine the history of standards-based reform and its corresponding reform movement within grading and reporting which is explicitly predicated on unlocking a formative function within these mediational means. I then examine literature informing how teachers use standards to design their units and then review how standards-based grades and report cards are intended to be constructed. I will close the chapter by reviewing the initial and limited research into standards-based grades and report cards.

In chapter four I will detail my study's mixed methods design, as dictated by my research questions. I begin with an overview of case studies and articulate how my robust description of the school site in chapter five, combined with the findings of my study, allow for analytic generalizations when situated within a broader body of theory and research. My Parallel Mixed Design allows for a meta-inference in later chapters to answer Research Question 3 in greater depth. After this, I consider research quality and issues of reliability and validity as they pertain to my mixed methods case study.

After an overview of mixed methods, I will detail my rationale for using focus groups to generate the data I will analyse to answer Research Questions 1

& 2. This section of the chapter provides an in-depth accounting of how I designed and implemented these focus groups. Next, I discuss the questionnaire used to investigate Research Question 3. I begin this section by discussing the rationale for the quasi-experimental design of this section of my research. Next, I describe how the Academic Self-regulation Questionnaire (SRQ-A) was distributed and the corresponding data collected. I then discuss the interview methods I used in gathering data from senior leadership at SIS, data intended to provide context to student perceptions and to facilitate my later analytic generalizations by providing a thicker description of SIS' standards-based reform initiative. I will close this chapter with a discussion of research ethics and the measures I put in place to ensure that my research was conducted ethically.

In chapter five I will provide a detailed overview of SIS and the middle school's standards reform initiative. After an overview of the school, I discuss the logistics and coding process of my interviews with three members of the senior leadership team who coordinated and led the standards-based reform at SIS. I then examine data from the school's 2016 self-study, conducted on the eve of my focus group and questionnaire data collection. This self-study paints a picture of a school in the midst of transition, with some teachers feeling unprepared for the upcoming SBG and SBRC shift. Next, I will analyse and discuss the results of my senior leadership interviews. This discussion reveals the conflicting nature of the rationale behind the standards-based reform, and the institutional, cultural and historical forces that influenced the school's decision to use of a hybrid SBRC. Finally, I will detail the specific processes the school expected teachers to follow when constructing SBGs and SBRCs.

Following my overview of SIS' standards-based reform, in chapter six I will answer Research Question 1: *What meanings do students construct from traditional letter grades and report cards?* I begin this chapter by providing an

overview of the iterative coding process I employed with focus group data collected over two years. Next, I examine the key themes that arose across focus groups with the application of the coding framework: student perceptions of parent influences with letter grades; the behavioural constructs that students perceived to both go into letter grade formulation and the feedback they took from letter grades; students' view of letter grades as fixed labels of intelligence and future success; and the norm-referenced ways students defined letter grades. I will close this chapter by situating these findings within broader theory and research. These analytic generalizations establish letter grades as traditional summative measures whose meaning continues to be heavily shaped by cultural, historical and institutional forces.

In chapter seven I will move to answering Research Question 2: *What meanings do students construct from standards-based grades and report cards?* Within this chapter I analyse and discuss key focus group themes in relation to the standards-based shift which arose from the application of the coding framework described in chapter six. Unlike the monolithic ways students perceived letter grades and report cards, their perceptions of SBGs and SBRCs revealed greater variation in meaning. One construction was of SBGs and SBRCs as formative feedback through a clearly articulated link between student work, the standards-based grade, and the corresponding rubric. Another construction was of SBGs and SBRCs as a frustrating and opaque extra step to get to the letter grade. After examining these two constructs, I will discuss how they were roughly mirrored in student perceptions of parent influence with standards-based grades and a division in their norm-referenced meaning of standards-based grades. Next, I compare the influence of society and media with standards-based grades to letter grades, before finally discussing how students perceived teachers' implementation of these new practices. I will close this chapter by situating these findings within broader theory and framework. The analytic

generalizations which follow suggest cautious hope for the formative potential of SBGs and SBRCs, while also considering why this potential was not reached for all students.

In chapter eight I will answer Research Question 3: *How does the standards-based grade and report card shift impact student motivation?* After an overview of questionnaire results, I detail my plan for analysis of questionnaire data using effect size (ES) and confidence interval (CI) measures. Next, I provide an overview of the process I used to compute these measures and share the results of the two years of questionnaire data collection: the standards-based shift in grades and report cards resulted in very small changes to student motivation when measured by ES. To provide a deeper understanding of these ES figures, I then analyse focus group data through an SDT lens to consider the main student constructs of standards-based grades and report cards as formative feedback and as an opaque extra step to their letter grade. This analysis is then integrated with the questionnaire data to form the meta-inference that the variations in students' meanings of standards-based grades and report cards led to offsetting effects on student motivation. I will close the chapter by considering the nature of quasi-experimental designed studies investigating assessment reform and the need for these studies to allow the time necessary for reform efforts to take hold.

I will close this thesis with a concluding chapter which summarizes my findings and then considers the challenges of my study and areas for future research.

Chapter Two: Theoretical Framework of Study

2.1 Introduction

In this chapter I will establish a broad theoretical framework of assessment, learning and motivation in order to better understand and contextualize the meanings that students construct from grades and report cards and the impact of the standards-based grades and report card shift on student motivation. I will start by examining assessment theory. After addressing summative assessment (SA), formative assessment (FA) and the notion of the learning gap, I will identify the FA processes this study theorizes SBGs and SBRCs to sit within: self-assessment with criteria rubrics aligned to grades and report cards. Following this, I will consider why the SA/FA dichotomy is unhelpful and look at criteria-based assessment systems in which the distinction between the two is blurred, with all assessment criteria-based and aligned to play a formative role. This assessment theory section will conclude with an examination of FA reform issues which call for FA to account for how students learn and student motivation.

To account for how students learn, I will use Vygotsky and Wertsch's work on the socially mediated nature of learning. In particular, I will focus on two key aspects of their work. First, I will use their anti-deterministic stance to theorize the two-way meaning-making process between students and assessment feedback through grades and report cards. Secondly, Wertsch's explication of the cultural, historical, and institutional forces which shape mediational means allows for a better understanding of how these means influence the ensuing mediated action of humans, which suggests that assessment tools should be examined for the sociocultural forces which shaped them. This framework is tied to my methodology I will discuss in chapter four,

and it provides a framework for analysing SIS' standards-based reform in chapter five and focus group results in chapters six, seven and eight.

In the last section of this chapter I will examine Self-Determination Theory (SDT) and discuss how FA is aligned to this theory of motivation. I will examine the basic tenets of SDT and its sub-theory, Organismic Integration Theory (OIT). This will establish how the support of students' basic psychological needs results in more autonomous forms of student motivation and will be essential for making sense of the questionnaire instrument used within my study to measure student motivation. Following this overview, I will consider SDT research which has investigated the specific role of autonomy and competence within students' motivation and learning. In general, this research shows that autonomy and competence supportive conditions in the classroom result in more self-determined forms of motivation in students and positive educational gains. Next, I will focus in on the SDT research which has examined assessment feedback and grades in particular. This research identifies assessment feedback as either informational or controlling. Informational feedback supports student autonomy and competence and is aligned to formative conceptions of assessment, such as self-assessment with rubrics, while controlling feedback is linked to traditionally summative forms of assessment, such as letter grades.

2.2 Assessment Theory

Assessment of student learning in education is "the process of gathering, interpreting, recording and using information about pupils' responses to educational tasks" (Lambert & Lines, 2000, p. 4). Within contemporary literature, assessment has been theorized as two distinct types: summative assessment (SA) and formative assessment (FA). SA, or Assessment of Learning, is the historically dominant model of assessment within schools (Shepard, 2000). A summative use of assessment data functions as an audit of student learning to verify and categorize what has been learned, usually occurring at the end of a

unit or teaching cycle. Because of this focus on past learnings, a summative use of assessment data traditionally does not guide future steps in learning (Cizek, 2010; Earl, 2003).

2.2.1 Formative Assessment and the Learning Gap

FA involves using assessment data to inform future steps of teaching and learning while that learning is still taking place (Harlen & James, 1997; Lambert & Lines, 2000; McMillan, 2010). The Assessment Reform Group (2002) defines FA as:

> …the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there.

Important to this definition is the identification of FA as a process. Within this process, any assessment data can function formatively if it is feeding back into the process of learning (Andrade, 2010b). Thus, the function assessment data serves within larger processes of teaching and learning is a key distinguishing characteristic in the SA/FA divide (Black & Wiliam, 2018; Harlen, 2006; Wiliam, 2010).

Embedded in the Assessment Reform Group's definition are the two generally recognized purposes that FA serves. First, FA uses data from assessments within the learning process to provide students with feedback while they still have further revision opportunities. Feedback is situated within the concept of the learning gap, which will be discussed shortly. Secondly, it provides teachers with data to make instructional adjustments based on the current understanding of students (Andrade, 2010a; McMillan, 2010).

While both teachers and students play important roles in the process of student learning, some conceptualizations of FA have lost sight of students and focused on FA as a process to help teachers make instructional adjustments (Flaitz, 2011; Stiggins, 2005; Taras, 2009). Within this study, I focus on the first purpose of FA because the role of students is essential to FA (Sadler, 1989;

Stiggins, 2010; Swaffield, 2011) as assessment data does not truly become formative unless it is used by students to close the gap between current and targeted learnings (Wiliam, 2010), which is why students are the key users of FA data (Cizek, 2010).

In response to the ways FA has been implemented in some contexts with a sole focus on teachers' instructional adjustments, the phrase Assessment for Learning (AfL) was used to identify the FA student-centred focus on the learning gap (Assessment Reform Group, 2002; Baird et al., 2017). Some have gone further to highlight the FA goal of students taking evaluative power and generating their own feedback through self and peer assessment (Sadler, 1989, 1998). This strand of FA is sometimes referred to as Assessment as Learning (AaL) (Dann, 2014; Earl, 2003). Because this study took place within an American international school and looks at SBGs and SBRCs, which are U.S. initiatives, I use the phrasing most commonly used in the U.S., FA. In doing so, I also evoke the student-centred learning gap focus of AfL and the AaL focus on student ownership of assessment processes.

A key concept central to FA is the learning gap metaphor. Sadler (1989) identified three elements for the FA feedback process to result in learning: 1) a concept of the learning target 2) the ability to compare current level of performance with the learning target 3) taking appropriate action to close the gap between current and desired standing. Black and Wiliam (1998) echoed Sadler in identifying three prerequisites for feedback to lead to learning: 1) recognition of the desired goal 2) evidence about present position 3) understanding of how to close the gap between 1 and 2. This gap metaphor was again reiterated by Hattie and Timperley (2007) who identified three questions that effective feedback must answer: 1) Where am I going? 2) How am I going? 3) Where to next? All of these conceptions of the gap metaphor are predicated on a feedback loop in which assessment data is contrasted against clear and fixed

learning criteria to identify students' current standing, from which next steps are formulated and acted upon to close the gap.

There are three key criticisms of the gap metaphor and assessment theory which are relevant to my study. First, FA is missing specific processes for application in the classroom which leaves it open to varied interpretations in practice. In response to this critique, I will provide the rationale for the specific FA self-assessment processes I theorize to take place in this study. Secondly, FA and SA have sometimes been constructed as a dichotomy which is unhelpful for creating healthy assessment systems. In response to this critique I will examine the blurred nature of FA and SA and consider how assessment systems, such as the standards-based system in this study discussed in the next chapter, can create cyclical processes of learning. Finally, the learning gap is overly simplistic and fails to account for broader processes of learning and motivation. I will examine how this critique takes relevance in varied FA implementation and then will proceed to situate FA within Vygotsky and Wertsch's theories on mediation to account for student learning, and SDT to account for student motivation. I will now address each of these critiques in turn.

2.2.2 Operationalizing FA as Self-Assessment with Criteria Rubrics

One criticism of FA is that it has not focused on everyday classroom processes, which has left FA implementation open to a wide variety of interpretations and led to calls for FA processes to be more clearly specified (Bennett, 2011; Taras, 2009; Torrance, 2012). Within this critique, both Bennett (2011) and Taras (2009) cited the implementation of five FA features identified by Black and Wiliam as an example of clarified FA processes. These five articulated FA processes are: 1) Shared Criteria 2) Developing class talk and questioning 3) Appropriate feedback 4) Peer-assessment 5) Self-assessment (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Black & Wiliam, 1998). These five features represent an important step towards the concretization of FA processes, yet even

these five can be combined and used in different ways and to different effects (Panadero & Jonsson, 2013).

Within this thesis I argue specifically for FA as a process where criteria are clearly articulated in a standards-based system with assessment rubrics, grades and report cards aligned to the criteria standards. Guided by teacher structures and feedback through rubrics, SBGs and SBRCs, students gain control over their learning gaps through self-assessment and corrective action. I will now examine the theoretical foundations of this process more closely, starting with an overview of broader self-assessment research before examining self-assessment with rubrics.

Self-assessment is "the evaluation or judgement of 'the worth' of one's performance and the identification of one's strengths and weaknesses with a view to improving one's learning outcomes" (Klenowski, 1995, p. 146), and it is widely recognized as an integral component at the heart of FA (Andrade, 2010a; Black & Wiliam, 1998, 2006; Brown & Harris, 2013; Lambert & Lines, 2000; Panadero & Romero, 2014; Sadler, 1989). While Klenowski's definition suggests an FA process with feedback, Andrade & Brown (2016) distinguished between summative self-assessment - the process of students' placing summative judgments on the merit their work, and formative self-assessment - making judgments which inform and lead to next steps in learning. Within this thesis I use the standard term, self-assessment, but with a focus on what Andrade & Brown identify as formative self-assessment.

Self-assessment requires evaluation skills which teachers should explicitly teach to transition students from relying solely on teacher-generated feedback to creating feedback of their own (Black et al., 2003; Wylie & Lyon, 2015). Doing this enables students to acquire the "guild knowledge" (Sadler, 1989, p. 126) of assessment evaluation traditionally held by teachers (Black & Wiliam, 1998; Earl, 2003; Sadler, 1998), and allows them to become both producers and consumers of

feedback (Andrade, 2010a). Self-assessment then is somewhat of a misnomer as the process typically requires structure and guidance from the teacher, and the degree to which it is a self-generated process is open to interpretation. Such is the reality of the mediated nature of learning where mediational means – such as self-assessment – are shaped by cultural, historical and institutional influences, which then in turn shape individuals who also use these mediational means in unique ways. I will return to discuss this mediated action view of sociocultural learning in more depth later in this chapter.

These dual roles of student and learning environment are also seen in self-assessment research findings. In general, studies have found a positive association between self-assessment and learning achievement, with teacher feedback to guide self-assessment judgments generally improving student performance (Brown & Harris, 2013). Other research suggests that the effect of self-assessment depends on the individuals with whom it is implemented, including students' understanding, culture, and relations with peers and the teacher (Andrade & Brown, 2016). Within understanding, a key unanswered question in research is if students need a minimum degree of understanding in the content of the subject in order to meaningfully assess themselves (Panadero, Brown, & Strijbos, 2016). While assessment mediational means can influence individuals and their learning, individuals also interact with those mediational means in particular ways based on their own unique identity.

These findings represent a broad field of self-assessment research, but it is important to recognize that much of the self-assessment research has focused on summative self-assessment investigating the accuracy of students' self-rating (Andrade & Brown, 2016). Similar to criticisms of FA's vagueness, student self-rating is just one example of a wide variety of tools and practices classified as self-assessment (Andrade & Brown, 2016; Brown, Andrade, & Chen, 2015) whose implementation takes place in a variety of ways (Brown & Harris, 2013). The

relationship between these different methods is unclear within previous studies (Brown et al., 2015) and there is no uniform agreement on a standard process in self-assessment (Panadero et al., 2016). Because self-assessment is an umbrella term, it is important for research to identify the particular forms being studied. I will now examine research related to the specific self-assessment process of feedback generated through self-assessment with rubrics linked to criteria standards and leading to further student revisions.

Clear criteria are needed for self-assessment (Brown et al., 2015) and with the rise of both FA and standards-based reform, rubrics[2] have become a common way teachers communicate criteria to students (Brookhart & Chen, 2015). A rubric is "…a coherent set of criteria for students' work that includes descriptions of levels of performance quality on the criteria" (Brookhart, 2013b, p. 4). Rubrics are recognized as an important FA tool (Andrade, 2010a; Brookhart, 2013b; Cizek, 2010; Panadero & Jonsson, 2013) which can be used to support self-assessment processes (Andrade et al., 2010).

This rubric/self-assessment combination allows students to consider the quality of their work against criteria objectives (Brown et al., 2015; Brown & Harris, 2013) and to use that data to inform next steps in closing learning gaps (Andrade, 2010a; Brookhart & Chen, 2015; Panadero & Romero, 2014; Ross, 2006). While this process requires teachers to provide feedback and instruction to guide student evaluative judgements and use of rubrics (Andrade, Du, & Wang, 2008; Brown & Harris, 2013), the rubric/self-assessment combination should allow students greater autonomy and a higher degree of control over their own learning gaps, a shift consistent with FA theorizations of centring the student and transforming traditional classroom roles and hierarchies (Black & Wiliam, 1998; Earl, 2003; Sadler, 1989; Swaffield, 2011).

[2] Please see Figure 3.2 in chapter three for an example.

While rubric-based self-assessment has been theorized to align and support a formative function of assessment, research on rubrics merits a degree of caution. Brookhart & Chen's (2015) review of the rubric literature recognized that as a body of research literature it "…is beyond its infancy but not yet mature" (p. 361), but that it suggests rubrics have a generally positive effect on student performance. They identified that rubrics are not an FA method by themselves, but instead a tool that needs to be combined within larger FA processes, and rubric studies often mix rubrics with varied FA techniques. Similarly, Panadero & Jonsson (2013) concluded in their review that rubrics may positively influence learning, but that it was unclear which self-assessment processes were needed in combination with rubrics to have the greatest effects.

Three studies in particular provide glimpses of the formative power of combining rubrics with self-assessment and have importance to my study for their theoretical generalizability. Andrade, Du & Myeck (2010) studied 162 middle school English students in the U.S. They found that students who looked at a model essay and discussed the rubric prior to self-assessing drafts had higher quality final writing products than those students who did not look at a model or self-assess with rubrics. Self-assessment with rubrics was guided by teachers who had students colour code key aspects of the rubric and then colour code the corresponding sections of their essay. If students did not have particular criteria in their writing, they wrote notes with action steps for later drafts. While the study's findings are promising, the treatment group's combination of viewing model essays along with the rubric-based self-assessment process raises questions as to the role of each variable in improving student writing. Also of importance, the detailed and specific nature of this taught self-assessment process with the rubric (colour-coding; notes for action steps) underscores the fact that self-assessment with rubrics can take many

idiosyncratic forms which represents a little explored corner of the black box of assessment.

In another study, Andrade & Du (2005) conducted focus groups with 14 U.S. undergraduate teacher education students to investigate their interaction with rubrics during the assessment process. Although the ages of the participants are different from the ages of the students in my study, the findings have direct importance to my theorization of self-assessment with criteria rubrics as an important FA process. Students reported using rubrics at all three steps of learning gap process: identifying criteria, comparing their work with the criteria, and taking action to close that gap. Some students also perceived their teacher's rubric-based evaluations as meaningful feedback to guide revisions towards those criteria on future assignments. Students contrasted teacher feedback through rubrics with receiving letter grades, which they felt did not provide meaningful formative information.

Finally, in another study of 100 undergraduates from a U.S. university, students who were given a rubric after a first draft of writing and asked to use it for evaluating and revising were able to self-evaluate and make meaningful revisions to their work. The authors suggested that learners at higher academic levels are more effective at self-assessment (Lipnevich, McCallen, Miles, & Smith, 2014), which raises questions about intrapersonal factors that influence the self-assessment process.

These findings imply that although rubrics can function summatively in providing final evaluations, they also can function formatively and offer informational feedback for future learning, a phenomenon called mixing (Brookhart, 2010). As I will discuss in the next section, criterion-referenced rubrics are essential for the alignment of traditionally summative assessment mediational means and formative assessment processes.

2.2.3 The Formative Function of Summative Assessment Mediational Means

To understand how summative mediational means can function formatively, it is helpful to consider criticism of dichotomous conceptions of FA and SA. This dichotomy oversimplifies the relationship between the two and implies that SA had no role to play in supporting learning (Biggs, 1998; Lau, 2016). Instead, within good assessment practice, SA should have a secondary learning support function (Bennett, 2011). Defining FA and SA through the purpose the assessment data is used for, or the inferences being drawn from that data, allows for this nuanced distinction (Black & Wiliam, 2018). This opens up the possibility that the same data could be used for both purposes, and that traditionally summative mediational means, such as grades and report cards, can function formatively (Black & Wiliam, 2018; Brookhart, 2010; Harlen, 2006). This notion of mixing FA and SA is supported by the classroom reality that some students don't distinguish between the two (Brookhart, 2001), and that some teachers find clear-cut distinctions unhelpful and unrealistic (Black et al., 2003). Even within the traditional gap conception of FA, closing the gap between current and desired learning requires teachers and students to first make an evaluative summative judgement about current standing in relation to the criteria. As such, FA includes SA (Taras, 2005, 2009).

What emerges in contrast to an FA/SA dichotomy is a much more complex relationship between these two functions of assessment data, a relationship that often blurs the lines between the two constructs and has the potential to create a cyclical process which aides student learning. Rather than rigid distinctions between the two constructs which are unhelpful for learning (Bennett, 2011; Harlen, 2006), assessment systems should make it possible for information gathered by teachers and students to be used for both FA and SA (Biggs, 1998). I will now examine research which sheds light on the mixing of FA and SA.

The King's -Medway-Oxfordshire Formative Assessment Project (KMOFAP) found that 24 math and science teachers in the U.K. were able to "use the aftermath of tests as an opportunity for formative work" (Black et al., 2003, p. 55). Students used completed tests to engage in peer marking, peer assessment, and self-assessment to identify next steps towards criteria objectives. This use of a traditionally summative assessment mediational means for formative purposes demonstrates that the same assessment can be used both formatively and summatively.

Brookhart (2001) interviewed 50 high achieving U.S. high school students from English and anatomy classes. She found that through a process of self-assessment, students used the same assessment data from essays and a science lab in both formative and summative ways as they worked to apply what they had learned in class to future areas of study, such as college or professional jobs. Brookhart recognized that an important condition for this mixing was that assessments were clearly criterion-referenced and that students understood the connection between the assessments and the curricular criteria. In the integration of these two assessment functions, students saw their learning as a continuous process where "…summative judgements were temporary stops along a learning path" (p. 167). Students' integration of FA and SA allowed them to construct a "cyclical process" (p.167) of learning for themselves, an example of formative self-assessment with the aid of criterion-based teacher feedback.

Is it possible for educational institutions and teachers to embed these cyclical processes of learning within assessment systems for all students to experience? Similar to Brookhart, Biggs (1998) recognized that "…there is a powerful interaction between FA and SA that could usefully be incorporated in an overall synthesis, so that …[feedback from both] are conceptualised within the same framework" (p. 106). Biggs outlined conditions for this FA/SA

synthesis. One necessary condition was for assessments to be "deeply criterion-referenced" (p. 107). Another condition was alignment between the criterion-referenced assignment and the *preceding* studied content and instruction. Biggs also acknowledged that summative grades can hold formative power when they are "…awarded based on a qualitatively derived hierarchy…" (p. 108), a hierarchy found in rubrics. Others have recognized the power of summative assessment aligned to the *proceeding* content to be studied by identifying that FA cycle lengths can take place across marking periods if the criteria of the assessments are continuous and give students the opportunity for further revision (Black et al., 2003; Wiliam, 2010).

In summary, to effectively mix FA and SA, assessments must be criterion referenced (Biggs, 1998; Black et al., 2003; Brookhart, 2001, 2010) and the timing and alignment of the different assessments within student learning cycles must be carefully considered (Biggs, 1998; Brookhart, 2010; Wiliam, 2010). All of these mixing requirements are present within the standards-based assessment, grading and reporting system that will be discussed in the next chapter. FA reform has the potential to reshape classroom assessment into cyclical processes of learning in which all assessment data, even that from traditionally summative mediational means, can function formatively and support students' self-assessment. While this potential is a hopeful thought, studies of FA reform suggest there is much work to be done to reach this goal.

2.2.4 Implementation of FA in the Classroom: The Need for Theories of Learning and Motivation

A final criticism of FA that has deep implications for my study is that the learning gap metaphor fails to account for the learner's role within the complex feedback learning process by reducing this process to a series of steps directed to students which they then fulfil (Dann, 2014; Hargreaves, 2011; Perrenoud, 1998; Torrance, 2012). This has led to the recognition that FA needs to account for

student learning and student motivation (Assessment Reform Group, 2002; Baird et al., 2017; Black & Wiliam, 2006; Wiliam, 2017). As Dann (2014) states, a fully articulated theory of FA "requires a more in-depth understanding of learning and the learner rather than merely recognising a communicative feedback loop within assessment" (p. 154). This lack of learning and motivational theory has resulted in teachers implementing FA processes according to pre-existing behaviouristic beliefs (Baird et al., 2017; Pryor & Crossouard, 2008; Torrance & Pryor, 2001), an issue I will now explore in more depth. When discussing FA reform studies in the following paragraphs, I have used research from the U.S. where possible to match the American international school setting of this study as closely as possible. However, because FA reform in the U.S. has been limited to a degree by the movement towards high-stakes accountability tests (Flaitz, 2011), I have also pulled from studies done in international contexts which have theoretical generalizability for my study.

To understand the fidelity of implementation issues with FA, it is helpful to reconsider the deep shift that FA represents. FA reform has implications across teaching and learning (Black, 2015; Black & Wiliam, 1998; Earl, 2003). Students are at the heart of the FA gap metaphor as active, self-determined agents in the construction of their own learning who ultimately take the steps to close their learning gaps (Dann, 2014; Harlen, 2006; Taras, 2009; Wylie & Lyon, 2015). The teacher's role shifts from being the centre point in traditional teaching models to being more of a coach on the side. This includes helping students gain knowledge, ownership and autonomy of assessment processes such as self-assessment with rubrics, with the aim of students closing their learning gaps (Cizek, 2010; Sadler, 1989, 1998; Shepard, 2000).

While FA may be theorized in this way, the work of Torrance, Pryor and Crossouard provides examples of the different forms that FA takes in implementation. Torrance & Pryor (2001) investigated teachers' FA practices at

primary schools in England and identified two main categories of classroom FA. Convergent assessment was behaviourist in nature and done by teachers to find out if students had learned predetermined objectives. Teachers held power through interactions formed by teachers initiating, students responding, and teachers evaluating (I-R-E). Divergent assessment, on the other hand, focused on what students understood and could do, and then used open-ended questions to push their thinking further. In this type of FA, students shared power and ownership of the learning process with the teacher and conversation was open and collaborative. This shared ownership created a space for student self-assessment which was not present with convergent FA.

In the work that followed, Torrance, Pryor and Crossouard have been careful to avoid presenting convergent and divergent assessment as a dichotomy, instead considering it as a continuum which teachers must move between during FA processes (Pryor & Crossouard, 2008). However, they found that convergent assessment was the dominant FA model in practice with some teachers viewing learning through transmission models, and teachers were often weak with divergent assessment practices (Pryor & Crossouard, 2008; Torrance & Pryor, 2001). Torrance, Pryor and Crossouard's findings that FA is typically implemented in behaviouristic ways which don't allow for the transfer of evaluative power to students is supported by other studies I will now discuss which have found that teachers struggle within FA reform to implement self-assessment processes.

Wylie & Lyon (2015) studied the breadth and quality of teacher FA reform using essential FA practices derived from Black & Wiliam (Black et al., 2003; Black & Wiliam, 1998): shared criteria, developing class talk and questioning, appropriate feedback, peer-assessment and self-assessment. Their two-year study included a sample of 202 high school math and science teachers from across the U.S. Participants attended an introductory two-day workshop where

they learned about FA research, participated in activities to develop understanding of the five FA strategies and concluded by formulating plans to implement the reform into their practice. There was a follow-up two-day workshop for teacher leaders on sustaining FA communities, and these teacher leaders then led monthly meetings with participants to reflect on and continue to develop FA practices. Wylie & Lyon found that in general, teachers were not able to integrate all five of the strategies in their practice, or to integrate critical aspects of the strategies. More specifically, they found that teachers' use of self-assessment (lumped together with metacognition and reflection) did not change and was rarely used in comparison to the other FA strategies over the two years of the study, despite training sessions. They concluded that teachers may need more support to integrate self-assessment into their practices, and more generally emphasized the amount of time and teacher support required to implement FA reform.

Similar to Wylie & Lyon's study, Lysaght & O'Leary (2013) also measured teachers' use of Black & Wiliam's five key FA characteristics by surveying 476 teachers across Ireland. It should be noted that within the study it is unclear what FA professional development these participants had undertaken, if any at all. They found that self-assessment practices (lumped together with peer-assessment) were only "sporadically" used, the most infrequent out of the other FA practices. Caution is warranted in the interpretation of these studies because although generally similar, both studies mixed Black & Wiliam's five essential elements together in different ways and included different techniques for each of the larger categories. While Black & Wiliam's five FA strategies may represent a positive step towards the concretization of FA practices, there is still much variation in how they can be implemented.

The issue is further complicated because the existing assessment practices of some teachers are engrained and rigid (Earl, 2003). A survey of 404 teachers

across Canada and the United States suggested that teachers hold established views on assessment that may be hard to change once they enter the teaching profession (DeLuca, Valiquette, Coombs, LaPointe-McEwan, & Luhanga, 2018). Without a thorough examination of these beliefs, it is probable that proposed FA processes which require far-reaching changes to teaching and learning (Swaffield, 2011) will come into conflict with pre-established teacher beliefs and will fundamentally warp the shape of FA reform (Lysaght & O'Leary, 2013), resulting in FA implementation that does not account for the active role of students within learning (Hayward, 2015).

These findings from research on the implementation of FA show the stark contrast in the forms that FA implementation can take in practice. They also suggest that some teachers struggle to get to FA as it has been conceptualized within this chapter with the development of student self-assessment with rubrics. I will now follow a long line of FA researchers who have used Vygotsky's theory of the socially mediated nature of learning to account for students' roles in FA (Baird et al., 2017; Black & Wiliam, 1998; Shepard, 2000; Torrance & Pryor, 2001).

2.3 The Mediated Nature of Student Learning

FA is theorized to serve as a tool which provides students with feedback about their current standing and the next steps in a broader learning progression (Sadler, 1989). This FA process is predicated on the belief that each learner constructs their own knowledge within a broader social context. Within this study, I will use the work of Vygotsky and Wertsch to establish the mediated nature of student learning and I will highlight two of their essential concepts: their anti-deterministic stance in showing how individuals actively construct meaning from mediational means in unique ways; and how Wertsch's consideration of the cultural, historical and institutional forces shaping mediational means built on Vygotsky's idea of word sense and can help to

understand the ways FA mediational means enable and constrain student learning.

2.3.1 The Agency of Individuals Within the Mediation of Learning

Vygotsky's General Genetic Law of Cultural Development stated that all higher mental functioning occurs first on the social intermental plane, and then secondly on the psychological intramental plane. He established this law through studying the developmental process in children as language moves from social speech to egocentric speech, before finally being internalized with inner speech (Vygotsky, 1986). Showing how language mediated the development of thinking laid the foundations for Vygotsky's theory of the mediated nature of higher psychological processes which distinguished him from contemporaries, such as Piaget (Cole & Wertsch, 1996). Vygotsky believed that humans only had access to the world indirectly through mediational means (Wertsch, 2007; Wertsch, del Rio, & Alvarez, 1995), and that to truly understand the mental processes of humans, one must first understand the means that mediate them (Wertsch, 1985).

Within his focus on mediation, Vygotsky avoided a deterministic view of human consciousness by considering the way humans can shape the very mediational means that shape them (Daniels, 2015). One instance of this was his example of a person tying a knot in a handkerchief as a reminder. In creating this mnemonic device, the individual was re-tooling an existing mediational means (Vygotsky, 1978). On a broader scale, every cultural environment is permeated with the actions and achievements of previous generations which ensures that meaning is built up in mediational means over the course of history (Cole & Wertsch, 1996; Daniels, 2001). Vygotsky believed that human thought was not uni-directionally impacted by the mediational means of our environment as we uniquely interact with and shape mediational means, and in doing so master our own higher mental functions (Bakhurst, 1996; Daniels,

1996a; Edwards, 2007). Because of this, the individual and the environment will always mutually shape each other, and we cannot define one independent from the other (van der Veer, 2007).

I will now give an overview of Wertsch's sociocultural extension of Vygotsky to establish his alignment with Vygotsky's anti-deterministic stance. Wertsch presupposed Vygotsky's mediated nature of higher mental functions (Daniels, 2001; Wertsch, 1998) and aimed to develop the theory by understanding how mental functioning was related to the cultural, institutional, and historical settings in which it occurs (Wertsch, 1991, 1998; Wertsch et al., 1995).

A key difference between Vygotsky and Wertsch was that to account for socioculturally situated means and the individuals using them, Wertsch agreed with Zinchenko (1985) that mediated action was the appropriate unit of analysis. This addressed a main critique of Vygotsky's work: he had designated word meaning as the unit of analysis of human consciousness when it was itself a tool that mediated human consciousness (Daniels, 1996a; Wertsch, 1985; Zinchenko, 1985, 2007). Wertsch's focus on mediated action has been criticized for not placing enough emphasis on the role of broader contexts and activity systems in higher mental functioning, a role which is brought to the forefront in Cultural-Historical Activity Theory (CHAT) (Cole, 1996; Daniels, 2001). This critique has roots in Leont'ev's criticism that Vygotsky's focus on meditational means had led him to overlook the activities being mediated (Cole & Gajdamasschko, 2007). Cole (1996) offers a middle ground in this issue by suggesting that activity and mediated action represent two points within the same process. As I will discuss, Wertsch's mediated action still accounts for broader social contexts but focuses more closely on mediational means and the human agents using them. This is particularly appropriate for my study which has foregrounded the impact of grades and report cards (mediational means) on students' (human agents)

construction of meaning (mediated action). For example, when a student views the mediational means of a grade they have received on an assignment or report card, that grade then mediates the meaning they construct from it. If that grade functions formatively, the ensuing mediated action of the student could involve self-assessment, with the student connecting that grade to the corresponding rubric to identify next steps in their learning.

I will now discuss how mediated action allows for individual agency within the process of mediation. Wertsch identified mediated action as the point of "irreducible tension" (Wertsch, 1998, p. 25) where socioculturally situated means and mental functioning meet. The mediated action carried out by individuals is always social because it involves the mediation of means that have come from cultural, historical and institutional contexts (Cole, 1997). Thus, mediated action transcends individual/social boundaries by virtue of being carried out by both the human agent and society (Wertsch, 1985, 1995, 1998; Wertsch & Tulviste, 1996).

Instead of studying mediational means and human agents in isolation, mediated action allows for a consideration of the active way each individual interacts with mediational means (Wertsch, 1994; Wertsch et al., 1995) which aligns it to Vygotsky's non-deterministic view (Bakhurst, 1996; Wertsch, 1991). The unique way individuals use mediational means results in the ability to shape the very mediational means that shape us (Daniels, 2001; Wertsch, 1998). Wertsch (1991) details this process further by using the metaphor of a tool kit to explain how individuals and cultures have access to their own set of mediational means. Because individuals don't have access to the same mediational means, their mediated action will be different. Ultimately, humans create themselves and their environments through mediated action (Wertsch, 1991) which represents a "continuous process of transformation and creativity" (Wertsch & Rupert, 1993, p. 230).

Because individuals consume mediational means uniquely, the intentions of means do not necessarily align to the mediated actions they produce, and studies of means production should be complemented by studies of their consumption (Wertsch, 1998). FA reform movements and the implementation of new FA mediational means, such as SBGs and SBRCs, must consider the unique ways students will use these means. This has implications for the design of my study that I will discuss in chapter four. If individuals and their environment are mutually shaping each other in an ongoing process, FA reform efforts should monitor and study the ways new assessment means are used by students with the expectation that the resulting mediated action won't always match the intentions of the means.

2.3.2 The Sociocultural Forces Which Shape the Production of Mediational Means

While using the mediational means from our environments in unique ways, we usually fail to recognize the role they play in our actions (Wertsch, 1998). This lack of awareness creates an illusion of perspective that often remains until a new mediational means frees human agents from pre-existing constraints, or until people go through the process of identifying the cultural, historical and institutional forces behind the means (Wertsch, 1991, 1998). This highlights the importance of examining the sociocultural forces which shape means to better understand the resulting mediated action (Wertsch & Rupert, 1993).

Vygotsky died before fully developing his theory and explaining the impact of environmental forces on higher psychological processes and mediational means (Bakhurst, 1996; Meshcheryakov, 2007; Scribner, 1985), but his late-in-life idea of word sense suggests the direction he would have gone had he lived longer (Daniels, 1996b; Kozulin, 1996; Wertsch & Tulviste, 1996; Zinchenko, 2007). He explained,

> A word acquires its sense from the context in which it appears. In different contexts, it changes its sense…The dictionary meaning of a word is no more than a stone in the edifice of sense, no more than a potentiality that finds diversified realization in speech (Vygotsky, 1986, p. 245).

With this shift to word sense, intermental functioning was repositioned within broader sociocultural contexts (Wertsch & Tulviste, 1996) and Vygotsky began to account for the cultural, historical and institutional factors that influenced meditational means (Kozulin, 1996; Wertsch, 1991).

While Vygotsky never elaborated on the formation of mediational means (Scribner, 1985), Wertsch attempted to explain the processes of their production. Mediational means are influenced, shaped and formed by institutional, cultural, and historical factors, and they become carriers of these forms of knowledge (Wertsch, 1994, 1995, 1998; Wertsch et al., 1995). Embedded in these sociocultural factors are the values and principles that regulate human behaviour (Daniels, 2006) which ensure that the cultural past will continue to influence the present (Cole, 1995). With these diverse societal forces behind their production, means rarely emerge solely in service of psychological considerations of individuals and groups (Wertsch et al., 1995; Wertsch & Rupert, 1993), or the demands of ensuing action (Wertsch, 1991). Further, mediational means are often created for one purpose and setting, and then adopted for use within another field, a process Wertsch termed "spin off" (Wertsch, 1998, p. 58). Relatedly, with the passage of time some tools continue in use because of historical precedent, even if they no longer ideally enable the people using them (Wertsch, 1998; Wertsch et al., 1995). This is partially because once tools become permanent through the actions of communities it becomes even harder to recognize their inherent constraints (Pea, 1993). The production of new mediational means usually focuses on the restrictions of previous means to free agents from constraints. Even if original constraints are addressed, these sociocultural forces make it difficult to organize new means in efficient ways for users and they often simultaneously introduce new and unforeseen constraints

of their own (Wertsch, 1998; Wertsch et al., 1995; Wertsch & Rupert, 1993). Because of spin off and precedence, the past must be viewed as inherent within the present through mediational means. With these dynamics of production, mediational means have the ability to both empower and constrain us (Wertsch, 2007).

Within my study, grades and report cards are situated as mediational means that mediate the learning of student agents. The meaning that students make of report cards represents mediated action. Wertsch's extension of Vygotsky details the importance of examining the sociocultural forces which shape mediational means. In the next chapter, I will use Wertsch's framework to examine the forces behind grades and report cards to better understand their impact on student agents. As Wertsch's theory suggests, both traditional letter grades and the new standards-based grades have been shaped by some forces which are not aligned to a formative conception of assessment, a point I will return to when analysing the meaning that students constructed of grades and report cards in this study.

2.4 Self-Determination Theory

I have just reviewed the work of Vygotsky and Wertsch to theorize the active role of students in the FA learning process and to establish a framework for examining the cultural, historical and institutional forces which shape FA means and the resulting mediated action from students. To examine how the SBG and SBRC shift impacted student motivation, I will now discuss Self-Determination Theory (SDT) to consider the reciprocal relationship between formative assessment practices and student motivation. This supports the Assessment Reform Group's (2002) call to account for the motivation that students bring to FA processes.

SDT is predicated on the belief that intrinsic motivation is innate within all humans at birth as we are inherently motivated to explore and interact with

the world around us. The degree to which the regulation of individuals' actions remains internal is dependent on environmental support of the three basic psychological needs: autonomy, relatedness and competence (Deci & Ryan, 1985; Ryan & Deci, 2000, 2017). A unique aspect of SDT is that in addition to considering the goal contents that human energy is directed towards, it also considers the goal motives that energize human behaviour in the first place (Deci & Ryan, 2000). While others have considered the constructs of intrinsic and extrinsic motivation, SDT has shown that there are different regulatory states within extrinsic motivation that have meaningful impacts on human behaviour, something not considered in other motivational theories (Deci, Vallerand, Pelletier, & Ryan, 1991).

2.4.1 Organismic Integration Theory

SDT consists of a subset of smaller theories that have been added on as the theory has been developed and refined through decades of empirical research (Vansteenkiste, Niemiec, & Soenens, 2010). The sub-theory of particular importance to this study's measurement of report card impact on student motivation is Organismic Integration Theory (OIT).

OIT considers how different behaviours and values are internalized within individuals across all types of motivation (see Figure 2.1 below). This continuum moves from amotivation - the complete lack of motivation, to more controlled forms of extrinsic motivation. It then continues towards more autonomous forms of extrinsic motivation that share similarities with intrinsic motivation, before arriving at intrinsic motivation (Ryan & Connell, 1989; Ryan & Deci, 2000). This nuanced view of extrinsic motivation is important in the context of schooling because many of the behaviours and actions that students take in school are extrinsic in nature (Deci et al., 1991; Niemiec & Ryan, 2009; Ryan & Deci, 2009).

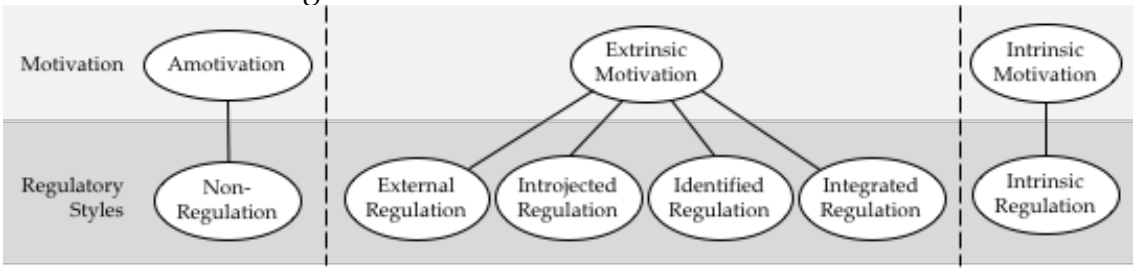
The different states of extrinsic regulation lie on a continuum of relative autonomy with each of these extrinsic regulations having its own unique and specific characteristics (Ryan & Connell, 1989). Within external regulation, individuals act because of controls put in place by others. While these external controls can lead to action, it comes from coercion that the individual does not identify with. Rewards are used to extrinsically motivate an individual, but if the reward is removed then the person will likely no longer attempt the action and move to a state of amotivation (Vansteenkiste, Ryan, & Deci, 2008).

The next controlled form of extrinsic motivation on the OIT continuum is introjected regulation. Individuals acting from this regulatory state have only partially internalized these regulations and do not accept them as their own. They act from strong feelings of anxiety to avoid guilt and shame as their self-worth is dependent on meeting these regulations of others (Deci & Ryan, 2016; Ryan & Deci, 2009).

Identified regulation is considered to be autonomous rather than controlled, as individuals feel that they are acting of their own volition when they identify with the values underlying a behaviour or norm. In this process, individuals begin to internalize beliefs and ideas and adopt them as their own because they recognize the usefulness and purposes of these regulations (Deci & Ryan, 2000; Ryan & Deci, 2009).

Finally, integrated regulation is a state that shares many qualities with intrinsic motivation as it occurs when individuals bring external behaviours into alignment with pre-existing values and behaviours already a part of their identity. (Deci & Ryan, 2000; Deci et al., 1991; Niemiec & Ryan, 2009).

Figure 2.1: OIT Continuum of Motivation



OIT posits that through the satisfaction of the three basic psychological needs of autonomy, competence and relatedness, behaviours and norms are further internalized within an individual. The higher the degree of need satisfaction, the more autonomous and internalized the form of regulation individuals experience (Deci & Ryan, 2016; Ryan & Deci, 2000; Sheldon, Ryan, Deci, & Kasser, 2004; Vansteenkiste, Lens, & Deci, 2006). Of particular relevance to this study because of their connection to an FA framework are the basic needs of autonomy and competence. Autonomy refers to "…the experience of behaviour as volitional and reflectively self-endorsed", while competence is the "…experience of behaviour as effectively enacted" (Niemiec & Ryan, 2009, p. 135). While both autonomy and competence will be examined within this study, autonomy in particular has an elevated role (Deci et al., 1991; Grolnick, Ryan, & Deci, 1991; Guay & Vallerand, 1997). According to Deci & Ryan (2000):

> …autonomy occupies a unique position in the set of three needs: being able to satisfy the needs for competence and relatedness may be enough for controlled behaviour, but being able to satisfy the need for autonomy is essential for the goal-directed behaviour to be self-determined and for many of the optimal outcomes associated with self-determination to accrue (p. 242).

Within assessment theory, there are links to how a formative function of assessment can support students' basic need for competence (Andrade, 2010a; Andrade et al., 2010; Black & Wiliam, 1998; Sadler, 1989, 1998), while much of the literature articulates the theme of student autonomy support (Cizek, 2010; Dann, 2014; Earl, 2003; Sadler, 1998). This suggests a strong potential for FA to nurture more autonomous forms of motivation within students, a point I will now outline in more detail.

2.4.2 Basic Needs Support, Autonomous Motivation and Positive Educational Outcomes

Across the SDT literature, there is empirical evidence from research in education validating the OIT tenet that support of the three basic psychological needs, and autonomy support in particular, results in more autonomous forms of student motivation and positive educational outcomes. In general, students benefit from autonomy support and suffer from controlling environments (Reeve, 2009). While there are many SDT studies measuring student motivation in the classroom, a limited number of these studies have investigated middle school age students. I will now review those studies with age groups matching or close to the middle school students of my study which establish the link between autonomy support, autonomous motivation and positive educational outcomes.

Deci, Schwartz, Sheinman, & Ryan (1981) surveyed 68 teachers from kindergarten through 6th grade to establish their classroom orientations as controlling vs. autonomy supportive and the corresponding impact on student motivation. Based on a smaller sub-sample of students from 35 teachers within the study, results showed that students with autonomy supportive teachers were more intrinsically motivated when compared with students of controlling teachers.

Soenens & Vansteenkiste (2005) surveyed students across an age range of 15-21 and found that teacher autonomy support was positively related to more autonomous and internalized forms of motivation towards school within students, which in turn was positively associated with students' grade point average. Similarly, students' perceived teacher autonomy support in grades 4-6 has been positively linked with perceived self-worth and cognitive competence (Ryan & Grolnick, 1986) and 11th and 12th grade students who perceived their teachers to be controlling had less autonomous forms of motivation and lower

academic achievement (Soenens, Sierens, Vansteenkiste, Dochy, & Goossens, 2012). Jang, Reeve, & Deci (2010) observed 133 teachers and their 2,523 students across grades 9-11 for teachers' classroom autonomy support and then surveyed 1,584 of those students for self-reported engagement. Results showed that teacher autonomy support predicted student engagement.

Ryan and Connell (1989) surveyed students in Grades 3-6 using the Academic Self-Regulation Questionnaire (SRQ-A) (see chapter four) and found that more autonomous forms of motivation were linked to enjoyment of school and positive coping strategies in response to perceived academic failure.

2.4.3 The SDT Implications of Assessment Feedback and Grades

In the previous section I reviewed research establishing that autonomy supportive classroom conditions result in students with more autonomous forms of motivation and positive educational outcomes. I will now examine the relationship in SDT between classroom assessment and students' basic psychological needs, with an emphasis on the corresponding impact on student motivation and learning. The studies reviewed will include summative letter grades as assessment mediational means, a topic I will discuss along with report cards in much greater depth in the next chapter.

Earlier in this section, I defined autonomy and competence. Now I will look at how they can be supported or thwarted in classroom assessment practices. Student autonomy is supported in the classroom when teachers lower evaluative pressures and forceful controls, and maximize student voice (Niemiec & Ryan, 2009). Competence support within educational contexts occurs when individuals are optimally challenged with tasks that are not too easy or hard (Deci & Ryan, 1985). When students are unable to develop mastery of criteria standards, their competence needs will be undermined (Ryan & Deci, 2017). These conditions for supporting the basic needs of competence and autonomy have deep implications for classroom assessment practices.

A key aspect of assessment feedback to students is whether it is experienced as informational and helpful, or as controlling (Deci & Ryan, 2016). In supporting both autonomy and competence, "…it is important that teachers provide students with the appropriate tools and feedback to promote success and feelings of efficacy" (Niemiec & Ryan, 2009, p. 139). Assessment and feedback represent important structures to facilitate student autonomy and competence. How teachers support student autonomy and competence within their classroom through the use of assessment mediational means has a direct impact on student motivation. FA through self-assessment with rubrics meets the needs-supportive conditions described within the SDT literature by providing students greater autonomy, direction and control in their learning.

SDT literature suggests that summative letter grades thwart students' autonomy and competence. Grolnick and Ryan's (1987) study of 5th grade students revealed that less controlling non-graded learning conditions led to greater conceptual learning in students, compared to increased pressure, and decreased interest and rote recall over time in more controlling graded conditions. They also found that students with more autonomous forms of motivation, as measured on the SRQ-A, had greater conceptual learning across all conditions of the study.

SDT has long theorized that the most detrimental rewards to motivation are those rewards given as a function of performance (Deci, Koestner, & Ryan, 1999) and that in educational contexts, letter grades are used as rewards and punishments and lead to less autonomous forms of motivation (Deci & Ryan, 1985; Grolnick & Ryan, 1987). As I will discuss in the next chapter, the research literature on letter grades has empirically confirmed that, in fact, many teachers use letter grades as rewards and punishments to control student behaviour (Bonner & Chen, 2009; Brookhart, 1994; Frary, Cross, & Weber, 1993; Guskey, 2009; Pilcher, 1994; Sun & Cheng, 2014). Because of this, it is reasonable to

conclude that the way letter grades and report cards have traditionally been used within the classroom works to decrease autonomous forms of motivation within students, something that in turn has a negative impact on student learning. On the other hand, the criteria-based and curriculum-aligned SBGs and SBRCs I will discuss in the next chapter have been theorized to function formatively (Guskey, 2015) and contain the essential requirements for FA/SA mixing (Biggs, 1998; Brookhart, 2001), which should result in them providing meaningful feedback to students and supporting their autonomy and competence.

2.5 Conclusion

In this study FA is operationalized through the use of student self-assessment with criteria rubrics aligned to past and future content to be studied. Within these cyclical learning opportunities, traditionally summative assessment mediational means have the opportunity to function formatively through providing feedback to guide future student learning. FA is contextualized within broader theories of learning and motivation to situate students as active and autonomous agents capable of using assessment mediational means to generate their own feedback to guide next steps in learning. Vygotsky and Wertsch's anti-deterministic framework informs us that students actively and uniquely use assessment means while also being shaped by them and the broader cultural, historical and institutional forces from which they come. This suggests that studies of assessment reform must consider students' perspectives and the unique ways they use and shape FA mediational means, a point I will return to in the methodology chapter and in my analysis of results in chapters six, seven, and eight. Wertsch's extension of Vygotsky provides a framework for my examination of the sociocultural factors that shape the formation of the assessment mediational means of grades and report cards. This framework suggests that the impact of these means on students will not match their

declared intention and points to the need to unpack the sociocultural forces that shape them to better understand their impact, a task I will focus on in the coming chapters. Finally, Self-Determination Theory further provides grounding for the FA notion of students as autonomous beings. SDT suggests that a formative use of assessment data with the specific FA process of self-assessment with rubrics should support students' basic psychological needs of autonomy and competence and thus lead to more self-determined forms of motivation within students. While summative letter grades should act as autonomy and competence thwarts, the theorized formative function of traditionally summative means, such as SBGs and SBRCs, should support these basic needs, although SDT research has not yet investigated this issue.

In the next chapter I will focus in on the empirical topic of this study, grades and report cards, beginning with a deep examination of letter grades and then turning my focus to SBGs and SBRCs.

Chapter Three: Grades and Report Cards

3.1 Introduction

In the previous chapter I established the theoretical framework for my study using assessment theory, the work of Vygotsky and Wertsch, and SDT. In this chapter, I will use this framework to examine the empirical element of this study: grades, report cards, and the shift to SBGs and SBRCs. Using Wertsch's framework, grades and report cards are situated as mediational means and the meanings student agents construct from grades and report cards represent mediated action. I will begin by examining the cultural, historical and institutional forces which have shaped traditional letter grades and report cards as a way to better understand how they constrain student thinking. Following this I will discuss the construction of letter grades and report cards to consider how these traditionally summative mediational means are constructed in ways that leave them bereft of formative information. Further, the construction and use of these letter grades have deep implications for student motivation as they represent classic autonomy and competence thwarts which SDT theorizes will result in students with more controlled forms of motivation.

After this examination of letter grades and report cards, I will review the sociocultural forces intertwined with the standards reform movement in U.S. education. Although dynamic and varied, a distinct strand of standards reform has been driven by a push for equity in education through criteria-referenced assessment, a stark contrast from the norm-referenced roots of letter grades. This strand in standards reform aligns this use of classroom standards with the FA processes discussed in the previous chapter, self-assessment with rubrics, and links it to the autonomy and competence supportive conditions theorized by SDT. Next, I will examine how standards-based curriculum design used within many internationals schools establishes the conditions needed for FA/SA mixing

through deep criteria-referencing and alignment of both preceding and proceeding learning activities and assessments. The SBG and SBRC literature has hypothesized these traditionally summative means to serve a formative function, but they are a new and under-researched topic. I will close this chapter by examining the limited SBG and SBRC research to establish the gaps in the research literature that my study aims to address.

3.2 Letter Grades and Report Cards

For the purposes of this thesis, grading is defined as teachers' evaluations of student performance, often represented by grade symbols, and reporting is how those results are communicated (Muñoz & Guskey, 2015).

3.2.1 The Sociocultural Forces that Shaped Letter Grades and Report Cards

Writing in 1913 about higher education grading systems in the United States, Isidor Finkelstein commented, "…few teachers stop to consider what the marking system under which they work really implies…" (Finkelstein, 1913, p. 5). Roughly a century after Finkelstein's comments, letter grades are one of the most entrenched traditions in U.S. education (Marzano, 2006) and are perceived by many as "…both fixed and inevitable – without origin or evolution" (Schneider & Hutt, 2013, p. 202). Investigating their history, however, reveals that far from being fixed and inevitable, grades and report cards have been continuously shaped and formed by cultural, historical and institutional forces throughout much of their existence.

Norm-referenced competition was present in grades from their outset: early grading systems in the United States followed the model of European universities and used grades as a method to rank students and motivate them through competition (Schneider & Hutt, 2013; Tocci, 2010). These university grading systems then became models for Kindergarten -12th grade (age 5-18) schools from the mid-17th century into the 19th century (Brookhart, 2004). The

common school movement of the 1800's saw the spread of free public schools paid for by taxpayers for all children. In these common schools percentage grading was widespread, and the first report cards were used to communicate student learning and behaviour to parents (Brookhart, 2004).

Great change came to grading practices with the development of a compulsory national school system, a development which roughly coincided with industrialization in the U.S. From 1870-1910, enrolments tripled in K-12 schools, and this phenomenon continued well into the 20th century. School leadership required communication tools when students moved or graduated to another level. Similarly, universities required an efficient method of communicating student achievement to deal with increased applications (Schneider & Hutt, 2013). To account for this new reality, the purpose of grades and report cards shifted from communicating student learning and behaviour, to sorting students for external institutional factors which prioritized efficiency (Schneider & Hutt, 2013; Tocci, 2010). With these changes, schools organized around the factory model with assessments and grades representing a set of gates to move to the next level (Earl, 2003). This shift in purpose from communicating student learning and behaviour to sorting and gate-keeping is a crucial moment which echoes to this day in the meaning students construct from grades, a point I will return to when analysing and discussing focus group results from my study in chapters six, seven and eight. The use of grades and report cards as tools to sort students and serve as gatekeepers to next levels of schooling represents a substantial obstacle towards feedback about learning. Far from communicating formative information, report cards and grades have roots as summative labels about student ability.

Psychological trends of the early 20th century also shaped the letter grading system. Within the mental testing movement, business and the military sought to categorize individuals through the use of precise measurements of

intelligence. These trends carried into education and led to attempts to standardize grading systems to precisely measure student achievement along tidy bell curves (Schneider & Hutt, 2013), a concept adapted from normal curve theory (Brookhart, 2004; Fendler & Muzaffar, 2008), and rooted in beliefs of fixed intelligence. By the 1940's, an A-F marking scheme embodying these psychological trends had become standardized within 80% of schools (Brookhart et al., 2016), with norm-referenced definitions (average, above average, etc.) still associated to these letter grades to this day (Guskey, 2004). The educational application of mental testing and normal curve theory within schools represents a clear example of Wertsch's spin-off (see chapter two). These means were established for non-formative purposes and without consideration for the ideal learning of students. Instead, the forces which shaped letter grade mediational means derive from competition amongst students with an emphasis on sorting rather than the belief that all students can meet learning criteria.

3.2.2 The Problematic Construction of Letter Grades

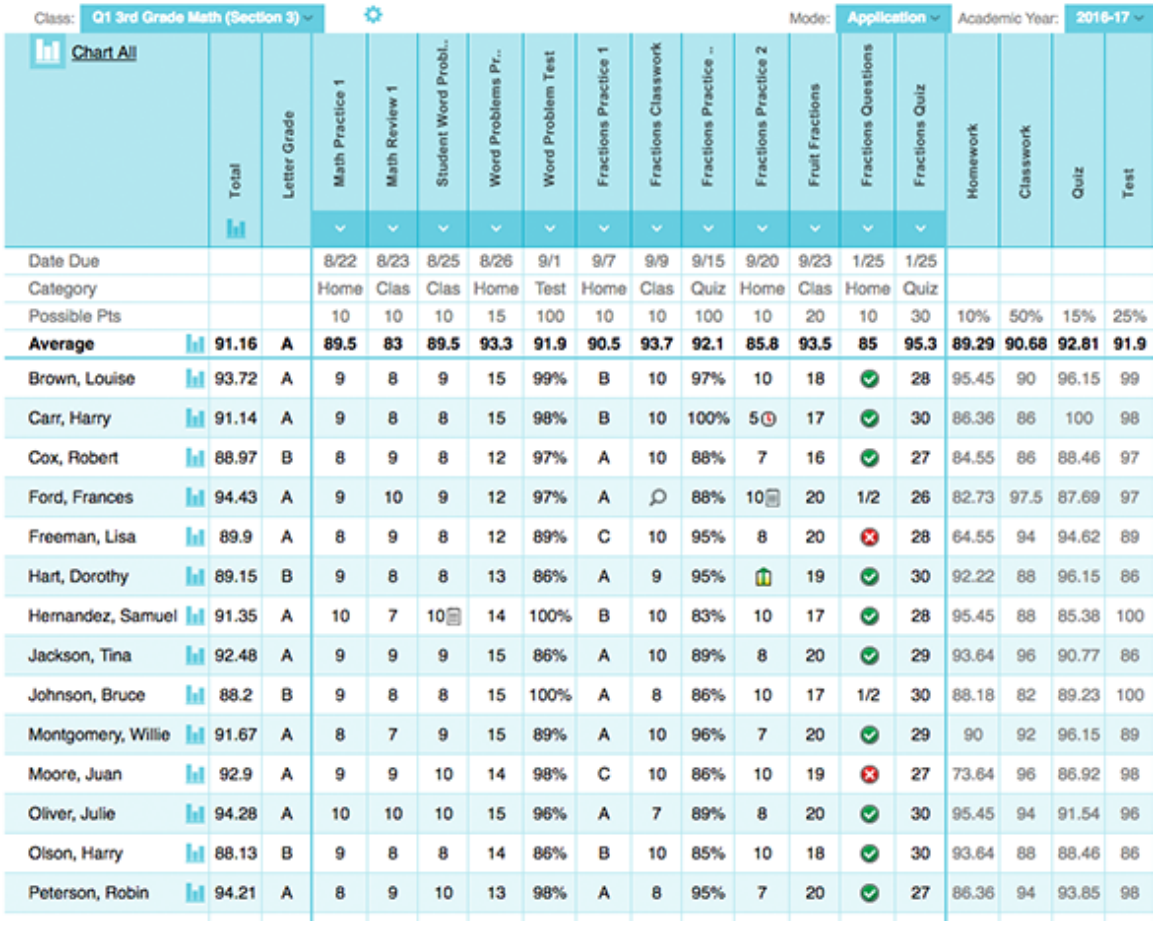
Just as most of the cultural, historical and institutional forces behind letter grades exclude a formative function, so too do the ways they have traditionally been constructed and used. I will begin this section by considering some of the general practices with letter grades construction, including norm-referencing and the process of using points from assignments to generate percentages which are then translated into summative letter grades for end of term report cards. Following this I will review empirical research into teachers' letter grade construction. This research reveals that teachers construct grades in idiosyncratic ways. While the research into these practices has considered different aspects of construction, it has focused disproportionately on investigating the elements teachers include when formulating grades, and more specifically, teachers' use of behavioural components when determining student grades.

Derived from the 20th century trends of normal curve theory and normative conceptions of intelligence discussed in the previous section, many teachers have traditionally used norm-referencing when grading students. Norm-referencing to evaluate and grade students means that student work is compared against that of other students to distinguish their overall grades, a practice which is historically more commonly used in secondary schools (Bailey & McTighe, 1996). Some teachers go as far as establishing the exact percentages of students who will receive the different letter grades associated with placement in the bell-curve, while for others, de facto curves often emerge to distinguish student performance even when they are not explicitly using bell curves to grade students (Guskey, 2015). While a common historical practice, the degree to which teachers use norm-referencing in their classrooms today is questionable. Guskey (2009) surveyed 513 teachers from across all school levels and found that 27% disagreed that grading and reporting should only be done in reference to criteria and never normed. McMillan's (2001) survey of 1,483 secondary teachers also suggested that the majority of teachers do not use norm-referencing to determine students' grades. When surveyed about the factors they used in determining grades, the average response to the item, "Performance compared to other students in the class" was "very little" (p. 24). The extent to which de facto curves remain in teachers minds even when explicit curves are not used, as Guskey (2015) claims, remains an unresearched topic. At minimum, norm-referencing appears to be used by a small but not insubstantial number of teachers, with implications for the formative function of grades which will be discussed later in the chapter with standards-based reform.

I will now describe the specific letter grading process that has most frequently been used in the formulation of grades within secondary schools (Guskey & Bailey, 2001) in order to establish a reference point for understanding SBG and SBRC reform and to also consider the SDT and FA implications. Within

traditional letter grading systems, teachers assign points to the different assignments students completed throughout a grading period (Iamarino, 2014). These points are entered into a gradebook throughout the semester, with these gradebooks organized by assignments and assignment types (Guskey & Bailey, 2001)(see Figure 3.1 below). These different assignment types are given different weighting which teachers use to generate an average for end of term overall percentages for each student (Campbell, 2012; Guskey, 1996a, 2009; Knight & Cooper, 2019; Miller, 2013)(see Table 3.1 below). A letter grade scale is then used to convert these overall percentages to a letter grade (see Table 3.2 below), which is represented on the report card for that grading period (Bailey & McTighe, 1996; Marzano, 2006). As will be discussed in the SBG section later in this chapter, this weighting and averaging by assignment type rather than learning objectives has meant that these traditional letter grades are opaque in the information they communicate about specific student standing and next steps for learning, thus limiting their ability to function formatively and thwarting student autonomy and competence.

Figure 3.1: Traditional Gradebook
(Common Goal Systems Inc., 2020)



Class: Q1 3rd Grade Math (Section 3) Mode: Application Academic Year: 2016-17

Chart All

	Total	Letter Grade	Math Practice 1	Math Review 1	Student Word Probl..	Word Problems Pr..	Word Problem Test	Fractions Practice 1	Fractions Classwork	Fractions Practice ..	Fractions Practice 2	Fruit Fractions	Fractions Questions	Fractions Quiz	Homework	Classwork	Quiz	Test
Date Due			8/22	8/23	8/25	8/26	9/1	9/7	9/9	9/15	9/20	9/23	1/25	1/25				
Category			Home	Clas	Clas	Home	Test	Home	Clas	Quiz	Home	Clas	Home	Quiz				
Possible Pts			10	10	10	15	100	10	10	100	10	20	10	30	10%	50%	15%	25%
Average	**91.16**	**A**	**89.5**	**83**	**89.5**	**93.3**	**91.9**	**90.5**	**93.7**	**92.1**	**85.8**	**93.5**	**85**	**95.3**	**89.29**	**90.68**	**92.81**	**91.9**
Brown, Louise	93.72	A	9	8	9	15	99%	B	10	97%	10	18	✓	28	95.45	90	96.15	99
Carr, Harry	91.14	A	9	8	8	15	98%	B	10	100%	5	17	✓	30	86.36	86	100	98
Cox, Robert	88.97	B	8	9	8	12	97%	A	10	88%	7	16	✓	27	84.55	86	88.46	97
Ford, Frances	94.43	A	9	10	9	12	97%	A		88%	10	20	1/2	26	82.73	97.5	87.69	97
Freeman, Lisa	89.9	A	8	9	8	12	89%	C	10	95%	8	20	✗	28	64.55	94	94.62	89
Hart, Dorothy	89.15	B	9	8	8	13	86%	A	9	95%		19	✓	30	92.22	88	96.15	86
Hernandez, Samuel	91.35	A	10	7	10	14	100%	B	10	83%	10	17	✓	28	95.45	88	85.38	100
Jackson, Tina	92.48	A	9	9	9	15	86%	A	10	89%	8	20	✓	29	93.64	96	90.77	86
Johnson, Bruce	88.2	B	9	8	8	15	100%	A	8	86%	10	17	1/2	30	88.18	82	89.23	100
Montgomery, Willie	91.67	A	8	7	9	15	89%	A	10	96%	7	20	✓	29	90	92	96.15	89
Moore, Juan	92.9	A	9	9	10	14	98%	C	10	86%	10	19	✗	27	73.64	96	86.92	98
Oliver, Julie	94.28	A	10	10	10	15	96%	A	7	89%	8	20	✓	30	95.45	94	91.54	96
Olson, Harry	88.13	B	9	8	8	14	86%	B	10	85%	10	18	✓	30	93.64	88	88.46	86
Peterson, Robin	94.21	A	8	9	10	13	98%	A	8	95%	7	20	✓	27	86.36	94	93.85	98

Table 3.1: Traditional Grading Category Weighting

Category	Weighting
Homework	15%
Participation	10%
Tests/Quizzes	30%
Essays/Projects	30%
Classwork	15%

Table 3.2: Percentage to Letter Grade Conversion Scale

Letter Grade	%
A+	98%+
A	93-97%
A-	90-92%
B+	88-89%
B	83-87%
B-	80-82%
C+	78-79%
C	73-77%
C-	70-72%
D+	68-69%
D	63-67%
D-	60-62%
F	59% and below

3.2.3 The Use of Behaviour Measures in Letter Grades

I have just described the general process many teachers use to arrive at letter grades for the end of term report cards. Amongst the issues with teacher construction of letter grades, most prominent in the research is how teachers use non-achievement behavioural criteria[3] in idiosyncratic ways when formulating a hodgepodge combined average for the final grades. Some have suggested that behaviour inclusion in letter grades is a common practice (Guskey, 2015; O'Connor, 1995, 2009, 2011) and this has been empirically found in research (Allen, 2005; Baron, 2000; Brookhart, 1993, 2013c; Brookhart et al., 2016; Campbell, 2012; Cox, 2011; Cross & Frary, 1999; McMillan, 2001; McMillan, Myran, & Workman, 2002; McMillan & Turner, 2014; Pilcher, 1994; Sun & Cheng, 2014). This inclusion of non-achievement criteria decreases the internal validity[4] of grades as an educational measurement of student learning towards criteria and contradicts the recommendation of some measurement specialists within education (Ebel & Frisbie, 1991; Oosterhof, 2001).

[3] For an example of non-achievement behavioural criteria, reference participation in Table 3.1 above.

[4] Please see chapter four for a discussion of internal validity.

McMillan et al. (2002) surveyed 901 teachers from grades 3-5 in Virginia and found that teachers include behavioural measures such as effort, participation, improvement and ability, behaviours McMillan termed, academic enabling. Behaviour inclusion is further complicated by teachers' use of different types of behaviour criteria (Cizek, Fitzgerald, & Rachor, 1995), with different weighting given to these criteria (McMillan, 2001; McMillan et al., 2002) in arriving at their final grades.

Other research suggests that behaviour measures are used by teachers as a secondary measure to help decide the case of borderline grades where a student's overall grade at the end of the semester is very close to the next grading category (Randall & Engelhard, 2009a). Randall & Engelhard (2010) surveyed 342 K-12 teachers and found that teachers were less likely to bump up a student's borderline grade if the teacher perceived the student to have poor behaviour or low motivation. Other studies suggest that teachers inflate grades with behavioural measures out of concern for their impact on students' futures (Brookhart, 1991, 1993; Randall & Engelhard, 2010; Sun & Cheng, 2014).

The most frequent finding, however, is that teachers include behaviour in summative grades as a way to control students through rewarding desirable behaviours and punishing students for undesirable behaviours (Bonner & Chen, 2009; Brookhart, 1994; Frary et al., 1993; Pilcher, 1994; Sun & Cheng, 2014). Guskey (2009) surveyed 299 secondary teachers and found that roughly half of them believed grades had some value as punishment. Further, compared to elementary teachers, secondary teachers were more likely to assign a mark of zero for missing and late work, take points off for undesirable behaviours, and to more generally use grades as a form of punishment.

Providing further insight into this use of grades as means of control, Brookhart (1993) surveyed 84 teachers from across all grade levels who were enrolled in masters level courses at Duquesne University. Brookhart found that

teachers viewed grades as the pay that students received for the work they did, and that grade rewards and punishments functioned as a key component of classroom behaviour management systems.

These findings suggest that letter grades are not simple one-dimensional measures of achievement (Brookhart, 2013a; Brookhart et al., 2016). Instead, they are complex mediational means used by teachers in a variety of ways, attempting to both help and control students. Because of these practices, the information conveyed in the same letter grade symbol can be entirely different from one class to another, and the clarity of information communicated to students about their learning becomes opaque.

Teachers' construction and use of letter grades have important theoretical implications to my study. SDT has long theorized that in educational contexts, letter grades are used as rewards and punishments and lead to less autonomous forms of motivation (Deci & Ryan, 1985; Grolnick & Ryan, 1987). Furthermore, the most detrimental rewards to motivation are those rewards given as a function of performance (Deci et al., 1999). Related to these SDT findings, the research literature on grading has empirically confirmed that, in fact, many teachers use grades as rewards and punishment as a method to control student behaviour (Bonner & Chen, 2009; Brookhart, 1994; Frary et al., 1993; Guskey, 2009; Pilcher, 1994; Sun & Cheng, 2014). Because of this, it is reasonable to conclude that letter grades and report cards, as they have traditionally been used within the classroom, typically work to decrease autonomous forms of motivation within students, something that in turn can have a negative impact on student learning.

From an FA perspective, this traditional use of grades as a behavioural control stands in stark contrast to feedback that communicates current standing to students and next steps for them to take towards meeting learning criteria. Based on the research discussed in this section, traditional letter grades have

been constructed in ways which obscure information about student learning rather than supporting it.

Attempts to reform classroom letter grading practices have been largely unsuccessful and echo the fidelity of implementation issues found with FA reform discussed in the last chapter. Many teachers in the U.S. are insufficiently trained on the best measures for increasing the internal validity of the letter grades they assign to students (Allen, 2005; Guskey et al., 2011; O'Connor, 2009). When teacher training has taken place, most studies have found that teachers who received the training did not follow the key recommendations when returning to their classrooms (Brookhart, 1993; Cross & Frary, 1999; Randall & Engelhard, 2009b). Bonner and Chen (2009) found that education measurement training led teacher candidates to soften in their support of including behaviour measures into letter grades, but the study did not follow these candidates to see the actual practices they implemented in the classroom, an important point given the poor fidelity of implementation found in other studies on grading reform. These findings provide further evidence that assessment reform is a complicated process, an important consideration for implementors of the SBG and SBRC reform which will be discussed in the next section, and a point I will return to when discussing standards-based reform at SIS in chapter five.

3.3 Standards-Based Grades and Report Cards

I will now discuss SBGs and SBRCs by providing a historical overview of the standards movement to consider the cultural, historical and institutional forces behind these grading and report card reforms in order to better understand their impact on students. After this overview, I will examine the formative implications of how teachers design units from these criteria standards, which has implications for how a standards-based system can allow traditionally summative mediational means, such as grades and report cards, to function formatively. Following this, I will review how SBGs and SBRCs are

constructed and look at the SDT and FA implications of these new means. Ultimately, SBGs and SBRCs are hypothesized to function formatively which should support students' basic psychological needs of competence and autonomy, and in doing so should result in more autonomous forms of motivation for students. This section will end by examining SBG and SBRC research.

3.3.1 The History of Standards Reform in the United States

The roots of the standards reform movement in the United States can be traced to 1983 when a federal committee issued a report titled, A Nation at Risk. This report declared that based on standardized testing measures, the U.S. was falling behind the rest of the world and facing a crisis in education (National Commission on Excellence in Education, 1983). At the heart of the issue was the lack of a coherent and rigorous curriculum across the nation's schools (Marzano, 1998), with great variation in what students were taught across different classrooms, schools and states (Marzano, Pickering, & McTighe, 1993). By 1989, President Bush and state governors agreed to a set of academic goals to be reached by the year 2000 which outlined the implementation of clear learning criteria across the nation, referred to as standards. These goals were formalized in 1994 when congress passed the legislation, Goals 2000: Educate America Act (McLaughlin & Shepard, 1995). These standards represented what students should know and be able to do (Guskey et al., 2011). By 1998, every state had standards or was in the process of developing them (Marzano & Kendall, 1998). Although standards do not represent a national curriculum, as each state has autonomy in which set of standards they adopt (Great Schools Staff, 2016), by 2004 standards had become the foundation of curriculum and assessment in all 50 states (Reeves, 2004a). The most frequently used set of curricular standards in the

U.S. are the Common Core State Standards (CCSS) for English and math[5]. In 2020, 41 U.S. states had adopted the Common Core ("Common Core State Standards Initiative," 2020). Standards for science are not as uniform as those for math and English, but in 2020, 20 states had adopted the Next Generation Science Standards (NGSS), with 24 other states opting to develop their own science standards ("National Science Teaching Association," 2020). I have highlighted these particular sets of standards because as I will discuss in chapter five, they are the sets of standards SIS adopted with its standards reform.

Standards were intended to allow students to see where they were going with their learning (Earl, 2003) which in turn was hoped to result in greater equity in education by allowing all students a greater opportunity to reach learning objectives (Marzano, 1998; Wiggins, 1993; Wiliam, 2011). Standards-based reform, with its emphasis on clear and fixed learning criteria represented a shift from the norm-referenced past of grading and assessment to a criteria-referenced future in which all students could reach learning expectations (Bailey & McTighe, 1996; Guskey, 2015).

The implementation of standards within U.S. education is a heavily politicized and controversial issue with opposing views on the form reform should take (Darling-Hammond, 2004). While there are many cultural, historical and institutional forces which have shaped the form standards have taken in practice, most publicly known is their association with high-stakes accountability tests and the No Child Left Behind Act of 2002 (Flaitz, 2011; Shepard, 2000). My purpose within this section is not to provide an overview of all of the forces behind standards reform, but instead to highlight a particular strand most applicable to American international schools which are mostly free of the pressures of high-stakes accountability tests. This strand is represented by a 1995 report issued by the National Academy of Education Panel on Standards-Based Reform and

[5] Please see chapter five for a specific example of these standards.

written by McLaughlin and Shepard. In this report, the authors detail that the foundation of standards reform was the belief that student achievement is impacted by expectations, and the call for educational equity. These beliefs were embodied in the movement's motto: high standards for all students. The paper called for new standards to move away from old instructional practices rooted in behaviourist principles and to focus on depth of understanding in student learning with real world applicability of skills needed for the 21st century. The standards were situated as criteria that would allow teachers to evaluate the current standing of students and identify next steps for improvement towards the criteria. With this view of standards, this strand of the reform movement aligned itself with formative conceptions of assessment, which in turn should support student autonomy and competence. The panel also recognized the unique way each student constructs understanding and called for standards reform to account for and be adaptable to the experiences of each student. This nuanced view of student learning aligns with Wertsch's view of mediated action as the irreducible tension between student agency and the mediational means of the classroom. It also set the foundation for standards-based assessment, grading and reporting practices to support student learning, a great shift from previous letter grading practices.

McLaughlin and Shepard (1995) also recognized that the reform movement's calls for greater equity were aspirational and not a given. The implementation of standards without the necessary time and resources, coupled with overemphasizing high-stakes accountability tests had the potential to create greater inequality amongst schools and students in the U.S. The authors cautioned that setting standards wouldn't improve education in and of itself but that it was a necessary step which required much work and teacher development beyond one-day trainings. These warnings of the time and immense effort needed for reform echo similar concerns that have arisen within FA reform (Sadler, 1998).

3.3.2 Standards-Based Curriculum Design

In the previous section I considered the cultural, historical and institutional influences at the heart of one particular strand of standards reform and the alignment of these forces to FA and SDT. I will now consider the processes of how these standards are used within many American international schools which lead to SBGs and SBRCs. While standards represent the key learning criteria of what students should know and be able to do (Guskey et al., 2011), from their onset it was recognized that standards were not a curriculum in and of themselves, but should be used to guide assessments and instruction with the goal of assessments that reflect important learning goals (McLaughlin & Shepard, 1995). Within many American international schools, teachers are expected to take the role of curriculum designers by using standards as the starting point from which they align assessments and daily instruction (Marzano, Heflebower, Grift, & Warrick, 2016). Many American international schools utilize the Understanding by Design (UbD) framework detailed by Wiggins and McTighe to "backwards design" curriculum from the standards (Wiggins & McTighe, 2005). I will now give a brief overview of this process, aided by work from other standards-based researchers. In doing so, my intention is not to give a detailed explication of UbD and all of its tenets, but to demonstrate that the standards-based curriculum design process used at SIS meets all of the requirements discussed in the previous chapter to allow for traditionally summative mediational means to function formatively.

UbD identifies three key stages within curriculum design. In stage one, teachers use standards to identify desired results from students by the end of the unit (Marzano & Kendall, 1998; Wiggins & McTighe, 2011). Because sets of standards typically include far more learning targets than is possible to reach in any given academic year, an important step for teachers is to identify priority standards within the larger set (Marzano, 2006; Marzano & Haystead, 2008;

McLaughlin & Shepard, 1995). These prioritized standards represent the essential content within a grade level or course that the majority of class time will be spent on (Heflebower et al., 2014; Marzano et al., 2016). If standards are written in such a way that they combine multiple skills, teachers must unpack the standard to identify individual skills, sometimes combining these with similar components found in other standards (Marzano & Haystead, 2008; Wiggins & McTighe, 2012). In stage two, teachers identify which assessments would allow students to show meaningful evidence of that expected learning and then proceed to construct those assessments (Fisher, Frey, & Pumpian, 2011; Wiggins & McTighe, 2011). Once priority standards and corresponding assessments are selected, teachers can begin to identify criteria levels for these standards within rubrics (Guskey, 2004; Guskey & Bailey, 2001). These rubrics are typically a four level learning progression with the goal for the standard as the third step in the progression, frequently termed "meeting" (Marzano et al., 2016) (see Figure 3.2 below). Because these rubrics are generalized and aligned to priority standards and not detailed to individual assignments, the same rubric can be used in later assignments which address the same standard (Scarlett, 2018; Wiggins & McTighe, 2011). This alignment in rubrics across different assignments allows students to use SBGs from one assignment to guide learning on future assignments encompassing that same standard. Finally, in the last stage, teachers plan daily learning experiences for students with materials which will allow them to develop desired understandings (Heflebower et al., 2014; Wiggins & McTighe, 2011).

Figure 3.2: Rubric Example from Social Studies

	Extending	Meeting	Progressing	Beginning
Gather Researched Information	I not only meet standards, but I extend my learning by demonstrating creativity and resourcefulness in collecting evidence beyond what is expected.	I gather a varied and sufficient amount of evidence related to my research questions, citing my sources.	I gather and cite my evidence, but several pieces of evidence are either unrelated to my research questions, insufficient, or could use more variety.	Several pieces of evidence are either unrelated to my research questions, insufficient, or could use more variety; In addition, I do not cite my sources.

By beginning with desired learning outcomes and working backwards to design assessments and learning activities around those standards (Marzano, 1998), the backwards design process ensures meaningful alignment within the curriculum to ensure opportunities for formative feedback which students can apply beyond one particular assignment (Fisher et al., 2011). This deeply criteria-referenced alignment of UbD meets the key requirements for FA/SA mixing discussed in chapter two (Biggs, 1998; Brookhart, 2001). If daily learning activities are not aligned to the larger end of unit assignment, or if that assignment is not aligned to the standards, FA opportunities are greatly diminished (Wiggins & McTighe, 2011). The changes required within the backwards design process to implement standards reform are far reaching across curriculum, assessment and instruction. They require teachers to reconceptualize the way they have traditionally designed curriculum by starting with daily activities and then working up towards end of unit assignments (Scarlett, 2018; Scriffiny, 2008).

The process of curriculum design described in this section requires teachers to make individual decisions at every step of the process. As such, even though classrooms or schools may use the same set of standards, these standards will never represent a standardized curriculum because context specific variables will result in a different curriculum design within every classroom.

3.3.3 Formulation of Standards-Based Grades and Report Cards

In the previous section I outlined the design process that many teachers use in American international schools to implement standards reform within their classroom. Building on this, I will now discuss the grading and reporting implications of this new way of designing curriculum. These changes will be included under the terms, standards-based grades (SBGs) and standards-based report cards (SBRCs). SBG and SBRC reform represent a later phase of the broader standards reform movement in U.S. education (Cox, 2011). As with most new mediational means that arise to address the constraints of previous means (Wertsch, 1998), much of standards-based grading reform has come in response to harmful traditional letter grading practices that obscure information about students learning towards specific criteria, such as the use of zeroes, averaging, norm-referencing, and combining behaviour measures with achievement in a single grade (Guskey, 2015; Hooper & Cowell, 2014; O'Connor, 2011; O'Connor & Wormeli, 2011; Wiggins, 1993). I will now discuss these practices in turn, before considering the broader SBG procedures that lead up to SBRCs.

The first issue from traditional grading systems that SBG and SBRC reform has sought to address is averaging assignments to determine final grades and the use of zeroes for missing or incomplete assignments. Averaging is problematic because outlier grades can disproportionately skew the average. Further, because learning is developmental, averaging points accumulated over a grading period punishes students for where they began in their learning progression, rather than recognizing their current standing at the end of the grading period (Guskey, 1996b; Marzano, 2006; O'Connor, 2011). Another troubling aspect of averaging is the common classroom practice of using of zeroes for incomplete or missing assignments (Campbell, 2012; Guskey, 2009; McMillan, 2001). These zeroes distort the overall average so that the final grade

does not accurately reflect student learning (Marzano, 2006; O'Connor, 2011; Reeves, 2004b), thus lowering the grade's internal validity (Guskey & Bailey, 2001). Averaging distorts the information delivered in report card grades from communicating students' current standing to providing a generalized symbol of achievement across a longer period of time, information that is not as helpful within an FA framework (Hooper & Cowell, 2014). As such, SBG and SBRC reform calls for teachers to eliminate the practice of using zeroes for missing or incomplete assignments, and for teachers to use other measures beyond mean for determining final grades (Reeves, Jung, & O'Connor, 2017).

The second predominant issue with traditional grading which SBG and SBRC reform has addressed is the practice of norm-referencing. Norm-referencing is problematic because it means that even if all students develop understanding of learning criteria, only a few can receive the top grades (Brookhart, 2004). It is also directly opposed to the criteria-referenced belief of standards-based education that all students can learn when measured against fixed learning outcomes (McLaughlin & Shepard, 1995). Although research suggests that norm-referencing is used by a small minority of teachers (Guskey, 2009; McMillan, 2001), this practice works to prevent grades and report cards from delivering feedback to students about their performance towards academic criteria, and instead tells them only about performance in relationship to their peers (Guskey, 1996b, 2015; O'Connor, 2011). The purpose of grades should not be competitive ranking, but to provide students with feedback to improve their learning (Reeves, 2007), and thus SBGs and SBRCs call for explicit criteria-referencing in place of norm-referencing (Guskey & Bailey, 2001; Marzano, 2006).

Lastly, SBG and SBRC reform has sought to separate achievement and behaviour factors in grades. Standards-based report cards at the end of term break achievement evaluations down by specific standards for each class, along with some form of process evaluation on student behavioural dispositions

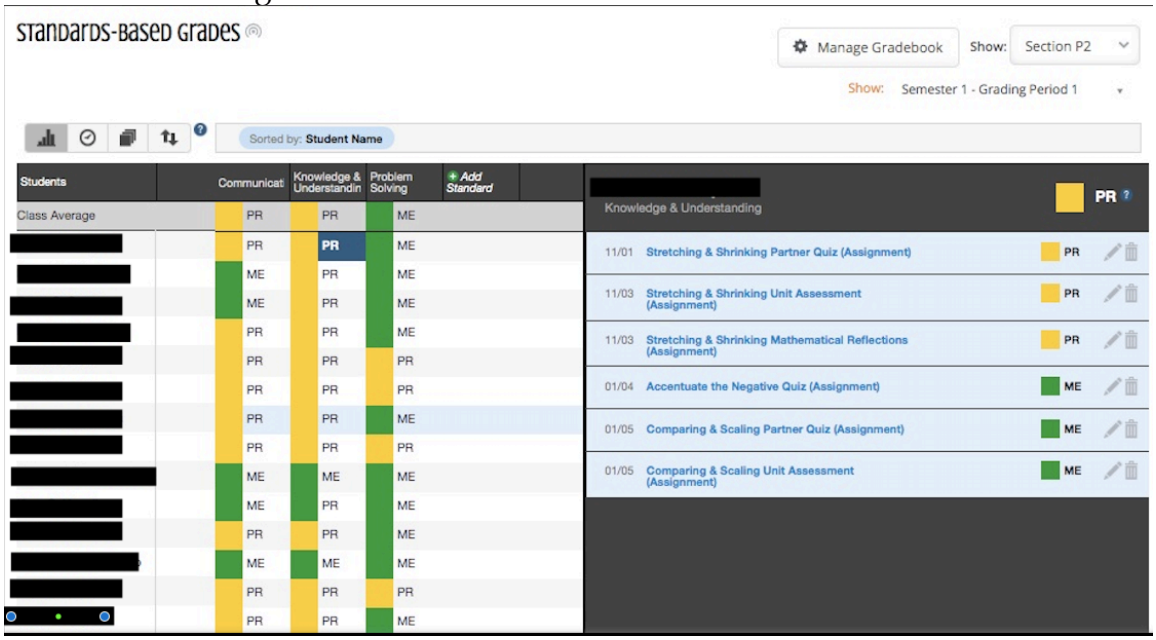
(Guskey, 1996b; Guskey & Bailey, 2001; Guskey et al., 2011; Muñoz & Guskey, 2015; Wormeli, 2006). This allows report card grades to represent achievement in regards to the course's learning outcomes (Fisher et al., 2011; O'Connor, 1995), and echoes the calls of some measurement experts from previous decades to exclude behavioural measures when computing grades (Ebel & Frisbie, 1991; Oosterhof, 2001). Further, separating behaviour from achievement helps to eliminate the practice of using grades as a behavioural punishment (Reeves et al., 2017).

These practices of eliminating zeroes, using other measures beyond averaging for determining course grades, criteria-referencing rather than norm-referencing, and disaggregating the achievement and behavioural information behind final grades are all intended to increase internal validity and address the opaqueness of previous letter grades which constrained students by providing very little feedback about learning (Guskey, 2015; O'Connor, 1995). As I will discuss later in this chapter, the clear intention within this reform was to allow SBGs and SBRCs to function formatively.

These responses to traditional grading practices form a key foundation to SBG and SBRC reform. I will now outline the process used by teachers to arrive at end of term SBRCs. With the new UbD approach to designing curriculum described earlier in this chapter, the gradebooks teachers use changes drastically. Rather than being organized by scores on individual assignments, the columns in standards-based gradebooks are organized by standards, with any given assignment containing scores for multiple standards (Marzano, 1998; Marzano & Kendall, 1998; Miller, 2013; Scarlett, 2018)(see Figure 3.3 below). For example, a project that targeted four standards would have four separate entries in the grade book for those respective standards. Many standards-based gradebooks will translate the standards performance level for each assignment to the corresponding 1-4 digit (Marzano & Kendall, 1998). By organizing gradebooks

by standards rather than individual assignments, students are given more information about their current standing in their learning (Scriffiny, 2008).

Figure 3.3: Standards-Based Gradebook for Math[6]



For end of grading term report cards, priority standards then feed into 3-5 larger reporting standards for each academic subject (Muñoz & Guskey, 2015). In creating these 3-5 reporting standards, teachers and schools must strike a delicate balance between not having too many standards on the report card so as to overwhelm students and parents, but not having too few so that the reporting standards are overly generalized and lose meaning (Great Schools Staff, 2016; Guskey & Bailey, 2001). Ideally, these reporting standards remain constant and extend from younger grade levels up through high school (Marzano, 2006). This alignment in reporting standards across grade levels supports parent and student understanding (Guskey & Bailey, 2001) and means that students receiving SBRCs at the end of one school year will then have opportunities to

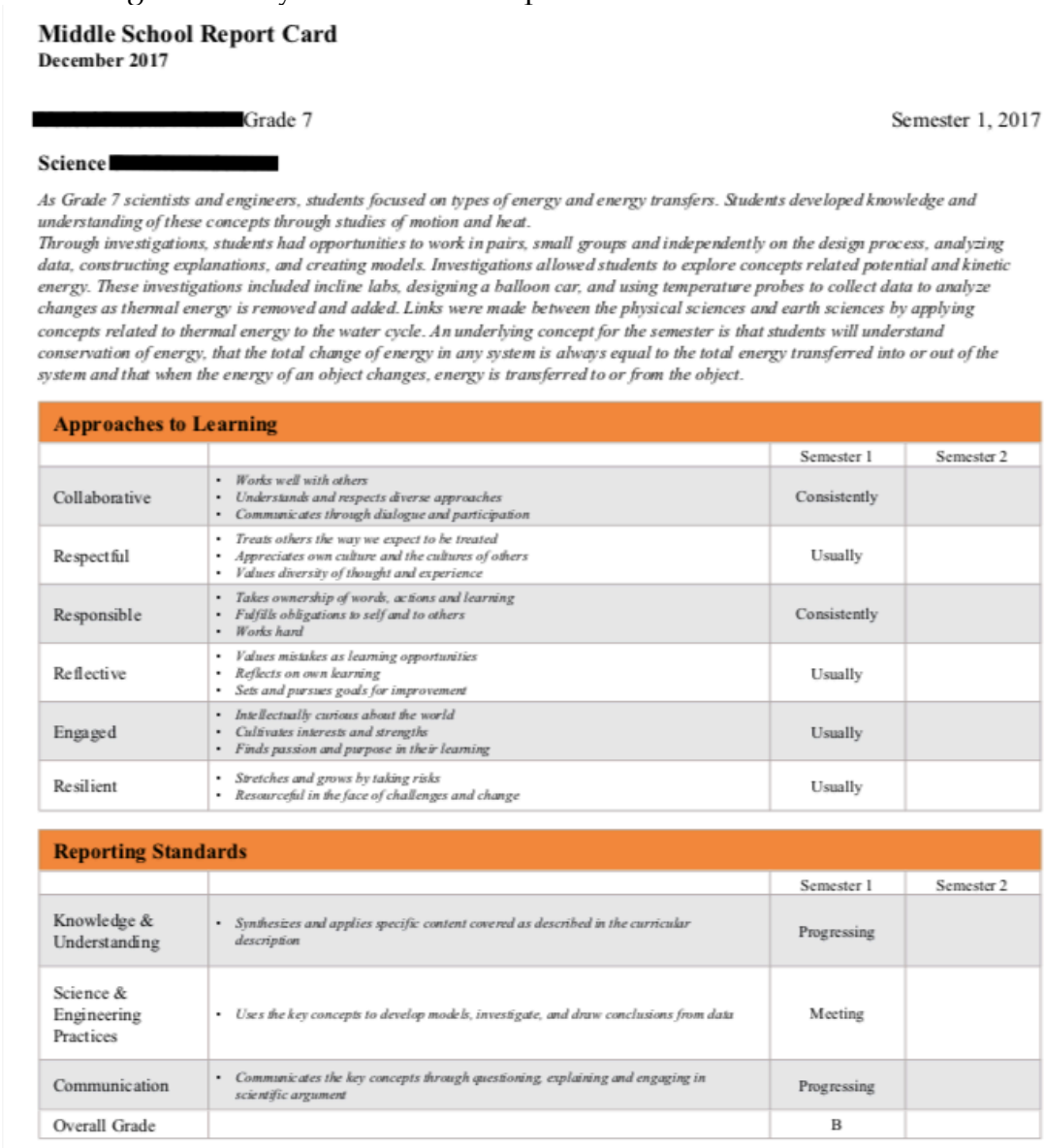
[6] Within this standards-based gradebook for a math class, the left pane has rows by student name (blacked out) and columns by the reporting standards for the class (Communication, Knowledge & Understanding, Problem Solving). The right pane displays all of the assignments for which the selected student received a standards-based grade for prioritized standards relating to the Knowledge & Understanding.

work towards those learning areas the following year, thus supporting the formative function of SBRCs.

On SBRCs, behaviour and achievement are separated with achievement communicated through grades on the reporting standards and behaviour communicated through separate descriptors using the school's predetermined behavioural categories. The terms used for the behavioural evaluations usually describe the frequency of the behaviour (Guskey & Bailey, 2001). Finally, SBRCs often include a course overview within the section for each class (Guskey & Bailey, 2001) (see Figure 3.4 below).

In determining the final grade for each reporting standard at the end of term, teachers must look beyond a simple average and determine the grade most representative of students' current understanding (Scarlett, 2018). Students should accumulate multiple grades for each reporting standard over a grading period (Marzano & Kendall, 1998). Often, the most recent evidence for a standard is given priority as an indication of students' development across a grading term (Marzano, 2010; Scarlett, 2018). When the most recent evidence does not accurately represent a student's learning, teachers are encouraged to consider other measures of central tendency, such as the most frequently occurring score (mode) and the pattern of learning over time (Marzano, 2010). Ultimately, teachers must rely on their professional judgement to determine students' final standing with the recognition that all evaluations are inherently subjective (Guskey, 1996b). With this emphasis on most recent evidence, the distinction between grades received for assignments and grades received on the final report card is lessened with SBGs and SBRCs much more than traditional grading systems. This fluidity allows for a greater possibility of students connecting the grades received on SBRCs with the corresponding rubrics used for those standards on individual assignments and accessing the formative information held within them.

Figure 3.4: Hybrid SBRC Example from Science Class



Middle School Report Card
December 2017

[redacted] Grade 7 Semester 1, 2017

Science [redacted]

As Grade 7 scientists and engineers, students focused on types of energy and energy transfers. Students developed knowledge and understanding of these concepts through studies of motion and heat.
Through investigations, students had opportunities to work in pairs, small groups and independently on the design process, analyzing data, constructing explanations, and creating models. Investigations allowed students to explore concepts related potential and kinetic energy. These investigations included incline labs, designing a balloon car, and using temperature probes to collect data to analyze changes as thermal energy is removed and added. Links were made between the physical sciences and earth sciences by applying concepts related to thermal energy to the water cycle. An underlying concept for the semester is that students will understand conservation of energy, that the total change of energy in any system is always equal to the total energy transferred into or out of the system and that when the energy of an object changes, energy is transferred to or from the object.

Approaches to Learning

		Semester 1	Semester 2
Collaborative	• *Works well with others* • *Understands and respects diverse approaches* • *Communicates through dialogue and participation*	Consistently	
Respectful	• *Treats others the way we expect to be treated* • *Appreciates own culture and the cultures of others* • *Values diversity of thought and experience*	Usually	
Responsible	• *Takes ownership of words, actions and learning* • *Fulfills obligations to self and to others* • *Works hard*	Consistently	
Reflective	• *Values mistakes as learning opportunities* • *Reflects on own learning* • *Sets and pursues goals for improvement*	Usually	
Engaged	• *Intellectually curious about the world* • *Cultivates interests and strengths* • *Finds passion and purpose in their learning*	Usually	
Resilient	• *Stretches and grows by taking risks* • *Resourceful in the face of challenges and change*	Usually	

Reporting Standards

		Semester 1	Semester 2
Knowledge & Understanding	• *Synthesizes and applies specific content covered as described in the curricular description*	Progressing	
Science & Engineering Practices	• *Uses the key concepts to develop models, investigate, and draw conclusions from data*	Meeting	
Communication	• *Communicates the key concepts through questioning, explaining and engaging in scientific argument*	Progressing	
Overall Grade		B	

3.3.3.1 The Hybrid SBRC

I will now discuss a final aspect of formulating SBRCs: the decision of some schools to use a hybrid SBRC which includes standards-based grades and an overall letter grade for each course. The decision to use a hybrid SBRC often stems from various cultural, historical and institutional forces, most of which have little to do with the ideal form of the report card for student learning. The transition to SBRCs represents a drastic shift for schools and families. Parents are often comfortable and familiar with the traditional letter grades they received as students which hold established meaning (Guskey & Bailey, 2001; Guskey & Jung, 2006). Some parents who grew up in norm-referenced grading

systems which served the purpose of comparing and sorting students may be confused and frustrated with criteria-referenced grading systems in which all students can meet standards (Bailey & McTighe, 1996). In addition to meeting parent needs, there are also broader concerns over how high schools and colleges will receive and interpret a report card which does not have letter grades (Boston, 2003; Guskey & Jung, 2006). Given letter grades' status as a "cultural icon" (Marzano, 2006, p. 125) with established meaning amongst parents and educational systems, some schools use a hybrid SBRC to avoid backlash against reform (Boston, 2003; Marzano, 1998; Scriffiny, 2008). The pressure to include a letter grade with the report card is especially strong in secondary schools (Scarlett, 2018). The benefit of this hybrid option is that it communicates more detailed information to students and parents through the SBGs than traditional report cards, while still providing them with the letter grades they are familiar with (Marzano, 2006; Marzano & Kendall, 1998). Given their problematic and opaque nature in communicating student achievement, it is with a tinge of irony that letter grades with a hybrid format can provide easier communication to parents (Bailey & McTighe, 1996).

To arrive at a letter grade for a hybrid SBRC, teachers convert final term SBGs to numbers (Beginning=1, Progressing=2, Meeting=3, Extending=4), weight reporting standards if necessary, and then determine an overall average which is converted to a letter grade using a conversion scale[7] (Marzano, 2006; Scarlett, 2018). Schools that use a hybrid report card should clearly communicate to students how these letter grades are formulated from the standards-based grades (Marzano & Kendall, 1998).

I close this section by noting that the guidelines for SBGs and SBRCs I have provided within this section represent rough guidelines teachers and schools follow within this reform movement. In reality, however, the form that

[7] Please see chapter five for a detailed account of this process with corresponding figures.

SBGs and SBRCs take varies across classrooms, grade levels and schools (Knight & Cooper, 2019; Reeves et al., 2017). This variation has implications for standards reform implementation which become clear in the SBG and SBRC research which I will discuss in the following section.

3.3.4 Initial SBG and SBRC Research

From the early stages of SBG and SBRC reform, it has been recognized that grading and reporting should serve dual SA/FA purposes by enhancing learning (Guskey & Bailey, 2001) through providing feedback to students about current standing and next steps towards learning goals (Guskey, 1996a; Guskey & Jung, 2006; Heflebower et al., 2014; Iamarino, 2014; Reeves et al., 2017). This formative function of SBRCs has been identified as their primary purpose (Muñoz & Guskey, 2015). As Guskey (2015) states:

> …rather than offering a culminating, final evaluation, the report card should be seen as part of a continuous and ongoing reporting process. Above all, the report card communicates information to facilitate improvements in student learning (p. 20).

This formative potential of SBGs and SBRCs should also work to support student autonomy and competence and lead to more autonomous forms of motivation in students. The belief in the formative function of SBGs and SBRCs has led to grand statements that SBRCs have "…the potential of altering K–12 education in the United States in dramatic and unprecedented ways" (Marzano, 2006, p. 125). It remains to be seen if such bold claims have come true, as empirical research on SBGs and SBRCs is still in its infancy and the FA rationale behind these reforms and their impact on student motivation is still largely uninvestigated (Brookhart, 2013c, 2013a; Brookhart et al., 2016; Guskey, 2015; Knight & Cooper, 2019; Scarlett, 2018). I will now review initial studies into SBGs and SBRCs and discuss their implications for this study.

Initial research that has focused on implementation within schools revealed that both parents and teachers prefer SBRCs to traditional letter grade

report cards, and that they have found them to provide high quality and clear information (Swan et al., 2014). Other research suggests that some of the old issues found with traditional letter grades and report cards are still present. These issues include separating behaviour and achievement from summative marks, including 0's for work not completed, averaging to determine final grades, and not clearly communicating to students how the teacher arrives at final grades (Brookhart, 2013a; McMunn, Schenck, & McColskey, 2003). SBRC grades depend on teacher judgement which is likely susceptible to considerations of student behaviour (Hooper & Cowell, 2014). Although experts recommend using the most recent evidence of learning and considering other measures of central tendency beyond the mean (Marzano, 2006; O'Connor, 2011), it is unclear which methods teachers currently use when deciding on summative grades with SBRCs. Another early finding is that some teachers are inconsistent in the methods they use to generate final SBRC grades and have ignored state recommendations and simply equated the standards-based grades with traditional percentages and the corresponding letter grades (Welsh et al., 2013). For example, these teachers would equate the second highest standards-based grade, *Meeting*, with the second highest letter grade, *B*, rather than recognizing that they were two different constructs. Other studies have shown that some teachers still include behavioural measures in determining final grades (Cox, 2011). These initial findings suggest fidelity of implementation issues consistent with issues present in FA and letter grade reform. Research investigating these reform initiatives should proceed cautiously and recognize that just as students will consume these new mediational means in unique and individual ways, so too will teachers.

Two recent studies in particular hold theoretical relevance to my study. Knight and Cooper (2019) investigated teacher perceptions of the SBG shift through interviewing seven teachers at separate schools at different phases of

SBG implementation. These teachers believed that with SBGs, their planning of assessment and instruction became more purposeful, and that SBGs resulted in clearer communication through their gradebook and clearer feedback to students. Some teachers elaborated that feedback in SBGs became clearer with the help of criteria rubrics for their assignments. The disaggregated nature of SBGs was reported to help students understand current standing and areas for improvement. With the alignment between grading, assignments and their corresponding rubrics, some teachers perceived students to have more control in their learning due to a clearer understanding of why they received their grade and how to improve their learning. On the other hand, all teachers in the study had to make compromises with the principles of SBGs and traditional letter grades. Further, teachers still mixed behaviour into SBGs when determining borderline grades, with some even including non-standards-based behaviour categories in their gradebook. Some teachers reported that the SBG transition was tough for students and that many parents did not understand the new system. Even within the new SBG system, teachers reported that some students were still grade driven and believed that the hybrid SBRC and parental pressure were the cause. Teachers emphasized the systemic nature of the SBG reform and stressed the time needed for successful implementation.

Scarlett (2018) switched to SBGs within two sections of his undergraduate university course on assessment. Although the level of schooling within this study is different from my study, the theoretical implications of Scarlett's findings are important to understanding if SBGs and SBRCs can function formatively. Over the course of the semester, Scarlett perceived that SBGs and SBRCs supported learning-focused communication with many students as grades became formative feedback to guide future learning. While Scarlett reported that some students liked the clarity of SBGs, there were others who were confused by it. Scarlett hypothesized that the confusion was due to

differences between SBGs and the traditional grading system students were familiar with and also because he was still learning how to implement SBG practices. While SBGs had many benefits, Scarlett recognized that it was hard to implement because it forces teachers to think about assessment differently, moving from traditional summative conceptions to a formative way of supporting student learning.

These two studies represent the tremendous potential of SBGs and SBRCs to unlock a formative function within these traditionally summative mediational means and to serve as valuable resources to guide student learning. At the same time, these studies also share a story of caution and the significant amount of time and effort needed to successfully implement this reform with the re-education of all parties involved (Watts, 1996). They also highlight the need for more studies of SBGs and SBRCs to investigate why some students are able use SBGs formatively, while others are confused and frustrated by it. Also relevant to my study and efforts to investigate student perceptions of SBGs and SBRCs, these studies are indicative of the lack of SBG and SBRC research which examines student perceptions (Brookhart et al., 2016).

3.4 Conclusion

Traditional letter grade report cards used in many schools today have been shaped by a wide range of cultural, historical and institutional forces, most of which are not aligned, or directly conflict, with the belief that all students can achieve. As Shepard (2000) stated, "Any attempt to change the form and purpose of classroom assessment to make it fundamentally a part of the learning process must acknowledge the power of these enduring and hidden beliefs" (p. 5). Beyond acknowledging the power of the beliefs embedded within letter grades, questions remain about the implementation of new grading and reporting mediational means. Can schools that use a hybrid SBRC which includes letter grades shift grading and reporting to a formative function if the

previous system which was firmly rooted in the summative function of norm-referencing and gatekeeping is still present? Further, Wertsch's myth of unidirectionality suggests that human agents uniquely use and shape mediational means, but if the agents have been socialized using the old means, how will each student's history influence the mediated action that ensues from interacting with these new grades and report cards? At the very least, this examination of the sociocultural forces which have shaped letter grades and standards reform suggests that shifting grades and report cards to a formative function will be a long and difficult task, a point I will return to later in my thesis.

Chapter Four: Methodology

4.1 Introduction

In the previous chapters I established the theoretical framework and empirical elements of my study using assessment theory, the work of Vygotsky and Wertsch, and Self-Determination Theory to better understand the rationale and implications of SBG and SBRC reform. In this chapter I will lay out the methodology for my mixed methods case study which has followed from the research questions, *What meanings do students construct from letter grades and report cards? What meanings do students construct from standards-based grades and report cards?* and *How does the standards-based grade and report card shift impact student motivation?* As I will discuss, because the first two of these questions suggest a qualitative approach and the final question suggests a quantitative approach, my study utilized a mixed methods design. This chapter will begin with discussing mixed methods studies, the analytic generalizations that can result from case studies, and an overview of research quality. Within my mixed methods case study, I specifically used focus groups and vignettes, questionnaires, and interviews. I will examine each of these methods in turn while outlining the preparations and procedures which were in place to implement data collection. Finally, I will close this chapter with a discussion of research ethics and the particular ethical issues related to my study.

4.2 Mixed Methods

On their own, quantitative and qualitative methods have strengths and weaknesses, but it is possible for an even stronger design to arise from their combination (Gorard & Taylor, 2004; Venkatesh, Brown, & Sullivan, 2016) with the two methods complementing and corroborating each other (Onwuegbuzie & Johnson, 2006; Tashakkori & Teddlie, 2009). Ultimately, a study's research questions should determine methodological decisions (Barbour, 2007; Gorard &

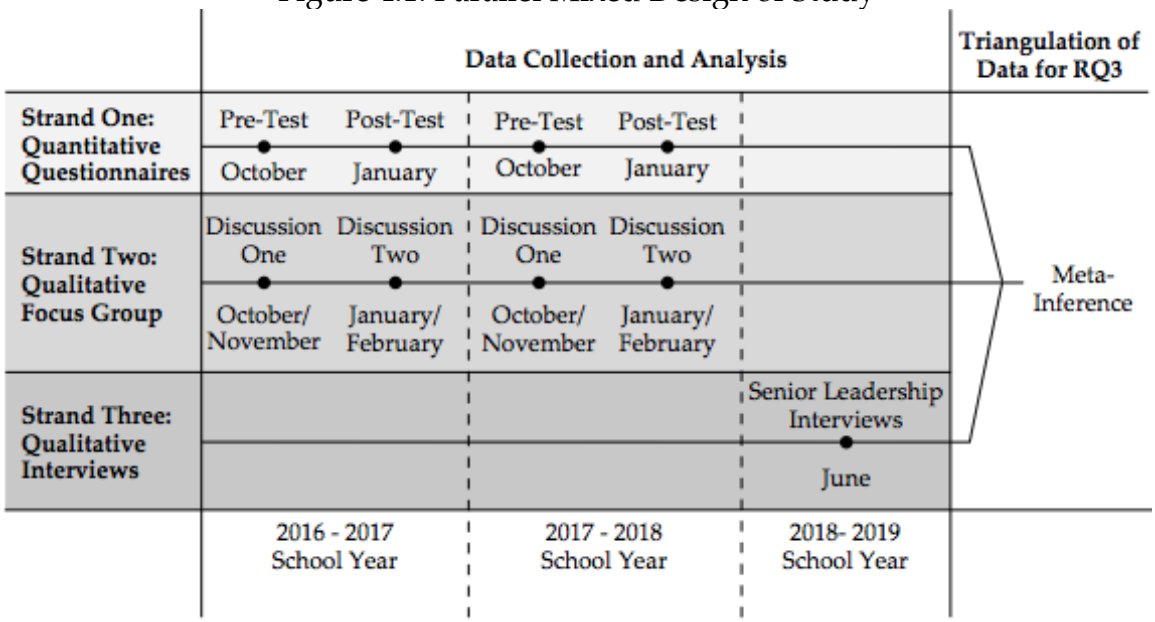
Taylor, 2004; Ponce & Pagán-Maldonado, 2015; Stewart, Shamdasani, & Rook, 2009; White, 2009). When research problems require both types of data, combining mixed methods data can lead to clear and deep understanding of the research questions (Creswell, 2012; Ponce & Pagán-Maldonado, 2015) with the quantitative data and analysis potentially offering a survey of larger patterns which can be explained in greater depth and complexity by qualitative data and analysis (Creswell, 2012; Spillman, 2014).

When warranted by the research questions, a mixed method approach should build on the strengths and weaknesses of the different methods (Ponce & Pagán-Maldonado, 2015; Wolff, Knodel, & Sittitrai, 1993) by considering where and how to mix them (Creswell & Garrett, 2008). This integration of the qualitative and quantitative strands at some stage in the research process is an essential element of a mixed design (Creswell, 2012; Teddlie & Tashakkori, 2006). Mixed method designs vary in the degree and timing at which the two strands of research intersect. My study used a Parallel Mixed Design (Tashakkori & Teddlie, 2009), a design which has alternatively been called Convergent Parallel Design (Creswell, 2012) and Concurrent Mixed Design (Teddlie & Tashakkori, 2006). This design involves at least two independent strands of research with both qualitative and quantitative research questions, data collection and analysis. The two strands may have a degree of time lag in the data collection and analysis, and the analysis is done independently of one another before being integrated at the end of the process to form a meta-inference to help answer research questions (Creswell, 2012; Tashakkori & Teddlie, 2009). Forming a meta-inference from the integration of strands requires the researcher to interpret if the two strands of inquiry contradict or complement each other, a process referred to as triangulation, which should lead to a better understanding of the results (Creswell, 2012; Gorard & Taylor, 2004; Spillman, 2014; Teddlie & Tashakkori, 2006; Wolff et al., 1993; Yin, 2013).

The design of my study stemmed from my research questions, listed in the introduction section of this chapter. The first two research questions, with their emphasis on meaning construction, suggest a qualitative method that allows for a deep investigation into the many layers of meaning that come with the socially-constructed understanding of different types of grades. The third question, with its emphasis on measuring the impact of the standards-based shift, suggests a quantitative method to measure specific changes in student motivation through an instrument aligned to Self-Determination Theory (see chapter two). While these first two questions and the third question operate as stand-alone problems and methodologies, the third question benefits greatly from applying the qualitative findings from the meanings students constructed to better understand the motivation measurements of the quantitative instrument.

Following the Parallel Mixed Design of my study, focus group, interview and questionnaire data were collected and analysed individually before integrating the respective analyses at the end of the study to form a meta-inference (Barbour, 2007; Caillaud & Flick, 2017; Ponce & Pagán-Maldonado, 2015). Figure 4.1 below diagrams the Parallel Mixed Design utilized in this study, and each strand of the research will be described in greater depth later in this chapter.

Figure 4.1: Parallel Mixed Design of Study



4.2.1 Case Studies

Yin (2018) defines a case study as a method used for in-depth investigation of a phenomenon within its authentic context. The case should be a concrete entity and the study is bounded by organizations and time periods. Case studies provide space to conduct the mixed methods studies described earlier in this chapter (Stake, 2005). In my study, the case investigated was the middle school within an American international school over the period in which it transitioned from traditional letter grades and report cards to standards-based grades and report cards. Because each case studied represents a dynamic and complex real world setting, it is important to describe the context of the case in as much detail as possible (Stake, 2005; Yin, 2013), a description I will provide in chapter five.

Researchers should prioritize selecting cases for research which provide the greatest opportunity for insight into the problem studied, as opposed to selecting a case for its convenience (Gomm, Hammersley, & Foster, 2000). While there was certainly a convenience element to selecting the school where I taught at to conduct my research, SIS also represented an "instrumental" case which was uniquely positioned for me to investigate the impact of the SBG and SBRC

shift (Stake, 2005). Similarly, because the formative rationale behind SBGs and SBRCs is largely untested (see chapter three), the school site also represented a "critical" case in which I could test theories through my findings of SBGs and SBRCs. If the preestablished theory did not hold true for this critical case, it would establish that the theory needs to be revised (Flyvbjerg, 2006).

Cases should not be viewed through a statistical analysis lens as sampling units and judged for their representativeness of a larger population (Yin, 2018). Instead, it should be recognized that every case is unique in its various contexts (Small, 2009). In this sense, SIS was not representative of some larger population of American international schools, but a unique setting in itself. This has implications for the generalization of the results from my mixed methods case study, which I will now discuss.

4.2.2 Generalizing

Lincoln and Guba (2000) argued that there can be no generalizations within qualitative research due to the uniqueness of every context across locations and time. Instead, they argued for case to case transfer in which researchers provide "thick descriptions" (Geertz, 1973) of the contexts in the case studied, which readers then use to determine the transferability of a case depending on its fittingness with readers' contexts. With this focus on transferability, they reasoned that qualitative research should produce working hypotheses, not generalized conclusions.

Others have argued that generalizing is unavoidable (Gobo, 2008), that qualitative research should generalize (Polit & Beck, 2010) and that it is too simplistic for researchers to reject generalization as a responsibility of the reader (Gomm et al., 2000; Payne & Williams, 2005). Lincoln and Guba's radical stance that there can be no generalizations in qualitative research can be contrasted with the more moderate stance of analytic generalization (Gobo, 2008; Yin, 2018),

with most qualitative studies using one of these two techniques (Onwuegbuzie & Leech, 2010).

Analytic generalizing involves making a generalization to a higher conceptual level, such as theory (Firestone, 1993; Yin, 2018). This process involves extracting researched ideas and linking these ideas to overlaps and gaps in the research literature (Yin, 2013), thus generalizing from the particulars of case studied to broader theory (Polit & Beck, 2010). Analytic generalizations should include an explanation of how the phenomena being studied produced the stated results (Yin, 2013). In this sense, an analytic generalization is linked with forming a working hypothesis (Lincoln & Guba, 2000; Yin, 2018) and can be supported by a thick description of the case context (Spillman, 2014). In making these analytic generalizations which can serve as a working hypothesis, case studies can help to improve theory (Spillman, 2014). Analytic generalizations also represent an opportunity to moderate claims by linking basic patterns to larger theories which helps researchers avoid overreaching with their findings (Payne & Williams, 2005).

Through analytic generalizations, it is possible for critical cases to serve as black swans which can falsify broader theory (Flyvbjerg, 2006). In chapter three I reviewed the new field of literature which forms SBG and SBRC theory and is predicated on the belief that SBGs and SBRCs should function formatively. If my case study reveals a dynamic in which SBGs and SBRCs function in other ways, this has the potential to provide an important analytic generalization to help revise the current theory behind SBGs and SBRCs. I will return to this point when discussing results and analysis in chapters seven and eight.

Within this thesis I will use Yin's (2013) analytic generalizations to the broader theories established in chapters two and three. With these analytic generalizations, I will also use the more moderate aspects of Lincoln and Guba's (2000) stance to frame these analytic generalizations as a working hypothesis

supported by thick description which will need further testing in later studies given the early stages of research into SBGs and SBRCs (Gobo, 2008).

4.2.3 Research Quality in Mixed Methods Case Studies

Within quantitative studies, discussions of reliability are concerned with the consistency of a study's results when measured over time, or the consistency of the items within the measure (Bryman, 2008; Howitt & Cramer, 2011). Within this study, the consistency of items in the Academic Self-Regulation Questionnaire (SRQ-A) was measured using Cronbach's Alpha, a process I discuss later in this chapter (Cohen, Manion, & Morrison, 2011; Howitt & Cramer, 2011).

Reliability differs between quantitative and qualitative methods (Cohen et al., 2011). Earlier in this chapter I discussed the implications of my mixed methods case study and situated my findings as unique to the particular case of SIS. In this sense, the findings of the study do not gain reliability in the traditional sense of being replicable in future studies (Howitt & Cramer, 2011). Instead, I aim to increase reliability in my study by providing thorough documentation showing consistency of the procedures and steps within my data collection process for focus groups and semi-structured interviews (Cohen et al., 2011). Others have argued that qualitative reliability is an issue of how a researcher's recorded data matches with what occurs in a natural setting (Creswell, 2009). Within this conceptualization of reliability, I can only claim that my results and findings are representative of the understanding about grades and report cards which students constructed and shared in a very particular focus group setting, a setting I will describe shortly. This same view is applied to students' questionnaire data and senior leadership interview responses. This view is aligned to the Vygotskyian underpinnings of my study, with the belief that sense is constructed in uniquely individual ways and socially mediated (Vygotsky, 1986).

Validity refers to the integrity of the conclusions drawn from research (Bryman, 2008). External validity is concerned with the generalizability of a study's findings to a larger population (Cohen et al., 2011; Coolican, 2009). Earlier in this chapter I situated SIS as a unique case, not intended to be representative of a larger population of schools. As such, external validity, in its traditional sense, is not a concern for my analytic generalizations in this thesis.

Internal validity is concerned with the accuracy of the research findings (Cohen et al., 2011; Coolican, 2009). The triangulation process I described earlier in this chapter in discussing meta-inferences is one way internal validity can be increased (Bryman, 2008; Cohen et al., 2011; Creswell, 2009; Howitt & Cramer, 2011). If a relationship between variables can be established using multiple measures, that supports the internal validity of that relationship (Howitt & Cramer, 2011). Another way internal validity can be increased is through an iterative process of coding and analysis of qualitative data (Creswell, 2009). In chapter six I will provide an in-depth overview of the procedures I had in place during coding and analysis to increase the internal validity of my focus group findings, such as defining coding terms and conducting five different rounds of the coding process.

One threat to internal validity is history - the possibility that over the course of the study, variables other than those being investigated were actually the cause of any measured changes (Creswell, 2009). When initially designing the quantitative side of my study, I planned to use a quasi-experimental design which did not include a pre-test for each year of the data collection. In switching to the design I ultimately used with a pre-test in each year (discussed later in this chapter), I was able to increase the internal validity of the study by accounting for differences between the two cohorts. However, history remained a threat to the internal validity of my findings as there was no way to account for variables

other than the form of grades and report cards which may have uniquely occurred in either year of the study (Shadish, Cook, & Campbell, 2002).

4.3 Focus Groups

Focus groups with vignettes were used in my study to investigate the research questions, *What meanings do students construct from traditional letter grades and report cards?* and *What meanings do students construct from standards-based grades and report cards?* Focus groups are used to explore the understanding that a group of participants construct about a given topic. A distinguishing characteristic of focus groups is that the source of data is participant interaction (Morgan, 1996, 2012; Smithson, 2000), and focus group facilitators must seek to enable this interaction (Barbour, 2007). Participant interaction provides moderators with insight into group consensus or disagreements, as participants can challenge each other and build on the ideas of others (Barbour, 2007; Litosseliti, 2003; Stewart et al., 2009). Participants can also encourage and support each other, with outgoing participants often breaking the ice for quiet or more reluctant participants (Kitzinger, 1994, 1995), a dynamic which should be followed closely as strong-voiced participants can silence the voices of others. Focus group interaction frequently includes participants clarifying and questioning the statements of others, which leads to deeper discussions (Kitzinger, 1995; Morgan, 1996).

By allowing participants to interact with a group of peers, rather than solely interacting with the researcher, focus groups have the potential to lessen the power hierarchy that exists between researcher and subject, teacher and student (Barbour, 2005; Belzile & Öberg, 2012; Kitzinger, 1994; Litosseliti, 2003) and to "bridge the gap between those in authority and the people they control" (Morgan, 1996, p. 149). These considerations were especially important given the inherent power imbalances with my role as both researcher and teacher at SIS, an issue I will discuss in more depth in the ethics section of this chapter.

A frequent criticism of focus groups is that participant responses are contaminated by interaction within the group setting, a critique driven by the divide between those that feel participants are sharing held truths, and those who see participants as socially co-constructing meaning (Belzile & Öberg, 2012). Following Vygotsky and Wertsch, however, all learning is social by virtue that it is mediated by means which are created within social, cultural and historical contexts. As such, all individual viewpoints are rooted in the social world (Vygotsky, 1986; Wertsch, 1991). Because the truths individuals hold have been socially co-constructed, the social nature of focus groups is one the method's great strengths and is directly aligned to the theoretical framework of my study. Within this framework, focus groups provide the opportunity to record socially constructed meaning making in action (Caillaud & Flick, 2017; Wibeck, Dahlgren, & Öberg, 2007). As Flores & Alonso (1995) stated, "The produced data are real because the participants reciprocally influence one another in the same way as happens in real life" (p. 99). Some argue that focus group data is indicative, not generalizable (Litosseliti, 2003; Stewart et al., 2009), as focus groups represent one type of meaning construction within a specific social and contextual setting (Litosseliti, 2003; Smithson, 2008). This indicative stance implies a plasticity to conclusions and overlaps with the concept of the working hypothesis. As I have already outlined earlier in this chapter, this working hypothesis view of generalizations is aligned to the analytic generalization (Yin, 2013) stance taken within this thesis.

Specific to the methods used within this study, focus groups have frequently been used within mixed methods designs (Bloor, Frankland, Thomas, & Robson, 2001) and can offer a deeper understanding of quantitative data (Barbour, 2007). Further, focus groups are helpful for interpreting questionnaires because they can potentially provide a deeper level of explanation for the questionnaire findings (Bloor et al., 2001; Caillaud & Flick,

2017; Flores & Alonso, 1995; Morgan, 1996; Smithson, 2008; Stewart et al., 2009; Wilmot & Ratcliffe, 2002).

4.3.1 Vignettes

Finch (1987) defined vignettes as "short stories about hypothetical characters in specified circumstances, to whose situation the interviewee is invited to respond" (p. 105), and researchers have increasingly integrated this method within focus groups (Barter & Renold, 1999; Brondani, MacEntee, Bryant, & O'Neill, 2008). Asking participants to respond to hypothetical characters and situations requires the use of their own experiences to articulate what is happening in the vignette (Jenkins, Bloor, Fischer, Berney, & Neale, 2010). I used vignettes (see Appendix A for examples within my focus group schedules) as a means in my study to help bridge students' personal experiences with the understanding they constructed with other focus group members.

I intentionally made vignettes simple and basic, while still providing enough context to make sense. The fewer details participants have for the scenario, the more they use their own experience to fill in the space which allows for insights into their perspective (Barter & Renold, 1999, 2000; Finch, 1987; West, 1982). This compliments Vygotsky's view that communication through mediational means is imprecise and requires individuals to draw upon their own background and experiences to form conceptual understanding (Vygotsky, 1986). Aligned with focus group interaction, how participants make sense of vignettes is very similar to how they make sense of events in their day-to-day lives (Jenkins et al., 2010).

Later in this chapter I will provide an overview of the ethical considerations of my study, but for now I would like to recognize that there are ethical allowances that come with the use of vignettes. The hypothetical scenario allows participant choice in either safely commenting on the third-person character, or choosing to divulge connections from personal experiences. The

third-person response option provides a safety net for participants who don't feel comfortable speaking about a sensitive topic, such as grades (Bradbury-Jones, Taylor, & Herber, 2014). This choice supports participant autonomy and control within the focus group process, and can help ease teacher-student and researcher-subject power imbalances (Barter & Renold, 2000). This also reduces participant fear of providing socially undesirable responses, which allows them to respond more truthfully and openly (Barter & Renold, 2000; Bradbury-Jones et al., 2014; Hughes & Huby, 2002).

Another ethical consideration with vignettes is that even though participants can direct their responses towards the third-person characters, many participants find themselves engaged in the vignette and end up sharing more of their personal experiences than they otherwise would have (Jenkins et al., 2010). This echoes a broader ethical concern of over-disclosure by participants within focus groups (Bloor et al., 2001). My experience with this limitation was that participants always had the option of switching their response back to focusing on the third-person character. Indeed, during my study some participants began responding to a vignette with their personal experiences, only to switch to referring to the character when they became uncomfortable with sharing their direct experience.

A frequent critique of vignettes is that participant responses do not match the actions that participants would take in real-life situations, thus questioning if participant beliefs are aligned to participant actions (Barter & Renold, 2000; Bradbury-Jones et al., 2014; O'Dell, Crafter, Abreu, & Cline, 2012). Every higher mental action humans take is mediated by the mediational means of our environment (Vygotsky, 1986; Wertsch, 1998). As such, the meaning constructed in any interaction should be viewed as contextual and rooted in the means which mediated the human action. Focus groups and vignettes represent one such

mediational means, and as researcher I can only claim that participant responses represent their ideas within this very specific research setting.

4.3.2 Facilitation

The persona and role of the focus group moderator will inevitably impact the data that is elicited through group interaction (Barbour, 2007; Smithson, 2000). Researchers should show reflexivity and acknowledge this impact (Smithson, 2008), and this is especially true in the case of this study because I was both the researcher and an insider at the institution where the group was held (Barbour, 2007). More specifically, because I was a teacher at the school, adult-child and teacher-student power imbalances were inevitably present within the group dynamics (Litosseliti, 2003). While these power imbalances were inevitable and must be acknowledged, as moderator I worked to deemphasize my status to allow participant voices and interaction to take pre-eminence (Barbour, 2007; Litosseliti, 2003; Smithson, 2008; Stewart et al., 2009).

There were two primary ways I attempted to deemphasize my status. The first way was through the introductory talk at the start of each session. The start of every focus group is an important time to set the tone and expectations for the discussion that follows, as how the group discussion is introduced can lead to discussions taking different forms (Wibeck et al., 2007). To help put participants at ease, we began with introductions, the purpose of the group, addressing that there were no right or wrong answers (Litosseliti, 2003; Wibeck et al., 2007), before finally encouraging participants to direct comments to each other and not me (Kitzinger, 1995). This was followed by a quick discussion guided by the prompt, "From your experience, what has to happen for a good group discussion to take place?" In most of the groups, participant responses covered all of the key points I wanted to address. These key points were: building on the ideas of others; questioning others; disagreeing when needed; stepping into the discussion space if they had ideas to share; and finally,

stepping back if they found they were talking disproportionately more than others. In cases where participant responses did not address one of these key points, I stepped into the discussion and shared. Front-loading these concepts prior to commencing group discussion was helpful in avoiding group dynamics harmful to the flow of conversation such as silent and dominating figures (Bloor et al., 2001; Litosseliti, 2003; Smithson, 2008) and the silencing of dissent (Smithson, 2000).

Setting discussion expectations at the start of the meeting through student generated responses went a long way towards participants regulating themselves during discussion. This aligned perfectly with the second way I attempted to deemphasize my status with my facilitation style: intervening as little as possible. Within focus groups, moderators must negotiate the tension of how much intervention is needed (Barbour, 2007; Smithson, 2000). While moderator intervention can be used to explore differences of opinion, encourage quieter participants to participate, address prolonged silences, and to redirect discussion when it gets off-track from the topic, the introductory talk eliminated the need for much intervention (Bryman, 2008; Flores & Alonso, 1995; Kitzinger, 1994; Morgan, 1996). Some of the best focus group discussions have very little moderator intervention and voice, as the moderator taking a back seat can empower participants to share ownership over the discussion (Barbour, 2007; Bloor et al., 2001; Litosseliti, 2003; Smithson, 2008). To transfer authority to the student participants, I tried to intervene as little as possible in the discussions, especially towards the start of the session (Bryman, 2008; Kitzinger, 1994). The few times I did intervene were when the discussion veered too far off track, during occasional unproductive silences, and on two occasions when individual participants dominated discussions for a length of time (Bryman, 2008; Wibeck et al., 2007).

By sharing authority for the discussions and enabling participants to moderate themselves, my hope was that my facilitation style and the larger atmosphere of the groups would lessen the power hierarchy that existed between myself and participants and would result in participants' open and active participation (Ryan, Gandha, Culbertson, & Carlson, 2014). My perception was that this facilitation style was successful in lessening these power imbalances. Regardless, my roles as primary researcher and teacher at the school represents a limitation as these power imbalances can never be eliminated (Smithson, 2000).

4.3.3 Physical Environment

The setting of a focus group should be determined by where participants are most likely to talk (Litosseliti, 2003), which requires researcher flexibility (Barbour, 2005, 2007). Focus groups were held in my classroom. I considered holding discussions in participants' classrooms where they might be more comfortable in a familiar setting, but I eventually decided that close proximity to their current teachers while discussing grades could limit discussion and cause distress. All sessions took place during the one-hour lunch block of the school day. By the time all participants arrived after getting their lunch, and leaving time at the end for travel to the next class, focus group sessions ran for approximately 40 minutes.

Upon entering, participants chose seats among desks set out in a U shape with the hope that facing each other would facilitate conversation. A U shape - rather than a circle - allowed student access to snacks I had placed in the centre of the U. This snack desk was where I placed the tape recorder to insure a clear recording (Barbour, 2007; Bloor et al., 2001), something worth noting because the recorder's position meant that it was clearly visible to students throughout all focus groups. I also placed my phone next to where I facilitated the discussion as a backup recording device.

I sat outside and behind the bend in the U to encourage direct interaction amongst participants and to lessen my presence in these discussions. Some participants in the groups did initially turn to respond to me, but in those cases, I, and other students, reminded them that they should direct comments towards their peers.

When students entered for the first focus group of each year they created name placards to place on their desk. My attempt to balance the similarities and differences of group members (Bloor et al., 2001; Flores & Alonso, 1995) resulted in groups where some students did not know each other, and I hoped that the name placards would allow them to achieve a degree of familiarity and friendliness that would be conducive towards collaborative discussion (Litosseliti, 2003).

4.3.4 Sampling

Purposive sampling involves sampling in strategic ways to gather a sample that is relevant to the topic being studied (Bryman, 2008). I used purposive sampling to strike a balance in the homogeneity and heterogeneity within the focus groups for discussing grades and report cards. Focus groups can benefit from using pre-acquainted groups as their participants and they are typically conducted with groups that share a degree of homogeneity (Barbour, 2007; Kitzinger, 1994; Litosseliti, 2003; Smithson, 2008). On the other hand, heterogeneity within the group can help to ensure participants with diverse opinions and experiences, so others suggest striking a balance with the similarities and differences of group members (Bloor et al., 2001; Flores & Alonso, 1995). In educational studies, mixed gender groups reflect the reality of the classroom and can be helpful for sparking discussion (Litosseliti, 2003; Smithson, 2008). Others suggest that mixed methodology studies using focus groups in conjunction with surveys should have focus group participants reflect

the sample of the survey (Bloor et al., 2001). If focus group participants don't reflect the sample of the survey, it can lead to invalid meta-inferences at the conclusion of a Parallel Mixed Design study (Onwuegbuzie & Johnson, 2006).

To create groups reflecting both classroom realities and the population sample of questionnaire participants, I established my sampling frame using grade level strata and gender (Cohen et al., 2011). Homogenizing by grade level with an equal number of boys and girls also allowed for a degree of comparison between the two years of the study with consideration for the different forms of report cards that were implemented at the different grade levels (Morgan, 1996). Students were then selected randomly from within their grade level strata and gender groups in the attempt to avoid friendship groups as participants often talk more freely if they don't know each other (Flores & Alonso, 1995).

On average, focus group literature recommends a range of five to nine participants (Barbour, 2007; Bloor et al., 2001; Flores & Alonso, 1995; Kitzinger, 1995; Litosseliti, 2003; Morgan, 1996; Smithson, 2008). Smaller groups are often better for sensitive topics (Bloor et al., 2001; Litosseliti, 2003; Morgan, 1996), while larger groups can result in some participants remaining silent (Smithson, 2008). Another factor I considered was overrecruiting in the anticipation of absences, as not all students who agree to take part are able to attend (Bloor et al., 2001). Using these considerations, I targeted 8-10 participants for each grade level focus group.

In assembling the Year One focus group participants, I used different approaches by grade level. Because Grade 5 and 6 students at SIS do not regularly check email, after randomly selecting five boys and five girls, I went to meet with potential participants during the start of the day advisory time[8] to

[8] Every grade in the SIS middle school began the day with a 30-minute advisory class. Advisories consisted of approximately 12 students and one teacher, with the purpose of meeting the social and emotional needs of students.

explain optional participation in the study and pass out consent forms to interested students.

Because Grade 7 & 8 students had their own laptops, I emailed selected participants to introduce myself and the study. Since I taught Grade 7, I drew my random sample for Grade 7 from students I did not teach. Emailing students was problematic because many students I contacted never replied, a common problem which often results in focus group participants being selected based on their availability (Smithson, 2008). The high non-response rate led to a prolonged period of selecting further names until I had received an acceptable number of signed consent forms. These non-responses and opt-outs represent a potential for bias within the data collected from the focus groups (Gorard, 2013). The students who did not feel comfortable speaking to teachers and peers about grades and report cards may have had valuable perspectives on the topic, but I was only able to ethically seek out those students who were interested and comfortable in discussing the topic within the described setting.

There were no repeat students who participated in focus groups during both years of data collection. It is worth noting that some of the students in the Grade 8 focus groups were students that I had taught the previous year when they were in Grade 7. I contemplated excluding previously taught students from sample selection, but concluded that as I was no longer teaching those students they would feel reasonably free to state their thoughts and opinions within the existing constraints of me still being a teacher at the school. Table 4.1 below details the Year One students I contacted and those who participated by grade level.

Table 4.1: Year One Focus Groups: Students Contacted and Student Participants

Grade	Students Selected and Contacted	Student Participants Who Returned Signed Consent Forms
5	10	8
6	10	10
7	23	7
8	30	8

Selection of Year Two focus group participants followed Year One procedures. The lone exception was that the poor Grades 7 & 8 email response rate during Year One led me to meet participants in-person to ask for participation during Year Two selection. To alleviate pressure to participate, I repeatedly stated that participation was entirely optional. Face-to-face meetings dramatically improved response rates - so dramatically that I reflected afterwards that even with repeated assurances of voluntary participation, some students likely felt pressured to participate. Table 4.2 below details the Year Two students I contacted and those who participated by grade level.

Table 4.2: Year Two Focus Groups: Students Contacted and Student Participants

Grade	Students Selected and Contacted	Student Participants Who Returned Signed Consent Forms
5	8	8
6	11	8
7	18	6
8	14	7

4.3.5 Frequency and Attendance

Holding a series of focus groups over a range of time allows for monitoring shifts in the meaning participants construct with the concept being studied (Barbour, 2007). I aligned frequency and timing of the focus group discussions with the quasi-experimental design of the questionnaire, discussed shortly. This allowed for potential longitudinal comparisons between grade levels across both years of the study to investigate students' constructed meanings from the different types of grades and report cards. For both years of

data collection, questionnaire distribution occurred once at the start of the school year in October, and then again in late January when students received their first semester report cards. My hope was that this alignment would allow the 16 focus group discussions to register any changes students felt as the year progressed, particularly following first semester report card distribution.

The initial Year One focus group meetings were delayed while waiting for school approval with letter of consent wording, and then waiting for students to return signed consent forms. This resulted in pushing back initial focus group meetings to late November and early December. For Year Two meeting dates, I attempted to replicate these times to allow for a degree of comparison of responses across the two years. The follow-up focus group meetings occurred both years in late January and early February, after the semester report card distribution in mid-January. Tables 4.3 and 4.4 below detail the meeting dates by grade level for both years of focus group data collection.

Table 4.3: Year One Focus Group Meeting Dates

Grade	First Meeting	Second Meeting
5	1 December 2016	1 February 2017
6	18 November 2016	27 January 2017
7	6 December 2016	6 February 2017
8	2 December 2016	9 February 2017

Table 4.4: Year Two Focus Group Meeting Dates

Grade	First Meeting	Second Meeting
5	28 November 2017	31 January 2018
6	29 November 2017	2 February 2018
7	7 December 2017	5 February 2018
8	6 December 2017	13 February 2018

Focus group attendance was strong as a high percentage of the students who agreed to participate in the study made it to the discussions. Attendance fell within the average recommended numbers (Barbour, 2007; Bloor et al., 2001; Flores & Alonso, 1995; Kitzinger, 1995; Litosseliti, 2003; Morgan, 1996; Smithson, 2008). One exception was the second meeting of the Year Two Grade 8 group. Only three students showed up for the discussion, with others forgetting or

dealing with other commitments. I rescheduled the meeting and conducted it the following week with perfect attendance. Overall, most absences resulted from student illness, forgetfulness, club commitments, and tests students had rescheduled due to absences. Tables 4.5 and 4.6 below detail focus group attendance across both years of data collection by grade level.

Table 4.5: Attendance at Year One Focus Group Meetings

Grade	November/December Meeting	January/February Meeting
5	8/8	7/8
6	6/10	9/10
7	7/7	5/7
8	8/8	7/8

Table 4.6: Attendance at Year Two Focus Group Meetings

Grade	November/December Meeting	January/February Meeting
5	8/8	5/8
6	7/8	8/8
7	6/6	5/6
8	5/7	7/7

4.3.6 Focus Group Schedule

Focus groups were semi-structured: because group discussions were different, follow-up questions were also different (Morgan, 1996). While direct comparison of exact questions was impossible, responses to the initial standardized questions and vignettes were compared for key similarities and differences.

Focus group schedules (see Appendix A) were formulated using three types of prompts to spur discussion. The first was an ice-breaker opener for the first session of each year to help participants establish a common ground, free-flowing conversation (Belzile & Öberg, 2012; Bryman, 2008; Kitzinger, 1994; Litosseliti, 2003), and to allow participants to begin discussion in a non-threatening way (Barter & Renold, 2000; Nygren & Oltedal, 2015). This was important because groups were purposively selected and did not necessarily

constitute friendship groups. I opened by calling out different letter grades and asking students to share the first thing that came to mind. This was followed by the vignettes described in the opening sections of this chapter, which constituted the majority of the focus group session. Finally, in the latter part of the session direct questions were asked.

All three of these prompt types were generated from multiple considerations. The first was following the research questions of this study to investigate the meaning students had constructed of grades and report cards. An example of this origin type was the ice-breaker activities described in the previous section. Another consideration was following the work of Wertsch to investigate the cultural, historical and institutional influences on the meaning that students construct from grades. This led to vignettes and direct questions about teacher, parent and peer influence. An example of this origin type was the vignette, *Billie is heading home to view the report card with parents. What do you think is going through Billie's head?* The next consideration for the schedule creation, particularly for Year Two focus groups, was investigating the school's transition from traditional letter grades and report cards to SBGs and SBRCs with prompts probing the similarities, differences, and the intersection of these two constructs. An example of this type of prompt was the vignette, *Taylor just looked at the report card online. What do you think Taylor is thinking about? Would Taylor's thoughts change if the report card only had standards-based marks on it?* A final source for focus group prompts was investigating interesting themes from previous focus groups further. An example of this was the vignette, *Izzy feels that letter grades can box people in and limit their learning. Why does Izzy feel this way? Will standards-based grades do the same thing?* In this instance, the concept of grades "boxing" a student in arose in an earlier focus group.

Using the origins just described, vignettes were crafted to make sure that they were plausible and contained hypothetical scenarios that were real enough

for participants to connect with them (Barter & Renold, 1999; Hughes & Huby, 2004; Jenkins et al., 2010). To ensure plausibility, vignette situations were based on experiences I had witnessed or talked with students about (Schoenberg & Ravdal, 2000). As discussed earlier in the chapter, vignettes were kept as simple and basic as possible, while still giving enough context to make sense (West, 1982). Relatedly, the vignette characters were given gender neutral names to allow participants the greatest chance to connect with vignette characters (Bradbury-Jones et al., 2014). Following the techniques used in previous studies (Brondani et al., 2008; Hughes & Huby, 2004; Schoenberg & Ravdal, 2000), participants were asked a series of open-ended questions after reading the vignette.

4.3.7 Piloting

Focus group piloting took place on 26 October 2016. The pilot was used to determine if the schedule led to interesting and substantive conversation, and to trial my facilitation style and the physical set-up (Barbour, 2005; Smithson, 2008). A purposive sample of students I had taught previously allowed me to focus on these issues without having the added concern of building new relationships. Further, since I already knew these students, I was able to get a better sense of their comfort and participation levels. After initial emails were sent, parent and participant consent and assent forms were signed and collected.

The focus group pilot showed that the interview schedule resulted in substantive discussion, suggesting that both the physical environment and my facilitation style achieved the desired effects. Coding of pilot discussion provided initial beginnings of the coding framework for later focus groups, which I will discuss in chapter six.

4.4 Questionnaire

4.4.1 Design

To investigate the impact of grade reports on middle school students' motivation, my mixed methods case study implemented a Quasi-Experimental Cohort Control Group Design with a pre-test from each cohort (Shadish et al., 2002). The Year One cohort served as a control group to provide an approximation of student motivation with business-as-usual letter grades and report cards. Repeating this process in Year Two with SBGs and SBRCs in place, the study compared the change over the course of Year One (control) with change over the course of Year Two (intervention). The pre-tests within this design accounted for differences between the two cohorts by providing baseline data for each cohort at the start of each year, something which helped to eliminate alternative explanations when considering if the new SBRC had been the cause of any change in student motivation (Gorard, 2013; Shadish & Luellen, 2005). It is worth noting that many of the students who participated as members of the control group in Year One participated as members of the intervention group during Year Two when they moved up to the next grade level. Table 4.7 below provides an overview of questionnaire data collection.

Table 4.7: Overview of Questionnaire Data Collection

Year One: 2016-2017 School Year (Pre-Test October 2016, Post-Test January 2017)	**Year Two**: 2017-2018 School Year (Pre-Test October 2017, Post-Test January 2018)
Grade 5 -No Standards-Based Grades /With Letter Grades (Control A1)	Grade 5 – Standards-Based Grades/No Letter Grades (Intervention A1)
Grade 6 - No Standards-Based Grades /With Letter Grades (Control B1)	Grade 6 - Standards-Based Grades /With Letter Grades (Intervention B1)
Grade 7 - No Standards-Based Grades /With Letter Grades (Control B2)	Grade 7 - Standards-Based Grades /With Letter Grades (Intervention B2)
Grade 8 - No Standards-Based Grades /With Letter Grades (Control B3)	Grade 8- Standards-Based Grades /With Letter Grades (Intervention B3)

As noted in the above table, not all grades in the middle school shifted to the same format of standards-based report cards. This opened up the opportunity to compare the results of the different SBRC form in Grade 5 with the results of the hybrid SBRC in Grades 6, 7, and 8. The design of this study in notation form is as follows in Figure 4.2:

Figure 4.2: Questionnaire Design in Notation Form

N	O	X_1	O				(Year One, Grade 5)
N	O	X_1	O				(Year One, Grades 6-8)
N				O	X_2	O	(Year Two, Grade 5)
N				O	X_3	O	(Year Two, Grades 6-8)

4.4.2 Questionnaire Instrument and Online Distribution

To measure the motivation students felt towards their academic work, I used Self-Determination Theory's Academic Self-Regulation Questionnaire (SRQ-A) (Ryan & Connell, 1989) (see Appendix B). Developed specifically for students in late primary and middle school, the SRQ-A uses the four subscales of external regulation, introjected regulation, identified regulation and intrinsic motivation to rate students' motivation on a continuum of relative autonomy (see chapter two). These subscale scores can then be combined to create a Relative Autonomy Index (RAI) score using the following formula: 2 X Intrinsic + Identified – Introjected – 2 X External. The SDT website (http://selfdeterminationtheory.org/) offers the questionnaire free of charge for use in non-commercial academic research.

One small change was made to the questionnaire. The second set of questions in the questionnaire asked students to consider, *Why do I work on my classwork?* The first statement for students to rate was, *So that the teacher won't yell at me*. Upon consulting with the middle school principal, teaching colleagues, and informally asking students, it was decided the statement was not applicable for the school as teachers rarely, if ever, yelled at students. The statement was changed to read, *So that the teacher won't get mad at me*.

The SRQ-A was distributed to students using the online software, Bristol Online Survey (BOS). After providing introductory pages for the questionnaire, I transcribed the questionnaire into BOS (see Appendix B). BOS was used to allow for ease of distribution and quicker analysis of the results.

To ensure that my analysis of results only included data from students who had participated in both pre-test and post-test, students were provided with a randomized PIN number to enter at the start of the questionnaire. Year One students were assigned a random number from 1-470, and Year Two students received a number from 500-971. A random PIN number was used rather than student names to provide students with a greater degree of anonymity. PINs were only matched to student names in the attempt to troubleshoot for duplicate numbers and to account for missing responses.

4.4.3 Pilot

The questionnaire was piloted with a group of 52 students in Grade 7 in late September 2016. Students shared that the BOS platform was easy to use and that the instructions and questions were easy to follow.

After pilot data collection, results were analysed using the software, Statistical Package for the Social Sciences (SPSS) version 22 to test the quality of the instrument for reliability. Reliability analysis was conducted by measuring Cronbach's Alpha for each of the four subscales, and for each scale with each question omitted. Cronbach's Alpha measures the extent to which the responses to the items cluster and help distinguish one person from another, with a strong reliability represented by Alpha scores in the range of .75 – 1.0 (Coolican, 2009). To test individual questions within a scale to see if they help or detract from the overall reliability, SPSS can measure Alpha for that scale if the item was deleted. If Alpha goes down, that item was increasing reliability, while if it goes up, that item was hindering reliability within that scale. These steps ultimately confirmed the reliability of the subscales within the questionnaire, as detailed

below.

Although the analysis of the pilot results go beyond the methodology of this chapter, they represent developmental results which confirmed my methodological decisions. As such, I have chosen to include them in this chapter rather than chapter eight with the results, analysis and discussion of Research Question 3.

Table 4.8: Cronbach's Alpha for External Regulation Subscale from Pilot

Cronbach's Alpha	N of Items
0.83	9
Cronbach's Alpha if Item Deleted	
Q3_2	0.80
Q3_6	0.80
Q4_1	0.77
Q4_6	0.81
Q5_4	0.80
Q5_8	0.83
Q6_1	0.79
Q6_4	0.82
Q6_8	0.85

As Table 4.8 above shows, Alpha for the External Regulation Subscale was within Coolican's (2009) recommended range. The only item that enhanced reliability if it was deleted was Q6_8. It was decided to keep this question because the change if deleted was small and there is value in keeping the original scale intact.

Table 4.9: Cronbach's Alpha for Introjected Regulation Subscale from Pilot

Cronbach's Alpha	N of Items
0.80	9
Cronbach's Alpha if Item Deleted	
Q3_1	0.77
Q3_4	0.75
Q4_2	0.76
Q4_4	0.78
Q5_1	0.78
Q5_2	0.81
Q6_2	0.76
Q6_5	0.78
Q6_7	0.82

As Table 4.9 above shows, Alpha for the Introjected Regulation Subscale was

within Coolican's (2009) recommended range. The only items that enhanced reliability if they were deleted were Q5_2 and Q6_7. It was decided to keep these questions because the change if deleted was small and to keep to the original scale.

Table 4.10: Cronbach's Alpha for Identified Regulation Subscale from Pilot

Cronbach's Alpha	N of Items
0.79	7
Cronbach's Alpha if Item Deleted	
Q3_5	0.78
Q3_8	0.75
Q4_3	0.75
Q4_8	0.74
Q5_5	0.82
Q5_7	0.73
Q6_6	0.74

As Table 4.10 above shows, Alpha for the Identified Regulation Subscale was within Coolican's (2009) recommended range. The only item that enhanced reliability if it was deleted was Q5_5. Again, it was decided to keep this question because the change if deleted was small.

Table 4.11: Cronbach's Alpha for Intrinsic Motivation Subscale from Pilot

Cronbach's Alpha	N of Items
0.82	7
Cronbach's Alpha if Item Deleted	
Q3_3	0.79
Q3_7	0.77
Q4_5	0.78
Q4_7	0.77
Q5_3	0.83
Q5_6	0.83
Q6_3	0.79

As Table 4.11 above shows, Alpha for the Intrinsic Motivation Subscale was within Coolican's (2009) recommended range. The only items that enhanced reliability if they were deleted were Q5_3 and Q5_6. In keeping with the decisions on the previous scales, it was decided to keep these questions because the change if deleted was a small 0.01.

Histograms of the individual scales on the pilot (see Figures 4.3-4.6 below) also revealed reasonable distributions of student scores, roughly resembling the

normal distribution of a bell curve. While the External Regulation, Introjected Regulation and Identified Regulation show a good spread, they are skewed towards the lowest score.

Figure 4.3: Histogram of External Regulation Subscale from Pilot

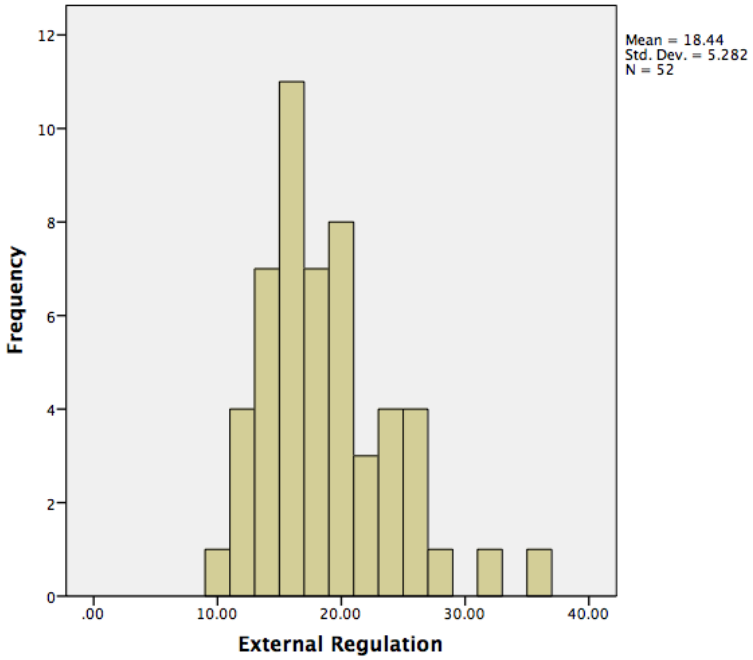


Figure 4.4: Histogram of Introjected Regulation Subscale from Pilot

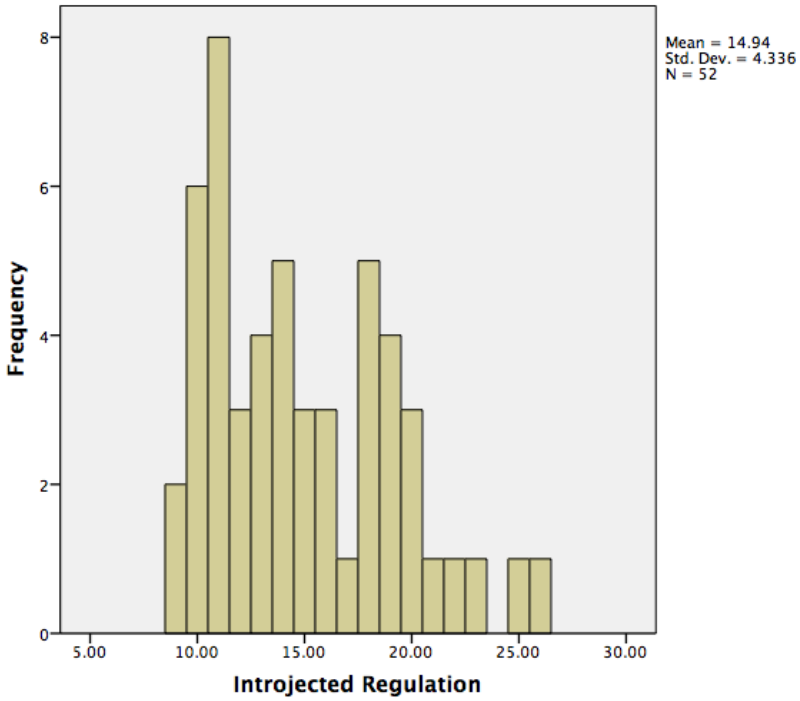


Figure 4.5: Histogram of Identified Regulation Subscale from Pilot

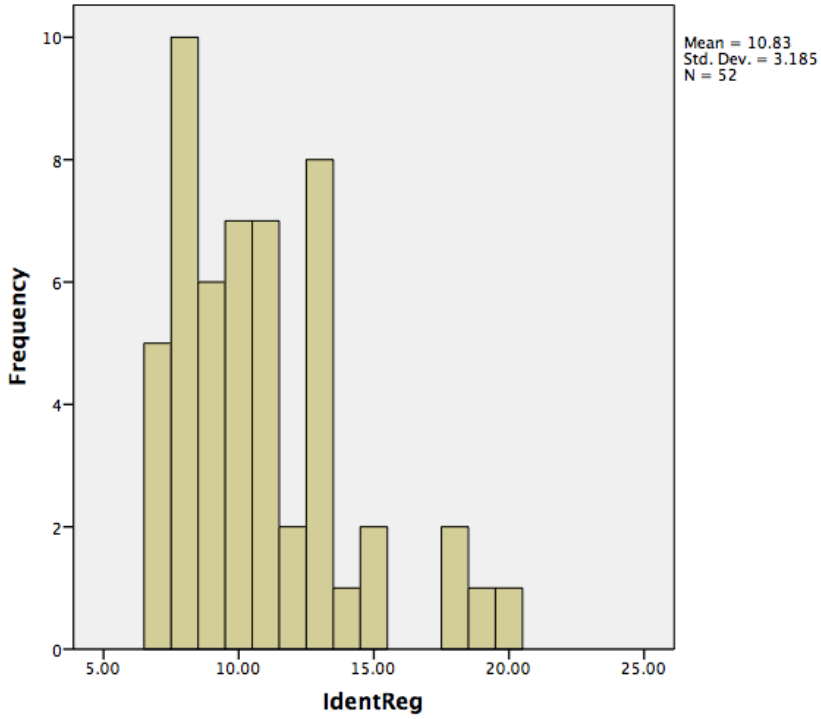
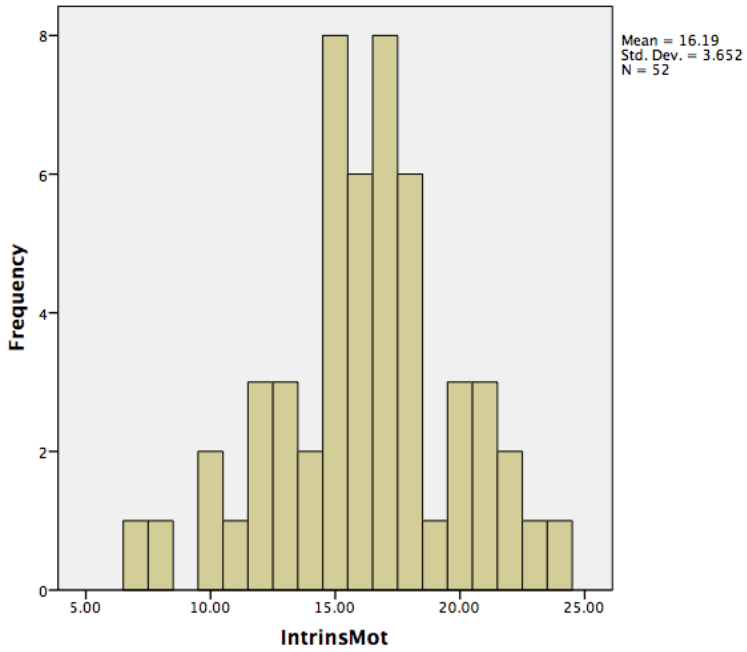


Figure 4.6: Histogram of Intrinsic Motivation Subscale from Pilot



4.4.4 Administration of Questionnaire

After using the pilot to confirm that the SRQ-A subscales were working well, data collection proceeded with the whole middle school. In the last section of this chapter I describe the ethical measures I took to ensure informed consent from students and parents. Following these measures, an email that contained initial directions and a link to the survey was sent to students during the start of the day advisory period (see Appendix C). Students followed this link and completed the questionnaire.

For students who arrived late to school or were absent on the day of the questionnaire, I emailed either their advisor (Grades 5 and 6) or the student directly (Grades 7 and 8) to ask them to take the survey the following day. I did not attempt to have these students retake the questionnaire after this point because the setting and conditions at school could have shifted enough from the initial response date to impact these later responses.

When taking the pre-test towards the start of the year, students went straight from my email to filling out the SRQ-A on BOS. To measure the impact of report cards on student motivation, the post-test directed students to first view their online report card for a few minutes before then proceeding to take the questionnaire.

4.4.5 Unforeseen Issues

It was initially intended to have all grades take the Year One post-test at the same time, but Grade 5 took the questionnaire two days later than Grades 6-8. Grade 5 teachers had not realized that the post-test required students to look at their semester report card before taking the questionnaire. The teachers thought it best to re-notify parents, as Grade 5 was the first time most students received report cards with letter grades and there was a chance that some

parents had not yet reviewed the report card with their children. Teachers worried that this had the potential to upset families, or to cause duress for students. An email was sent out to the Grade 5 parents notifying them of the questionnaire procedure (see Appendix D) and giving them another opportunity to opt out. Following the email, no parents chose to opt out. This was not an issue during Year Two of the study because the Grade 5 report cards no longer had letter grades and were very similar to what students had received the previous year.

During the Year One pre-test there were minor issues with slow internet connectivity, presumably due to so many students being online at the same time. A couple of teachers reported that a small number of students had to wait while pages loaded through the BOS website, but this problem was not widespread. This issue did not arise during the other sessions.

During the administration of Year One questionnaires, two separate teachers reported their surprise at student concerns of responses being matched to student names and then some form of reprisal taking place. These students expressed concerns that teachers were monitoring their responses in real time, despite assurances from the facilitating teacher this was not the case. One teacher did share that students felt reassured upon seeing that their random PIN numbers were truly random. These student concerns speak to larger issues with questionnaire response bias and students responding with what they think is socially desirable, as opposed to responding as they truly feel (Van De Mortel, 2008).

4.5 Interviews

To provide greater depth and contextualize the data I collected and analysed from students' perspectives about SBG and SBRC reform at SIS, I conducted one-off semi-structured interviews with three members of SIS' senior leadership team who coordinated and implemented the reform. These

interviews were not intended to be comprehensive in investigating senior leadership's perspective of SBG and SBRC reform at SIS. Instead, I hoped that they would add another dimension to my thick description of the school site and strengthen my analytic generalizations (Yin, 2018).

Interviews are a specific form of conversation which allow a researcher to understand the world from a participant's perspective through questioning techniques (Kvale, 1996). This conversation is not one-sided but an active interaction between researcher and subject, as understanding is co-constructed (Fontana & Frey, 2003; Kvale, 1996). This view aligns interviews with the Vygotsky and Wertsch underpinnings of my study which situate meaning construction as a complex interaction between individuals' unique backgrounds and the mediational means of the environment (Vygotsky, 1986; Wertsch, 1991).

Researchers must decide the degree of structure they use in interviews, with less structure theorized to place fewer constraints on the subject and result in discussions of more depth (Coolican, 2009; Mercer, 2007). I ultimately settled on conducting semi-structured rather than structured interviews because there were certain topics of standards-based grading and reporting reform at SIS that I wanted to make sure to explore with senior leadership, while also leaving space for them to share perspectives on issues I was not unaware of.

Semi-structured interview guides (see Appendix E) were created using key themes which arose in focus group discussions and the research literature. In particular, I aimed to investigate the reform process at SIS, the decision to use a hybrid SBRC, and parent and teacher education about standards reform.

Prior to beginning the interviews, I obtained consent from senior leadership to record the interviews and informed them that any data I reported to a wider audience would be anonymized. During the process of writing up responses as part of chapter five, I became more aware of internal confidentiality risks (Floyd & Arthur, 2012). Because these three members of senior leadership

represented almost the entire team of senior leadership who led the standards-based reform in the middle school of SIS, there was the potential for other SIS constituents who read my thesis to identify the senior leadership members. As such, in chapter five I tried to minimize my use of direct quotes and to paraphrase as much as possible to decrease the likelihood of these senior leadership members being identified.

Because I was a teacher at SIS, these individuals from senior leadership were both my research subjects and supervisors. During interviews my dual roles as inside researcher played out in particular ways. As a teacher, I had a slight sense of not wanting to overstep boundaries and pressure senior leadership into disclosing information through follow-up questions. At the same time, with my role as researcher, I was keen to clarify statements and follow-up on lines of discussion that were helpful towards understanding the school's reform (Bryman, 2007).

When conducting these interviews, I also wondered at times if the senior leadership members were disclosing more to me than they would to an outsider because of our pre-existing relationship. At other times, I felt they were possibly holding back and disclosing less because of my role as a researcher who would be sharing these conversations with a wider audience. In this sense, I was both insider and outsider during these interviews (Mercer, 2007). I was also aware when analysing and writing up the interviews within chapter five of this thesis that a critical stance towards the school and senior leadership's reform process could undermine my relationships with these individuals who were also friends and mentors (Floyd & Arthur, 2012). Later in this chapter, I will further discuss these ethical issues stemming from my role as inside researcher at SIS.

4.5.1 Logistics

My semi-structured interview guide was established with a priori issues which arose from my analysis of student focus group comments and my

theoretical framework (see Appendix E). Interviews took place in the respective offices of senior leadership members in June 2019 at the end of the 2018-2019 school year, more than one year after I had finished data collection for student focus groups and questionnaires (see figure 4.1). I conducted one-off interviews with each of the three individuals, ranging in length from 15-45 minutes. The range in time was a result of squeezing in the interviews when their busy schedules allowed. Following interviews, I transcribed the interviews into NVivo. In chapter five I will discuss the coding framework I used to analyse these interviews and then integrate that data within my detailed description of SIS' standards reform initiative.

4.6 Research Ethics

4.6.1 Insider Research

While conducting my study I was uniquely positioned as both a researcher and teacher at SIS. Insider research is often more complex than an insider/outsider dichotomy, and depending on the context there were ways that I was both insider and outsider during the research process (Mercer, 2007). As I mentioned previously in this chapter, in my interviews with senior leadership they likely shared more information with me as a result of our relationships which had developed over a number of years, a benefit of my insider status as a teacher at the school. At the same time, I was not a member of their senior leadership team, and as a researcher who would be sharing their responses to a wider audience, they likely modified responses due to this aspect of my outsider status.

Being an insider likely helped me to understand some of the dynamics of standards reform at SIS in greater complexity, but it is also possible that being acculturated to the school site also left me blind to other aspects of the reform's

impact (Mercer, 2007). My insider status at SIS led to particular ethical issues in the power imbalances that existed between myself and the student and senior leadership participants in my study (Floyd & Arthur, 2012). Attempts to lessen these imbalances drove some of the methodological decisions I have discussed in this chapter, such as using focus groups with vignettes, and my facilitation style within these focus groups. (British Educational Research Association [BERA], 2018).

Although my insider status granted me access to conducting research within the institution, it also placed certain ethical constraints as I gathered data. For example, I had previously made my opinions known about standards reform and its implementation at SIS during faculty meetings and informal conversations with the very senior leadership members I interviewed. In focus groups with students, I was also constantly aware that my presence as a teacher may dissuade students from honestly addressing the way they felt about their teachers' grading practices. Because I was a teacher at the school, my very presence likely had a high degree of intrusiveness during focus group discussions, regardless of measures I adopted to remove myself from the conversation. My dual role also likely led to distortion in administrator responses because they were aware that our relationship would need to continue long after the interviews had ended (Mercer, 2007). These complex relationship dynamics which arose as a result of my insider and outsider status at SIS represent an important aspect of the data I collected across my study and yet another way in which my findings are unique to this particular case study.

4.6.2 Informed Consent

Obtaining informed consent from participants was a coordinated effort of working through Durham University's School of Education Ethics Committee and with senior leadership at SIS. Initial ethics clearance for my study was granted by Durham University's School of Education in the summer of 2016.

Leading up to this, I consulted with the principal, the head of school and the director of teaching and learning at SIS. It was decided that using an opt-out method for the questionnaire would be appropriate because students regularly took surveys during their advisory block, and the degree of anonymity with the questionnaire resulted in low risk for student duress (British Educational Research Association [BERA], 2018). For this opt-out method, a note was sent out at the start of each year in the school's weekly email newsletter providing parents with the opportunity to opt their children out of taking the survey (see Appendix F).

To inform students in both years of the focus groups and questionnaire, I delivered five-minute presentations at the start of the year school assemblies introducing myself and the purpose of the study. Each year, one assembly consisted of Grades 5 and 6, and another of Grades 7 and 8[9]. I attempted to strike a balance between giving students enough information to ethically inform their participation decisions, while also keeping my presentation brief to avoid contamination of data (Litosseliti, 2003). I communicated that the school was collecting feedback from students as it moved to SBGs and SBRCs and that my study would examine how the shift impacted on student learning and motivation. All students would have the chance to participate in upcoming questionnaires if they wanted to, and some students would have the option to participate in grade-level focus group discussions.

At one of the opening faculty meetings for each year of the study, I gave a five-minute presentation to the teachers of the middle school to reiterate the information provided to students at the earlier assembly. Part of my intention with this was for teachers to be equipped to answer any student questions which arose leading up to and during questionnaires and focus groups.

[9] See chapter five for detailed description of the school's structure.

Prior to participation in focus groups, I debriefed selected students about the purpose of the study and what it would involve as a participant. I also stressed that participation was optional, aware that some students may feel pressured to participate, in part because of my role as a teacher at the school (British Educational Research Association [BERA], 2018). Those who chose to participate were then required to sign assent forms and ask their parents to sign consent forms prior to the initial focus group sessions (see Appendix G). These assent and consent forms included an informational letter and reiterated what I had communicated during the introductory conversations, including students' ability to withdraw from the study at any time.

4.6.3 Privacy

Maintaining anonymity and confidentiality for participants was a priority throughout all phases of my research process (British Educational Research Association [BERA], 2018; Bryman, 2008). Focus group and interview participants were informed that their names would not be used when reporting out collected data. When I transcribed focus group and interview discussions, names were substituted with codes and names were never recorded. Students were given unique PIN numbers to enter with their questionnaires rather than their names, and these PINs were only matched to student names to account for missing responses.

In addition to maintaining a high degree of anonymity and confidentiality, I also took measures to ensure that data was stored securely (Bryman, 2008). Assent and consent forms were kept in a locked cabinet and audio recordings were stored in password protected files.

4.7 Conclusion

Within this chapter I have laid out the rationale for my study's design. Answering my study's research questions required a mixed methods approach

which led me to adopting a Parallel Mixed Design utilizing focus groups with vignettes, interviews and a questionnaire. Focus groups with vignettes and interviews have alignment to the Vygotsky and Wertsch theoretical underpinnings of my study, and the SDT questionnaire allowed me to measure the impact of the SBRC shift on student motivation. In this chapter I also explained why making analytic generalizations from my mixed methods case study was appropriate, a topic I will return to in much greater depth in chapters six, seven and eight which explore the results, analysis and discussion of my three research questions.

Part of a researcher's responsibility when making the type of analytic generalizations I have argued for in this chapter is to offer a thick description of the research setting. In the next chapter, I will detail Sutter International School and the climate in which my study took place.

Chapter Five: Standards Reform at Sutter International School

5.1 Introduction

In this chapter I will provide an in-depth description of standards reform at SIS, which will help to contextualize the student perspectives I analyse and discuss in chapters six through eight. In providing this rich backdrop I also hope to increase the internal validity of the analytic generalizations I make in those later chapters.

In chapters two and three, I applied the theoretical framework of my study to the empirical element being measured, SBGs and SBRCs. This application suggests that SBGs and SBRCs are a form of formative assessment. The process of how SBGs and SBRCs are implemented is complex, however, and the form they can take often varies between schools, classrooms and teachers. Of particular note, some schools use a hybrid SBRC that includes a traditional letter grade to go with the SBGs (see chapter three). Because grading mediational means are often used in idiosyncratic ways (see chapter three), it is important for SBG and SBRC research to clearly explicate the school context and form of the mediational means being measured.

Within this chapter, I will provide the details of the shape that SBG and SBRC reform took at SIS, the site of my study. I will begin by describing the school and its student population. I then discuss my process of coding data from interviews with three senior leadership members who led the reform at SIS. Next, I will review the timeline and broader shape of standards reform at SIS relying heavily on my interviews of senior leadership and the school's 2016 self-study for accreditation. This review sheds light on the historical, institutional and cultural forces which shaped key decisions of the reform, with a particular focus on the school's use of a hybrid SBRC. Another key aspect highlighted in this section is senior leaderships' perspective of teachers' wide range of

understanding about standards-based reform, even once the reform had been implemented. Following this broader overview of the school reform context, I will examine the specific SBG and SBRC processes senior leadership established for teachers to assess and report on standards.

Within this chapter I will draw from official school documents which guided the standards reform process, data from the school's website, the 2016 self-study from the school's most recent accreditation process, and my three semi-structured interviews with the senior leadership members who oversaw the SBG and SBRC reform process.

5.2 Overview of SIS

Sutter International School (SIS) is an American international school[10]. The school is a non-profit independent school founded in the 1950's that goes from kindergarten (age 5) to 12th grade (age 18). Although SIS is non-profit, in the U.S. it would be considered a tuition-charging private school. Tuition increases between the three divisions of the school (discussed shortly) with families of middle school students paying approximately $38,000 per year. Similar to many other American international schools, for some SIS families, a parent's employer pays this tuition as part of their expat compensation package.

Approximately 1,350 students attend SIS, and these students represent over 60 nationalities and speak more than 30 languages in addition to English. Typically, 75% of students hold a U.S. passport and more than 40% have a passport from two or more countries. 85% of graduating seniors attend U.S. universities which frequently include the top universities in the country, while the other 15% attend universities outside of the U.S., meaning that all graduating seniors go on to attend university.

[10] Location and region of the school has not been included to protect the anonymity of participants.

SIS has 150 full time teachers with 130 holding a master's degree. There are 45 full-time teachers in the middle school division. Teachers tend to stay at the school on average for a little over eight years, with changeover rates each year averaging 13% of the teaching population.

SIS is organized into three divisions within the school, consisting of the Elementary School (Kindergarten – 4th grade, ages 4-9), the Middle School (Grades 5-8, ages 9-14), and the High School (Grades 9-12, ages 14-18). Each division's leadership structure consists of a principal and a deputy principal. Overseeing the entire school is the head of school. The school also has a teaching and learning department with a director and deputy director overseeing the curriculum and teaching practices in all three divisions. The director and deputy director of teaching and learning work closely with the head of school and divisional leadership on all reform efforts related to classroom learning.

The three divisions have an approximately equal distribution of the 1,350 students, with the middle school where my study was conducted holding an average of 470 students aged 10-14. International schools typically have a high degree of transiency in their student population, as new job postings lead families to move to different countries. At SIS, however, approximately 85% of students return for the following school year. This high rate is part of the reason there is frequently a waiting list for students to attend the school. Recent years have also seen an increase in both local and international applicants, adding increased pressure on enrolment spots.

Within SIS, curriculum at all levels and departments is established and reviewed using a curriculum review cycle. By 2016, each subject area at SIS had adopted sets of standards (see chapter three) during these review cycles. According to the school's 2016 self-study for accreditation which will be examined in the next section of this chapter, the purpose of these curriculum reviews was "… to review and update curriculum, standards, assessment and

instructional practices in alignment with current research..." These cycles occurred every four to five years for each subject area and were led and facilitated by the teaching and learning department who worked closely with division leadership and the heads of the department undergoing the review from all three divisions within the school. Because all three divisions of the school are present in the curriculum review process, it provides the school with the opportunity to ensure that the curriculum and standards for each department is aligned as students move upwards through each grade of the school, alignment the school describes as "vertical." This vertical alignment is supported by adopting sets of standards that progress from Kindergarten to 12th grade. Further, standards and curricula are documented using the software, Atlas Rubicon, to map curriculum at all grade levels and allow teachers to see the progression of standards, curricula and assessments across the school. In theory, all of these structures for vertical alignment represent essential components for traditionally summative assessment mediational means to function formatively (Biggs, 1998). As I will discuss, however, the school's 2016 self-study makes clear that vertical alignment of standards-based assessment practices at SIS was aspirational and a work in-progress.

5.3 Interviews of SIS Senior Leadership

In chapter four I provided an overview of the rationale and considerations which guided my one-off semi-structured interviews with three members of the senior leadership team at SIS who led standards reform in the middle school. Before analysing and discussing these interviews in the next section, I will first share my coding process for the data from these interviews. As discussed in chapter four, these interviews were not intended for an in-depth investigation into senior leadership perspectives on standards reform at SIS, but to enhance my thick description of the broader context within which students' perspectives were shared through focus group and questionnaire data. This thick description

will lend support to my analytic generalizations in chapters six, seven and eight (Yin, 2018).

Soon after their completion, I transcribed audio recordings of the three senior leadership interviews into NVivo. At this time, I anonymized the senior leaders' names with the codes, SL-A, SL-B, SL-C.[11] Once transcribed, I used a simple a priori coding framework established from the key themes of the interview guide (see Appendix E). For an overview of the rationale behind these separate codes and their corresponding interview guide questions, please see Appendix H. I conducted one coding round, with the results represented in Table 5.1 below. Because one of the interviews was only 15 minutes long I was unable to address all of the items on the interview guide. This accounts for why some codes in the table below were only found in two sources.

Table 5.1: Senior Leadership Interview Coding

Code	**Sources**	**Occurrences**
Hybrid	2	17
Implementation Steps	3	31
Implementation Missteps	3	30
Parent Community	2	9
Teacher Idiosyncrasy	3	31

Once coding was completed, I used this data in the following section to further analyse and discuss the context of standards reform at SIS.

5.4 Standards Reform at SIS

The decision to shift to standards-based grades and report cards at the SIS middle school has its immediate roots in discussions that took place during the 2012-2013 school year. According to SL-C[12], the impetus for this change came from the observations of teachers that students' hyper letter grade focus was detracting from attention to the actual learning objectives. Further, leadership believed that there were varied and outdated assessment and reporting practices

[11] Please refer to chapter four for a discussion of my concerns about internal confidentiality.
[12] From this point forward, I will refer to senior leadership members by the codes discussed earlier in this chapter.

in place within the school. Because of this variation in teacher practices, the school hoped to implement standards reform to provide students with consistent assessment practices in all of their classes, with a focus on formative feedback towards learning goals. Within these assessment practices, standards reform was seen as an opportunity to prioritize depth of understanding in students' learning. As SL-C stated,

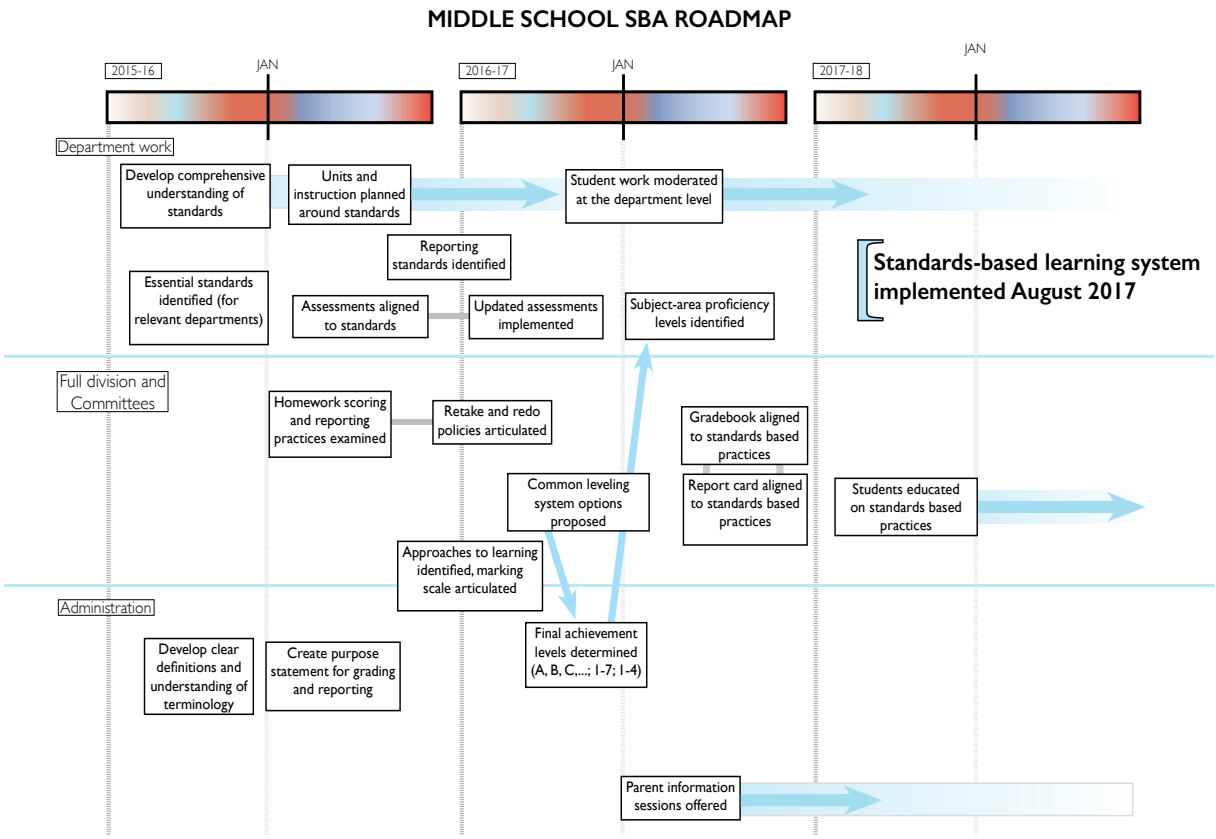
> Having those standards be central allow us to position deeper understanding as more integral to student learning experience…On a day-to-day basis, I think the quality of the thinking that the students need to do to be successful can be more sophisticated, can be deeper.

Thus, the rationale behind the decision to make the shift to SBGs and SBRCs was firmly situated in enhancing student learning through providing students with feedback about meaningful learning objectives. I will return to this rationale later in this chapter to contrast it with the cultural, historical and institutional forces which dictated the shape that the SBRC ended up taking.

5.4.1 Standards-Based Assessment Roadmap

With this rationale behind standards reform at SIS, at the start of the 2015-2016 school year, middle school administrators, the head of school, and the teaching and learning department created the Middle School Standards-Based Assessment Roadmap (see Figure 5.1 below). This roadmap outlined the key steps in the standards reform process leading up to the full implementation of a standards-based grading and reporting system at the start of the 2017-2018 school year.

Figure 5.1: Planned Timeline of Standards Reform in SIS Middle School



The three years within this reform timeline incorporate the year before data collection with my focus groups and questionnaire began (2015-2016), the last year with traditional letter grades and report cards (Year One of my data collection, 2016-2017) and the first year with standards-based grades and report cards (Year Two of my data collection, 2017-2018). The three rows of the Roadmap represent where the work would take place within the school. Department work in row one refers to work done by teachers within separate disciplinary departments, such as English, social studies, math, science, etc. Full division and committees in row two refer to work done by the entire middle school and separate committees that were established, such as an SBRC committee. These committees were made up of a middle school administrator, an administrator from the teaching and learning department, and teachers from across the middle school division who were selected for their expertise in that particular area. Administration in row three refers to the head of school, the teaching and learning department, and the middle school principal and deputy

principal. In general, the roadmap begins with establishing a theoretical understanding of standards-based assessment, grading and reporting. Both SL-C and SL-A shared that this stage of the reform focused on building a common understanding amongst teachers about key aspects of assessment such as FA and SA, criteria-referencing vs. norm-referencing, and authentic demonstrations of understanding.

Following this initial phase, the roadmap moves into the particulars of aligning curriculum, assessments and grading practices with the standards. The final steps of year two in the road map and the start of year three identify educating students and parents on standards-based reform, coinciding with the 2017-2018 school year, the first year SBRCs were in place. While the roadmap had a flow from the theoretical to the practical, SL-A and SL-C reflected later that senior leadership had waited too long to have teachers make standards-based grading changes in their classroom practices, which they felt resulted in some teachers feeling overwhelmed and unprepared to make the standards-based shift, even after SBGs and SBRCs were implemented.

While the reform efforts captured in the roadmap were led by the groups previously mentioned, the school also brought in external consultants during the 2016-2017 school year to lead instructional sessions with teachers to guide them on the principals of standards-based grading discussed in chapter three.

5.4.2 The State of Standards Reform on the Eve of Data Collection

In the paragraphs that follow I will provide a snapshot of reform at SIS in early 2016 on the eve of collecting data for my study. In addition to senior leadership interviews, this snapshot is constructed from the school's 2016 self-study for accreditation. This accreditation process is standard for American international schools, occurs every five years, and at SIS it was conducted by the external agency, the Middle States Association (MSA).

Within the school's 2016 self-study, an important theme amongst teacher comments was varied understanding and implementation levels with the standards-based shift. Some teachers expressed that the reform efforts had already created a shift at the school and the focus on standards had led students to more critical thinking in many classes. Others agreed with the reform but described feeling unprepared for the standards-based shift and needing more training and time to understand the reform. Specifically, some identified UbD and standards-based grading practices (see chapter three) as areas where they needed more training. Other teachers stressed that there was not enough time in their daily schedules to implement these new reform measures. Some expressed concern that some departments had adopted standards but that teachers had not actually applied them in practice to their courses. Other teachers felt that that not all teachers were committed to the school's standards reform efforts and perceived that many teachers were still hanging on to traditional letter grade practices, with a focus on the letter grade rather than the learning outcomes it should represent. The 2016 self-study went on to summarize these issues in recognizing that the school was in the early stages of standards implementation with significant work remaining.

In addition to the varied levels of teacher understanding, another issue which arose in teacher comments from the self-study was educating the parent community on the standards reform. Some teacher comments recognized the need to have a clear plan for communicating to parents about the new standards-based system, and more specifically to help parents understand the implications of SBGs and SBRCs. While the reform roadmap (see figure 5.1) included final steps of educating students and parents about the standards-based shift, there were later questions about how effective these efforts were. Reflecting back on the reform process in 2019, SL-A recognized,

> We really, I don't feel, educated [parents] enough about why and what does [standards-based assessment] mean…and kids. It just didn't happen and so it was a bit of a band-aid that first year. Kids were just like…"what does this mean?"

The school did take steps to educate parents. In the latter part of the 2016-2017 school year, administrators recorded video tutorials which were sent out to the entire parent community through a weekly newsletter. These coincided with a couple of in-person tutorials at monthly parent meetings. Educating students about the SBG and SBRC shift was left to classroom teachers and it is unclear the extent to which they explicitly addressed the issue and the form that these lessons took.

In addition to considering issues of parent education, the 2016 self-study also highlighted the school's use of rubrics. In chapters two and three, I articulated the specific processes of formative assessment I theorize to take place in my study: student self-assessment through SBGs and SBRCs linked to criteria rubrics. The 2016 self-study makes it clear that teachers, parents and administrators perceived that rubrics were in place within the school's assessment practices. While there was agreement that rubrics were in place within the school, comments from the self-study regarding some teachers' lack of understanding about standards raise questions about whether these rubrics were aligned to the standards, a key requirement for SBGs and SBRCs to function formatively through self-assessment processes.[13]

This question of alignment between the standards and classroom assessment practices across the school was a broader theme in the self-study, which concluded that there was still a great deal of work to be done in vertically aligning curricula and assessment practices across classrooms and grade levels. The study identified implementing a school-wide, standards-based assessment

[13] Refer to FA/SA mixing in chapter three for full discussion.

system with alignment between standards, rubrics, and classroom assignments as a major next action step for the school.

This self-study is a snap-shot of where the school was at in the early part of 2016. While there were signs of cultural changes in the school related to the standards-based shift, the self-study reveals a great deal of work to be done and teachers in very different places with their understanding of standards reform. In the next section I will examine senior leaders' perspectives from 2019 interviews on teachers' implementation of standards-based grading practices during the reform process. These comments reveal the perception of senior leadership that many of the issues captured in the 2016 self-study continued well into the 2017-2018 school year when standards-based grades and report cards were implemented in the middle school.

5.4.3 Teacher Education of Standards Reform at SIS

While the 2016 self-study represented a snap-shot of the school in early 2016, during semi-structured interviews in 2019 senior leadership members reflected back on the 2016-2017 and 2017-2018 school years during which I collected focus group and questionnaire data. In leading reform efforts, SL-A, SL-B and SL-C all spoke about the variation they perceived in teachers' understanding and implementation of SBGs and SBRCs. SL-A and SL-C shared that some teachers were eager to implement the new grading practices, while for others, the transition at start the 2017-2018 school year to standards-based grading and reporting was a source of great anxiety and frustration. SL-B believed that some frustrated teachers were still heavily rooted in a letter grade, points-based grading system, while SL-A felt that some of teachers' anxiety came from concerns about the lack of student and parent education regarding the new standards-based system.

Because of perceived variation in teacher understanding about standards-based assessment practices, senior leadership differentiated their methods to

support teacher development. In interviews, SL-A and SL-C described how they had initially attempted to educate teachers through division-wide faculty meetings. They eventually found that the most effective way to meet individual teacher needs was to provide instruction and feedback in smaller department meetings, meetings with teaching teams, meetings with individual teachers, and through optional drop-in sessions. SL-C summarized these efforts in sharing, "narrowing the focus for our support was one of the ways that we found it to be more effective to bring about institutional change." SL-C perceived that these sessions were taxing for teachers

> … because while doing the full-time job of teaching their students on a day to day basis, they needed to layer on top of that how to more substantially change their ways in which they were providing feedback to students and then reporting out.

Senior leadership perceived much teacher progress over the course of the reform efforts. With the initial implementation of SBGs and SBRCs during the 2017-2018 school year, however, all three of them recognized the varied practices that were in use by some teachers that ran counter to standards-based grading recommendations, such as giving zeroes for missing assignments (see chapter three). By the second year of SBG and SBRC implementation (2018-2019, one year after my focus groups and questionnaire data collection had ended), however, SL-A and SL-B believed that some teachers who had struggled the most with the standards-based shift had made personal breakthroughs in reforming their practices.

The perspectives of administrators captured in this section, along with the teacher perspectives identified earlier in the 2016 self-study suggest that the standards reform at the school was a messy process with many teachers at their own unique place of understanding and implementation with SBG and SBRC practices, particularly during the initial implementation in the 2017-2018 school year.

5.4.4 The Hybrid SBRC

A key step in the implementation of standards-based grading and reporting at SIS was determining the form of the SBRC. Similar to many schools initiating SBRC reform, SIS decided to use a hybrid SBRC for the majority of the middle school grade levels (see chapter three for hybrid overview). I will now examine the details and rationale of this decision as shared by senior leadership in interviews.

The school decided initially to use a hybrid SBRC (see Figure 3.4) with both letter grades and standards-based grades for 7th and 8th grade, while 5th and 6th grade would have an SBRC with only standards-based grades. Upon further deliberation, it was decided for 6th grade to have a hybrid report card for the first year of the SBRCs (2017-2018) because without a one-year hybrid for 6th grade, that student cohort would have had letter grades in 5th grade (2016-2017 school year), gone to just standards-based grades in 6th grade, (2017-2018 school year) and then transitioned back to a letter grade hybrid SBRC in 7th grade (2018-2019 school year). By having a hybrid SBRC the first year of the SBRC shift (2017-2018) for 6th grade, and then switching 6th grade to a standards-based report card without letter grades for the 2018-2019 school year, this cohort was provided with a smoother progression in grading systems through their middle school experience.

Examining the cultural, historical and institutional forces that shaped the form that the SBRC ultimately took is illuminating for understanding the mediated action students shared through focus group comments that I will discuss in chapters six, seven and eight. There was much deliberation that led to the decision to have a hybrid SBRC in some grades and an SBRC with no letter grades in others. SL-C articulated the need to make sure that there was community support for the SBG and SBRC reforms. This larger community consisted first of parents, but also included external considerations such as

students who would need to transfer to other schools, and students' paths to universities. SL-C summarized these considerations as "Everything outside just the scope of the student in our 7th grade classrooms on a day to day basis."

Adhering to perceived expectations of the parent community also drove the decision to use the hybrid SBRC. SL-C shared that the choice of using a hybrid SBRC for some of the grade levels

> …was a recognition of how shifting away from something [letter grades] that has such a long history and feels like it has got such a concrete gold standard currency to it, will provide uncertainty to parents in terms of understanding how their students are doing.

Both SL-A and SL-C articulated this rationale for the hybrid SBRC further: going too far with the grade and report card reform by not incorporating the "gold standard" letter grades could cause backlash from the parent community which potentially could force the school to have to backtrack on the reform measures it had taken. SL-C further shared that part of this potential backlash came from concerns within in the broader community that with a shift to SBGs, students would not be prepared for next levels in their schooling when they would encounter letter grades. Ultimately, the decision to use a hybrid SBRC was "…more just for transition and also in recognition of the history of what the community, ours and many others, were used to" (SL-C).

Seen from another perspective, the hybrid system was a compromise solution that contained inherent contradictions. As SL-A commented, "…we were trying to move away from numbers, letters, and percentages, and [a hybrid SBRC] still is bound by that in a way." Viewed through a Wertschian lens, the historical, institutional and cultural forces which shaped the form of the new hybrid SBRC had very little to do with the ideal design for student learning. These institutional and historical forces behind the hybrid SBRC at SIS can be contrasted with those behind the rationale for the larger standards reform at the school: empowering deeper understanding in students through formative feedback.

As mentioned earlier, following Year Two of my data collection, the school shifted the 6th grade SBRC to not include letter grades. The following school year (2019-2020) it removed letter grades from the 7th grade SBRC. While these changes came after my focus group and questionnaire data had been collected, they reflect issues the school encountered with the hybrid SBRC during my second year of data collection. SL-A and SL-C shared that the decision to move 7th grade to an SBRC with no letter grades came partially from the recognition that letter grades were a "clunky translation" (SL-C) of standards-based grades and the larger recognition that the standards-based grading system worked best on its own. This "clunky translation" speaks to a common theme from the student focus groups discussed later in chapters six, seven, and eight: these two separate grading and reporting systems hold conflicting foundations with one rooted in norm-referencing and the other in criteria-referencing. Combining the two together at SIS, while understandable for practical external considerations and very consistent with other schools' rationales for using a hybrid SBRC (see chapter three), led to conflicting messages sent to students about their learning.

5.5 SBG and SBRC Processes at SIS

Having described the broader context of standards-based grades and report card reform at SIS, I will now look at the specific SBG and SBRC processes the school instructed teachers to use in their classrooms. In the cases where senior leadership reflected on teacher deviations from the expected norm, I have included their comments and summarized the corresponding implementation issue. Because I was a teacher at the school during the years of the study, I will use examples from my social studies class to clarify these processes.

By Year One of my data collection (2016-2017), the school had adopted standards for all of the core class subjects through curriculum review cycles. English, math, and social studies used the Common Core State Standards

(CCSS), while science adopted the Next Generation Science Standards (NGSS) (see chapter three). As I discussed in chapter three, sets of standards include more learning objectives than can be taught in a school year, and it is common for any given standard to combine multiple learning objectives. Because of this, teachers must unpack standards and prioritize which standards to address in their classes (Marzano & Kendall, 1998). As departments at SIS engaged in the shift to SBGs and SBRCs, one of their first steps was unpacking and prioritizing standards.

Within the CCSS, a key skill my teaching partner and I identified was analysing evidence with logical reasoning. As such, we prioritized the CCSS standard, CCSS.ELA-LITERACY.W.7.1.B: "Support claim(s) with logical reasoning and relevant evidence, using accurate, credible sources and demonstrating an understanding of the topic or text." (Common Core State Standards Initiative, 2019). While the other elements of this standard are important, we unpacked it with analysis in mind to arrive at "Support claim(s) with logical reasoning."

Next, we created a rubric for this standard (see Figure 5.2 below). We made the descriptions of each achievement level generalized and not specific to the parameters of individual assignments. This allowed us to use the same analysis rubric on different assignments across the school year. Our intention with this alignment was to allow the rubric to function summatively for assignments that had ended and formatively for future work students would do with the skill of analysis[14].

[14] Please refer back to chapter two for a full discussion on mixing formative and summative assessment.

Figure 5.2: Rubric for Analysis

	Extending	Meeting	Progressing	Beginning
Analyse Evidence	In addition to meeting the standard, I strive for greater depth by analysing evidence through multiple lenses and by corroborating them with my previous analyses.	I analyse evidence with logical reasoning, explaining how/why the evidence links to the claims and continually asking (and trying to answer) "so what?", "who cares?" and "why does this matter?"	I begin to explain how/why evidence links to the claim, but the reasoning lacks clarity, depth, or contains logical fallacies.	I summarize evidence that could be related to my claim, but I do not explain how/why the evidence links back to the claim.

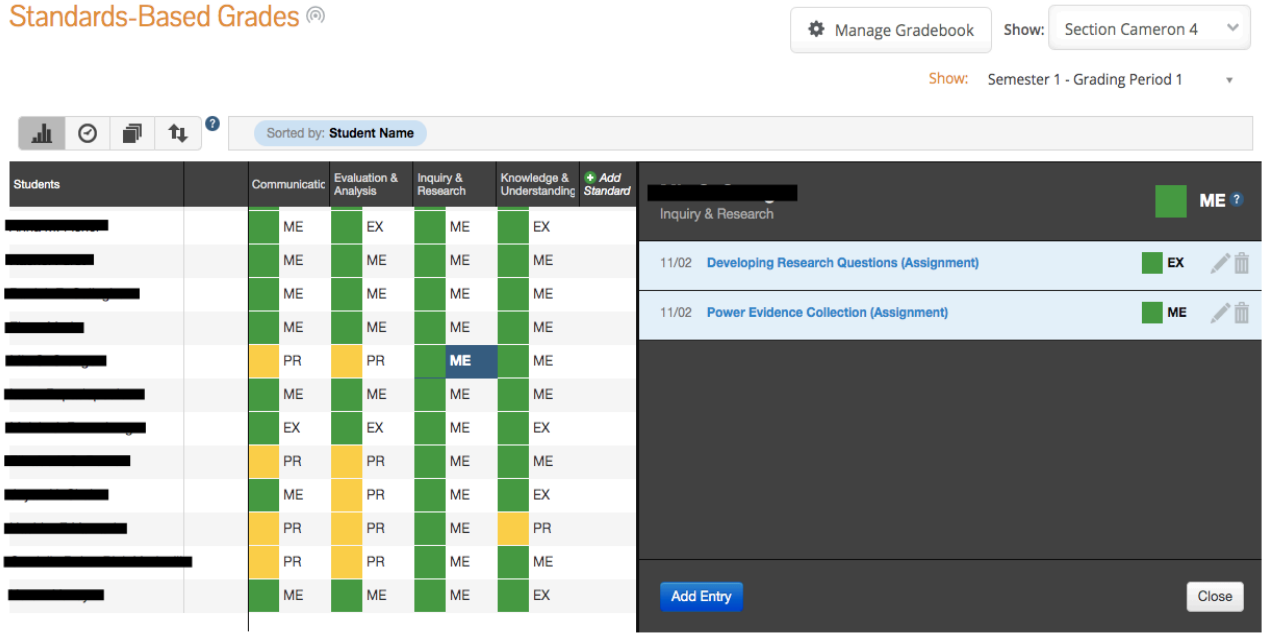
SIS had to decide how to provide meaningful information about student achievement on the SBRC without it becoming overwhelming with too much information. Because classes typically assess too many prioritized standards over the course of a semester to report out on all of them on the SBRC, the school decided to have approximately four broad reporting standards for each class (Guskey & Bailey, 2001; Muñoz & Guskey, 2015). The reporting standards for social studies were Inquiry & Research, Evaluation & Analysis, Knowledge & Understanding, and Communication.

As students received standards-based grades for prioritized standards on assignments during the semester, these would then be entered into the standards-based gradebook (see Figure 5.3 below). The gradebook was organized by larger reporting standards in the columns. For example, in Figure 5.3 below, the Inquiry & Research reporting standard holds assignment grades for the prioritized standards Developing Research Questions and Evidence Collection.

To determine the final standing for each reporting standard at the end of the semester, the school recommended that teachers use the most recent evidence from assignments to determine the standards-based grade. Teachers

were allowed autonomy to use other measures if they felt that the most recent evidence was not representative of a student's learning.

Figure 5.3: SIS Standards-Based Gradebook



For teachers in grades 6-8 who needed to arrive at an overall letter grade for the hybrid SBRC, senior leadership provided a conversion chart (Figure 5.4 below) which was also published in the student handbook[15]. Teachers were instructed to first convert the standards-based grades for the reporting standards into numerical values (Extending = 4, Meeting = 3, Progressing = 2, Beginning = 1), find the average of these numerical values, and then convert this average to a letter grade using the conversion chart. Teachers were also provided with an Excel calculator (see Figure 5.5 below) to facilitate this process.

[15] This was given to all students at the start of every year. It contained weekly calendars for writing down homework and a section which contained logistical information, such as school rules and policies.

Figure 5.4: Standards-Based Grade to Letter Grade Conversion Chart

Reporting standards

[redacted] is a standards-based school. We have selected standards and benchmarks that reflect the latest research on learning outcomes for our students. Students receive feedback based on targets for different learning areas in each subject. For example, in Humanities or English, students will receive separate proficiency marks for Reading, Writing, Listening & Speaking.

The proficiency levels are as follows:

Beginning	Progressing	Meeting	Extending
• is not yet meeting [redacted] grade-level standards and benchmarks • requires support, reinforcement and additional instruction to understand grade level skills and concepts • numerical equivalent = 1	• is working towards meeting the [redacted] grade-level standards and benchmarks • demonstrates a partial or inconsistent understanding of grade-level skills and concepts • numerical equivalent = 2	• consistently and independently meets [redacted] grade level standards and benchmarks • demonstrates a solid understanding of grade-level skills and concepts • numerical equivalent = 3	• in addition to what is required to meet the standard, the student goes beyond what has been taught and applies their learning in a more sophisticated way • numerical equivalent = 4

In Grade 5, students will receive only proficiency level marks on their report card. In Grades 6-8, in addition to the proficiency levels for each main standard, students will also receive an overall letter grade. This letter grade will be determined after taking an average of the main reporting standards and using the scale below:

Numerical Scores to Letter Grades	Numerical Equivalent
3.75-4.0	A+
3.26-3.74	A
3.0-3.25	A–
2.75-2.99	B+
2.26-2.74	B
2.0-2.25	B–
1.75-1.99	C+
1.26-1.74	C
1.0-1.25	C–
0.75-0.99	D+
0.30-0.74	D
0.00-0.29	F

Figure 5.5: Excel Letter Grade Conversion Calculator

Overall Grade Averaging

Enter the numerical score for each reporting standard to see the average to be converted to a letter grade.

	Reporting standard score
	Reporting standard score
	Reporting standard score
	Reporting standard score
	Reporting standard score
	Reporting standard score
#DIV/0!	#DIV/0!

Proficiency Level Score Average	Letter Grade
3.75-4.0	A+
3.26-3.74	A
3.0-3.25	A-
2.75-2.99	B+
2.26-2.74	B
2.0-2.25	B-
1.75-1.99	C+
1.26-1.74	C
1.0-1.25	C-
0.75-.099	D+
0.3-0.74	D
0.00-0.29	F

For the final steps of the SBRC (see Figure 5.6), teachers provided a narrative course overview and gave behaviour marks for the students' "approaches to learning." This followed recommendations from experts such as Guskey (2015) to separate behaviour from achievement on the SBRC. SIS used the following behaviour categories: collaborative, respectful, responsible, reflective, engaged, and resilient. Students were then given marks based on the frequency they exhibited these traits.

Figure 5.6: SIS SBRC

Middle School Report Card
December 2017

[redacted] Grade 7 Semester 1, 2017

Science [redacted]

As Grade 7 scientists and engineers, students focused on types of energy and energy transfers. Students developed knowledge and understanding of these concepts through studies of motion and heat.
Through investigations, students had opportunities to work in pairs, small groups and independently on the design process, analyzing data, constructing explanations, and creating models. Investigations allowed students to explore concepts related potential and kinetic energy. These investigations included incline labs, designing a balloon car, and using temperature probes to collect data to analyze changes as thermal energy is removed and added. Links were made between the physical sciences and earth sciences by applying concepts related to thermal energy to the water cycle. An underlying concept for the semester is that students will understand conservation of energy, that the total change of energy in any system is always equal to the total energy transferred into or out of the system and that when the energy of an object changes, energy is transferred to or from the object.

Approaches to Learning

		Semester 1	Semester 2
Collaborative	• *Works well with others* • *Understands and respects diverse approaches* • *Communicates through dialogue and participation*	Consistently	
Respectful	• *Treats others the way we expect to be treated* • *Appreciates own culture and the cultures of others* • *Values diversity of thought and experience*	Usually	
Responsible	• *Takes ownership of words, actions and learning* • *Fulfills obligations to self and to others* • *Works hard*	Consistently	
Reflective	• *Values mistakes as learning opportunities* • *Reflects on own learning* • *Sets and pursues goals for improvement*	Usually	
Engaged	• *Intellectually curious about the world* • *Cultivates interests and strengths* • *Finds passion and purpose in their learning*	Usually	
Resilient	• *Stretches and grows by taking risks* • *Resourceful in the face of challenges and change*	Usually	

Reporting Standards

		Semester 1	Semester 2
Knowledge & Understanding	• *Synthesizes and applies specific content covered as described in the curricular description*	Progressing	
Science & Engineering Practices	• *Uses the key concepts to develop models, investigate, and draw conclusions from data*	Meeting	
Communication	• *Communicates the key concepts through questioning, explaining and engaging in scientific argument*	Progressing	
Overall Grade		B	

5.5.1 Teacher Implementation Issues

The conversion charts to arrive at letter grades for the hybrid SBRC (see Figure 5.6 above) were created by senior leadership with the sole intention of

teachers using them to calculate end of term final letter grades for the hybrid SBRC in grades 6-8. However, in the initial year of SBG and SBRC implementation (2017-2018), SL-A and SL-C shared that some teachers began using this chart on individual assignments, even specifying letter grades on rubrics along with the corresponding numerical figures down to the decimal points from the conversion chart. For example, in the illustrative model in Figure 5.7 below, rather than identifying that a student was *Meeting* for their analysis, the teacher would give the student a distinct 3.6 score for their analysis, sometimes even including a letter grade *A* with this number.

Figure 5.7: Rubric with Number Distinction

	Extends	**Meets**	**Progressing**	**Beginning**
Analyse Evidence	In addition to meeting the standard, I strive for greater depth by analysing evidence through multiple lenses and by corroborating them with my previous analyses.	I analyse evidence with logical reasoning, explaining how/why the evidence links to the claims and continually asking (and trying to answer) "so what?", "who cares?" and "why does this matter?" 3.6	I begin to explain how/why evidence links to the claim, but the reasoning lacks clarity, depth, or contains logical fallacies.	I summarize evidence that could be related to my claim, but I do not explain how/why the evidence links back to the claim.

SL-A and SL-C also described some teachers wanting to delineate between a "*high Meeting*" and a "*low Meeting*." For example, in the illustrative model in Figure 5.8 below, the teacher has circled the right side of the meeting category to signify that the student is a low meeting, or only just barely *Meeting* the standard as it is described in the rubric.

Figure 5.8: Rubric with High/Low Distinction

	Extends	Meets	Progressing	Beginning
Analyse Evidence	In addition to meeting the standard, I strive for greater depth by analysing evidence through multiple lenses and by corroborating them with my previous analyses.	I analyse evidence with logical reasoning, explaining how/why th evidence link to the claims an continually asking (and trying to answ er) "so what?", "who cares?" and "why does this matter?"	I begin to explain how/why evidence links o the claim, ut the easoning acks clarity, epth, or ontains ogical fallacies.	I summarize evidence that could be related to my claim, but I do not explain how/why the evidence links back to the claim.

Speaking on these idiosyncratic teacher grading practices, SL-C perceived that some teachers found it was hard to identify only four levels of performance for a standard and they felt the need to distinguish areas of subtlety. It was unclear how widespread these practices were, but it was occurring enough for the senior leaders to address the issue at a division-wide faculty meeting during the second semester of Year Two of my data collection (2017-2018). SL-B felt that some teachers' continued practice of putting a letter grade on rubrics after SBGs and SBRCs had been implemented indicated a larger lack of understanding about the principles of standards-based assessment and grading. Teachers issuing numerical scores with letter grades and distinguishing between a high and low standards-based grade raises the question of how different the SBG and SBRC system was for students in the 2017-2018 school year when compared with the letter grade system of 2016-2017. I will return to these variations in teachers' grading practices when I discuss them from students' perspectives in chapter seven.

Other issues arose with determining the final letter grade for the hybrid SBRC. One was that teachers veered from the official school policy in arriving at the overall letter grade, a process described earlier in this chapter. SL-C shared that "you might have a student with a *Meeting*, *Meeting*, *Progressing* [on the

reporting standards] and they're getting a *B+*, and another student with a *Meeting*, *Meeting*, *Progressing* and they're getting a *B*." This issue was widespread during the 2018-2019 school year and SL-C believed that the official policy of using the Excel calculator was rarely followed. If most teachers were not following the official school policy as SL-C perceived, it raises the questions of how teachers were arriving at the final letter grade standing for the hybrid SBRC, and how varied their practices were in this calculation. The possible implications for students of these practices were that the traditional issues of opaque and idiosyncratic grading practices in determining final grades were alive and well with the hybrid SBRC. These variations in teachers' standards-based grading practices, as perceived by senior leadership, suggest that even during the second year of the SBRC implementation - one year after I had completed gathering focus group and questionnaire data - many teachers at the school were still struggling to reconcile the two grading systems being combined in the hybrid SBRC.

5.6 Conclusion

Reflecting on standards-based reform at the end of the 2018-19 school year, senior leadership recognized many of the positive changes that had occurred within assessment practices at the school. Some of these changes reflected technical aspects of SBGs: "People don't grade homework, we don't have extra credit anymore…we've separated behaviour from achievement" (SL-B). Other changes identified were at a broader level: "I hear more dialogue that's more concretely rooted in [students'] performance towards a learning standard as opposed to generalities about they received this grade on the test. So I think that link is more clear" (SL-C).

While senior leadership was quick to give credit for these changes, they also recognized that there was still a lot of work to be done. SL-B perceived a deeper "cultural shift" that still needed to occur for some teachers in

understanding "the why" behind the standards reform and shifting evaluative power traditionally held by teachers over to the students.

This chapter has provided a picture of the complicated process of SBG and SBRC reform at SIS and the many components which led to the final end of term grades. The transitory state of SIS, particularly during the first year of SBG and SBRC implementation, sets the backdrop from which students shared their perceptions about SBGs and SBRCs in chapter seven. First, however, in the next chapter I will discuss the iterative process I used to code focus group data, and then I will examine the meaning that students constructed from traditional letter grades and report cards. This examination will corroborate SL-C's perception, stated earlier in this chapter, that students' hyper letter grade focus directed their attention away from learning criteria, a phenomenon which likely contributed to letter grades functioning in purely summative ways.

Chapter Six: Results, Analysis and Discussion of Research Question 1: *What meanings do students construct from traditional letter grades and report cards?*

6.1 Introduction

In this chapter I will use focus group data from my study to answer the research question, *What meanings do students construct from traditional letter grades and report cards?* I will begin by describing the iterative process of developing a coding framework and assigning those codes to focus group responses. Following this overview of the coding process I will analyse the key themes which became apparent during the coding process. These themes were parents' influence on the meanings students constructed of letter grades, letter grade formulation, feedback from letter grades, grades as labels, and norm-referencing. Next, I will situate this analysis within the broader framework of my study by connecting the focus group findings to assessment theory, the work of Vygotsky and Wertsch, and the relevant grade and report card research. In discussing these findings within this framework I will identify gaps and overlaps between the two to formulate my analytical generalizations.

6.2 Coding

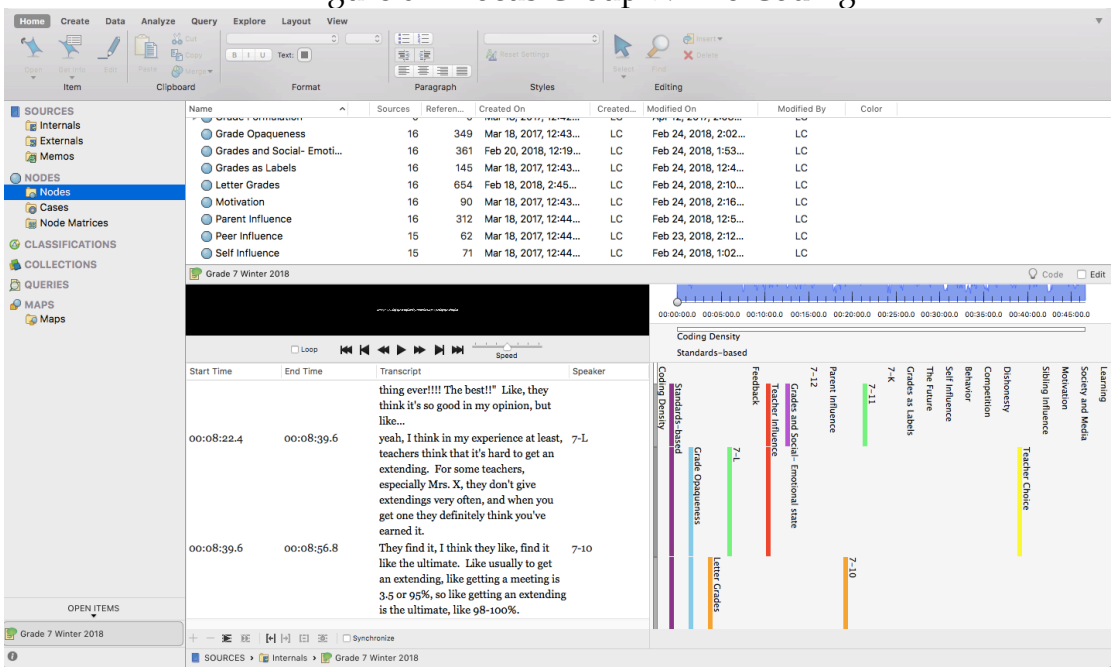
6.2.1 Initial Coding Framework

The initial coding framework (see Table 6.1 below) was established after transcribing the pilot focus group from the audio recording into the coding software, NVivo (see Figure 6.1 below).

Table 6.1: Initial Coding Framework and Occurrences for Focus Group Pilot

Code	Occurrences
Family Influence	19
Grade Meaning:	
Behaviour	11
Learning	7
Teacher's Choice	9
Grade Opaqueness	14
Grades as Labels	18
Media Influence	2
Peer Influence	4
Self Influence	9
Standards-based	14
Teacher Influence	5
The Future	10

Figure 6.1: Focus Group NVivo Coding



Following the Vygotsky and Wertsch underpinnings of my study, I established a priori codes related to how the social milieu influences the meaning students construct from grades. These codes were *Family Influence, Peer Influence, Media Influence* and *Teacher Influence*. While coding, some student comments were conspicuously absent of any reference to these outside influences and revealed a more intrinsic standard in how students thought of grades. From this observation, I added the code of *Self Influence.*

To investigate the meanings students construct from grades and report cards, the a priori codes of *Behaviour, Learning*, and *Teacher's Choice* were added under the heading of *Grade Meaning*. *Behaviour* referred to linking grades with student behaviours such as effort and participation, while *Learning* referred to the recognition of grades representing student learning achievement, and finally, *Teacher's Choice* was for instances of teachers using methods to generate grades that were beyond student control. *Grade Opaqueness* was added to investigate instances of student confusion in the meaning of the grade they received.

Once coding began, three other themes became quickly apparent. First was the idea that grades represented a permanent identifier of student success and intelligence. This theme was identified with the code, *Grades as Labels*. Relatedly, many students thought of letter grades as the key to future success and happiness. To capture this theme, I added the code, *The Future*. Finally, the code, *Standards-based* was added because the topic came up naturally, even during student discussion about prompts that had only referred to letter grades.

6.2.2 Revising the Initial Pilot Codes

The first round of Year One (2016-2017) coding followed the completion and transcription of each of the Year One focus group meetings – a period of time that spanned four months. The second and third rounds of coding were done over a week-long school holiday, approximately two months after the final Year

One focus group had been conducted. For an overview of the process of revising the coding framework, please see Appendix I.

Through this process, I arrived at the following coding framework for my Year One focus groups:

Table 6.2: Final Year One Focus Group Coding Framework

Competition
Dishonesty
Feedback
Grade Formulation
Behaviour
Learning
Teacher Choice
Grade Opaqueness
Grades as Labels
Motivation
Parent Influence
Peer Influence
Self Influence
Sibling Influence
Society and Media
Standards-based
Teacher Influence
The Future

6.2.3 Year One Coding

I have just described the evolution of my coding framework for the Year One focus groups and I will now describe the process I used for assigning these codes to transcripts of focus group discussions.

Data was coded on an iterative basis. Two rounds of coding were planned to increase internal validity of my assigned codes by double checking them for accuracy (Cohen et al., 2011; Coolican, 2009), and also because new codes had been added during the first round of coding. This second round of coding also identified problematic issues of unnecessary overlap between codes. For example, *Behaviour* was a sub-code underneath *Grade Formulation*. During the second coding round, I realized that this category needed to distinguish grade formulation from behavioural factors (e.g., participation, effort). Other references to *Behaviour* should have been coded as the *Feedback* students took

away from letter grades (e.g., I only got a C, and therefore I need to start putting in more effort to this class).

Another example of coding overlap resulted from me coding all comments about grades representing more than student learning as *Grades as Labels*. This left the category too broad and with too much overlap. I narrowed this category to specify how grades became aspects of student identity with a degree of permanency.

Because of issues like those described above, I decided that a third round of Year One focus group coding would be necessary to revise previous codings. Prior to beginning this third round, I spent time defining each code with the aim of further increasing the validity of my codings through greater accuracy (see Appendix J).

During the third coding round I reviewed previously assigned codes, checking them against definitions to clarify ambiguities, incorrect codings, and overlap between codes. Table 6.3 below represents final Year One coding occurrences, with the sources column referring to the number of focus groups where codes were identified. With four focus groups across the four grade levels meeting twice, there were eight possible sources in which a code could occur.

Table 6.3: Final Year One Focus Group Coding Occurrences

Code	Sources	Occurrences
Competition	8	28
Dishonesty	7	15
Feedback	8	73
Grade Formulation		
Behaviour	8	53
Learning	7	18
Teacher Choice	7	35
Grade Opaqueness	8	109
Grades as Labels	8	82
Motivation	8	43
Parent Influence	8	127
Peer Influence	8	26
Self Influence	6	23
Sibling Influence	4	7
Society and Media	5	20
Standards-based	8	47
Teacher Influence	8	49
The Future	7	55

6.2.4 Year Two Coding

My coding of Year Two data (2017-2018, see Figure 4.1 for timeline of study), rounds four and five overall, took place over a week-long school holiday one week after completion and transcription of the final focus groups in February 2018. I will now describe this process of coding and revising the coding framework.

Topics of Year Two focus group discussions switched to exploring SBGs and SBRCs. Round four of coding established the pre-existing framework's robustness for sorting the new data, as codes were general enough to apply to letter grades and standards-based grades. For example, *Grade Opaqueness* in Year Two still applied to the ways in which the meaning of standards-based grades was unclear to some students. To distinguish between which types of grades comments referred to, the new code, *Letter Grade* was added to go along with the pre-existing code, *Standards-Based.*

Round Four of coding involved adding *Letter Grade* codes to Year One transcripts, reviewing Year One codings for accuracy with evolving code definitions, and completing an initial coding of the Year Two focus group data. During this fourth round of coding, a new theme became apparent related to grades' impact on students' social-emotional state. Some of these emotional aspects had initially been coded under *Grades as Labels*, but they lacked the permanence associated with that code. Other references to these emotions had been left uncoded. *Grades and Social-Emotional* was added to the coding framework.

My fifth and final coding round reviewed Year Two codings to increase their accuracy. This round also coded the first half of the focus groups for missed *Grades as Social-Emotional* codes and recoded some *Grades as Labels* to *Grades as Social-Emotional*. Final coding occurrences from both years of data collection are represented in Table 6.4 below.

During Year Two coding, combinations of existing codes were utilized to represent new phenomena, rather than creating additional, more specific codes. For example, a repeated theme of parent confusion about the meaning of the new standards-based grades and report cards was represented with *Grade Opaqueness + Parent Influence + Standards-based*. Alternatively, students talked of parents in these situations wanting to know how SBGs translated into letter grades. This was represented by adding the *Letter Grade* code to the combination above. My ability to create these code combinations signified an effective and robust coding framework.

Table 6.4: Final Focus Group Coding Occurrences

Code	Sources	Occurrences
Competition	16	76
Dishonesty	10	23
Feedback	16	262
Grade Formulation		
Behaviour	16	100
Learning	15	52
Teacher Choice	13	51
Grade Opaqueness	16	349
Grades and Social-Emotional	16	361
Grades as Labels	16	145
Letter Grades	16	654
Motivation	16	90
Parent Influence	16	312
Peer Influence	15	62
Self Influence	15	71
Sibling Influence	6	11
Society and Media	9	29
Standards-based	16	586
Teacher Influence	16	160
The Future	14	81

After completing this iterative coding process, major codes were then collapsed to allow for broader analysis of focus group data to answer corresponding research questions. For example, letter grade comments coded under *Grades as Labels*, *The Future*, and some *Grades and Social-Emotional* were all included in the analysis sub-heading: *Grades as labels.*

6.3 Analysis of Research Question 1

Using the coded focus group data I described in the previous section, I will now analyse focus group data related to the first research question of my study.

6.3.1 Parent Influences on Students' Letter Grade Meaning

Participants in focus groups frequently referenced social influences that impacted the way they thought of letter grades. While there were occasional

comments relating to teachers and peers, for a substantial majority of students, the strongest social influence on their constructed meanings from letter grades was their parents. A small minority of these students referenced parents that created non-threatening interactions around letter grades, represented by comments such as, "My parents, they don't really care how good my grades are. They just care that I work hard... " (Grade 5 Winter 2017, 5-5[16]) and "My parents on the other hand, they say if I got a *B*, 'great job, you'll do better next time, don't worry about it'" (Grade 7 Autumn 2016, 7-2). These types of comments reflected parents as allies in students' success at school: "…if parents want to talk to you about a grade that you got, they're trying to figure out what happened, they're trying to help you" (Grade 5 Winter 2017, 5-1).

Student comments about letter grades and their parents frequently described parental emotions around letter grades. Some of these occurrences included positive parent emotions, usually within the context of students receiving high letter grades. These occurrences were represented by comments about parents feeling "really happy" (Grade 5 Winter 2018, 5-K) and "excited" (Grade 7 Autumn 2016, 7-3). The vast majority of these comments, however, referenced parent emotions with letter grades as a negative reaction. Some students described this as parents "being on the offensive on grades" (Grade 8 Winter 2018, 8-L), and getting a low grade of a *C* on a report card would lead parents to getting "very upset" (Grade 5 Autumn 2016, 5-B). Another student commented, "…even if I get a *B* my parents won't be happy, they always want me to get an *A*" (Grade 6 Autumn 2016, 6-2). Students framed the high stakes of parents' emotional responses to letter grades with comments like, "*C* stands for you're dead if you come home with that" (Grade 6 Autumn 2017, 6-10).

[16] As discussed in chapter four, I anonymized student focus group comments during the transcription process by giving each student a reference code with the first number identifying their grade level and the second letter or number specific to the individual student. When citing student comments, I also have included the focus group in which the comment occurred.

One theme within student comments about parents' negative reactions to letter grades was that even when parents tried to minimize the focus on letter grades and emphasize the importance of students putting forth their best effort, what really mattered in the end was the letter grade:

> For me, my mom tells me that she doesn't care that much about the letter grade, she cares why I have that grade, or what I'm doing to have that grade. I'm not sure it's always true though. I still remember the first time I got a *B* or *C* she FREAKED out and started yelling, and then later she was like, 'oh no, I only care about why you have it' (Grade 8 Autumn 2016, 8-C).

> See in my house, the logic is if you try your best it doesn't matter what you get on a test. But if you got a *C*, *F*, or *D*, well you evidently didn't try too hard! So, that's kind of how that works (Grade 7 Autumn 2017, 7-12).

As students discussed their parents' influence, they often shared that parents' reactions to undesirable grades led to negative effects on students' social-emotional well-being. These comments are representative of such occurrences:

> If [the vignette character[17]] has a bad grade and his parents see it then his parents will get mad. That's why I get scared (Grade 8 Autumn 2016, 8-B).

> I'm nervous every time my report card comes out because if I get a bad one my parents will get mad at me (Grade 6 Autumn 2016, 6-1).

> …when you get a *C* or a *D* you feel terrible to yourself and really stressed out, and you look at it with your parents and you're scared of what they'll say about it (Grade 5 Winter 2018, 5-K).

It was not just lower letter grades of *C*, *D* and *F* that students associated with parental pressure and negative consequences for their well-being, however. For some students, any grade below an *A* led to pressure from parents with repercussions for students' well-being:

> I can get upset with myself cause my mom always gets mad at me if I don't get an *A* or above because she feels that I can do a lot better (Grade 5 Autumn 2016, 5-C).

[17] As discussed in chapter four, focus group participants responded directly to the vignettes in varying degrees. For example, this student began with a direct response to the vignette before then sharing from a first-person perspective.

> But sometimes if [your parents] tell you, "oh, you did awful, come on, you gotta get an *A*," sometimes that basically makes you worry too much about it and brings you back to where you get nervous about your grades (Grade 6 Winter 2017, 6-2).

There were a variety of ways that students reacted to the stress, anxiety and pressure from parent responses to letter grades. Some students shared that the stress and pressure they felt from parents to achieve the highest letter grades resulted in self-described states of amotivation:

> …the grown-ups say, "you need to do better." And the child thinks, "but this was my very best, how can I do any better?" And they might feel like they're not good enough, so they feel like they have to stop trying (Grade 5 Autumn 2016, 5-2).

> [With a low grade] you feel so terrible that you don't want to do anything, you don't want to do any more, and you start really disliking the class (Grade 5 Winter 2018, 5-K).

Other students shared that parental pressure to get desirable grades drove some of them to lies and dishonesty with their parents about the grades they received, as seen in this response, "…if I got a *B*, I would be tempted to tell my parents that I got an *A*- or an *A* or something" (Grade 5 Autumn 2016, 5-3), and this comment, "If I got a *C*, I'd just be stressed out about it the whole rest of the day, I'd just be like, 'How can I make sure that my parents don't find out about this?'" (Grade 7 Autumn 2017, 7-K). One student summarized this issue in sharing:

> I think that's a problem for a lot of people in the grade… if they get a *C* they completely freak out and they go into really dire things to not show their parents what grade they got (Grade 7 Autumn 2017, 7-L).

Students' concerns associated with parent reactions to grades led to a culture of stress and anxiety, something that was not always alleviated by receiving A's, as represented in the following response:

> If you've got parents that are really pushy and they want you to do really well, then you're going to feel more stressed, even if you know everything by heart. Even if you'll get an *A+* on it, you're going to feel stressed, just because, at school you'll still feel stressed but then at home as well. It's meant to be calm, but it will still be really stressful (Grade 7 Winter 2017, 7-C).

6.3.2 Letter Grade Formulation

Within student comments about letter grade formulation, only a few identified that their grades were indicators of understanding about learning objectives. This sentiment was expressed in comments of grades showing "you've got the main logistics of it" (Grade 8 Autumn 2017, 8-L), or the recognition that "some of your letter grade is about how well you're performing, not how much work you've put in" (Grade 6 Autumn 2017, 6-10).

Another smaller theme in students' perceptions of letter grade formulation was that letter grades were the result of teacher choice. Within this theme of *Grade Formulation-Teacher Choice*, students lacked control and influence in the assessment process leading up to the letter grade. A recurring aspect of grades being formulated by teacher choice was that students thought they were meeting learning targets, only to find out through evaluations with letter grades that there was a mismatch between their own concepts of quality for that learning target and that of their teacher. This sentiment was captured in comments such as these:

> … you might think you're doing really good in that topic, but after all, the teachers assign what your grade is, so they might think that you did not do as well as you thought (Grade 6 Winter 2017, 6-1).

> And I think the teacher has some say in what [the vignette character] gets. Cause she could do it by how hard she tries, but also, the teacher is like overseeing her. So if she thinks that she's doing a good job, like what her best might be, maybe it isn't what the teacher thinks is meeting 5th or middle school levels (Grade 5 Winter 2017, 5-B).

Other student comments coded under *Grade Formulation-Teacher Choice* more directly addressed the lack of agency and autonomy alluded to in the previous comments. This was represented in this 6th grade student's statement:

> At the end, it's not [the student's] grade to give, but the teacher's grade. And [the student] can try to change it by working hard and stuff, but he can't really switch the whole outcome his way" (Grade 6 Winter 2017, 6-2).

A 7th grade student explained this lack of autonomy with letter grades through the lens of authority and power:

> With grades in general and the school system, I think it's widely revolved around authority. And I think that authority and power isn't spread enough that it reaches the students" (Grade 7 Winter 2017, 7-1).

These comments were indicative of some students' perception that they lacked the power and ability to shape and influence the formulation of the grades that they received.

Above all else, students equated the formulation of their grades with the behaviours that went into them. In contrast to the *Grade Formulation-Teacher Choice* theme in which student comments revealed a lack of autonomy and control, *Grade Formulation-Behaviour* comments reflected that students felt they could control the grade they received through hard work:

> Well, if in a way [the vignette character] does control what turns up on his report card because he has to be able to control himself in class. He has to accept goals and he has to want to do well to get a good grade, so he has to work hard (Grade 6 Winter 2017, 6-B).

> I think that [the vignette character's] own work habits and how hard he tries will determine what his grades are. He can't beg his teacher to change his grades. He has to work hard to earn grades that he thinks are good for him (Grade 8 Winter 2017, 8-D).

Within these types of responses, getting an undesirable grade was because "You didn't do enough, or you didn't put enough effort into this…" (Grade 8 Fall 2016, 8-C). Similarly, other students described grades of *B* and *C* resulting from:

"You didn't work hard" (Grade 5 Autumn 2017, 5-13).
"You worked hard but didn't focus" (Grade 5 Autumn 2017, 5-11).
"You could've worked harder" (Grade 5 Autumn 2017, 5-12).

On the other hand, many students described the positive behaviours that resulted in desirable grades. To get an *A*, a student "...has to behave well" (Grade 6 Winter 2017, 6-C). Further, getting an *A* "means that you were working hard…" (Grade 6 Autumn 2016, 6-B) and "…you should be very proud of yourself because you tried the hardest you could" (Grade 5 Autumn 2016, 5-1).

6.3.3 Feedback from Letter Grades: Behaviour and Opaqueness

Just as many student comments associated the letter grades they received with the behaviours that went into them, focus group responses suggested that the feedback students constructed from letter grades was related to behaviours they should exhibit in the future. Occasionally, some comments linked this behavioural feedback to a positive letter grade which was "… telling you keep on working hard on this particular thing, that you're doing good, and just keep it up" (Grade 6 Autumn 2016, 6-1). Most frequently, behavioural feedback was associated with undesirable grades, often described as anything less than an *A*. Any grade lower than an *A* led one student to "…understand that next time I need to try even harder" (Grade 7 Autumn 2017, 7-K). Similarly, another student commented that a vignette character who received less than an *A* "…might be thinking, 'I worked really hard, but maybe I should try harder next time'" (Grade 8 Autumn 2017, 8-L) which echoed another student for whom a grade of *B* "… would probably mean that you're okay but you need to try harder" (Grade 8 Autumn 2016, 8-C).

Paradoxically, the letter grade feedback of trying harder sometimes had a deleterious effect on student motivation and was recognized as flawed:

> I don't really think that you should be compared to a letter. It's putting you in a certain box. Like, let's say you got a *B*-, and you tried your hardest, but you only got a *B*-. So then you just stop trying because you know that you're always going to get a *B*- (Grade 6 Autumn 2017, 6-K).

> Well, if [the vignette character gets] really, really discouraged and they don't think that they're going to get a good grade then they might put less effort into it because they don't think that they're going to get a good grade anyways. So then they put even less effort into it (Grade 5 Winter 2017, 5-1).

While some student comments revealed a theme of letter grade feedback related to behaviour, other comments described letter grade feedback with a degree of opaqueness. For a student who wanted to understand which specific learning areas led to the grade they received and which learning gaps to work on going forward and how to work on them, letter grades offered little help. Within

this theme, some student comments reflected that letter grades did not inform them of their standing with larger learning criteria:

> Yeah, with letter grades they don't really give you any description… and you don't know what you did well on and what you need to work on" (Grade 5 Autumn 2017, 5-M).

> [A letter grade] doesn't help you out, it just tells you what you got. And the grades don't tell you what you need to improve on and what you're good at. They're saying that if you're past the levels or below the levels or meeting the levels… they're almost a bit lazy in some cases to not really describe what's going on. (Grade 5 Winter 2018, 5-10).

Other student comments went one step further to articulate that in addition to lacking feedback about what students were doing well in their learning and where they had specific learning gaps, letter grades also did not help with identifying steps to take going forward to close those gaps:

> I don't really like the overall grade system…They're like, "In humanities...in social studies, you got an *A*." Right, but where's my indicator of what I need to do to get better, and what's my indicator of what I did well, and what's my indicator of what I did bad? (Grade 6 Autumn 2017, 6-10)

> I think that [the letter grade] wouldn't tell what [the vignette character] did right and what he did wrong. It just says, overall what happened. So I don't think that he would know what to fix for the future (Grade 8 Autumn 2017, 8-K).

6.3.4 Grades as Labels

Many students talked about letter grades as labels with a degree of permanency. The permanency of these labels had the potential to put students in "a certain box" (Grade 6 Autumn 2017, 6-K). The box metaphor is a telling representation of the permanency that comes from not knowing how to progress in one's learning, particularly if the feedback the letter grade gives you is to work harder, you have already put forth your best effort, and the main thing you associate with the formulation of the letter grade is also behaviour. The degree to which students sensed this permanency and lack of accessible progression between letter grades was articulated by many focus group students in many forms, but was best represented in this comment:

> I kind of feel like when I get a letter grade, somebody sort of just like, there's like a big stone wall and somebody's chipped in my letter grade, and it goes up with all my other letter grades and sort of just...they're there, but not really doing anything, but like everyone, the teach…it's in full view... (Autumn 2017 Grade 7, 7-10).

One aspect of the permanency of these labels was student intelligence. As a student shared, "I feel like grades and report cards they show who you are, like they can say, 'oh she's smart…'" (Grade 8 Autumn 2016, 8-B). Similarly, another student commented,

> For me, you're supposed to think about [letter grades] as an update in how you're doing in your classes, but then you really think about your intelligence and how, it's kind of like you're setting these letter grades as part of your identity (Grade 7 Winter 2017, 7-1).

Other students recognized that there was supposed to be a formative aspect to their grades and report cards, but that they were unable to access that feedback and instead saw them as permanent markers of ability, as represented by this comment:

> I guess report cards are meant to be there for guidance and show you your progress and how you're doing and how you need to improve for the next semester or next year. But I feel like everyone kind of thinks a bit like, "oh this is all I can do. This is what I usually get in this class, that's it. I can't get higher than that. I'm not good at this class" (Grade 8 Winter 2017, 8-A).

Another characteristic of letter grade labels was indicating future success in life. This attitude was represented in this comment:

> …if you're a straight *A* student then you're a smart student and you're going to do great things, and you're going to have a good future. If you're a straight *C* student, then you're average, you haven't got a bright future, which I think is all the labels (Grade 7 Winter 2017, 7-A).

Even for the youngest students from the Grade 5 and 6 focus groups, there was an awareness of letter grades serving as gatekeepers for acceptance into elite universities:

> If you want to get into a good university...The best university on earth, it's not like it can keep seven billion people in one university. If you're in the best university, they kind of want you to earn your way in, not just say, "hey look, I can pay, just let me in." It doesn't really work like that -they mostly see your grades… (Grade 5 Winter 2017, 5-4).

> …you want to get good grades when you get closer to college because you want to get into a good college. So, every time you're getting closer and closer, so you're really scared somehow because you want good grades for college (Grade 6 Autumn 2016, 6-1).

Beyond just acceptance into elite universities, the slippery slope of letter grades and the future included wealth and success in life:

> I think that secretly, or even subconsciously, everyone wants to get an *A*+, be successful, and have as much money as possible, because I think that's sort of what school was about. It's about becoming successful, and then from that you can get money (Grade 7 Autumn 2016, 7-A).

Often, students clearly linked this slippery slope with the social influence of their parents:

> I have a friend and her family is like, very big on all these grades. They want her to get to Stanford, Harvard, like all those huge fat colleges. And if she gets a bad grade, she got a *B*- on something, I don't remember what, and her parents got so mad at her. They were like, 'the colleges are going to look back, they're going to see that *B*-, and that's going to be a minus for the college, like a negative. So you won't get into a good job, you won't have such a good pay' (Grade 5 Winter 2017, 5-B).

> [Grades] put you under so much pressure, and also like your parents say, "when your report card comes out," they're like, "this means everything, you know. You won't get into a good college if you get bad grades." Some parents say that. And then if you do fail, you just feel very bad and you feel like you're not really going to be successful… (Grade 7 Winter 2017, 7-4).

Similar to this last quote, other student comments articulated that the permanency and loaded meaning of letter grade labels had deep implications for students' social-emotional well-being. Across the Year One focus groups, students spoke of how their constructed meanings of letter grades were sources of stress, anxiety, and self-doubt:

> Everyone just thinks that reports just matter in how smart you are, and it kind of changes your perspective in how you feel about yourself (Grade 7 Winter 2017, 7-C).

> You're supposed to think like, it's supposed to like help you with what you need to improve, but sometimes, you kind of think if you got a *B* or an *A* or a *C*, you'd kind of think, that's not a good grade. So, it's meant to help you, but then sometimes it can just make you upset, and like sad too (Grade 5 Winter 2017, 5-A).

So much was riding on letter grades that when asked what students were supposed to think about report card letter grades and what they actually think about them, a student replied:

> I think that you're really supposed to see it as a way to get better. These are all things that can help you improve and succeed in your class. For me personally, I see it as the holy grail of life or death (Grade 6 Winter 2017, 6-5).

6.3.5 Norm-Referencing

Another theme which arose in student discussions about letter grades was identifying which letter grade was the "average." Many students felt that the average grade at SIS middle school was a *B*, as represented by this comment:

> But just a *B*, it doesn't tell you [about next steps to take], it just says that you're average (Grade 5 Autumn 2017, 5-11).

Not only was a *B* identified as average, student comments also revealed that average and below average were undesirable places to be and they strove to be above average:

> So, I think an average score is like a *B*, *B*-...like, *B*, *B*-, *B*+. I think an *A* is a really good score, and, I think it's a good goal to have, to get an *A* in your class because that means your over average and doing well (Grade 6 Autumn 2016, 6-2).

> But *C*, it's like you're under average so it just like disappoints you when you get a *C*, because it's not as good as a *B* or an *A* (Grade 6 Autumn 2016, 6-1).

While most students felt that a *B* was average, confusion crept in when trying to reconcile that average with the historical norm-referenced definition of *C* as average (as discussed in chapter three). This theme was represented when the Grade 8 Autumn 2016 focus group discussed how they thought of the letter grade, *C*:

-It's supposed to be average, but it's more like below average (Grade 8 Autumn 2016, 8-3).
-They say it's average, but really, it's not (Grade 8 Autumn 2016, 8-1).
-It's not the average for [our school], it's the average for across schools in America (Grade 8 Autumn 2016, 8-4).
-I feel like *B's* and *B+'s* are kind of the average (Grade 8 Autumn 2016, 8-1).
-It's not an average to my parents. When my mom found out that I had a *C* in Algebra...and I said, "it's average", and she said, "it doesn't matter, it's a *C*" (Grade 8 Autumn 2016, 8-D).

This discrepancy between the historical definitions of letter grades and what students perceived as the average grade at SIS resulted in paradoxical comments, such as, "Yeah, pretty much everyone in this school is above average" (Grade 7 Winter 2018, 7-L).

Some students struggled to reconcile the alternative historical definition of a *C* as average with the *B* average they had constructed from experience. When citing the historical letter grade definitions of the average grade in the examples below, students referenced the school's student handbook, given to every student at the start of the school year:

> Actually, isn't *C* in the handbook, isn't it that *C* is the average grade and *B* is above average, and *A* is like, really good? (Grade 7 Autumn 2017, 7-K)

> But when you look at the handbook, it says that *B* range is above average and the *C's* are average. (chorus of "yeah's") So technically you're above average [if you get a *B*] (Grade 7 Winter 2018, 7-11).

Interestingly, the handbook did not actually include these norm-referenced definitions for any of the years during my data collection.

Regardless of which grade represented average, norm-referenced definitions were pervasive in student comments about letter grades and reflected a culture of using norm-referencing to identify achievement. The act of comparing students' performance against each other was also seen in comments coded under *Competition*. Competition through norm-referencing student achievement was accepted as immutable by some students, as represented by these comments:

> I think there's competition between honestly, everyone. It's something that everyone does. It's a feeling that someone always wants to be better than someone else. So, people always want to do better than them, so lying may come into consideration just to get first…It's just a competition that goes through everyone's mind, and it's not something that you can really stop, but you can hold it off a little bit (Grade 8 Autumn 2016, 8-2).
>
> At the end of the day, grades are grades and whether they're there or not, people are always gonna find different ways to compare intelligence. It's just how it is and how grades have been popularized in media… (Grade 8 Winter 2017, 8-A).

Other students echoed this idea that norm-referenced competition was embedded within broader society:

> I think it's all about really just the society, and the fact that there are so many things compared to it. Like, you can get Grade *A* meat, and you can get Grade *B* meat. I think the entire system is based on who wants to be at the top rather than how we work as a group (Grade 7 Autumn 2016, 7-A).

Some student comments reflected that norm-referenced competition negatively impacted students' social emotional state:

> Cause everything we compare somebody with somebody else. If they have a higher grade then it makes you feel like, "oh my gosh, I'm way less superior than this guy, he's much smarter than me." And it just makes you, without realizing, it just makes you feel really sad. It makes you feel horrible (Grade 5 Winter 2017, 5-4).
>
> [The vignette character] may not want, if she got a bad score, she may feel bad, because other kids may have gotten better scores than her, and she may not want to be at the bottom (Grade 5 Autumn 2017, 5-J).

6.4 Analytic Generalizations in Answering Research Question 1

I will now situate my analysis of focus group letter grade findings within the framework of Wertsch's sociocultural extension of Vygotsky, assessment theory, and the relevant grading research discussed in chapter three. Yin (2013) calls for analytical generalizations to extract empirical findings from case studies to broader overlaps and gaps within theory and research (see chapter four). As I will discuss, the meanings students constructed from letter grades in the focus groups of my study largely overlapped with the existing FA and letter grade literature. This student perspective on letter grades is under-researched and

helps to fill a gap within the research literature, a point I will discuss further in my conclusion chapter.

6.4.1 Letter Grade Report Cards as Traditional Summative Measures Lacking in Formative Function

A summative function of assessment audits student learning and represents an end point in a progression of study (Andrade, 2010; Earl, 2003; Stiggins, 2010; Torrance & Pryor, 2001), and letter grades and report cards have long been the ultimate form of SA in the U.S. (Wiggins, 1993). Focus group findings support this broader research view of letter grades as strictly summative measures. Students consistently described letter grades as labels with a degree of permanency, so permanent in fact that they were "etched in stone". The permanency of letter grades as fixed labels of ability was consistent with a summative use of assessment data where the progression of learning has ended. This finding overlaps with previous qualitative findings that letter grades function summatively as labels of students' intelligence and identity (Thomas & Oldfather, 1997).

Assessment systems should allow for assessment data to be used for both SA and FA (Biggs, 1998) and even SA should have a secondary function to support learning (Bennett, 2011). As traditionally summative assessment mediational means have the ability to function formatively (Black et al., 2003; Brookhart, 2001), I will now consider the degree to which students comments revealed a formative function within traditionally summative letter grades.

FA is represented by a learning gap metaphor and requires an understanding of learning criteria, awareness about current standing of learning in relation to the criteria, and appropriate steps to close the gap between current standing and the criteria (Black & Wiliam, 1998; Hattie & Timperley, 2007; Sadler, 1989). Focus group responses expressed that letter grades lacked major aspects of the FA learning gap with a letter grade's inability to communicate

what they were doing well, what they were doing poorly, and what they needed to work on going forward towards criteria. These student frustrations with the lack of accessible formative feedback in letter grades are consistent with grade literature which has hypothesized that the opaque ways letter grades are constructed through averaging points accumulated in weighted categories and the inclusion of behaviour measures obscures meaningful feedback to students (Guskey & Bailey, 2001; Hooper & Cowell, 2014; O'Connor, 2011).

The specific FA process I have identified in this thesis is teachers transferring evaluative power to students through self-assessment with the aid of rubrics to identify current standing in relation to clear criteria and then taking action to close that gap (Andrade et al., 2010; Earl, 2003; Sadler, 1989). A key aspect of this self-assessment process is for students to "…hold a concept of quality roughly similar to that held by the teacher…[and] to monitor continuously the quality of what is being produced during the act of production itself…" (Sadler, 1989, p. 121). In this regard, focus group responses coded under *Grade Formulation-Teacher Choice* reflected a mismatch between student beliefs about quality and those of the teacher. This mismatch was indicative of students who had not developed the "guild knowledge" (Sadler, 1989) of teachers' expertise with evaluation. Some student comments went even further to identify that the school's grading processes kept the power of evaluation concentrated with the teacher and not shared with students. Thus, student perceptions of how letter grades were constructed revealed key aspects of FA processes to be missing, namely the transfer of autonomy and power with evaluation through the successful application of self-assessment (Black & Wiliam, 1998; Swaffield, 2011). The end result of this was that students were unable to produce their own feedback through self-assessment processes (Andrade, 2010a).

Another sign letter grades were devoid of a formative function was the almost unanimous degree to which students did not connect letter grades to rubrics, a key aspect of FA and self-assessment (Andrade et al., 2010; Panadero & Jonsson, 2013). These rubrics were in place within the school, however, as explained in the 2016 self-study (see chapter five). This disconnect between letter grades and the assignment rubrics that led to those grades is interesting because there were student comments which identified that letter grades and reports cards were *supposed* to offer formative feedback about their learning gaps. In place of this formative feedback, what students constructed instead were meanings about behaviour.

Many comments revealed that students saw grades as the result of the behaviours that went into learning, and they viewed letter grade feedback in behaviour terms, such as trying harder. Previous research has found that many teachers include behaviour measures when formulating letter grades (Allen, 2005; Baron, 2000; Brookhart, 1993, 2013c; Brookhart et al., 2016; Cox, 2011; Cross & Frary, 1999; McMillan, 2001; McMillan et al., 2002; McMillan & Turner, 2014; Pilcher, 1994; Sun & Cheng, 2014), particularly for the purpose of controlling student behaviour (Bonner & Chen, 2009; Brookhart, 1994; Frary et al., 1993; Guskey, 2009; Pilcher, 1994; Sun & Cheng, 2014). This practice runs contrary to the recommendations of some measurement experts (Ebel & Frisbie, 1991; Oosterhof, 2001) and obscures the degree to which grades can communicate achievement information (Fisher et al., 2011; O'Connor, 1995; Reeves et al., 2017). The findings from my study suggest that teachers' practices of including behavioural measures in grades to control student behaviour are quite successful. Students in this study perceived behaviour as the essential component behind their letter grades to the exclusion of formative meanings related to their learning gaps, seen by their inability to link letter grades to the rubrics. Behaviour is an important part of learning and something that students

should get feedback about, but it becomes problematic when it is ubiquitous in the meaning students construct from their grades. Student comments revealed behaviour feedback to be further problematic: the message to try harder does not help to close a learning gap if one is already trying hard.

6.4.2 The Sociocultural Influences which Continue to Shape Letter Grade Meaning

Wertsch's sociocultural lens further helps to understand why students in my study were unable to obtain formative feedback from letter grades, highlighted by their almost universal inability to connect a letter grade with the information in the corresponding rubrics. In chapter three, I examined the cultural, historical and institutional forces which shaped the formation of letter grades. Using Wertsch's framework revealed letter grades as mediational means used for identifying fixed conceptions of intelligence through norm-referenced comparison with peers which sorted students and served as a gatekeeper to higher levels of schooling (Brookhart, 2004; Earl, 2003; Schneider & Hutt, 2013). Student focus group comments revealed a mediated action consistent with these historical and institutional forces which shaped letter grades. Echoing 20th century beliefs in fixed intelligence, focus group comments described letter grades as labels of intelligence with a degree of permanency and deeply intertwined with students' identity. Other student comments mirrored the gatekeeping and sorting function of letter grades in viewing them as keys to unlocking a successful future by getting into a good college which would then lead to a good job. Finally, although SIS had no formal policy of norm-referencing, focus group comments continuously revealed that students thought of their letter grades in norm-referenced ways in accordance with the historical roots of these mediational means. This was reflected in comments which referenced letter grades definitions as "average", "above average", and "below average", and went as far as stating that these definitions were listed in the

school's student handbook. The fact that these definitions had never been in the student handbook during the years of data collection reveals the historical meaning embedded within them. In this norm-referenced view, student success was defined through competition and receiving better grades than one's peers, a dynamic which further removes student focus from performance towards academic criteria (Guskey, 2015; Reeves, 2007).

These hidden historical influences embedded within letter grades and the meaning students constructed of them represent an example of what Wertsch (1998) termed the illusion of perspective. Humans are often unaware of the constraints mediational means place on our thinking and it usually is not until the means has been replaced that we become aware of these constraints (Wertsch, 1998). In this case, letter grades represent tools of a bygone era infused with beliefs from the past about fixed labels of ability representing normed standing amongst peers and singular in their role as gatekeepers to higher levels of schooling. It is likely that these embedded historical meanings in letter grades played a further role in preventing students from connecting their letter grade standing with the formative feedback potential of the rubric. Students were largely unaware of these constraints, represented by how they falsely assumed norm-referenced meanings had come from the school's student handbook. There were a minority of students who had broken through this illusion of perspective and identified the lack of FA in letter grades. Interestingly, even these students who recognized this constraint of letter grades were almost entirely unable to then connect letter grades with the corresponding rubrics and thus remained constrained by these mediational means.

An important aspect of understanding the meanings students constructed from letter grades within my study is their perceived parental influence. Vygotsky's general genetic law stated that all higher mental functions occur first on the social plane before being internalized in the individual (Vygotsky, 1986).

Because parents are a primary social force in student lives, it follows that how parents construct meaning around grades will influence how their children construct meaning, and focus group discussion confirmed this. Student comments consistently described parents stressing a summative function of letter grades, emphasizing the need to get A's as the key to unlocking a successful future, often in high-stakes contexts with the well-being of both parent and child hanging in the balance. These same views were also consistently expressed by students as their own perspective. Thus, parental influence on the meaning students constructed from letter grades revealed a reciprocal force of development, one in which parents perpetuated the same historical and institutional influences embedded in letter grades onto their students which had likely been forced on to them.

Wertsch avoided a deterministic stance by using mediated action as a unit of analysis and accounting for the unique way we as humans hold agency to shape our environments and the very mediational means which shape us (Wertsch & Rupert, 1993). In this regard, the confined scope within which students felt they had agency in shaping the letter grades they received is noteworthy. Most students shared that they could influence their grades through effort, but they also recognized that this tactic was a dead-end if they were already trying hard. Students' perceived lack of agency extended to how the grade was formulated, with mismatches between student and teacher evaluations. Using Wertsch's (1991) tool kit metaphor which described the unique tool kit each individual brings to interaction with mediational means, focus group comments were almost universal in revealing that students were missing the tools from their kit to be able to hold agency in unlocking the formative function of letter grades through connecting them to rubrics.

A key tenet of Wertsch's work was the need to go beyond studying the production of means to also study their consumption (Wertsch, 1998). Because

mediational means are embedded with cultural, historical and institutional forces, he believed this was necessary because the intentions of means do not equate to the mediated action they produce. In this sense, the mediated action of students in this study represents a curious finding. My examination of the cultural, historical and institutional forces revealed letter grades to be a means used to identify students' fixed intelligence and then to serve as gatekeepers to higher levels of schooling. Further, teachers have historically used these means as a way to control student behaviour. In this sense, the findings from my study reveal that letter grades produced exactly the mediated action from students they were intended to: fixed labels of intelligence identified through norming with peers and a hyper focus on behaviour. If on the other hand, one believes that all students can learn and it is the role of assessment and reporting processes to aid in their development rather than simply sorting and controlling them (Guskey, 2015), then these mediational means produced concerning and unintended results.

6.5 Conclusion

I began this chapter by reviewing the iterative process I used in establishing a coding framework to analyse focus group results. These results were then analysed to answer my research question, *What meanings do students construct from letter grades and report cards?* This analysis revealed that students' constructed meanings were intertwined with the influences of their parents and that they saw letter grades as behavioural measures with little formative feedback and indicative of future success and norm-referenced standing amongst peers. Next, I situated this analysis within the broader theoretical framework of this study to inform the analytical generalizations of my findings: letter grades function summatively with little formative feedback. Further, the historical forces embedded in these mediational means work to limit their formative function and are reinforced by parents.

These letter grade findings overlap with a key rationale for standards-reform at SIS that senior leadership shared in chapter five: students' hyper letter grade focus removed their attention from learning criteria. In the next chapter, I will detail student perceptions of SBGs and SBRCs at SIS through analysis and discussion of focus group results related to my research question, *What meanings do students construct from standards-based grades and report cards?* Doing so will reveal the possibility that SBGs and SBRCs can function formatively for students, while also showing that they have the potential to create an additional layer of opaqueness for students about their current standing in classes.

Chapter Seven: Analysis and Discussion of Research Question 2: *What meanings do students construct from standards-based grades and report cards?*

7.1 Introduction

In the previous chapter I detailed the iterative process of establishing and applying a coding framework to my focus group data to answer the research question, *What meanings do students construct from letter grades and report cards?* Using this same coding framework and focus group data, within this chapter I will answer the research question, *What meanings do students construct from standards-based grades and report cards*? Whereas letter grades had been perceived by students with a high degree of uniformity, my analysis of focus group data will reveal contrasting themes in the meanings students constructed from SBGs and SBRCs. One perspective students articulated was a link between SBGs, SBRCs and rubrics, which allowed students access to formative feedback about their learning gaps. Relatedly, some students felt that their parents were also aware of this formative function in these new mediational means. Some students shared that with this formative function they benefitted from improved well-being when compared to their mediated action from letter grades. Another perspective shared by focus group participants was that standards-based grades represented a frustrating and opaque extra step in identifying their letter grade. Similarly, some students shared that their parents did not understand the new grading system and only wanted to know their child's letter grade. Following this analysis, I will then discuss other key standards-based themes which arose in student comments related to norm-referencing, society & media, and teachers' implementation of standards-based reform.

In the later part of this chapter, I will take time to situate these key themes within formative assessment theory, the work of Vygotsky and Wertsch, and the relevant standards-based grades and report card research literature. I will argue that SBGs and SBRCs can function formatively in ways traditional letter grades

and report cards cannot. Schools hoping to maximize the formative function of SBGs and SBRCs should avoid using a hybrid SBRC and take efforts to provide continuing teacher and parent education.

7.2 Analysis of Research Question 2

Using the coded focus group data that I described in the opening sections of chapter six, I will now analyse focus group data related to the second research question of my study.

7.2.1 Formative Feedback from Standards-Based Grades

In the previous chapter I analysed and discussed the way students largely saw letter grades as the product of behavioural measures with similar behavioural themes in feedback about future learning. This behavioural theme appeared minimally in comments about the formulation of SBGs, represented by observations that a grade of *Meeting* meant that "you worked hard and focused" (Grade 5 Autumn 2017, 5-11), while a *Progressing* resulted from a student "…not listening, or talking too much" (Grade 5 Winter 2018, 5-K). Conversely, while students rarely expressed that their letter grades were made up of learning constructs, they frequently associated learning with the formulation of SBGs. Students expressed that a *Beginning* meant that "You don't really understand the topic…(Grade 8 Autumn 2017, 8-L), "You're missing some big understanding" (Grade 8 Autumn 2017, 8-K) and that a student "…really needs a lot of help because they're probably not really understanding what they're trying to do" (Grade 5 Winter 2018, 5-K). Similarly, a *Progressing* grade meant that you "…haven't learned everything" (Grade 5 Autumn 2017, 5-10) and caused one student to "…get worried because I obviously didn't learn enough and I should be at *Meeting* but I'm not" (Grade 7 Autumn 2017, 7-K).

Beyond associating learning with the formulation of standards-based grades, many students also interpreted SBGs as feedback about their learning.

One aspect of this was the recognition that the names of SBGs were definitionally linked to their current standing in a learning progression. This sentiment was expressed in comments of "it's all in the name" (Grade 8 Winter 2018, 8-10), and

> The standards actually have the meaning. *Meeting* means meeting the expectations, *Extending* is doing above, *Beginning* is you're kind of beginning to get it, and *Progressing* is also you're progressing to do it. So basically, the standards are the definition (Grade 5 Winter 2018, 5-K).

Connected to these definitions of standards, many students expressed that they represented a year-long progression of learning. In essence, a standards-based grade and report card "shows how you've progressed" (Grade 5 Autumn 2017, 5-J). When asked about the grade of *Progressing*, a student commented, "… at the beginning of the year I'm okay with *Progressing*, like, I know I'll still have time to grow and get to *Meeting*, which is where I want to be" (Grade 8 Winter 2018, 8-11). Other students echoed that the end goal of this progression was a grade of *Meeting*, which "is where you want to be at the end of the unit, or the end of the curriculum" (Grade 7 Winter 2018, 7-K).

In the previous chapter I discussed the disconnect between letter grades and the corresponding rubrics, even though those rubrics were systemically in use at SIS. A key element of SBGs as formative feedback was that many students associated SBGs with assessment rubrics, often to such a degree that they amalgamated the two constructs. As one student stated, "The standards are the rubric" (Grade 7 Autumn 2017, 7-J). Students who made this connection commonly recognized that the rubrics held information to identify their current standing and also guide their next steps in learning. This link was represented by a student who shared, "One thing I find very useful about standards-based reporting is they give you the rubrics…and you can understand how you can achieve all of the standards" (Grade 8 Autumn 2017, 8-10). Another student commented that a vignette character looking at a rubric containing standards-

based grades "…knows exactly what he needs to do in the future and what he has already mastered" (Grade 8 Autumn 2017 8-M).

Often, students who recognized that SBGs and rubrics held valuable information about their current standing and next learning steps contrasted this dynamic with static letter grades, which they saw as missing this formative information:

> I kind of think that having *Extending*/*Progressing*/*Meeting* thing, I think it's actually really helpful when you have a rubric because it tells you what you need to do to achieve that goal, instead of just saying, "you need to get a 35% to get a D or something (agreement) (Grade 5 Winter 2017, 5-1).

> I don't think a letter grade means anything. When you see a letter, that doesn't necessarily represent something. When you have standards-based grades, it explains kind of what fits in this category [on the rubric] so that you know what's the difference between what I'm doing and what I can do to improve. And so I think that using a standards-based grading system is better overall (Grade 6 Winter 2017, 6-A).

As represented in the following quote, some student comments revealed the extent to which standards-based grades were associated with the rubric and letter grades were not:

> Well with standards-based, say you want to meet, then it says what you have to do exactly. Whereas with a letter grade, there's no description for each letter" (Grade 8 Autumn 2017, 8-K).

Some students who associated SBGs with the corresponding rubrics were able to reconcile the school's hybrid SBRC which combined letter grades and standards-based grades (see chapters three and five). They accepted that these were two different grading systems with SBGs "separated from letter grades – they're not the same thing" (Grade 6 Winter 2018, 6-11). Other students articulated this reconciliation further, best represented by this comment:

> The rubric is structured differently than just giving *A's*, *B's*, *C's*, *D's*...you can't compare the two together because you're more, you're specializing more, you're saying more with where [students] are on the rubric, whereas the *A* you're just giving a grade, so I don't think that you can compare the two, it's kind of like comparing a drink and food. You just can't compare the two because one's a solid, one's a liquid, it just doesn't make sense (Grade 5 Winter 2018, 5-10).

Within this new formative framework some students had constructed of SBGs linked to rubrics, SBGs were much "more precise and help you grow more" than their "vague" (Grade 7 Autumn 2017, 7-J) letter grade counterparts, which caused SBGs to be seen as fluid and malleable when compared to the permanent labels of letter grades. As a result, some students felt that receiving a standards-based grade "makes it less feel like you're boxed in" (Grade 6 Autumn 2017, 6-12) than receiving a letter grade.

Student comments revealed positive social-emotional benefits for those who were able to focus on the formative feedback of SBGs through rubrics. These students believed that feedback from standards came in a "positive light" (Grade 8 Winter 2018, 8-10), and that compared to receiving letter grades, it was not "so rough" (Grade 8 Winter 2017, 8-B). Another student commented that standards-based feedback through a rubric "boosts your confidence…[and would] encourage kids to keep going" (Grade 5 Winter 2018, 5-10) a lot more than letter grades. Whereas undesirable letter grades had been identified as high stakes labels of ability associated with negative repercussions for well-being, some students shared that "*Beginning* means you're starting to understand it, it doesn't mean that oh no you definitely failed this, you're the worst" (Grade 6 Winter 2018, 6-11) and "I've got things to learn, it's okay" (Grade 7 Autumn 2017, 7-12). Students comments extended this standards-based well-being upgrade to other grades: "…when you see *Meeting* you're still proud of yourself, whereas when you get a *B*, you feel like you're not as smart as when you see *Meeting*" (Grade 5 Autumn 2016, 5-C). Likewise, with *Progressing*:

> Well, if you compare a *C* to a *Progressing*, doesn't a *C* sound a lot worse than *Progressing*? (CHORUS OF YES'S) Like *Progressing* is like you're working on it. But if you think *C*, they put you on academic probation (Oh crepes!!! This isn't going to work out for me!). But really, you're just working on the subject and you're not that bad at it. But from the point of view of a letter grade it sounds a lot worse than it actually is (Grade 6 Autumn 2017, 6-12).

Another student elaborated on the relationship between well-being and the access to formative feedback - or lack thereof - found in the two different grading systems:

> …[Standards-based grades] would definitely help [the vignette character] understand and maybe get some more stress off of them, because they'll actually know exactly what to do instead of just like, "I know exactly what to do but I always get a mediocre grade and not a good one" (Grade 8 Winter 2017, 8-C).

7.2.2 Standards-Based Grade Opaqueness

In contrast to student perceptions of SBGs as formative feedback connected to rubrics, some students expressed meanings of SBGs deeply rooted in the previous letter grade system. Within this theme, standards-based grades were seen as imprecise and opaque in providing the information that really mattered: the student's letter grade standing. Some students felt that standards-based grades were "sort of a trick" (Grade 6 Winter 2017, 6-5) that "sugar coats" (Grade 6 Winter 2017, 6-2) feedback by "shoving the [letter] grade under the rug" (Grade 6 Winter 2017, 6-5).

In the previous section I discussed that some students had been able to reconcile the school's hybrid grading system and gain access to the formative potential of standards-based grades and rubrics. In contrast to this were expressions of frustration in trying to merge standards-based grades into the letter grade system. A large part of this frustration came from trying to identify how standards-based grades translated into letter grades. As one student shared, "We don't really know where is conforming to the letter grade, so I think that's all kind of confusing when we convert it" (Grade 8 Winter 2018, 8-10). Another student detailed this translation confusion further in commenting on a vignette character:

> Like if she got a *Progressing*, maybe she would be upset with that because she doesn't know what that means really. Like, you know it's in the *B* range, but it's anywhere from low to high. So it really doesn't give you that much information…(Grade 8 Autumn 2017, 8-J).

At the heart of the translation issue was that there were only four standards-based grades compared to 13 different letter grades (A+, A, A-, etc.) that corresponded to 100 points on the percentage scale. Because of this, there was a "wider range that you can get" (Grade 7 Autumn 2017, 7-10) with the letter grade system. This sentiment was present in many comments, such as this one:

> And the rubric only has four things: *Extending* which is really good, *Meeting* which is just meeting, *Progressing* is not good really, and *Beginning* is bad. But what if it's in between? What if it's just okay-kind of good? There's no thing for that, like that would be a *B* (Grade 5 Winter 2018, 5-12).

This type of thinking led students to feel like "it's easy to like get stuck in the web of *Progressing*" (Grade 8 Winter 2018, 8-J) and that "…with letter grades you can be a little more precise…" (Grade 7 Winter 2018, 7-12). In this sense, standards-based grades could take on the same static features that some students felt towards letter grades.

While some student comments had detailed frustration with converting imprecise standards-based grades into letter grades, other students offered suggestions of why they remained rooted in the letter grade system. One student shared, "letter grades are kind of something I grew up with and I kind of know them better, and [standards-based grading] was unfamiliar" (Grade 7 Winter 2018, 7-11). Within the hybrid system, some students found themselves resorting to the grading system with which they had the most experience and established frameworks of understanding. Another student articulated this danger of the hybrid model:

> I feel like we've all grown up with the letter grade, like it's basically ancient, so everybody knows what it means and they know what's good, and what's bad. So it's just easy to look at the letter grade without even glancing at the rubric, and then you're never going to get good at the skill (Grade 7 Winter 2018, 7-L).

This dependence on letter grades came to a logical conclusion with the end of term hybrid report card. One student had predicted that with the new standards-based grading and reporting the following year,

> people will still find a way to get that connection to the *A*, or *C*, or whatever. So I don't think it's going to make a difference whether they change it to *Meets* and *Progressing*" (Grade 8 Winter 2017, 8-A).

As predicted, the following year a student recalled looking at their SBRC and shared:

> Yeah, I think that, um, basically, I skimmed through all of the standards and looked at the letter and decided that it was my grade and that's what really mattered to me at the end of it (Grade 8 Winter 2018, 8-11).

Similarly, when asked what was going through the mind of a vignette character who was looking at an SBRC, a student remarked:

> She's not looking at the, definitely not looking at the standards. She's probably looking at the letter at the bottom. I know I personally did the same thing and I won't be able to tell you if I got, like, on this social studies standard whether I got a *Meeting*, *Exceeding*, etc. I'll tell you my grade in social studies though! (Grade 7 Winter 2018, 7-12)

7.2.3 Parent Influence

For some students, their interactions with parents about standards-based grades represented an upgrade from the stressful conversations about letter grades. This sentiment was represented by a student who shared that if parents look at a report card and "see all *Meets*, they'll congratulate you, but if they saw all *B's*, or all *C's*, it seems a lot worse than getting a *Meeting*" (Grade 8 Winter 2017, 8-3).

Some students felt that parents would recognize the formative feedback embedded within standards-based grades and that this would drive student-parent conversations. For example, a student shared that a vignette character's SBRC conversation with parents would be more formatively productive than previous letter grade discussions because

> when the parents look at the report card they wouldn't go straight to the letter as it has changed. But it would go straight to the, how she could again, improve, and where she's at there" (Grade 6 Winter 2018, 6-K).

Another student shared that

> …a letter grade doesn't tell you anything, but on the other hand, the other one [standards-based grade] tells you what you've done wrong and stuff like that. So if you talk to your parents about it, they might be like, "oh now you have to work on this", while like an A or a B, they'll be like, "okay, you got this" (Grade 5 Autumn 2017, 5-J).

Other students, represented by the following comment, elaborated further that a key aspect of these positive interactions was parents understanding the year-long learning progression of standards and taking this into account when reacting to standards-based grades:

> For my parents it depends when in the year it is. So if I'm at the beginning and I get a *Progressing* it's okay for there to be room for improvement. But if at the end of the year I'm still getting *Progressing's*, I haven't improved as much as I should have in the year (Grade 8 Winter 2018, 8-11).

Another student felt that students could explain this year-long standards-based learning progression to parents who did not understand, reflecting a renewed agency which had been absent in discussions of letter grades with parents:

> I think *Progressing*, *Meeting*, *Extending* have a softened blow than *C's*, *B's*, and *A's*, so I think it will be a much calmer conversation. The student can explain that if he got *Progressing* and the parents weren't very happy about it maybe because they don't know what it means, he could explain that it just means that he's working towards something and he doesn't know quite enough to actually be comfortable with meeting standard (Grade 8 Autumn 2017, 8-10).

Student-parent interactions within the framework of standards-based grades as formative feedback were marked by a focus on developing learning and an improved social-emotional state when compared to letter grades and other constructions of standards-based grades, which I will now discuss.

Similar to students who saw a standards-based grade as an added layer of opaqueness in determining their letter grade, some students perceived that parents just wanted to know their child's letter grade. These students felt that their parents were from a different generation during which "they had old-fashioned grades" (Grade 5 Winter 2018, 5-K) that they had become "used to"

(Grade 8 Winter 2018, 8-K). Other students expressed that part of the problem with standards-based grades was that "parents are not very educated on what everything means [and] just concentrate on the letter grade" (Grade 8 Winter 2018, 8-J). Some students felt that parents would want the standards-based grades translated into a letter when students brought home graded assignments:

> Older parents…might be like, "I just want to see how well you're doing. Are you doing an *A, B* or an *F*? I don't want to know about this *Meeting* or *Extending*, I just want to know if you're getting an *A*, *B*, *C* or an *F*?" (Grade 5 Winter 2018, 5-10)

Relatedly, when the hybrid standards-based report card came out, they felt that parents focused only on the letter grades:

> I think they should make the letter grades less obvious [on the report card]. Cause your parents don't care if you're *Extending, Meeting, Progressing, Beginning* - their eyes go straight to that letter grade (yeah) (uh huh). So you might get an *A*- or a *B*+, and they don't care about the explanation, they just say, "you did bad" or "you did good" or "you did average" or "you need to do better" (Grade 6 Winter 2018, 6-10).

While some student comments expressed that parents prioritized letter grades over standards-based grades, others described high-stakes interactions with parents who had adopted standards-based grades in name only. These parents set expectations of acceptable grades and emotional well-being was contingent upon receiving these grades. Students shared that parents were "disappointed in me if I get a *Progressing*" (Grade 8 Winter 2018, 8-K), and receiving a *Beginning* on assignments could lead a student to think, "oh no, let's hide it away from my parents" (Grade 6 Winter 2018, 6-13). These high-stakes student-parent SBG interactions and the corresponding negative impact on students' social-emotional state were exemplified in this comment:

> Like in my experience, when my friend got a *Beginning*, she started crying because she was like, "my parents are going to kill me!" And I think it's, [students] don't care as much as their parents do, so it's sort of becoming, "what are my parents going to think of this [letter] grade?" (Grade 7 Winter 2018, 7-L)

For some students, their interactions with parents around their standards-based grades functioned very similarly to how other students had described letter grade interactions.

7.2.4 Norm-Referencing

Students' concepts of norm-referencing within standards-based grades were varied to a much greater degree than the norm-referenced meaning they applied to letter grades. Some students spoke of a new conception of average that was devoid of norm-referenced meaning. They acknowledged that within a standards-based learning progression, the goal was for all students to meet standards. Within this framework, the average grade carried positive connotations through associations with students meeting their learning goals:

> Yeah, I feel like if you get a *Meeting*, I feel like it's what most people get, it's the average, so you feel like you're *Meeting* the standard and you're doing well. You're right where you should be. And, I feel like that's a good feeling, so if you get a meeting, that's sort of what you shoot for, and sometimes if you get a *Meeting* then you go for an *Extending* (Grade 7 Autumn 2017, 7-L).

This new construction of average was particularly clear in comments comparing the old letter grade system to the new standards-based system. In the following quote, the student recognized the moving targets of bell-curve norm-referencing embedded in the letter grade system and how standards-based grading would break free of those constraints:

> Although it will probably be easier for every child in the class to get a *Meeting*, because if everyone got the standard of an *A*, maybe [the teacher] would lower that standard so that only some people that did maybe slightly better than the people who just got a lower *A*…they would maybe get a *B*, so it would be harder for everyone to get an *A* than a *Meeting* (Grade 8 Autumn 2017, 8-10).

Other students identified other benefits of this new standards-based conception of average. One student spoke of how the elimination of a norm-referenced average with a standards-based grading system would have positive social-emotional benefits for students:

> I agree, it would also release some of the stress on you because with the standards-based system, with the standard and *Meeting* or *Progressing* - I feel like, there's, in that thing, there would never be an "average" grade on that (Grade 8 Winter 2017, 8-1).

Another student perceived that parents would grasp this new conception of average:

> I think [parents are] generally happy with *Meeting*, because that's where, that's the average where everyone's supposed to be. I don't think they really think too much about it. They're just like, "okay, you got a *Meeting*." There's still room for improvement, but overall that's a pretty good score (Grade 8 Winter 2018, 8-L).

Contrasting the view that *Meeting* was the average standards-based grade representing positive, secure standing, other students still used norm-referencing to understand standards-based grades. Students in the Grade 5 Winter 2018 group shared that *Extending* means "…you're better, you're exceeding both your teacher and what the other students are getting, so you're the smartest" (Grade 5 Winter 2018, 5-13), while another student shared that if a student got a *Progressing*,

> They think that mostly they could be sad and really angry with themselves cause they think that so many people are so much better than them and they wanted to have more (Grade 5 Winter 2018, 5-K).

Other students elaborated on this norm-referenced meaning of standards-based grades. One student shared that "most students' grades only count as opposed to others" (Grade 6 Winter 2018, 6-13). Another student continued:

> Yeah like, you base grades off of other people. Say someone gets a *Progressing* and you get a *Meeting*, you're like, "oh, I'm really good!" But if you get a *Meeting*, and they get a *Meeting*, it's like you're not as good anymore (Grade 6 Winter 2018, 6-12).

Some students also felt that their parents reinforced this norm-referenced conceptualization of standards-based grades:

> Well my parents are like 7-K's[18]: "Work harder. You are average right now. Below average! You must work harder. *Progressing* is not good, you have to be *Meeting*!" (Grade 7 Winter 2018, 7-12)

[18] In instances when participants referred to other participants by name, I inserted that person's coded pseudonym when transcribing.

7.2.5 Society and Media

Comments related to *Society and Media* occurred infrequently in student discussions about SBGs and SBRCs. Nonetheless, the contrast between letter grades and standards-based grades within this coding theme was particularly noteworthy for the sociocultural implications I will discuss later in this chapter. Student comments about letter grades were marked by an awareness that their understanding was shaped by the ways letter grade constructs were portrayed in broader society and the media, represented by these observations:

> If you see a movie and it's a movie that you really like, and this person says, "oh DANGIT, I got a B, or I got an F," and it's so bad and you're like, well if a B is bad to them then it's probably bad to everyone else, so I don't want to get a B (Grade 5 Autumn 2017, 5-L).

> …there is those stereotypes in the media, like nerd, or geek, or jock. And they kind of, once you see them, once you look closely, the nerds get *A's*, the geeks get decent grades, and jocks get *F's*. And that standard is kind of put out there and if you're this certain type of person - cause those characters kind of do live in our school, and I think those standards are put out there by the media (Grade 8 Winter 2017, 8-2).

While comments on letter grades acknowledged the influence of society and media in shaping meaning, the references to standards-based grades were notable for their observed lack of media influence and students' curiosity in how these mediational means would function in larger society. Students in the Grade 5 Autumn 2017 group recognized that standards-based grades did not have points of reference within media:

> In movies you see them going, "oh no, I got a *C*!" and that seems horrible. You say, "oh no, I got *Progressing*," that wouldn't seem so bad because the number grades are more old-fashioned and people use them and they seem more scary than the rubric and what it says (Grade 5 Autumn 2017, 5-M).

> Yeah, like in preschool...do you play enough? *Progressing*! (laughs) (Grade 5 Autumn 2017, 5-11)

> Yeah that's what I was saying. Like in movies or TV shows, they don't really have, "Ahhh man, I got *Progressing* on a test!", they'd say, "Ahhh man, I got a C." Like I said, they're more old-fashioned and they're more well-known (Grade 5 Autumn 2017, 5-M).

The lack of representation in media for standards-based grades left some students in the Grade 5 Winter 2017 focus group feeling let down with the new grading system:

> It's what you see in like, the movies! (Grade 5 Winter 2017, 5-5)
>
> *A+* on your paper, circled! (Grade 5 Winter 2017, 5-2)
>
> Yeah, and when you just see *Extending* and *Progressing* it's kind of boring (agreement). It's like black and white, and then colour came to town (Grade 5 Winter 2017, 5-C).

Within broader society, student responses about letter grades reflected a sense of letters as keys to unlocking future success. The function of standards-based grades as societal gate-keepers, on the other hand, was a point of curiosity and confusion in how they would function:

> The colleges are going to look at the letters usually because that's what college standards are usually. So if the colleges go and see, "oh, 6-1 got *Extending* in his English class", they're going to be, "oh, what's an *Extending*?" (Grade 6 Winter 2017, 6-4)

This confusion extended to thinking about applying to future jobs:

> You go to get a job and you're like, 'I got a *Meeting* in school" and they're like, "what the heck does that mean?" (Grade 7 Winter 2018, 7-12)

7.2.6 Student Perceptions of Teachers' Standards-Based Implementation

As SIS converted to the use of standards-based grading and reporting, students shared that there were differences in how teachers implemented the new grading system. Amongst these comments, there were two key variations students identified: some teachers used numbers with decimals to represent grades on rubrics; and some teachers distinguished between high and low levels for each respective standards-based grade on the rubric.

The first issue students identified with teacher implementation was that some teachers evaluated students' work on rubrics by translating the standards-based grade for each learning objective into an overall number on a four-point scale down with decimals listed to the tenths (see Figure 5.7, chapter five). Students then converted these numbers to letter grades through the conversion

scale that was in the back of the school-provided homework diary (see Figure 5.4, chapter five). The conversion scale was intended for end of semester grade conversions for students to understand how teachers converted multiple reporting standards to a letter grade for the hybrid SBRC[19]. Students' perceptions of this process were detailed in the following comments:

> On my [subject omitted] test, [the teacher] writes down a decimal, and we can convert them into letter grades…on the little charts in our homework diaries (Grade 6 Winter 2018, 6-L).

> I also got to talk about the *Extending/Progressing* thing, I like that, but that's not how they grade you. They show you that, but then they give you a bunch of numbers that you're supposed to decipher. To be completely honest, I don't think most people go by the numbers, I think a kid goes home, their parents see the numbers, and they just whip out their homework diaries and they translate it to a letter (Grade 6 Autumn 2017, 6-10).

As the last quote suggests, for some students and parents who were rooted in a traditional grading system with letter grades linked to a 100 point scale, the use of these numbers on the rubrics caused them to "pay more attention to the numbers" (Grade 8 Autumn 2017, 8-K) and left them feeling that teachers had "…just taken a number and put it into a smaller number" (Grade 8 Autumn 2017, 8-J), leaving them to "…figure out how to translate it back into letter grades"(Grade 8 Winter 2018, 8-12). For some, the end result of this focus on the numbers instead of the standards-based grade and corresponding formative information in the rubric was that "…if you get a decimal and a really specific number you can't really improve as much" (Grade 6 Autumn 2017, 6-K). Teachers use of numbers on a four-point scale with decimals worked to transform informational year-long progression feedback into a fixed label, quite similar to the meaning many students had constructed of letter grades.

A second issue students identified within teachers' SBG implementation was that on rubrics some teachers distinguished between whether students were at a high level for a particular standards-based grade, or a low level (see Figure

[19] Refer to chapter five for a full articulation of this process.

5.8, chapter five). One student remarked, "in [Teacher X's] class, she sometimes circles different areas of the box, so then I sort of think differently about - that's kind of different" (Grade 7 Autumn 2017, 7-10). With this practice, students expressed confusion over teacher inconsistencies from class to class and their exact numeric and corresponding letter grade standing with comments like the following:

> So they'll circle a certain area on the rubric and you won't know, like, in English you could have gotten a high *Meeting*, and then in social studies, you're like, did I get a high *Meeting*, a middle *Meeting*, or a low *Meeting*? And that causes confusion without the numbers or the grades (Grade 7 Autumn 2017, 7-L).

Ultimately, students wanted all of their teachers to "be more consistent…[and] on the same page" (Grade 8 Winter 2018, 8-K) in how they implemented the new standards-based grading system "…so that kids can really understand what they're getting and not have to, like, know all these different things and convert it" (Grade 7 Autumn 2017, 7-J).

7.3 Analytic Generalizations in Answering Research Question 2

In comparison to the more uniform way in which students conceptualized letter grades, the meanings students constructed of standards-based grades and report cards were filled with sharp contrasts. I will now situate these contrasting findings within broader formative assessment and standards-based theory, the work of Vygotsky and Wertsch, and relevant research findings.

7.3.1 The Formative Function of SBGs and SBRCs

FA is centred around a gap metaphor with three key steps for students: clear learning criteria; understanding of current standing; and taking action to close learning gaps (Black & Wiliam, 1998; Hattie & Timperley, 2007; Sadler, 1989). Many student focus group comments about standards-based grades and report cards described all three of these key FA gap steps. Further, these descriptions revealed a relationship in which standards-based grades and the

corresponding rubrics were inseparably connected. Beyond simply making this connection, student comments reinforced previous research on self-assessment by describing a process of comparing current standing against criteria objectives on the rubric (Brown et al., 2015; Brown & Harris, 2013) and then using that data to inform their next action steps to close the gap (Andrade, 2010a; Brookhart & Chen, 2015; Panadero & Romero, 2014; Ross, 2006). This finding also overlaps with the findings of previous studies that students are able to self-assess with rubrics to generate formative feedback (Lipnevich et al., 2014) in ways which letter grades don't allow (Andrade & Du, 2005).

In becoming both producers and consumers of formative feedback (Andrade, 2010a), student descriptions of linking their standards-based grades and report cards to rubrics to generate formative feedback represented a heightened degree of autonomy compared to their behavioural constructions of letter grade feedback, a point I will discuss in the next chapter for its SDT implications. Further, these students demonstrated the "guild knowledge" (Sadler, 1989, p. 126) of evaluative processes traditionally held by teachers, an important power shift within FA reform (Black & Wiliam, 1998; Dann, 2014; Earl, 2003; Sadler, 1989).

Through their descriptions of using standards-based grades and report cards formatively, these students provided further evidence that traditionally summative mediational means can function formatively (Black, Harrison, Lee, Marshall, & Wiliam, 2004; Brookhart, 2001, 2010; Harlen, 2006). FA and SA mixing requires criterion referenced assignments aligned to both the preceding and proceeding content to be studied (Biggs, 1998; Brookhart, 2010), elements present in the standards reform and UbD curriculum design process at SIS (see chapter five). Similar to previous studies on mixing, these focus group students described learning as a continuous and fluid process where standards-based grades "…were temporary stops along a learning path" (Brookhart, 2001, p. 167).

This finding supports literature which has theorized SBGs and SBRCs to function both summatively and formatively in guiding student learning by providing formative feedback to the learner (Guskey, 1996b, 2015; Heflebower et al., 2014; Muñoz & Guskey, 2015; Reeves et al., 2017).

The production of standards-based grades and report cards for formative purposes is an example of how new mediational means are often created in response to the constraints of previous means (Wertsch, 1998). For some, these new standards-based mediational means addressed the FA deficit constraints of letter grades by contributing to a mediated action from students that was in-line with a formative conception of assessment. Mediational means are carriers of the forms of knowledge which shaped their construction (Wertsch, 1994, 1995) and distinct strands within standards-based reform intended for fixed and clear criteria to allow for formative feedback and the opportunity for all students to meet standards, a departure from grading's norm-referenced past (Bailey & McTighe, 1996; Guskey, 2015; McLaughlin & Shepard, 1995). In addition to some students' formative use of standards-based grades, these beliefs manifested in the new definition of average described by some focus group participants, one in which all students could achieve and meet standards without diminishing the accomplishments of peers. This represented a significant upgrade over the norm-referenced meanings students associated with letter grades and had corresponding benefits to students' well-being.

Another aspect of standards-based grades and report cards related to FA was that whereas students saw letter grades as formulated from behaviour and also took behavioural feedback from them, students rarely associated standards-based grades and report cards with behavioural measures. A key tenet of SBGs and SBRCs is separating behaviour and achievement measures and reporting on them separately, as the inclusion of behaviour measures obscures formative information (Guskey, 1996b; Guskey & Bailey, 2001; Guskey et al., 2011; Muñoz

& Guskey, 2015; O'Connor, 2011; Wormeli, 2006). These findings suggest that this separation is largely successful in removing behavioural meanings from standards-based grades.

Because mediated action is produced through the reciprocal influences of the individual and the social and cultural contexts of their environment (Vygotsky, 1986; Wertsch, 1998), findings from focus group data in relation to the code, *Society & Media* suggested another way that standards-based grades were freed from letter grade constraints and enabled to function formatively. Whereas media reinforced students' understanding of letter grades as high-stakes norm-referenced labels, in a small number of instances student comments revealed that the absence of standards-based grades in media and broader society freed them from the "cultural icon" (Marzano, 2006, p. 125) status of letter grades and presented a comparatively blank slate for re-defining standards-based grades as formative means. The degree to which this was true for all students with the school's hybrid SBRC is a question I will return to in the coming sections of this chapter.

I will close this section on a note of caution in addressing how the findings of my study fit within the broader research literature. As I have established in chapter two, self-assessment is a misnomer and the transfer of evaluative power in the classroom through student self-assessment with rubrics usually requires explicit instruction from teachers (Black et al., 2003; Cizek, 2010; Wylie & Lyon, 2015). Within self-assessment with rubrics, there are also nuanced and varied ways teachers instruct students on this process (Andrade et al., 2010). More broadly, teachers struggle to transfer evaluative power to students (Pryor & Crossouard, 2008; Torrance & Pryor, 2001) and to implement student self-assessment within their classrooms (Lysaght & O'Leary, 2013; Wylie & Lyon, 2015). How learners' backgrounds and identities impact their ability to self-assess is also an unknown and under-researched area of self-assessment

literature (Andrade & Brown, 2016; Lipnevich et al., 2014; Panadero et al., 2016). Thus, the self-assessment with rubric process of FA I have theorized within this study is an immensely complex process, one with many variables my study cannot account for. Many students within my study described a formative use of standards-based grades and report cards through their link to rubrics, a formative use that was almost entirely missing with letter grades. This suggests that standards-based grades and report cards can function formatively in ways traditional letter grades and report cards cannot. This analytic generalization is supported by the FA and standards-based literature I have discussed in the preceding paragraphs and represents a preliminary finding for the relatively young research field of standards-based grades and report cards.

7.3.2 Variations in Student Responses to SBGs and SBRCs

Consistent with Wertsch's (1998) theory that the production of new means can overcome constraints of previous means while also introducing unforeseen constraints of their own, some students only saw standards-based grades and report cards as an opaque extra step to get to their letter grades. The standards-based grade became a frustrating symbol, one that only obscured the letter grade information they wanted. Some students also discussed SBGs and SBRCs in norm-referenced ways, signifying the letter grade lens through which they viewed these new mediational means. For these students, letter grades were the engrained and dominant symbols from their past and seemed to block the formative function of standards-based grades. This is consistent with findings from this study (see previous chapter) and others which show that letter grades fail to provide formative feedback to students (Brookhart, 2013c; Thomas & Oldfather, 1997).

These differing student perceptions of standards-based grades and report cards as formative feedback and as an opaque extra step stand in stark contrast

to each other. It would have been surprising if there was uniformity in how students constructed meanings from these new means, however, and the comparative uniformity of students' constructed meaning of letter grades speaks to their standing as the "concrete gold standard" (SL-C[20]) and "cultural icons" (Marzano, 2006, p. 125) with deeply engrained meaning across our society. The unique way individuals construct meaning from the mediational means in our environment is exemplified by Vygotsky's concept of word sense. He described sense as the "sum of all psychological events aroused in our consciousness by the world. It's a dynamic, fluid, complex whole…" (Vygotsky, 1986, pp. 244–245). Wertsch, del Rio, and Alvarez (1995) elaborated further that mediational means are uniquely used by individual agents and referred to the misconception that tools shape human thinking uniformly as the myth of unidirectionality. While the means of our environment mediate human thinking, each individual uniquely brings her/his experiences to construct meaning with those mediational means. Those experiences, in turn, have been shaped by the social influences within each person's life. The unique experiences and social influences within each student's history contributed to the multifaceted ways they viewed the new standards-based means. I will now discuss two major social influences which shaped the meaning students constructed of standards-based grades and report cards: their parents and teachers.

Some students shared that their parents viewed the new standards-based grading system as formative feedback while others expressed that their parents viewed it to varying degrees through the lens of the letter grade system. Early SBG and SBRC research identified parent education as a key component of standards-based reform initiatives because many parents interpret standards-based grades using a letter grade framework (Guskey, 2004). Some teachers in the 2016 self-study along with some senior leadership members at SIS in 2019 felt

[20] SIS senior leadership member; see chapter five.

that parent education about SBGs and SBRCs remained a pressing need (see chapter five). Some focus group participants reinforced these findings in stating that their parents were uneducated about the new system and called for SIS to educate parents about the new system. Student experiences with parent interpretations of the new grading system likely shaped the meaning students themselves constructed, as the perceived parent divide in SBG and SBRC meaning roughly matched the same divide in student meaning.

Another major social influence in children's lives is their teachers. Participant comments revealed the perception of some that teachers were implementing standards-based grading inconsistently, with some using specific numbers with decimals to delineate standing on the rubric, and others distinguishing between high and low for each standards-based grade with where they circled on the rubric (see Figures 5.7 and 5.8). In interviews, senior leadership also identified these particular variations in teachers' implementation practices (see chapter five). Because individuals uniquely interact with mediational means, the consumption of means can result in a form of means production (Wertsch, 1998). The unique experiences teachers brought to their own consumption of standards-based grading resulted in the production of new forms of standards-based grading in their classes, according to senior leadership and student comments. These practices deviated from the school's intended implementation of standards-based grading, a discrepancy consistent with fidelity of implementation issues found in early research on standards-based reform (Brookhart, 2013a; McMunn et al., 2003; Welsh et al., 2013) and broader FA initiatives (Flórez Petour, 2015; Hayward, 2015; Livingston & Hutchinson, 2017). Similar to previous findings about SBG and FA reform, this finding also suggests that standards-based reform is a messy process that requires a great deal of time and teacher training (Knight & Cooper, 2019; Lysaght & O'Leary, 2013; Wylie & Lyon, 2015). These examples signified the varied ways teachers

implemented the new grading system and shaped SBGs and SBRCs to get them to conform to the previous grading system. The idiosyncratic ways teachers attempted to implement the new SBGs and SBRCs also speaks to the myth of unidirectionality. Just as students were bringing their own perspective to make meaning from SGBs and SBRCs, so too were teachers bringing unique experiences and backgrounds in trying to understand these new mediational means. Student perceptions of these varied implementation practices represent a powerful social influence which likely inhibited some students from utilizing SBGs and SBRCs formatively.

In addition to the influences of parents and teachers, a closer examination of the school's decision to use a hybrid SBRC helps to understand the varying meanings students constructed of standards-based grades and report cards. The historical, institutional and cultural forces which influence means production often make it difficult to organize new means in efficient ways. As a result, new means often unintentionally introduce new and unforeseen constraints (Wertsch et al., 1995; Wertsch & Rupert, 1993). In chapter five I discussed the parental and external factors SIS considered in deciding to use a hybrid SBRC. These factors were consistent with the findings of standards-based research that schools use hybrid SBRCs as a compromise to avoid backlash against standards-based reform (Marzano, 1998; Scriffiny, 2008). This feared backlash stems from the recognition that letter grades hold established meaning for the parent community who can be confused and frustrated by the shift to a criteria-based standards system (Bailey & McTighe, 1996; Guskey & Bailey, 2001; Guskey & Jung, 2006), and out of concerns for how a new standards-based system will be received by external institutions, such as high schools and universities (Boston, 2003; Guskey & Jung, 2006). These influences which shaped the report card's final form were a departure from SIS' original rationale of shifting to a standards-based system to support deep student understanding through a

formative construction of assessment and reporting (see chapter five). The variations in student responses about SBGs and SBRCs speak to these conflicting influences in the forces which shaped standards reform and the hybrid SBRC at SIS.

Of course, the mediated action of humans is not unidirectionally influenced by mediational means. Individuals interact uniquely with the means of their environment and have the ability to shape the means which in turn shape them (Daniels, 2001; Wertsch, 1998). Within this process, Wertsch's tool kit metaphor describes how individuals and groups have access to their own unique set of tools which results in differentiated mediated action (Wertsch, 1991). Some students had the tools to reconcile the two concurrent grading systems within the hybrid model and access formative feedback through the standards-based grade and rubric link. Other students, however, lacked this linking tool and standards-based grades were transmuted and foreign symbols they were forced to translate back to letters with imperfect and often frustrating results.

Wertsch's (1998) illusion of perspective describes a phenomenon in which individuals are unaware of the ways mediational means can constrain mediated action, often until a new means frees them of the previous constraints. Students who had the tools to link standards-based grades to rubrics and gain formative feedback often broke through this illusion and recognized the constraints and detriments of the letter grade system which they saw as devoid of this information. Other students who lacked these tools remained firmly within this illusion, wanting only their letter grade standing and largely unaware of the formative potential of the new system.

The variation in meaning students constructed from standards-based grades and report cards within my study overlaps with similar variations found in initial standards-based grading research, with teacher perceptions that some

students gained access to meaningful formative feedback and others were left frustrated and confused (Knight & Cooper, 2019; Scarlett, 2018). In line with sentiments expressed by focus group participants, Knight & Cooper (2019) reported that teachers believed hybrid report cards and uneducated parents contributed to some students' frustrations with the standards-based shift and their inability to access formative feedback from the SBRC. While there is no set form for SBRCs to take (Guskey et al., 2011), these findings suggest that schools hoping to increase formative feedback to students should avoid using a hybrid SBRC that includes letter grades.

7.4 Conclusion

This chapter analysed focus group responses and contextualized them within broader theory and research to answer the research question, *What meanings do students construct from standards-based grades and report cards*? Some student comments linked standards-based grades and report cards to rubrics to generate formative feedback, suggesting that these new mediational means can function formatively in ways that traditional letter grades cannot and supporting previous research which has established that traditionally summative means can function formatively. Other student comments, however, reflected a view of standards-based grades as an opaque extra step towards letter grade standing, disconnected from rubrics.

These diverse student interpretations are to be expected when considering Wertsch's myth of unidirectionality and they were further perpetuated by the influence of teachers, parents and the school's choice of using a hybrid SBRC. These findings suggest that schools should avoid a hybrid SBRC if they want to maximize the formative function of standards-based grades and report cards, and that continued teacher and parent education is an essential element of a standards-based reform initiative.

In the next chapter I will answer my third research question, *How does the standards-based grade and report card shift impact student motivation?* I will begin by analysing questionnaire data from my study and then re-examining focus group data from an SDT perspective. I will then discuss my meta-inference from triangulating this data within the context of SDT research.

Chapter Eight: Results, Analysis and Discussion of Research Question 3: *How does the standards-based grade and report card shift impact student motivation?*

8.1 Introduction

In this chapter I will use questionnaire, focus group and unstructured interview data to answer Research Question 3: *How does the standards-based grade and report card shift impact student motivation?* I will begin the chapter with an overview of results across both years of my questionnaire data collection, using SDT's Academic Self-Regulation Questionnaire (SRQ-A) (as detailed in chapter four). Following this, I will analyse questionnaire data to estimate the size of impact of the intervention in effect sizes with bootstrapped confidence intervals. This analysis will reveal that the shift to standards-based grades and report cards resulted in very little overall impact on student motivation, represented by very small effect sizes. Next, I will analyse focus group data to examine students who described a formative function in standards-based grades and report cards with their link to rubrics. This self-assessment process supported these students' autonomy and competence, and should have resulted in more autonomous forms of motivation. On the other hand, focus group analysis also showed that some students, perceiving standards-based grades as an opaque extra step to identifying their letter grade, felt a loss of autonomy as they no longer were able to identify their standing. The offsetting effects on motivation of these two prominent themes from focus groups likely resulted in the very small effect sizes found in the questionnaire analysis. I will close this chapter by considering temporal aspects of assessment reform through research and data from my study which suggest that standards-based reform needs to be given time to develop in order to effect change.

8.2 Questionnaire Results

I will now discuss the questionnaire results from my study, beginning with Year One of the data collection and then moving to Year Two. Within each year, I will begin with results from the pre-test and then review post-test results.

8.2.1 Year One Questionnaire Results

8.2.1.1 Year One Pre-Test Responses

Of the 468 students in the middle school at the start of the 2016-2017 school year, 453 successfully participated in the Year One Pre-test, representing a 97% response rate. Table 8.1 below breaks these figures down by grade level:

Table 8.1: Year One Pre-Test Response Rate by Grade Level

Grade Level	Number of Students in Grade	Number of Students Who Successfully Completed Survey	Response Rate
5	107	106	99%
6	117	114	97%
7	117	112	96%
8	127	121	95%

The 15 missing responses included seven questionnaires that I manually excluded, and four students who were absent or late in arriving to school and unable to complete the questionnaire, as detailed in the Tables 8.2 and 8.3 below.

Table 8.2: Year One Pre-Test Excluded Responses

Number of Responses	Reason for Exclusion
3	Duplicate PIN number. I included the response that aligned with the Grade Level and Gender of the assigned PIN, and excluded the duplicate that did not match.
2	Duplicate PIN number. Both responses matched the Grade Level and Gender of the assigned PIN, and thus both were excluded.
1	Invalid PIN number entered.
1	Response excluded due to computer glitch and scrambled responses.
4	Unknown

Table 8.3: Year One Pre-Test Late Arrivals and Absent Students

Late-arrivals	Absent	Late-arrival Successful Completion	Absent Successful Completion	Total Unable to Complete
4	10	3	7	4

8.2.1.2 Year One Post-Test Responses

Of the 468 students in the middle school at the start of the year, five students moved at the end of the first semester, leaving 463 possible students to take the Year One Post-test. Although new students arrived at the start of the second semester, they were excluded from the post-test because they were not present for the pre-test. Of these 463 students, 446 successfully completed the post-test resulting in a 96% response rate. Table 8.4 below breaks these figures down by grade level:

Table 8.4: Year One Post-Test Response Rate by Grade Level

Grade Level	Number of Students in Grade	Number of Students Who Successfully Completed Survey	Response Rate
5	107	101	94%
6	117	111	95%
7	113	110	97%
8	126	124	98%

The 17 missing responses included 7 questionnaires that I manually excluded and 7 students who were absent or late arrivals unable to make-up the questionnaire, as detailed in Tables 8.5 and 8.6 below.

Table 8.5: Year One Post-Test Excluded Responses

Number of Responses	Reason for Exclusion
3	Duplicate PIN number. I included the response that aligned with the Grade Level and Gender of the assigned PIN, and excluded the duplicate that did not match.
2	Invalid PIN number entered.
2	No PIN number or gender/grade information entered.
3	Unknown

Table 8.6: Year One Post-Test Late Arrivals and Absent Students

Late-arrivals	Absent	Late-arrival Successful Completion	Absent Successful Completion	Total Unable to Complete
5	9	3	4	7

Because Grade 5 students were taking the questionnaire on a Friday (see chapter four), two days after the other grades, I decided not to have absent and late-arrival Grade 5 students make up the questionnaire the following Monday because it was too many days removed from when the other grades took the post-test and when the report card was originally distributed.

8.2.1.3 Overall Year One Participation

There were 463 students who attended the middle school for the duration of Year One. Of these 463 students, there were 30 who did not successfully complete either or both the pre-test and the post-test and were removed prior to the analysis of results. The 433 students who successfully completed both pre and post-test represent a 94% response rate for Year One of the study.

8.2.2 Year Two Questionnaire Results

8.2.2.1 Year Two Pre-Test Responses

Of the 472 students in the Year Two middle school cohort at the start of the year, one opted out, leaving 471 possible students to take the questionnaire. 456 successfully participated in the Year Two Pre-test, representing a 97% response rate. Table 8.7 below breaks these figures down by grade level:

Table 8.7: Year Two Pre-Test Response Rate by Grade Level

Grade Level	Number of Students in Grade	Number of Students Who Successfully Completed Survey	Response Rate
5	101	101	100%
6	122	119	98%
7	120	119	99%
8	128	117	91%

The 15 missing responses included four questionnaires that I manually excluded, and 10 students who were absent or arrived late that were unable to complete the questionnaire, as detailed in Tables 8.8 and 8.9 below:

Table 8.8: Year Two Pre-Test Excluded Responses

Number of Responses	Reason for Exclusion
2	Duplicate PIN number. I included the responses that aligned with the Grade Level and Gender of the assigned PIN, and excluded the duplicate that did not match.
2	Duplicate PIN number. Both responses matched the Grade Level and Gender of the assigned PIN, and thus both were excluded.
1	Unknown

Table 8.9: Year Two Pre-Test Late Arrivals and Absent Students

Late-arrivals	Absent	Late-arrival Successful Completion	Absent Successful Completion	Total Unable to Complete
10	9	7	1	10

8.2.2.2 Year Two Post-Test Responses

Of the 472 students in the middle school at the start of the year, six left at the end of the first semester and one opted out of the survey, leaving 465 possible students to take the Year Two Post-test. 433 successfully participated in the Year Two Post-test, representing a 93% response rate. Table 8.10 below breaks these figures down by grade level:

Table 8.10: Year Two Post-Test Response Rate by Grade Level

Grade Level	Number of Students in Grade	Number of Students Who Successfully Completed Survey	Response Rate
5	98	93	95%
6	122	115	94%
7	118	108	92%
8	127	117	92%

The 32 missing responses included the following six questionnaires that I manually excluded:

Table 8.11: Year Two Post-Test Excluded Responses

Number of Responses	Reason For Exclusion
1	Duplicate PIN number. I included the response that aligned with the Grade Level and Gender of the assigned PIN, and excluded the duplicate that did not match.
4	Duplicate PIN number. Both responses matched the Grade Level and Gender of the assigned PIN, and thus both were excluded.
1	Typo on PIN number.
6	Unknown

The number of late arrival and absent students displayed in the Table 8.12 below were the result of a common illness. Ultimately, 17 absentees and three late students were unable to complete the questionnaire.

Table 8.12: Year Two Post-Test Late Arrivals and Absent Students

Late-arrivals	Absent	Late-arrival Successful Completion	Absent Successful Completion	Total Unable to Complete
18	31	15	14	20

8.2.2.3 Overall Year Two Participation

In Year Two, there were 465 students who attended the middle school for the entire school year. Of the 465, there were 46 who did not successfully complete either the pre or post-test. The 419 students who successfully completed both the pre and post-test represent a 90% response rate for Year Two of the questionnaire.

8.2.3 Overall Participation in Study

Table 8.13 below represents participation numbers across both years of questionnaire data collection:

Table 8.13: Participation Across both Years of Questionnaire Data Collection

	Total Students at Start of Year	Students Who Left at Semester or Opted Out	Total Possible Students to Participate in Entire Year	Students Who Didn't Complete Pre-Test and/or Post-Test	Students Who Successfully Completed Both Pre-Test and Post-Test	% Successful Participation of Possible Respondents
Year One	468	5	463	30	433	94%
Year Two	472	7	465	46	419	90%

Participation in the study was strong, with the number of students who completed both pre-test and post-test at 90% or higher for both years.

8.3 Analysis of Questionnaire Data

8.3.1 Questionnaire Analysis Plan

To ensure I would not be tempted to try different analytical approaches to find desirable results, I established a plan for analysis prior to looking further into the questionnaire results. A key consideration was that the two years of data collection came from a population sample of an entire middle school, and not a randomly selected sample. Population samples have many affordances, which include greater participation rates and limiting selection bias (Thygesen & Ersbøll, 2014), but the ensuing analysis of the data can potentially be problematic, as inferential statistics are dependent on random sampling (Berk & Freedman, 2001; Gorard, 2006). Following careful deliberation, I arrived at a plan for analysis that I will now outline.

8.3.1.1 Statistical Significance Testing

Null Hypothesis Significance Testing (NHST) has traditionally been used to indicate the probability that observed results would be obtained by random sampling assuming that there was no actual difference between the measured conditions (Rowntree, 1981), or, put another way, to indicate "the likelihood that

the difference between the two groups could just be an accident of sampling" (Coe, 2002, p. 7).

NHST measures (i.e., p-values) have been misused in scientific research as a whole (Wasserstein, 2016; Wright, 2003), and educational research specifically (Gibbs, Shafer, & Miles, 2017). Additionally, p-values lack meaningful information about the size of the effect for experimental conditions (Cumming, 2012; Wasserstein, 2016) and are dependent on sample size (Coe, 2002). I decided to exclude statistical significance measures from my analysis due to these issues, and also because my sample was not randomly chosen and inferential statistics are dependent on random sampling (Berk & Freedman, 2001).

8.3.1.2 Effect Sizes and Confidence Intervals

A suggested alternative to NHST is analysing results with effect sizes (ES) to show the size of the effect, and with confidence intervals (CI) to measure the preciseness of these effect estimations (Banjanovic & Osborne, 2016; Coe, 2002; Cumming, 2012; Wilkinson, 1999; Wright, 2003). One benefit of using these measures within statistical analysis is that they are conducive towards meta-analytic thinking (Thompson, 2002). ESs are also useful to measure effects on unfamiliar scales, such as those in the SRQ-A used in this study (Coe, 2002).

Although I rejected the use of NHST in part on the basis of my sample not being randomly selected, I still chose to include CI measures, even though these inferential statistics measures are also predicated on a randomized sample. This followed the call for study results to be accompanied by measures of uncertainty, such as CIs (Bird et al., 2005).

Some researchers have made the case that a population sample, such as the one from my study, can be conceptualized as just one sample of a larger hypothetical super-population that stretches across time and geography (Thygesen & Ersbøll, 2014). Others have argued that conceptualising a finite

population as a sample of a hypothetical infinite super-population is potentially reasonable (Gibbs et al., 2017; Hartley & Sielken, 1975). Although some feel that a super-population approach is fallacious reasoning used by researchers to justify the incorrect application of inferential statistics (Berk & Freedman, 2001), I utilized this approach in my statistical analysis to justify my use of CIs to provide error measurements on the ES measurements.

8.3.1.2.1 Effect Size for Pre-Test Post-Test with Control

Cohen's *d* is an effective and recommended measure for determining ES (Cumming, 2012). The basic formula for *d* is the change in mean score divided by the standard deviation (Cohen, 1992). A key consideration within this formula is which standard deviation to use – that of the pre-test, post-test, or a pooled combination of the two. This study used the pooled standard deviation of the pre-tests (Morris, 2008). Morris (2008) suggested the following formula for calculating effect sizes within a study that used a pre-test post-test-control group design:

Figure 8.1: ES Formula

$$d = \frac{(M_{post,T} - M_{pre,T}) - (M_{post,C} - M_{pre,C})}{SD_{pooledpre}}$$

where T is the treatment group, and C is the control group.
Within this formula, the equation for the pooled standard deviation of the pre-tests is defined as:

Figure 8.2: Pooled SD of Pre-Tests Formula

$$SD_{pooledpre} = \sqrt{\frac{(n_T - 1)SD_{pre,T}^2 + (n_C - 1)SD_{pre,C}^2}{n_T + n_C - 2}}$$

8.3.1.2.2 Bootstrapped Confidence Intervals

Bootstrapping works by replacing one set of responses from a data set with another randomly selected set of responses, from the same full set. This process is repeated over a chosen number of times to generate estimates of the

error which occurred in a given study's sample (Banjanovic & Osborne, 2016; Kirby & Gerlanc, 2013; Wood, 2005). In doing this, bootstrapping mimics the error measurements of inferential statistics which attempt to measure how accurate an observed data set is within a potentially infinite number of other samples which could have been drawn from any given population. Thus, bootstrapping is an effective way to measure error within population samples and make statistical inferences using the measured data set (Thygesen & Ersbøll, 2014). This study followed Wood's (2005) recommendation to compute bootstrap figures using 10,000 resamples, as this number was found to produce almost identical CIs to using 100,000 resamples.

8.3.2 Computing Effect Sizes

Results from all questionnaire sessions were calculated in SPSS. After the response sessions, results were exported from BOS in an Excel spreadsheet, which was then imported into SPSS.

Results for each year were first calculated by scale and grade level for the descriptive statistics of mean and standard deviation (SD), as represented in Tables 8.14-8.23 below. The SD measures in particular were needed to determine the pooled SD of the pre-tests outlined by Morris (2008) as the denominator in the ES formula (see Figure 8.1 above).

Table 8.14: Year One External Regulation Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	99	2.83	0.56	2.75	0.62
6	109	2.93	0.57	2.88	0.62
7	105	2.93	0.58	2.95	0.51
8	120	2.99	0.53	2.98	0.52

Table 8.15: Year One Introjected Regulation Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	99	2.87	0.60	2.85	0.58
6	109	3.07	0.48	2.96	0.55
7	105	3.05	0.49	3.00	0.53
8	120	3.06	0.47	3.04	0.53

Table 8.16: Year One Identified Regulation Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	99	3.53	0.45	3.48	0.45
6	109	3.36	0.45	3.32	0.52
7	105	3.30	0.51	3.26	0.58
8	120	3.32	0.50	3.28	0.56

Table 8.17: Year One Intrinsic Motivation Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	99	2.99	0.66	2.86	0.66
6	109	2.62	0.68	2.43	0.68
7	105	2.51	0.63	2.46	0.70
8	120	2.52	0.65	2.50	0.66

Table 8.18: Year One RAI Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	99	0.98	2.02	0.85	2.44
6	109	-0.34	2.37	-0.56	2.55
7	105	-0.57	2.46	-0.72	2.26
8	120	-0.69	2.13	-0.72	2.05

Table 8.19: Year Two External Regulation Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	95	2.77	0.68	2.72	0.69
6	110	2.74	0.60	2.76	0.63
7	107	2.90	0.58	2.99	0.54
8	107	2.97	0.47	2.86	0.52

Table 8.20: Year Two Introjected Regulation Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	95	2.87	0.65	2.90	0.60
6	110	2.84	0.63	2.91	0.59
7	107	2.99	0.52	2.98	0.56
8	107	2.98	0.48	2.90	0.60

Table 8.21: Year Two Identified Regulation Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	95	3.44	0.51	3.44	0.49
6	110	3.38	0.47	3.36	0.49
7	107	3.29	0.52	3.20	0.58
8	107	3.24	0.52	3.16	0.65

Table 8.22: Year Two Intrinsic Motivation Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	95	2.91	0.78	2.86	0.66
6	110	2.71	0.66	2.51	0.63
7	107	2.40	0.67	2.36	0.66
8	107	2.41	0.63	2.37	0.64

Table 8.23: Year Two RAI Mean and SD

Grade Level	N	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
5	95	0.86	2.57	0.83	2.56
6	110	0.49	2.33	-0.04	2.26
7	107	-0.70	2.35	-1.03	2.11
8	107	-0.86	2.06	-0.74	1.99

Next, for the numerator in the ES formula, changes in mean across each year of the study were determined within SPSS using the mean figures listed in the tables above. At the same time, SPSS also computed the bootstrapped (10,000 resamples) standard error (SE). At this point, pre-test SDs from each year were also computed along with the bootstrapped (10,000 resamples) standard error (SE). This analysis is detailed by SRQ-A scale in Tables 8.24-8.28 below.

Table 8.24: External Regulation Change in Mean with SD and SE

	Y1 Change	Std. Error	Y2 Change	Std. Error	Y1Pre SD	Std. Error	Y2Pre SD	Std. Error
Grade 5	-0.08	0.05	-0.05	0.05	0.56	0.04	0.68	0.05
Grade 6	-0.05	0.04	0.02	0.04	0.57	0.04	0.60	0.03
Grade 7	0.03	0.04	0.09	0.03	0.58	0.03	0.58	0.04
Grade 8	-0.01	0.04	-0.11	0.04	0.53	0.04	0.47	0.03

Table 8.25: Introjected Regulation Change in Mean with SD and SE

	Y1 Change	Std. Error	Y2 Change	Std. Error	Y1Pre SD	Std. Error	Y2Pre SD	Std. Error
Grade 5	-0.02	0.05	0.03	0.04	0.60	0.04	0.65	0.05
Grade 6	-0.11	0.04	0.07	0.05	0.48	0.03	0.63	0.04
Grade 7	-0.05	0.03	-0.01	0.04	0.49	0.04	0.52	0.04
Grade 8	-0.02	0.03	-0.08	0.04	0.47	0.03	0.48	0.04

Table 8.26: Identified Regulation Change in Mean with SD and SE

	Y1 Change	Std. Error	Y2 Change	Std. Error	Y1Pre SD	Std. Error	Y2Pre SD	Std. Error
Grade 5	-0.05	0.04	0	0.04	0.45	0.04	0.51	0.05
Grade 6	-0.05	0.05	-0.02	0.04	0.45	0.04	0.47	0.04
Grade 7	-0.04	0.04	-0.09	0.04	0.51	0.06	0.52	0.04
Grade 8	-0.04	0.04	-0.08	0.04	0.50	0.03	0.52	0.05

Table 8.27: Intrinsic Motivation Change in Mean with SD and SE

	Y1 Change	Std. Error	Y2 Change	Std. Error	Y1Pre SD	Std. Error	Y2Pre SD	Std. Error
Grade 5	-0.13	0.05	-0.05	0.05	0.66	0.04	0.78	0.05
Grade 6	-0.19	0.05	-0.20	0.05	0.68	0.04	0.66	0.04
Grade 7	-0.06	0.05	-0.03	0.05	0.63	0.05	0.67	0.04
Grade 8	-0.02	0.04	-0.05	0.04	0.65	0.04	0.63	0.04

Table 8.28: RAI Change in Mean with SD and SE

	Y1 Change	Std. Error	Y2 Change	Std. Error	Y1Pre SD	Std. Error	Y2Pre SD	Std. Error
Grade 5	-0.13	0.17	-0.03	0.15	2.02	0.14	2.57	0.18
Grade 6	-0.21	0.15	-0.53	0.15	2.37	0.17	2.33	0.13
Grade 7	-0.16	0.16	-0.33	0.14	2.46	0.16	2.35	0.23
Grade 8	-0.03	0.15	0.12	0.13	2.13	0.14	2.06	0.13

These two calculation sets represented in the tables above provided all of the components I needed to calculate ESs using the Morris (2008) formula. The necessary data from these tables was then transferred into an Excel spreadsheet to compute the final ESs, represented in Table 8.29 below.

Table 8.29: Final Effect Size Measurements

	External Regulation	Introjected Regulation	Identified Regulation	Intrinsic Motivation	RAI
Grade 5	0.05	0.08	0.10	0.12	0.04
Grade 6	0.13	0.31	0.07	-0.02	-0.14
Grade 7	0.12	0.07	-0.10	0.04	-0.07
Grade 8	-0.19	-0.13	-0.09	-0.04	0.07

For clarity, Figure 8.3 below provides one example of the ES calculation with corresponding figures from the study inputted:

Figure 8.3: ES Formula for Grade 5 Introjected Regulation

$$\frac{(.03) - (-.02)}{\sqrt{\frac{(95-1).65^2 + (99-1).6^2}{95+99-2}}} = .08$$

8.3.3 Computing 95% Confidence Intervals

Computing 95% CIs entailed combining the bootstrapped 68% errors from the figures used to compute the ESs. Where the original ES formula called for adding or subtracting, the root mean square of the errors was used. Where the original formula called for multiplying and dividing, fractional errors were used and combined with the same root mean square calculation (Hogan, 2006). Following this rationale, the formula for obtaining the standard error (SE) figures for each ES measurement was:

Figure 8.4: SE for ES Formula

$$SE = ES\left(\frac{\sqrt{\left(SE_{MeanChange_T})^2\right)+\left(SE_{MeanChange_C})^2\right)}}{(MeanChange_T)-(MeanChange_C)}+\frac{\sqrt{\left(SE_{preSD,T}\right)^2+\left(SE_{preSD,C}\right)^2}}{\frac{\left(SD_{pre,T}\right)+\left(SD_{pre,C}\right)}{2}}\right)$$

Excel was used to first compute these SE measurements. The SE was then multiplied by 1.96 and either added to, or subtracted from, the ES to arrive at the 95% CI range for each ES measurement (Cumming, 2012). Final CI figures are provided by scale in Tables 8.30-8.34 below, followed by plots of final ES and CI figures by scale in Figures 8.5-8.9.

Table 8.30: External Regulation SE and 95% CIs for ES Figure

	External Regulation			
	ES	SE	95%+	95%-
Grade 5	0.05	0.01	0.08	0.03
Grade 6	0.13	0.01	0.15	0.10
Grade 7	0.12	0.01	0.14	0.10
Grade 8	-0.19	0.01	-0.17	-0.21

Table 8.31: Introjected Regulation SE and 95% CIs for ES Figures

	Introjected Regulation			
	ES	SE	95%+	95%-
Grade 5	0.08	0.01	0.11	0.06
Grade 6	0.31	0.01	0.34	0.29
Grade 7	0.07	0.01	0.10	0.05
Grade 8	-0.13	0.01	-0.10	-0.15

Table 8.32: Identified Regulation SE and 95% CIs for ES Figures

	Identified Regulation			
	ES	SE	95%+	95%-
Grade 5	0.10	0.02	0.14	0.07
Grade 6	0.07	0.02	0.10	0.03
Grade 7	-0.10	0.01	-0.07	-0.13
Grade 8	-0.09	0.02	-0.06	-0.12

Table 8.33: Intrinsic Motivation SE and 95% CIs for ES Figures

	Intrinsic Motivation			
	ES	SE	95%+	95%-
Grade 5	0.12	0.02	0.15	0.09
Grade 6	-0.02	0.02	0.01	-0.05
Grade 7	0.04	0.02	0.07	0.01
Grade 8	-0.04	0.01	-0.01	-0.07

Table 8.34: RAI SE and 95% CIs for ES Figures

	RAI			
	ES	SE	95%+	95%-
Grade 5	0.04	0.1	0.24	-0.15
Grade 6	-0.14	0.08	0.03	-0.30
Grade 7	-0.07	0.08	0.09	-0.23
Grade 8	0.07	0.1	0.26	-0.12

Figure 8.5: External Regulation Plot of Final ES and 95% CI

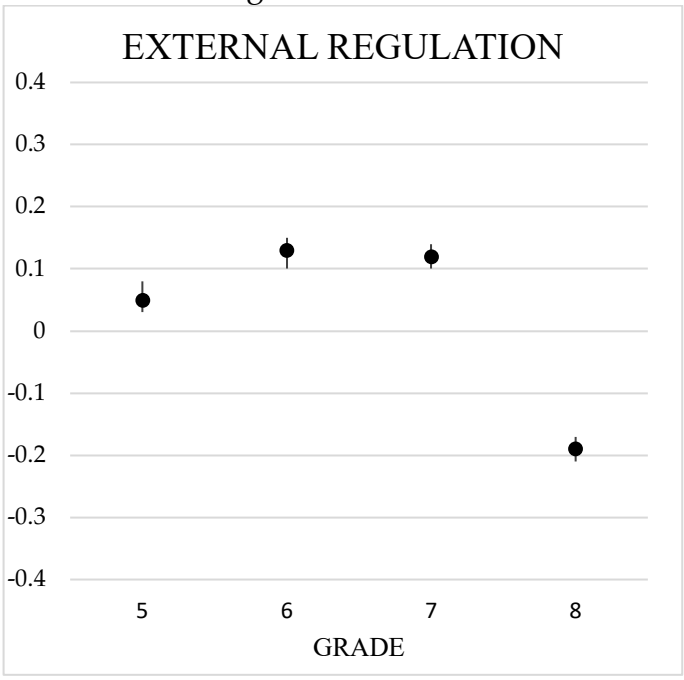


Figure 8.6: Introjected Regulation Plot of Final ES and 95% CI

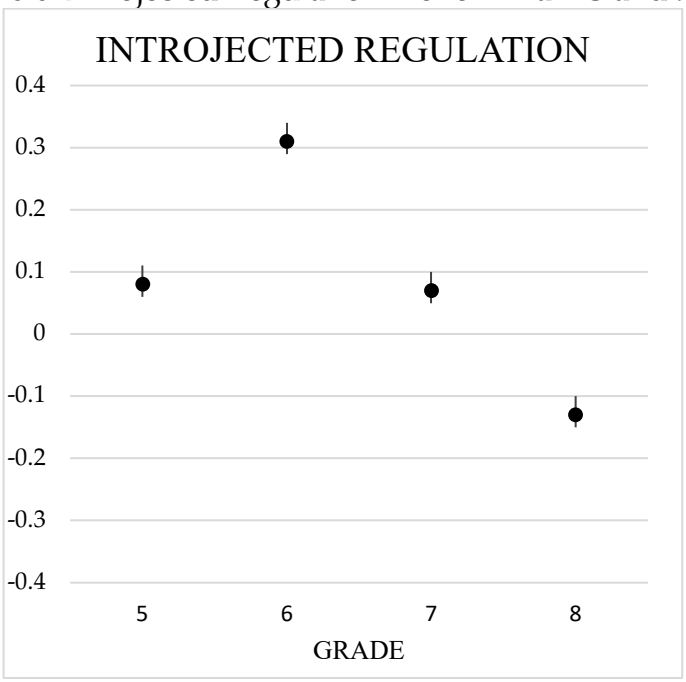


Figure 8.7: Identified Regulation Plot of Final ES and 95% CI

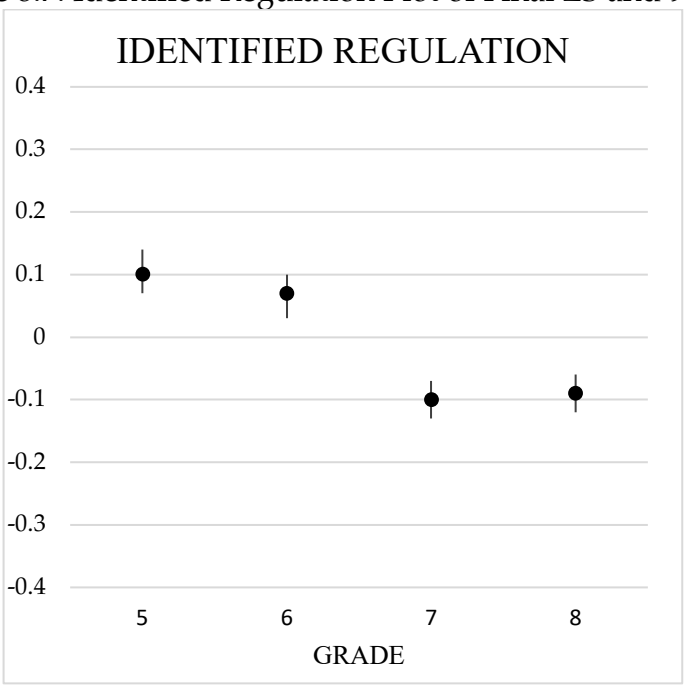


Figure 8.8: Intrinsic Motivation Plot of Final ES and 95% CI

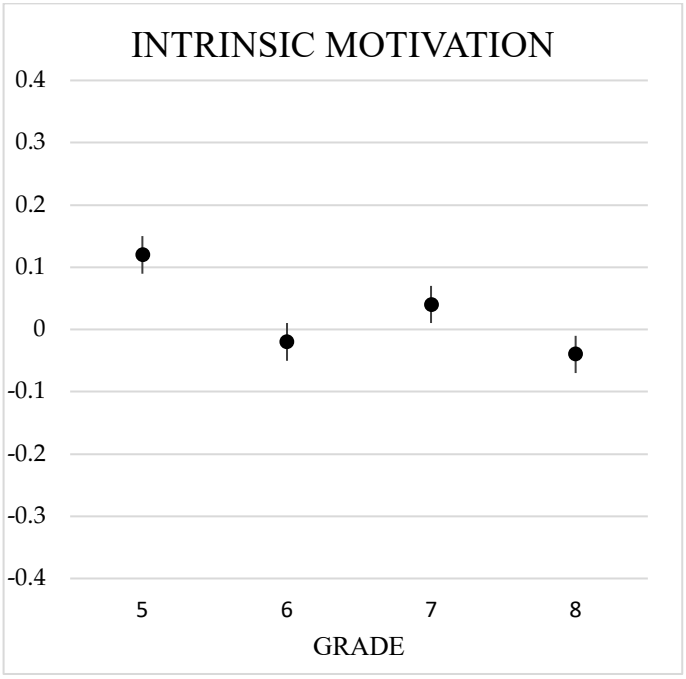
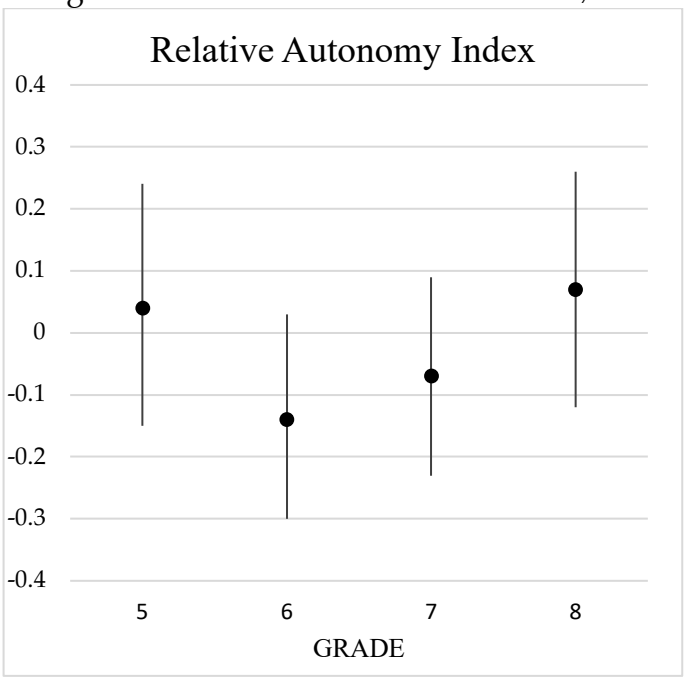


Figure 8.9: RAI Plot of Final ES and 95% CI



8.3.4 Summary of Questionnaire Findings

ES and CI results revealed that there was very small effect on motivation across each grade level and scale as the result of the shift to standards-based grades and report cards. Using conventional definitions of ES as 0.20 = small, 0.50 = medium, 0.80 = large (Cohen, 1988), only Grade 6 Introjected Regulation had an ES that would be considered small. All other ES were below 0.20, a range

considered a weak effect (Cohen et al., 2011). Further, CI results indicate that I can be confident that within a larger superpopulation in exactly the same circumstances, only two of the measured scales (Grade 8 External Regulation, Grade 6 Introjected Regulation), along with the Relative Autonomy Index (RAI) scales would possibly fall within an ES considered to have a small effect of higher than 0.20, while none would reach the medium ES of 0.50.

Because these scales represent a continuum of motivation, if standards-based grades and report cards had a truly positive impact on autonomous and self-determined motivation, one would expect a trend going from a very small (or even negative) ES for External Regulation and Introjected Regulation, and then growing larger on the positive side for Identified Regulation and Intrinsic Motivation. This trend would be reflected with a larger and positive RAI figure (see chapter four). Both Grades 5 and 8 show this trend on a very small scale; they have a RAI ES that is positive, albeit very small.

The comparatively large Grade 6 Introjected Regulation ES is a curious result. Introjected Regulation is characterized by having regulations which guide one's actions that are not actually integrated into one's beliefs and identity. This type of regulation is marked by doing things due to pressure and guilt avoidance (Ryan & Deci, 2000). It is likely that the implementation of the SBRC took different shapes within different grade levels, and the way it was implemented in this instance may have resulted in students feeling pressured to adopt this new standards-based way of thinking about report cards.

8.4 Focus Group Analysis Related to Research Question 3

In chapter seven I analysed and discussed the varied ways students perceived the new grading system: standards-based grades and report cards as formative feedback through self-assessment with rubrics; and standards-based grades and report cards as an opaque extra step towards the letter grade. I will now re-examine these key focus group findings for their SDT implications,

looking first at student comments about the formative function of standards-based grades and report cards and secondly at standards-based grades as an opaque extra step. Following this analysis I will triangulate these findings with questionnaire findings and situate them within SDT research literature to support my analytic generalizations that the formative function of standards-based grades and report cards likely supported students' basic psychological needs of autonomy and competence and led to increased autonomous motivation, while standards-based grades as an opaque extra step likely thwarted students' need for autonomy and led to an offsetting decrease in autonomous motivation.

8.4.1 The Formative Function of Standards-Based Grades as Autonomy Supportive

As I discussed in chapter seven, some focus group students described a formative function of standards-based grades and report cards through a self-assessment process of linking standards-based grades and rubrics. For some, this formative function of the standards-based grading system began before they had received their standards-based grade from their teacher. These students articulated a process in which they felt they were able to generate their own formative feedback while working on assignments, prior to turning the assignments in. A key aspect of this process was that students identified clear learning targets within the rubrics to guide their next steps while working on an assignment, represented by this comment:

> I think that with the *progressing/meeting* thing, you sort of know what's expected of you. For like a paper, if you look at the rubric, you know what you need to do to get a certain section, like *meeting*, *exceeds*[21] and stuff. So it's more helpful because then you really know, it's more guided as to what to do, it's not so rough (Grade 8 Winter 2017, 8-B).

[21] In the Quasi-Experimental Design Questions section at the end of this chapter I will discuss the implications of students who still had not internalized the correct terminology for some of the standards-based grades.

Similarly, another student shared that a vignette character[22] working on an assignment

> … knows what is required if he's doing the [standards-based grade], not the letter grade. Because he can see what he has to do and then he can apply his knowledge to fit that (Grade 8 Autumn 2017, 8-L).

Once students received their standards-based grade through a teacher's evaluation on a rubric, the relatively autonomous process of self-assessment continued as some students described triangulating their standards-based grade, the rubric descriptions, and their work:

> When you [look at a standards-based grade] there's always a definition underneath, so there's always why you got this grade or not. So when you read it, when you actually read the definition you can understand and you can relate to things that happened that was related to that definition (Grade 5 Winter 2018, 5-K).

From this step of reading the description of the standards-based grade attained on the rubric and linking it to past performance, other students shared how they were then able to identify the next steps in their learning progression on the rubric:

> I think [standards-based grades are] easier to understand for a student because it's all there for you on the rubric, you can just look at your section and look down the line (Grade 8 Winter 2017, 8-4).

Another student summarized this self-assessment process of triangulating a standards-based grade from the teacher with rubric information and previous work to generate formative feedback:

> So you're a *Beginning* let's say, then [the rubric] has what you did wrong and then the next box you can look over to see why you didn't get it. So you can read over your essay and you can see, "Yeah, I can see, I didn't include descriptive terms, I didn't use similes or metaphors." You can see exactly what you did wrong so you can improve next time (Grade 7 Autumn 2017, 7-L).

Other students shared that the formative function of standards-based grades linked to rubrics relieved the confusion of not knowing how to take further steps in their learning:

[22] For a full discussion of the vignettes I used within focus groups, please refer back to chapter four.

> …[Standards-based grades with rubrics] tells you what you're doing now, but it [also] tells you what you can do to get the next thing. So you're not just stuck there thinking, "what do I do now, cause I did all that I could have." You actually know what you can do (That's a good point!) (Grade 5 Winter 2017, 5-1).

Some students frequently contrasted the formative process of self-assessment with standards-based grades and rubrics to the old letter grade system, which these students saw lacking information to help them develop in their learning. This sentiment was represented in this comment:

> …the standards are better to grow and improve, because with the percentages [linked to letter grades], you just get the percentages, but with the standards, it has, in each of the boxes it actually has words…But I feel like percentages it's kind of more vague…you know that it could be improved on but you don't know exactly what you did wrong...So I think that the standards are more precise and help you grow more (Grade 7 Autumn 2017, 7-J).

Another student expressed similar confusion and also frustration with the lack of meaningful information about their current standing in letter grades:

> …with *Beginning/Approaching/Meetings* and *Exceeding*, you can see how much you've developed since the beginning of the year and what you need to improve on. But with just letter grades, like if they just give you *B*, you ask questions like, "what for? I didn't do anything wrong." (Grade 7 Winter 2017, 7-4).

The SBGs provided formative information that students were able to process autonomously about current standing and the direction of next steps. Letter grades were lacking this information, leaving students in the dark. In this context, even desirable *A* letter grades were notable for their lack of formative information, with students uncertain of specific details about their current standing:

> I agree with 6-1[23], because a letter grade, if they give you an *A*, yeah, you're happy, but you don't really know what it means. You don't really know what things you did well on - maybe you missed just one little thing (Grade 6 Winter 2017, 6-C).

From this perspective, the formative function of standards-based grades and report cards linked to rubrics represented an upgrade over the previous

[23] In anonymizing focus group comments, where one participant referred to another by name, I inserted that individuals coded pseudonym.

letter grade system, allowing students to take next steps in their learning with a greater degree of autonomy.

8.4.2 The Autonomy Thwarting Nature of Standards-Based Grades as an Opaque Extra Step

The other predominant perception of standards-based grades that students shared was standards-based grades as a frustrating and opaque extra step towards identifying their letter grades. A key aspect of this perspective was interpreting the standards-based grades through a letter grade framework. Leading up to the implementation of standards-based grades and report cards, when asked if the new grading system would change the way students saw grades, one student responded:

> I don't think it will. Cause I know in 7th grade a lot of teachers look for that approach. But I notice that everyone else and me, we'd ask the teacher, "Oh, is this kind of like an *A*, or like a *B*?" So people will still find a way to get that connection to the *A*, or *C*, or whatever. So I don't think it's going to make a difference whether they change it to *Meets* and *Progressing* (Grade 8 Winter 2017, 8-A).

Once standards-based grades and report cards were officially implemented at the school, this student's prediction was realized by some. For these students, the standards-based grade was simply an extra-step that required them to translate it back into a letter grade. One focus group participant recognized how another student instinctively translated standards-based grades back into letters:

> 7-10, you're already translating it into *A*,*B*,*C* letter grades. And that's what I think is the thing about the standards: people will always translate. You weren't saying *Progressing* is a *Progressing* or a *Meeting* is a *Meeting*. You were saying, "*Progressing* is in the *B* range" and I think that's kind of what happens with people…(Grade 7 Winter 2018, 7-12).

Some students shared the steps they had to take to translate a standards-based grade back into a letter grade. This process was detailed by one participant who described what a vignette character would do upon receiving a standards-based grade for an assignment:

> I think Alex might be very confused at first, but then he'll gradually figure out how to translate it back into letter grades. Like [SIS] gave us the little thing in the homework diary[24] to see where a 1.5 or a 2.6 is on the marks. So, he might be confused but he can figure it out (Grade 8 Winter 2018, 8-12).

For many, however, this extra-step of translation was opaque and imprecise, and they expressed confusion with their standards-based grade and how to conduct this translation. One student explained that

> A lot of kids are used to, especially schools in America, are used to the letter grades. So in a way, for a long time I was getting all the standard *Extending*/*Meeting*/*Progressing*/*Beginning*, and I did not actually know what my grade was. Like if I got all *Meetings*, I would think I would be good. And that pretty much is good. But I would never actually know, so in a way, it's kind of just shoving the grade under the rug. (I agree) (Grade 6 Winter 2017, 6-5).

As discussed in chapter seven, part of this translation confusion was that the four standards-based grades imprecisely translated into 13 letter grade categories corresponding to a 100-point scale:

> I feel like there needs to be a balance because I feel like the *Meeting*, *Progressing*, *Beginning* doesn't tell you where you're at because a *Beginning* can be anywhere from a 0-79, and it doesn't really tell you what your grade actually is (Grade 8 Winter 2018, 8-J).

> But just *Extending*, *Meeting*, *Progressing*; that there's only 4 so it's kind of a bit harsh, cause some can be overreacting- *Progressing* can be like overreacting (Grade 5 Winter 2018, 5-11).

For some students, one result of receiving a standards-based grade and no longer being able to identify their letter grade standing was a negative impact on their well-being:

> It's kind of like the *Meeting* thing, where you have a lot of different things that it could be. It just kind of makes, personally, it just kind of makes me worry about whether I got the worst grade, the best grade, a not good grade. It just kind of feels bad (Grade 7 Autumn 2017, 7-11).

[24] The standards-based /letter grade conversion chart in the homework diary. See Figure 5.4 and chapter five for a full discussion. Also of note, as discussed in the previous chapter, this student used numbers with decimals as a substitute for the actual standards-based grade, suggesting that the variation in teacher practices resulted in some students seeing the two as synonymous.

Some students who viewed standards-based grades as an opaque extra-step recognized that it was hard to shift away from letter grades because these were the engrained symbols that they had grown up with:

> You spend three years understanding letter grades, so I kind of just need to know how that compares. And I don't know if that's just a bad habit, but, it's like, I need to understand it (Grade 8 Autumn 2017, 8-K).

> I think that the people who were here last year or have been here for a while are kind of used to the letter grade system and it's kind of big drastic grade to go to standards and instead of letters go to 4.0's and things like that [25](Grade 7 Autumn 2017, 7-K).

Others reiterated the severity of the reform and believed that the shift to the new standards-based grading system was going to take time:

> I think that it will take, cause over time the schooling system was built off of the grades that it's going to take so much time. It can't just take one switch to fix everything. It's going to take years, and it's going to take a lot of thought (Grade 7 Autumn 2016, 7-1).

As this student recognized, the engrained nature of letter grades makes standards-based grading reform a complicated task requiring patience, effort, and time. I will return to this point at the end of this chapter when considering its implications for quasi-experimental studies aiming to measure assessment reform.

8.5 Meta-Inference for Research Question 3

Qualitative and quantitative methods can be combined in mixed methods studies to corroborate and complement each other (Onwuegbuzie & Johnson, 2006; Tashakkori & Teddlie, 2009). In particular, the larger patterns found in quantitative data can be explained in more depth with qualitative data and analysis (Creswell, 2012; Spillman, 2014). Within this chapter, I utilized my study's Parallel Mixed Design by analysing separate qualitative and quantitative strands of data which I will now integrate to form a larger meta-inference in

[25] Again, notice this student's use of a number with a decimal in place of referring to the standards-based grade. For a full discussion on this issue, refer to the chapter seven section, Teacher Implementation Issues.

identifying how these strands overlap (Tashakkori & Teddlie, 2009). In forming these meta-inferences, I also situate them within theory and research to support my analytic generalizations (Yin, 2018).

8.5.1 The Negating Impact of Variations in Students' Standards-Based Grade Meaning

Theoretically the degree to which individuals are autonomously motivated depends on environmental support of their three basic psychological needs: relatedness, competence, and autonomy. The more these needs are satisfied, the more autonomous and internalized the form of regulation individuals experience (Deci & Ryan, 1985; Ryan & Deci, 2000, 2017). SDT defines competence as understanding how to reach given outcomes (Deci et al., 1991) and a key component of competence support is the ability to understand where one is at and how to move along a learning progression with optimally challenging tasks that are not too easy or hard (Deci & Ryan, 1985; Ryan & Deci, 2009). Students who described self-assessment processes in linking their standards-based grades to rubrics consistently had their competence supported with key formative information of how to move from their current standing to desired learning outcomes. Further, they expressed their preference for standards-based grades over the previous letter grade system which they saw as lacking a formative function, leaving them confused with how to move forward in their learning. As such, SGBs and SBRCs supported some students' competence in ways letter grades did not.

According to SDT, autonomy is "…the experience of behaviour as volitional and reflectively self-endorsed" (Niemiec & Ryan, 2009, p. 135). Focus group participants who felt standards-based grades and report cards functioned formatively frequently described having the power to self-assess and make informed decisions about the next steps in their learning through their triangulation of their standards-based grade, the rubric, and their completed

work. This process constituted a heightened autonomy when compared to letter grades where student descriptions almost never described this autonomous self-assessment process in identifying next steps in learning. These findings of the autonomy and competence supporting nature of the formative function of standards-based grades and report cards are consistent with assessment theory research which has posited that the formative function of assessment can support students' competence (Andrade, 2010a; Andrade et al., 2010; Black & Wiliam, 1998; Sadler, 1989, 1998) and autonomy (Cizek, 2010; Dann, 2014; Earl, 2003; Sadler, 1998).

However, not all students experienced the standards-based grading and reporting initiative formatively. For other students, standards-based grades represented an added layer of confusion in determining their letter grade standing. SDT has long stated that in educational contexts, grades represent autonomy thwarting rewards and punishments, leading to less autonomous forms of motivation (Deci & Ryan, 1985; Grolnick & Ryan, 1987). This has been confirmed in letter grade research with the finding that teachers include behaviour measures in letter grades to control student behaviour (Bonner & Chen, 2009; Brookhart, 1994; Frary et al., 1993; Pilcher, 1994; Sun & Cheng, 2014). Students rooted in traditional letter grades now had the standards-based grade as an imprecise extra step to determine their letter grade, and this opaqueness represented a further autonomy thwart as they were no longer able to identify precise letter grade meaning. These students were very clear: the new standards-based grading system with the hybrid SBRC did not represent business as usual with letter grades, but instead it was an additional autonomy thwart interpreted in controlling ways. This was seen in comments that viewed standards-based grades as a "trick" that shoved the letter grade "under the rug." According to SDT, autonomy plays a heightened role among the three basic psychological needs in supporting more autonomous forms of motivation (Deci

& Ryan, 2000), and limitations on student autonomy should result in less autonomous forms of motivation (Ryan & Deci, 2017). The autonomy thwarts represented by standards-based grades and report cards were more deleterious to some students' motivation than if the original letter grade system had remained untouched.

It should be noted that students who saw standards-based grades and report cards as an opaque extra step likely did not experience these mediational means as an increased competence thwart. As discussed in chapter six, students almost never associated letter grades with next steps in their learning. As such, their fixation on letter grades within the standards-based shift and their inability to realize the formative function of standards-based grades and report cards by linking them to rubrics would have been very similar to how they experienced letter grades, from a competence perspective.

Given the variation in students' meanings of SBGs and SBRCs, it is unsurprising that questionnaire results revealed very small effect sizes on student motivation with the standards-based shift. It is probable that the end result of these different viewpoints of the new grading system was that gains in autonomous motivation for some students were negated by losses for others.

8.5.2 Quasi-Experimental Design Questions

One explanation of the varied meanings students constructed of SBGs and SBRCs and the very small ES results is the design of the study. The Year Two treatment group took their post-test at the end of the first semester, after one semester with SBGs and just after receiving the SBRC for first time. According to Sadler (1998), quasi-experiments can struggle to measure the effectiveness of new feedback constructs because assessment reform can take a long time to effect change. This is perhaps why even Grade 5, who did not have letter grades in Year Two of the study, had no meaningful ES with the standards-based grading and reporting shift: not enough time had elapsed for the new method to

effect systemic change and these Grade 5 students viewed the new standards-based grades just as all the other students did – with variations between formative feedback and opaque extra steps. One sign that data was collected while the students were still in transition to the new standards-based system was that some students in focus groups referred to the standards-based grades of *Progressing* as *Approaching*, or *Extending* as *Exceeding*. In this sense, these new mediational means had not yet concretized in meaning and some students were still in the process of understanding their basic elements.

Senior leaderships' comments in chapter five also help to understand the process of reform the school was going through during my study. The school year after the final questionnaire and focus group data had been collected, leadership recognized that there was still work to do and shared their perception that teachers were in different places of understanding and implementation during the initial year of standards-based grades and report cards at the school. Within the school's 2016 self-study, some teachers also expressed the sentiment that they had not had enough time to understand and adopt standards-based grading practices in their classroom. Overall, the 2016 self-study suggested varied understanding and implementation levels of the school's teachers with standards-based practices, corroborating senior leadership interview comments and student focus group comments about the different shapes standards-based grading took in their classes (see previous chapter). The findings from this study that standards-based reform is challenging and requires time to take systemic effect are also similar to FA research which found that teachers need a good deal of time and support to successfully implement FA reform, particularly self-assessment processes (Wylie & Lyon, 2015). Initial SBG research has also found that SBG reform is challenging for teachers and requires time for successful implementation (Knight & Cooper, 2019; Scarlett, 2018).

The transitional state of the reform at SIS during the two years when questionnaire and focus group data were collected helps to explain the varied meanings students constructed of standards-based grades and report cards and the very small effect sizes found in my questionnaire analysis.

8.6 Conclusion

Within this chapter I have sought to answer my research question, *How does the shift to standards-based grades and report cards impact student motivation?* I began by reviewing the results of my questionnaire and then analysing those results according to my predetermined analysis plan. Analysis of the questionnaire results revealed that the shift to standards-based grades and report cards had little overall impact on middle school students' motivation at SIS, determined through effect sizes with bootstrapped confidence intervals. Triangulating these findings with focus group data and senior leadership interviews provided a deeper understanding of the small effect sizes. It is possible that the shift to standards-based grades and report cards did enhance autonomous motivation for some students through supporting their basic psychological needs of autonomy and competence. These gains could have been offset by constructions of standards-based grades and report cards as autonomy thwarting extra steps to the letter grade. I finished this chapter by considering the need for quasi-experimental studies measuring assessment reform, such as this one, to account for the long amount of time needed for reform changes to take effect. In the following conclusion chapter, I will return to this challenge, while also considering the research contributions of my study and areas for future research to build on these findings.

Chapter Nine: Conclusion

I have stressed throughout this thesis that the findings I discussed in the previous three chapters come from my case study of one middle school within an American international school. In these previous chapters, I have situated my findings within a broader body of theory and research to form analytic generalizations which represent preliminary findings for initial research into SBGs and SBRCs. I will now review these findings before discussing challenges with my study, suggesting future areas for research, and providing closing thoughts on my work.

9.1 Summary of Findings

Analysis and discussion of my three research questions in the three preceding chapters revealed six key findings from my study: letter grades as purely summative measures; the formative function of standards-based grades and report cards; standards-based grades and report cards as an opaque extra step towards the letter grade; avoiding hybrid SBRCs to maximize FA potential; the importance of continued parent and teacher education on standards reform; and that ES measurements revealed the shift to SBRCs to have a very small impact on student motivation overall, possibly the result of the negating motivational impacts of varied student constructions from standards-based grades and report cards. I shall now discuss each of these findings in turn.

My first significant finding was that middle school students at SIS perceived letter grades in purely summative ways, distinctly lacking a formative function. Student responses did not exhibit formative self-assessment processes with letter grades, highlighted by the disconnect between their letter grades and the corresponding rubrics. It is likely that a major cause of this disconnect was that students perceived letter grades as constructs of behaviour measures and fixed ability. I have drawn on Wertsch's framework to reveal that many of the cultural, institutional and historical influences which shaped letter grades are

still embedded in these means. Students constructed meanings of letter grades as fixed labels of intelligence derived from norm-referenced standing with peers, and as gatekeepers to higher levels of schooling and future success. Students also perceived their parents to reinforce these meanings. It is likely that these embedded meanings further prevented students from realizing the formative link between their letter grades and the corresponding rubrics.

My next significant finding was that many middle school students at SIS articulated a formative function of their standards-based grades and report cards. Students described using SBGs and SBRCs for all three of the FA gap steps: identifying the learning target; identifying current standing; and formulating and acting on next steps to close the gap. An essential aspect of this was linking their standards-based grade to the criteria rubrics. This mixing of traditionally summative means for formative purposes supports the FA rationale behind standards-based grade and report card reform. This formative function of standards-based grades was supported by students' views that standards-based grades were largely free from constructs of behaviour measures. Further, some students perceived that these mediational means held a new definition of average - one in which the average standards-based grade represented success for all. Another supporting factor was some students' perception that parents understood and reinforced the formative function of these mediational means.

In contrast to some students' perception that standards-based grades and report cards functioned formatively, another important finding from my study was that some students described standards-based grades and report cards as a confusing and frustrating extra step towards identifying their letter grade standing. This perception was in-line with the shared view of some that standards-based grades held the same norm-referenced meaning as letter grades. Students identified factors that made it hard to move away from letter grades and embrace the new reform, namely, inconsistent teacher implementation,

parents who were uneducated in the new standards-based system, and the school's use of a hybrid SBRC. Wertsch's (1991) theory states that new mediational means, in trying to address constraints of the previous means they replace, often introduce new and unforeseen constraints of their own. Some students' constructed meaning of standards-based grades and report cards as an opaque extra-step to their letter grade is a major such constraint of these new means.

Another important finding from my study was that schools attempting to maximize the formative potential of SBGs and SBRCs should avoid using a hybrid SBRC. An exploration of the cultural, institutional and historical forces which shaped SIS' decision to use a hybrid report card revealed influences which contrasted with the school's FA rationale to begin standards-based reform. Although the decision to include a hybrid SBRC was practical and done to avoid reform backlash, it ultimately had little to do with supporting student learning through enhancing formative feedback. In this sense, SIS was very typical of schools which opt to use a hybrid SBRC. The inclusion of the letter grade in the SBRC made it more difficult for some students to unlock the formative potential of these new mediational means.

My fifth key finding was the importance of continued teacher and parent education on standards reform. Given letter grades' dominant and engrained cultural status, it is likely that they will remain an embedded historical and institutional presence in any new form of grading and report cards. As such, even schools that avoid a hybrid model and do away with letter grades altogether will have to carefully educate teachers and parents alike to reconcile the two grading systems as separate entities, an understanding some focus group students articulated.

My final key finding was that ES measurements revealed the shift to SBRCs to have a very small impact on student motivation overall. Student

autonomy and competence were supported through the formative function of standards-based grades and report cards, while student autonomy was thwarted through seeing standards-based grades and report cards as an opaque extra step to discovering letter grade standing. It is possible that these two conflicting dynamics of the new mediational means resulted in potential gains in autonomous motivation being negated and that more time was needed for the reform to take hold.

9.2 Challenges with my Study

One challenge with my study was navigating my dual roles as both researcher and teacher at SIS. I took precautionary steps to lessen the power imbalances that existed between myself and participants. Nevertheless, my roles certainly impacted participant responses. In chapters two and four I recognized that all forms of human knowledge are contextual. Mediated action represents what Wertsch (1991) referred to as the irreducible tension between the agency of humans and the mediational means of our environment, means which are fluid, dynamic and shaped by contextual factors. In this sense, the data I collected was contextually specific just like that of every other social science investigation, and within this study I have tried to acknowledge and describe these contextual factors to the greatest extent I could.

Another challenge with my study was the length of time that data was collected to measure the impact of standards-based grades and report cards. Almost all data was collected over a two-year period at SIS: the final school year with letter grades and report cards, and the first year of standards-based grade and report card implementation. While two years represents a substantial amount of time for PhD data collection, one year is a relatively short period to measure the impact of assessment reform and it was likely an insufficient amount of time for the reform to take hold. This was evidenced by students' focus group comments, my interviews with senior leadership, and the school's

2016 self-study, all of which suggest that during this second year of data collection many teachers still needed time to come to terms with the new grading system. In hindsight, my data collection to measure the impact of the reform initiative likely ended too early. Future studies into standards-based grading reform, and FA reform in general, should take the lengthy amount of time assessment reform requires into consideration.

An additional challenge within my study was my lack of control in the implementation of standards-based reform at SIS. In chapter two I detailed the fidelity of implementation issues with FA reform, and in chapter three I discussed similar findings in initial research on standards-based reform. Student and senior leadership perceptions from my study revealed that there were varied ways teachers implemented SBGs and SBRCs within their classroom. Investigating teacher implementation was beyond the scope of my study, but the varied implementation in different classrooms surely impacted student perceptions of the reform.

Finally, in this thesis I have stressed the need for my case study to be viewed as an investigation into one particular case which was not intended to be representative of a broader population of schools or students, but whose findings were capable of supporting broader analytic generalizations (Yin, 2018). I have tried to provide an in-depth description of SIS and detail how the school's affluence and international population constitute a unique population and setting. In many ways this case was exceptional when compared to typical U.S. schools and my findings should be treated accordingly.

9.3 Areas of Future Research

In reflecting on my study and this larger thesis, I have eight recommendations for future research which I will now discuss in turn.

Research into grades and report cards should prioritize investigating student perspectives. It has been my experience during a career in education

that adults in schools all too frequently minimize or overlook the valuable insights students are capable of contributing regarding their own education. This becomes very problematic when considering that the intentions of mediational means rarely match the mediated action they produce (Wertsch, 1998). It is imperative that schools and researchers investigate student perspectives to better understand the ultimate impact of reform initiatives on student learning.

Parent education continues to be an important area for future research into standards-based reform. Because letter grades appear to be etched in stone within US society and seen as the gold standard in communicating student achievement, schools hoping to unlock the formative potential of standards-based grades and report cards must help parents understand this function as well. Research should investigate the most effective ways to do this and continue to identify obstacles to these ends. Guskey's work (E.g. Swan, Guskey, & Jung, 2014) represents a good starting point in this regard.

In this thesis have tried to communicate the vastly complex process which takes place in going from an initial set of standards to arriving at an end of term SBRC. These different processes include prioritizing standards and then aligning prioritized standards to rubrics, assignments and daily instructional lessons. UbD (Wiggins & McTighe, 2005) provides an essential framework for this process, but very little research has investigated how teachers implement it within the classroom. As I have argued, this alignment is an essential building block for FA/SA mixing, and it is one worth investigating further. Future studies should consider the ways in which teachers design and implement their standards-aligned units.

Findings from my study revealed the student perception that teachers implement standards-based grading practices in idiosyncratic ways, suggesting issues similar to variations found with teachers' implementation of letter grade

practices. Future studies should investigate how teachers implement standards-based grading practices with a particular focus on which practices are more supportive of enhancing students' formative use of assessment data.

Within this thesis I identified the specific FA process which I argued to take place with standards-based grades and report cards: student self-assessment with rubrics. While many students in my focus groups clearly described this process, it is immensely complex and future research should investigate why some students conceptualized this FA process with SBGs and others did not. This investigation should include, but not be limited to the following sub-topics: What meanings do students construct in the process of self-assessment with SBGs, SBRCs and rubrics? What are the most effective ways for teachers to support students in the process of using SBGs and SBRCs formatively through self-assessment with rubrics? How do parental influences impact students' formative use of SBGs and SBRCs?

My study considered the impact of varying forms of SBRCs with a particular emphasis on the hybrid model. My initial results suggest that the hybrid inhibits a formative use of SBRCs, but more research is needed into the impact of these different SBRC forms.

During the initial phases of my data analysis I believed that student perceptions of standards-based grades and report cards represented a dichotomy, with some students shifting to a formative use and some remaining rooted in viewing them through a traditional letter grade lens. Further analysis revealed greater complexity as there were some students who articulated aspects of both of these constructs of SBGs and SBRCs. This suggests a continuum of perceptions with some students lying in a middle ground between the polar opposites. Further research should attempt to develop this continuum through items on a questionnaire, measure where students are on this continuum, and

identify variables which facilitate student movement towards a formative use of grades and report cards.

Finally, Wertsch's theory on mediational means was a powerful lens to examine assessment practices in general, and grades and report cards more specifically. Wertsch's theory has been underutilized in educational research and I recommend more researchers apply it, particularly when investigating educational reform.

9.4 Entering New Illusions of Perspective

In part, many of these areas for future research I have recommended are essential because standards-based grades and report cards are a relatively young reform movement. Beyond filling the broad gaps in the research base which I first identified in the introduction of this thesis, the predictive potential of Wertsch's theory offers further motivation for this research: the illusion of perspective holds just as true for new mediational means as it does for old ones. There is hope that standards-based grades and report cards, as my study suggests, can unlock a formative potential in these traditionally summative means. However, new means often introduce new and unforeseen constraints of their own, as seen by students in my study who were left frustrated and disempowered by SBGs and SBRCs. Researchers and educators should continue to understand this reality in more complexity with the hope of identifying ways to shape standards-based grades and report cards further to allow more students access to their formative potential.

APPENDIX A: Focus Group Schedules

Fall 2016 Schedule

Ice Breaker

Physical grade sheets. What do you think these grades mean? What do they stand for/represent to you? Why would someone get this grade?

B. What would your parent think?
C. Teachers?
D. Peers?

Vignettes

1)Pat's report card is about to come out and Pat is extremely nervous. Why do you think Pat is nervous?

2)Jo is very happy when viewing the new report card. Why do you think Jo is so happy?

3)Sam has just looked at the report card online. It leaves Sam confused. Why do you think Sam is confused?

4)Billie go to look at the report card online at the end of the school day. Now Billie is heading home to view the report card with parents. What do you think is going through Billie's mind?

Questions

1. Tell me about your first memories of getting Report cards. How about the last couple of years?

2. What did your parents/teachers/peers think about them?

3. Does your report card represent you?

4. What role do report cards play in the process of learning?

5. Do you feel that you have control over the grades you get?

Winter 2017 Focus Group Schedule

I)Vignettes

1. Alex's report card is about to come out. Does Alex have control over the grades that come on the report card?

Follow up – Who holds the power/control with this report card that Alex gets?

2.Charlie feels like having to get good grades on the report card is really piling on a lot of pressure and stress. What effect does this have on Charlie? Do you think there are ways Charlie will try to resist all of this?

Follow up – Do you think Charlie's parents and teachers try to help Charlie with the constraints of the report card? How?

3.Jordan's report card just came out and Jordan's parents want to talk about it. How do you think this talk will go?

Follow up - Why do you think that Jordan's parents care about grades so much? How does Jordan feel about this?

-The next day, Jordan's teachers also want to talk about the report card. How do you think this talk will go?

-At the end of the day, Jordan gets to talk with friends about the report card. How do you think this talk will go?

Follow up – After all this, do you think grades/report cards should hold this significance/importance in Jordan's life?

4. Jamie's teachers are trying to move from just using letter grades to using a standards-based approach with specific standards and Extends, Meets, Progressing, and Beginning. Do you think this shift is going to help Jamie's learning?

Follow up – Will it help the way Jamie feels towards school?

II)Direct Questions, if time:

1. How are you supposed to think about report cards? How do you really think about them?

What message does a grade send on the surface? Beneath the surface?

2.Do you think the report card reflects your learning? Does it reflect you?

3.Do students sometimes play the game of school to get decent grades and get teachers and parents off their back, even if their grade in the end doesn't reflect their learning? How is it possible to play this game?

5.How do RC's impact the way you think about your learning?

Fall 2017 Focus Group Schedule

Ice Breaker:

1.Letter grade read outs with meaning, then sb grades. Say what comes to mind.

Vignettes:

2. Nicky just got back a project and it has a letter grade on it. What do you think Nicky is thinking about?

Possible Follow-up: Do you think the letter grade will help Nicky's learning going forward?

3. Alex looks at the SB marks on a rubric after a project. What is going through Alex's head?

Possible Follow-up: : Do you think the SB Mark will help Alex's learning going forward?

5. Izzy wants to get the highest letter grade possible on a big assessment for class. Why do you think this is?

In a different class, Jamie wants to get the highest standards-based mark possible on an assessment. Why do you think this is? Is the reason any different?

6. Taylor just looked at the report card online. What do you think Taylor is thinking about? Would Taylor's thoughts change if the report card only had standards based marks on it?

7. Bobby's parents want sit down and talk about Bobby's letter grades. How do you think this talk will go?

Why do Bobby's parents care so much about grades?

8. Mel's parents want sit down and talk about and Mel's Standards-based marks. How do you think this talk will go?

9.Nat is not sure to about whether to share a letter grade with friends. Why is Nat not sure? Would this situation change for Nat if it wasn't a letter grade but a standards-based grade?

Direct Questions:

10. Is it possible for every student to get an A? Is it possible for every student to meet standards?

11. What is the purpose of report cards? Do you think schools should have them? Who do you think is the intended audience for report cards?

12.How do SB grades relate/connect to your learning? How do letter grades relate to your learning?

13.What role do report cards play in your learning? What role are they SUPPOSED to play?

14.What role do grades and report cards play in your future? Standards-based grades?

15. Have the meaning of grades changed as you've gotten older?

16.How do society and the media affect the way we think about grades?

17. Does this report card allow a student to recognize strengths and areas for improvement? (Use old school one)

18.Why do students worry about their letter grades? Is it the same with SB marks?

Winter 2018 Focus Group Schedule

Icebreaker:
How do parents/teachers/students talk/think about Extending/Meeting/Progressing/Beginning?

Vignettes:

Nicky just got his first semester SBRC. What is going through Nicky's mind when looking at this SBRC?

Nicky's new SBRC just came out and when he gets home his parents want to talk. How will this talk go? What will Nicky's parents think about the SB marks?

Alex just found out that the school won't be using letter grades on assessments or report cards anymore, just teacher feedback and SB marks. What does Alex probably think of this? How will this impact Alex?
What motivates a student besides grades?

Izzy feels that letter grades can box people in and limit their learning. Why does Izzy feel this way? Will SB grades do the same thing?

Taylor just got back the rubric for a big assessment in one of her classes and the teacher circled where she was and also wrote down a number. What do you think Taylor is thinking about? Could the teacher have done this differently to help Taylor think about learning? Is there a better system?

Bobby's teacher asks the class to do a self assessment of their learning before turning in an assessment and getting final feedback from the teacher. Does a standards-based system (standards/rubrics,etc) help the class to have a better idea of where they stand? Will the self-assessments be accurate with the teacher's assessment?

Mel gets this report card. What do you think Mel is thinking about?

Nat's teacher has told the class that Nat's school switched to SBRC's so that there was more meaningful information with the learning achievement separated from behaviors. Before letter grades mixed everything together. Do you think that the teachers at Nat's school will keep behaviour separated from learning when they figure out the final marks for the new SBRC's?

Direct Questions:
If you could give one suggestion to the school and the teachers to improve the way grading and report cards work, what would it be? (Give think time)

Extending = A, Meeting = B, Progressing= C, Beginning =D and below. Do you agree? Why/why not?

What ways do SBRC's help you? What way do they hinder you?

Do students understand the way that teachers arrive at the final SB mark for the SBRC reporting standards? The final letter grade?

APPENDIX B: SRQ-A Questionnaire in BOS and Items by Subscale

Academic Motivation Survey - Winter 2017

0% complete

Page 1: Review of Step One

Before taking this survey, please make sure that you have completed the first step of viewing your report card online in Veracross.

If you have already completed this step, please proceed with the survey.

Next >

Powered by online surveys | copyright | survey contact details

Academic Motivation Survey - Winter 2017

14% complete

Page 2: Welcome

Welcome to the winter portion of the Academic Motivation Survey.

Some key points:

- Purpose - This survey will be used in Mr. Cameron's doctoral study, and it is also intended to provide feedback to the school about the impact of grade reports on the motivation students feel towards their academic efforts.
- Answers - There are no "correct" answers to this survey, so please answer honestly.
- Privacy - The Random Pin Number you are given is only used to confirm that a person takes both the pre-test (October) and post-test (January). **Your responses to this survey will never be matched to your name,** so in this sense, the survey is anonymous.
- This survey should take you about 5 minutes.

Thank you for taking the time to do this survey!

< Previous

Next >

Page 3: Student Information

What grade are you in?

- ○ 5
- ○ 6
- ○ 7
- ○ 8

Gender

- ○ Male
- ○ Female
- ○ Prefer to self describe
- ○ Prefer not to say

Your Random Pin Number should be projected on the classroom whiteboard. Please enter your Random Pin Number below:

Page 4: Why do I do my homework?

This part of the survey uses a table of questions, view as separate questions instead?

Why do I do my homework?

	Very true	Sort of true	Not very true	Not at all true
1.Because I want the teacher to think I'm a good student.	☐	☐	☐	☐
2.Because I'll get in trouble if I don't.	☐	☐	☐	☐
3.Because it's fun.	☐	☐	☐	☐
4.Because I will feel bad about myself if I don't.	☐	☐	☐	☐
5. Because I want to understand the subject.	☐	☐	☐	☐
6.Because that's what I'm supposed to do.	☐	☐	☐	☐
7.Because I enjoy doing my homework.	☐	☐	☐	☐
8.Because it's important to me to do my homework.	☐	☐	☐	☐

< Previous Next >

Page 5: Why do I do my work in class?

This part of the survey uses a table of questions, view as separate questions instead?

Why do I do my work in class?

	Very true	Sort of true	Not very true	Not at all true
9.So that the teacher won't get mad at me.	☐	☐	☐	☐
10.Because I want the teacher to think I'm a good student.	☐	☐	☐	☐
11.Because I want to learn new things.	☐	☐	☐	☐
12.Because I'll be ashamed of myself if it didn't get done.	☐	☐	☐	☐
13.Because it's fun.	☐	☐	☐	☐
14.Because that's the rule.	☐	☐	☐	☐
15.Because I enjoy doing my classwork.	☐	☐	☐	☐
16.Because it's important to me to work on my classwork.	☐	☐	☐	☐

‹ Previous Next ›

Page 6: Why do I try to answer hard questions in class?

This part of the survey uses a table of questions, view as separate questions instead?

Why do I try to answer hard questions in class?

	Very true	Sort of true	Not very true	Not at all true
17.Because I want the other students to think I'm smart.	☐	☐	☐	☐
18.Because I feel ashamed of myself when I don't try.	☐	☐	☐	☐
19.Because I enjoy answering hard questions.	☐	☐	☐	☐
20.Because that's what I'm supposed to do.	☐	☐	☐	☐
21.To find out if I'm right or wrong.	☐	☐	☐	☐
22.Because it's fun to answer hard questions.	☐	☐	☐	☐
23.Because it's important to me to try to answer hard questions in class.	☐	☐	☐	☐
24.Because I want the teacher to say nice things about me.	☐	☐	☐	☐

‹ Previous Next ›

Page 7: Why do I try to do well in school?

This part of the survey uses a table of questions, view as separate questions instead?

Why do I try to do well in school?

	Very true	Sort of true	Not very true	Not at all true
25. Because that's what I'm supposed to do.	☐	☐	☐	☐
26. So my teachers will think I'm a good student.	☐	☐	☐	☐
27. Because I enjoy doing my school work well.	☐	☐	☐	☐
28. Because I will get in trouble if I don't do well.	☐	☐	☐	☐
29. Because I'll feel really bad about myself if I don't do well.	☐	☐	☐	☐
30. Because it's important to me to try to do well in school.	☐	☐	☐	☐
31. Because I will feel really proud of myself if I do well.	☐	☐	☐	☐
32. Because I might get a reward if I do well.	☐	☐	☐	☐

Academic Motivation Survey - Winter 2017

100% complete

Thank You!

Your time and honesty are much appreciated. Have a fantastic day!

Powered by online surveys | copyright | survey contact details

SRQ-A Items by OIT Subscale

Subscale	SRQ-A Items
External Regulation	2, 6, 9, 14, 20, 24, 25, 28, 32
Introjected Regulation	1, 4, 10, 12, 17, 18, 26, 29, 31
Identified Regulation	5, 8, 11, 16, 21, 23, 30
Intrinsic Motivation	3, 7, 13, 15, 19, 22, 27

APPENDIX C: Questionnaire Email to Students

Fall 2017 Email

Dear Middle School Students,

Thanks for taking 5 minutes to take this survey on academic motivation.

This survey should be completed independently, and the explanations are on the first page when you follow the link.

Here is the link: https://durham.onlinesurveys.ac.uk/academic-motivation-survey-fall-2017

Thanks for your time!

Mr. Cameron

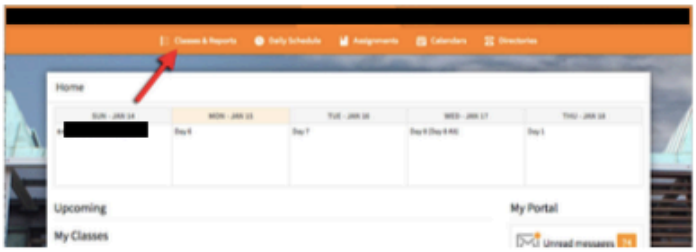
Winter 2018 Email

Dear Student,

This email has all the links you need to complete the final Academic Motivation Survey. The purpose of this survey is to provide feedback to the school about the impact of report cards on the motivation you feel towards academics.

Step One - View Your Report Card
Please take a couple of minutes to look over your first semester report card. You should do this even if you have previously viewed your report card. Here is the link to the Veracross portal: https://portals.veracross.eu/asl/login.

Once you login, on the menu bar at the top, click on "Classes & Reports" link (indicated by the arrow).



Next, click on the link to access the Semester 1 report cards (indicated by the arrow).

Step Two - Complete the Winter Academic Motivation Survey

After looking over your report card, please click on the following link to take the survey: https://durham.onlinesurveys.ac.uk/academic-motivation-survey-winter-2018-2

Thank you for your time,

Mr. Cameron

APPENDIX D: Winter 2017 Grade 5 Parent Email

Hi Grade 5 parents-

If you would, please see the following email below from my colleague, Latham Cameron. It is in reference to a survey the Grade 5 students will be taking in advisory.

Many thanks, and please let me know if you have any questions.

Take care,
[redacted]

_____

Dear Grade 5 Parents,

As a reminder of an earlier [redacted] entry, in advisory this Friday, Grade 5 students will have a chance to fill out a survey looking at the impact of grade reports on student motivation. Survey results will be used as feedback for Latham Cameron's (Grade 7 social studies) doctoral studies and for the Middle School. The procedure for this survey involves students first viewing their semester report card prior to taking the survey. If you do not want your child to participate, please contact me (latham_cameron@[redacted]) to let me know by Thursday afternoon. Also, please don't hesitate to let me know if you have any questions.

Thanks,

Latham Cameron

APPENDIX E: Senior Leadership Interview Guide

1. What were the larger objectives and purpose of standards reform at SIS?

2. What were the key steps in implementing this reform?

3. What issues arose during implementation?
   -Could you talk about the school's decision to use a hybrid SBRC?
   -Could you talk about the issue of parent education (steps taken, where it stands, etc)?
   -Could you talk about the issue of teacher education (steps taken, where it stands, etc)?

4. Now that were are almost two years in to standards-based grades and report cards, what are the positive outcomes you have noticed? Negative outcomes?

5. What are the next steps in the process for the school?

APPENDIX F: Questionnaire Opt-Out Parent Communication

SURVEY ON STUDENTS' MOTIVATION

In advisory this October and February, middle school students have a chance to fill out a survey on student motivation, as the middle school implements the use of standards in the curriculum. Responses will remain confidential, and survey results will be used as feedback for the Middle School, and in social studies teacher Latham Cameron's doctoral study through Durham University. If you do not want your child to participate in this survey, contact Latham Cameron by Thursday, 29 September. Questions to Latham or Director of Teaching and Learning.

Durham University
School of Education

1 November 2016

Participant Information Sheet

Dear Student:

You are invited to take part in a research study of how students make meaning of report cards. Please read this form carefully and ask any questions you may have before agreeing to be in the study.

The study is conducted by Latham Cameron ([redacted], Grade 7 social studies) as part of his PhD studies at Durham University. This research project is supervised by Jonathan Tummons (jonathan.tummons@durham.ac.uk) from the School of Education at Durham University, and [redacted]), the Director of Teaching and Learning at [redacted].

The purpose of this study is to see how students make meaning from report cards, and how [redacted] implementation of a standards-based curriculum impacts student motivation and student learning.

If you agree to be in this study, you will be asked to partake in some focus group discussions, which will take place at lunch for 30-40 minutes. These focus group discussions will involve you and some of your peers discussing the way you view report cards, and the impact they have on your life.

Your participation in this study will take two 30-40 minute lunchtime sessions .

You are free to decide whether or not to participate. If you decide to participate, you are free to withdraw at any time without any negative consequences for you.

All responses you give or other data collected will be kept confidential. The records of this study will be kept secure and private. All files containing any information you give are password protected. In any research report that may be published, no information will be included that will make it possible to identify you individually. There will be no way to connect your name to your responses at any time during or after the study. There will be an audio recording of the focus groups. This recording will be erased once the discussions are transcribed. When transcribing, all names will be removed to protect your anonymity.

If you have any questions, requests or concerns regarding this research, please contact me via email at [redacted] or come and see me at my classroom in [redacted].

This study has been reviewed and approved by the School of Education Ethics Sub-Committee at Durham University (date of approval: 08/06/16), and [redacted] Director of Teaching and Learning.

Thanks,
Latham Cameron

Leazes Road
Durham City, DH1 1TA

Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311
www.durham.ac.uk
Durham University is the trading name of the University of Durham

Durham University
School of Education

Shaped by the past, creating the future

Declaration of Assent

- I agree to participate in this study, the purpose of which is to see how students make meaning from report cards, and how [redacted] implementation of a standards-based curriculum impacts student motivation and student learning.
- I have read the participant information sheet and understand the information provided.
- I have been informed that I may decline to answer any questions or withdraw from the study without penalty of any kind.
- I have been informed that data collection will involve the use of recording devices.
- I have been informed that all of my responses will be kept confidential and secure, and that I will not be identified in any report or other publication resulting from this research.
- I have been informed that the investigator will answer any questions regarding the study and its procedures. Latham Cameron, School of Education, Durham University can be contacted via email: latham.cameron@durham.ac.uk or telephone: [redacted]
- I will be provided with a copy of this form for my records.

Any concerns about this study should be addressed to the School of Education Ethics Sub-Committee, Durham University via email to ed.ethics@durham.ac.uk.

_____

Date Participant Name (please print) Participant Signature

I certify that I have presented the above information to the participant and secured his or her consent.

_____

Date Signature of Investigator

Leazes Road
Durham City, DH1 1TA

Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311
www.durham.ac.uk
Durham University is the trading name of the University of Durham

Durham University
School of Education

Shaped by the past, creating the future

1 November 2016

Parent Information Sheet

Dear Parent/Guardian:

Your son/daughter is invited to take part in a research study of how students make meaning of report cards. Please read this form carefully and ask any questions you may have before agreeing for them to be in the study.

The study is conducted by Latham Cameron ([redacted] teacher, Grade 7 social studies) as part of his PhD studies at Durham University. This research project is supervised by Jonathan Tummons (jonathan.tummons@durham.ac.uk) from the School of Education at Durham University, and [redacted]), the Director of Teaching and Learning at [redacted].

The purpose of this study is to see how students make meaning from report cards, and how [redacted] implementation of a standards-based curriculum impacts student motivation and student learning.

If you agree for your son/daughter to be in the pilot of this study, they will be asked to partake in two 35-minute group discussions. These focus group discussions will take place at lunch, and the students will be asked about how they view report cards, and the impact they have on the students' lives.

Your child's participation in this study will take approximately 70 minutes.

You and your child are free to decide whether or not to participate. If you decide to participate, your child is free to withdraw at any time without any negative consequences.

All responses your child gives or other data collected will be kept confidential. The records of this study will be kept secure and private. All files containing any information they give are password protected. In any research report that may be published, no information will be included that will make it possible to identify your child individually. There will be no way to connect their name to their responses at any time during or after the study. There will be an audio recording of the focus groups. This recording will be erased once the focus groups are transcribed. When transcribing, all names will be removed to protect your child's anonymity.

If you have any questions, requests or concerns regarding this research, please contact me via email at [redacted].

This study has been reviewed and approved by the School of Education Ethics Sub-Committee at Durham University (date of approval: 08/06/16), and [redacted] Director of Teaching and Learning.

Thanks,
Latham Cameron

Leazes Road
Durham City, DH1 1TA

Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311
www.durham.ac.uk
Durham University is the trading name of the University of Durham

Parental Declaration of Informed Consent

- I agree to participate in this study, the purpose of which is to see how students make meaning from report cards, and how [redacted] implementation of a standards-based curriculum impacts student motivation and student learning.
- I have read the participant information sheet and understand the information provided.
- I have been informed that my child may decline to answer any questions or withdraw from the study without penalty of any kind.
- I have been informed that data collection will involve the use of recording devices.
- I have been informed that all of my child's responses will be kept confidential and secure, and that they will not be identified in any report or other publication resulting from this research.
- I have been informed that the investigator will answer any questions regarding the study and its procedures. Latham Cameron, School of Education, Durham University can be contacted via email[redacted]
- I will be provided with a copy of this form for my records.

Any concerns about this study should be addressed to the School of Education Ethics Sub-Committee, Durham University via email to ed.ethics@durham.ac.uk.

_____

Date Parent Name (please print) Parent Signature

_____

Name of Son/Daughter

I certify that I have presented the above information to the parent and secured his or her consent.

_____

Date Signature of Investigator

Leazes Road
Durham City, DH1 1TA

Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311
www.durham.ac.uk
Durham University is the trading name of the University of Durham

APPENDIX H: Rationale for Interview Guide and Codes

Code	Rationale
Hybrid	Given the complex issue of the hybrid report card and the tension students described in focus groups between letter grades and standards-based grades, I hoped to better understand the rationale for the school's decision to use this SBRC form.
Implementation Steps	Standards-based grade and report card reform is a complicated process. I hoped to detail the steps SIS had taken during this process.
Implementation Missteps	Because this reform is complicated, I wanted to hear more about what senior leadership would do differently if they could start over, in the hopes that this would also help to understand what had taken place during the course of the reform.
Parent Education	Because students referenced parent influence so heavily in focus groups, with many feeling parents were uneducated on the reform, I wanted to explore senior leaderships' perspectives on this topic.
Teacher Education	Students in focus groups discussed the varied ways they perceived that teachers had implemented standards-based grades. I hoped to gain an additional perspective on this issue.

APPENDIX I: Focus Group Code Revisions

Code	Explanation
Family Influence	Two family member groupings– parental influence and sibling influence – became clear while coding the first focus group meetings, thus *Family Influence* was replaced with *Parent Influence* and *Peer Influence*.
Competition	*Competition* was added when coding the first meetings of Year One focus groups revealed a theme of competition between participants and their peers and siblings.
Dishonesty	During early coding sessions, I noticed a repeated theme that grade-related pressures led students to cheat and lie. I added the code, *Cheat and Lie*, but eventually settled on *Dishonesty* as a slightly broader term.
Society and Media	Beyond just *Media*, during coding round two the larger societal influences in how students thought about grades arose leading to broadening the code to *Society and Media*.
Grade Formulation	*Grade Meaning* was not specific enough and had too much overlap with the *influence* codes. It was changed to *Grade Formulation* to specify how students thought of the key factors leading to their final grade.
Feedback	SBRCs are theorized to function as formative feedback. *Feedback* was added because during the second focus group sessions, students referred to the clarity of feedback from standards-based grades.
Motivation	The theme of *Motivation* became apparent during round two of coding and was added in the hopes of supporting data generated through the questionnaire.

APPENDIX J: Focus Group Code Definitions

Competition – References to an individual's standing in comparison to that of his/her peers or siblings, including associations of grade meaning with norm-referencing.

Dishonesty – All instances where students are, or consider being dishonest in activities related to grades.

Feedback – Information students receive from grades and report cards that identifies where they are at and what they need to work on going forward. Whereas *Grade Formulation* is looking back at what went into the formation of the grade symbol, *Feedback* refers to the messages sent of current standing and future steps to take. This code also includes information received from grades about learning progress towards a standard. Contrast with Grade Opaqueness below.

Grade Formulation – References from students regarding how their grade is formulated, and what goes in to the making of that grade.

> **Behaviour** – References to the teacher counting behaviour in the grade, or the recognition that behaviour (e.g., effort) led to them receiving the grade. Not to be confused with *Parent Influence* and *Feedback* telling kids to put in more effort as they move forward.
>
> **Learning** – References to student learning and understanding of the curricula resulting in the grade she/he received.
>
> **Teacher Choice** – Could also be thought of as teacher control or teacher judgment. References to final grade resulting from teacher decisions or judgments that were out of the student's control. Includes discrepancies in teacher grading methods, often identified by students as a teacher being a "harsh" grader.

Grade Opaqueness – Any reference to not understanding why the student received a grade, not understanding what the grade represents, or the recognition that grade does not reflect her/his actual learning. This Opaqueness could be referencing both student and parent confusion about the grade meaning, and works to inhibit constructive *Feedback*.

Grades as Labels – Comments of students recognizing that the grade they receive identifies some (semi-) permanent part or characteristic of who they are. This may reference their abilities as a learner, but it could also be in reference to their future success. These *Labels* sometimes come in the form of direct titles (e.g., Straight-A Student)

Grades and Social-Emotional – Mentions of students' social-emotional state depending on, or being affected by the grades they receive. This code could also apply to the grade-related social-emotional state of parents and teachers. Whereas Grades as Labels has an element of permanency, this code refers to an emotional state in the moment of thinking of and interacting with the concept of grades.

Letter Grade – References to letter grades of an A-F marking system. Not specific to letter grade report cards, but could include this topic.

Motivation – Either direct or indirect references to grades motivating or demotivating students. Also includes key aspects related to Self-Determination Theory, such as references to rewards and punishments with grades.

Parent Influence (on the sense that students make of grades/report cards) – Any reference to interactions between students and parents or references to how parents impact the sense students make of grades.

Peer Influence – See parent influence above.

Sibling Influence – See parent influence above.

Self Influence – Instances where students refer to expectations and behaviours related to grades and report cards as their own.

Society and Media – References to how the larger society and media frame grades and report cards, with direct or implied influence on students' sense of grades.

Standards-based – References to standards-based assessment and grading. Not specific to standards-based report cards, but could include this topic.

Teacher Influence – See *Parent Influence*. Potential overlap with *Grade Formulation – Teacher Choice*. This code should be about how teachers talk about grades and use grades in the classroom with students.

The Future – Any reference to the role report cards and grades play in a student's future life.

Bibliography

Allen, J. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House*, *78*(5), 218–223.

Andrade, H. L. (2010a). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). London: Routledge.

Andrade, H. L. (2010b). Summing up and moving forward. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 344–351). London: Routledge.

Andrade, H. L., & Brown, G. T. L. (2016). Student self-assessment in the classroom. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 319–334). New York: Routledge.

Andrade, H. L., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research, and Evaluation*, *10*(3).

Andrade, H. L., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice*, *17*(2), 199–214.

Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, *27*(2), 3–13.

Assessment Reform Group. (2002). Assessment for Learning: 10 principles. Retrieved July 20, 2019, from http://www.assessment-reform-group.org.uk

Bailey, J., & McTighe, J. (1996). Reporting achievement at the secondary level: What and how. In T. R. Guskey (Ed.), *Communicating student learning: 1996 ASCD yearbook* (pp. 119–140). Alexandria, VA: ASCD.

Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment

and learning: fields apart? *Assessment in Education: Principles, Policy & Practice*, *24*(3), 317–350.

Bakhurst, D. (1996). Social memory in Soviet thought. In H. Daniels (Ed.), *An introduction to Vygotsky* (pp. 196–218). London: Routledge.

Banjanovic, E., & Osborne, J. (2016). Confidence intervals for effect sizes: Applying bootstrap resampling. *Practical Assessment, Research & Evaluation*, *21*(5), 1–20.

Barbour, R. (2005). Making sense of focus groups. *Medical Education*, *39*(7), 742–750.

Barbour, R. (2007). *Doing focus groups*. London: Sage.

Baron, P. A. B. (2000). Consequential validity for high school grades: What is the meaning of grades for senders and receivers? *Paper presented at the Annual Meeting of the American Educational Research Association*. New Orleans, LA.

Barter, C., & Renold, E. (1999). The use of vignettes in qualitative research. *Social Research Update*, (25), 1–5.

Barter, C., & Renold, E. (2000). "I wanna tell you a story": Exploring the application of vignettes in qualitative research with children and young people. *International Journal of Social Research Methodology*, *3*(4), 307–323.

Belzile, J., & Öberg, G. (2012). Where to begin? Grappling with how to use participant interaction in focus group design. *Qualitative Research*, *12*(4), 459–472.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25.

Berk, R., & Freedman, D. A. (2001). Statistical assumptions as empirical commitments. *UCLA: Department of Statistics*. Retrieved from https://escholarship.org/uc/item/0zj8s368

Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education: Principles, Policy & Practice*, *5*(1), 103–

110.

Bird, S. M., Cox, D., Farewell, V. T., Goldstein, H., Holt, T., & Smith, P. C. (2005). Performance indicators: Good, bad, and ugly. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(1), 1–27.

Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 161–177.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box. *Phi Delta Kappan*, *86*(1), 8–21.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139–148.

Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81–100). London: Sage.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575.

Bloor, M., Frankland, J., Thomas, M., & Robson, K. (2001). *Focus groups in social research*. London: Sage.

Bonner, S. M., & Chen, P. P. (2009). Teacher candidates' perceptions about grading and constructivist teaching. *Educational Assessment*, *14*(2), 57–77.

Boston, C. (2003). High school report cards. *ERIC Digest*.

Bradbury-Jones, C., Taylor, J., & Herber, O. R. (2014). Vignette development and administration: A framework for protecting research participants. *International Journal of Social Research Methodology*, *17*(4), 427–440.

British Educational Research Association [BERA]. (2018). Ethical Guidlines for Educational Research. London: BERA. Retrieved from https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018

Brondani, M. A., MacEntee, M. I., Bryant, S. R., & O'Neill, B. (2008). Using written vignettes in focus groups among older adults to discuss oral health as a sensitive topic. *Qualitative Health Research*, *18*(8), 1145–1153.

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, *10*(1), 35–36.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, *30*(2), 123–142.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, *7*(4), 279–301.

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice*, *8*(2), 153–169.

Brookhart, S. M. (2004). *Grading*. Upper Saddle River, New Jersey: Pearson.

Brookhart, S. M. (2010). Mixing it up: Combining sources of classroom achievement information for formative and summative purposes. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 279–296). London: Routledge.

Brookhart, S. M. (2013a). Grading. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 257–271). London: Sage.

Brookhart, S. M. (2013b). *How to create and use rubrics for formative assessment and grading*. Alexandria, VA: ASCD.

Brookhart, S. M. (2013c). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 69–90.

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, *67*(3), 343–368.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., … Welsh, M. E. (2016). A century of grading research: Meaning and

value in the most common educational measure. *Review of Educational Research*, *86*(4), 803–848.

Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, *22*(4), 444–457.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks, California: Sage.

Bryman, A. (2007). The research question in social research: What is its role? *International Journal of Social Research Methodology*, *10*(1), 5–20.

Bryman, A. (2008). *Social research methods* (3rd ed.). New York, NY: Oxford University Press.

Caillaud, S., & Flick, U. (2017). Focus groups in triangulation contexts. In R. Barbour & D. L. Morgan (Eds.), *A new era in focus group research* (pp. 155–177). London: Palgrave Macmillan.

Campbell, C. (2012). Learning-centered grading practices. *Leadership*, *41*(5), 30–33.

Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3–17). London: Routledge.

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, *3*(2), 152–179.

Coe, R. (2002). It's the effect size, stupid. What effect size is and why it is important. *Paper Presented at Annual Conference of the British Educational Research Association*. University of Exeter, England. Retrieved from http://www.cem.org/attachments/ebe/ESguide.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.).

Hillsdale, N.J.: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). London: Routledge.

Cole, M. (1995). Socio-cultural-historical psychology: Some general remarks and a proposal for a new kind of cultural-genetic methodology. In J. V. Wertsch, P. del Rio, & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 187–214). Cambridge, MA: Cambridge University Press.

Cole, M. (1996). *Cultural psychology : A once and future discipline*. Cambridge, MA: Belknap Press of Harvard University Press.

Cole, M. (1997). Culture and cognitive science. *Outlines. Critical Practice Studies*, *5*(1), 3–15.

Cole, M., & Gajdamasschko, N. (2007). Vygotsky and culture. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 193–211). Cambridge, MA: Cambridge University Press.

Cole, M., & Wertsch, J. V. (1996). Beyond the individual-social antinomy in discussions of Piaget and Vygotsky. *Human Development*, *39*(5), 250–256.

Common Core State Standards Initiative. (2019). Retrieved November 3, 2019, from http://www.corestandards.org/ELA-Literacy/W/7/

Common Core State Standards Initiative. (2020). Retrieved March 8, 2020, from http://www.corestandards.org/

Common Goal Systems Inc. (2020). Online Gradebook - TeacherEase. Retrieved March 16, 2020, from https://www.common-goal.com/online-gradebook.aspx

Coolican, H. (2009). *Research methods and statistics in psychology* (5th ed.). London: Hodder Education.

Cox, K. B. (2011). Putting classroom grading on the table: A reform in progress. *American Secondary Education*, *40*(1), 67–87.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Boston: Pearson.

Creswell, J. W., & Garrett, A. (2008). The "movement" of mixed methods research and the role of educators. *South African Journal of Education*, *28*(3), 321–333.

Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, *12*(1), 53–72.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Daniels, H. (1996a). Introduction : Psychology in a social world. In H. Daniels (Ed.), *An Introduction to Vygotsky* (pp. 13–39). London: Routledge.

Daniels, H. (1996b). Psychology in a social world. In H. Daniels (Ed.), *An Introduction to Vygotsky* (pp. 1–27). London: Routledge.

Daniels, H. (2001). *Vygotsky and pedagogy*. London: Routledge.

Daniels, H. (2006). The 'social' in post-Vygotskian theory. *Theory & Psychology*, *16*(1), 37–49.

Daniels, H. (2015). Mediation: An exapansion of the socio-cultural gaze. *History of the Human Sciences*, *28*(2), 34–50.

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 149–166.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, *106*(6), 1047–1085.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268.

Deci, E. L., & Ryan, R. M. (2016). Optimizing students' motivation in the era of testing and pressure: A self-determination theory perspective. In W. C. Liu, J. C. K. Wang, & R. M. Ryan (Eds.), *Building autonomous learners* (pp. 9–29). Singapore: Springer.

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, *26*(3–4), 325–346.

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, *25*(4), 355–375.

Earl, L. M. (2003). *Assessment as learning : Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.

Ebel, R., & Frisbie, D. (1991). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice Hall.

Edwards, A. (2007). An interesting resemblance: Vygotsky, Mead, and American pragmatism. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 77–100). Cambridge, MA: Cambridge University Press.

Fendler, L., & Muzaffar, I. (2008). The history of the bell curve: Sorting and the idea of normal. *Educational Theory*, *58*(1), 63–82.

Finch, J. (1987). The vignette technique in survey research. *Sociology*, *21*(1), 105–114.

Finkelstein, I. (1913). *The marking system in theory and practice*. Baltimore:

Warwick & York.

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher*, *22*(4), 16–23.

Fisher, D., Frey, N., & Pumpian, I. (2011). No penalties for practice. *Educational Leadership*, *69*(3), 46–51.

Flaitz, J. (2011). Assessment for learning: US perspectives. In R. Berry & B. Adamson (Eds.), *Assessment reform in education. Education in the Asia-Pacific region: Issues, concerns and prospects, vol 14* (pp. 33–47). Dordrecht: Springer.

Flores, J. G., & Alonso, C. G. (1995). Using focus groups in educational research: Exploring teachers' perspectives on educational change. *Evaluation Review*, *19*(1), 84–101.

Flórez Petour, M. T. (2015). Systems, ideologies and history: A three-dimensional absence in the study of assessment reform processes. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 3–26.

Floyd, A., & Arthur, L. (2012). Researching from within: External and internal ethical engagement. *International Journal of Research and Method in Education*, *35*(2), 171–180.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*(2), 219–245.

Fontana, A., & Frey, J. H. (2003). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (2nd Ed., pp. 61–106). London: Sage.

Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice*, *12*(3), 23–30.

Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books, Inc.

Gibbs, B. G., Shafer, K., & Miles, A. (2017). Inferential statistics and the use of

administrative data in US educational research. *International Journal of Research & Method in Education*, *40*(2), 214–220.

Gobo, G. (2008). Re-conceptualizing generalization: Old issues in a new frame. In P. Alasuutari, L. Bickman, & J. Brannen (Eds.), *The SAGE handbook of social research methods* (pp. 193–213). London: Sage.

Gomm, R., Hammersley, M., & Foster, P. (2000). Case study and generalization. In R. Gomm, M. Hammersley, & P. Foster (Eds.), *Case study method: Key issues, key texts*. London: Sage.

Gorard, S. (2006). *Using everyday numbers effectively in research*. London: Continuum.

Gorard, S. (2013). *Research design: Creating robust approaches for the social sciences*. London: Sage.

Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research*. Maidenhead: Open University Press.

Great Schools Staff. (2016). Rethinking report cards. Retrieved September 13, 2019, from https://www.greatschools.org/gk/articles/rethinking-report-cards/

Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology*, *52*(5), 890–898.

Grolnick, W. S., Ryan, R. M., & Deci, E. L. (1991). Inner resources for school achievement: Motivational mediators of children's perceptions of their parents. *Journal of Educational Psychology*, *83*(4), 508–517.

Guay, F., & Vallerand, R. J. R. J. (1997). Social context, student's motivation, and academic achievement: Toward a process model. *Social Psychology of Education*, *1*(3), 211–233.

Guskey, T. R. (1996a). Introduction. In T. R. Guskey (Ed.), *Communicating student learning: 1996 ASCD yearbook* (pp. 1–5). Alexandria, VA: ASCD.

Guskey, T. R. (1996b). Reporting on student learning: Lessons from the past - prescriptions for the future. In T. R. Guskey (Ed.), *Communicating student learning: 1996 ASCD yearbook* (pp. 13–24). Alexandria, VA: ASCD.

Guskey, T. R. (2004). The communication challenge of standards-based reporting. *Phi Delta Kappan*, *86*(4), 326–329.

Guskey, T. R. (2009). Bound by tradition: Teachers' views of crucial grading and reporting issues. *Paper presented at the Annual Meeting of the American Educational Research Association*. San Francisco.

Guskey, T. R. (2015). *On your mark : Challenging the conventions of grading and reporting*. Bloomington, IN: Solution Tree Press.

Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press.

Guskey, T. R., & Jung, L. A. (2006). The challenges of standards-based grading. *Leadership Compass*, *4*(2), 1–4.

Guskey, T. R., Jung, L. A., & Swan, G. M. (2011). Grades that mean something. *Phi Delta Kappan*, *93*(2), 52–57.

Hargreaves, E. (2011). Teachers' feedback to pupils: "Like so many bottles thrown out to sea"? In R. Berry & B. Adamson (Eds.), *Assessment reform in education* (pp. 121–133). Dordrecht: Springer Netherlands.

Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp. 103–117). London: Sage.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, *4*(3), 365–379.

Hartley, H. O., & Sielken, R. L. (1975). A "super-population viewpoint" for finite population sampling. *Biometrics*, *31*(2), 411–422.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational*

Research, *77*(1), 81–112.

Hayward, L. (2015). Assessment is learning: The preposition vanishes. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 27–43.

Heflebower, T., Hoegh, J. K., & Warrick, P. B. (2014). *A school leader's guide to standards-based grading*. Bloomington, IN: Marzano Reasearch Laboratory.

Hogan, R. (2006). How to combine errors. Retrieved March 16, 2018, from http://www.met.rdg.ac.uk/~swrhgnrj/combining_errors.pdf

Hooper, J., & Cowell, R. (2014). Standards-based grading: History adjusted true score. *Educational Assessment*, *19*(1), 58–76.

Howitt, D., & Cramer, D. (2011). *Introduction to research methods in psychology* (3rd ed.). Harlow: Pearson.

Hughes, R., & Huby, M. (2002). The application of vignettes in social and nursing research. *Journal of Advanced Nursing*, *37*(4), 382–386.

Hughes, R., & Huby, M. (2004). The construction and interpretation of vignettes in social research. *Social Work & Social Sciences Review*, *11*(1), 36–51.

Iamarino, D. L. (2014). The benefits of standards-based grading: A critical evaluation of modern grading practices. *Current Issues in Education*, *17*(2), 1–12.

Jenkins, N., Bloor, M. J., Fischer, J., Berney, L., & Neale, J. (2010). Putting it in context: The use of vignettes in qualitative interviewing. *Qualitative Research*, *10*(2), 175–198.

Kirby, K. N., & Gerlanc, D. (2013). BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods*, *45*(5), 905–927.

Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health & Illness*, *16*(1), 103–121.

Kitzinger, J. (1995). Qualitative research: Introducing focus groups. *BMJ : British Medical Journal*, *311*, 299–302.

Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education: Principles, Policy & Practice*, *2*(2), 145–163.

Knight, M., & Cooper, R. (2019). Taking on a new grading system: The interconnected effects of standards-based grading on teaching, learning, assessment, and student behavior. *NASSP Bulletin*, *103*(1), 65–92.

Kozulin, A. (1996). The concept of activity in Soviet psychology: Vygotsky, his disciples and critics. In H. Daniels (Ed.), *An introduction to Vygotsky* (pp. 99–122). London: Routledge.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.

Lambert, D., & Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. London: Routledge.

Lau, A. M. S. (2016). 'Formative good, summative bad?' – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, *40*(4), 509–525.

Lincoln, Y. S., & Guba, E. G. (2000). The only generalization is: There is no generalization. In R. Gomm, M. Hammersley, & P. Foster (Eds.), *Case study method: Key issues, key texts* (pp. 27–44). London: Sage.

Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science*, *42*(4), 539–559.

Litosseliti, L. (2003). *Using focus groups in research*. London: Continuum.

Livingston, K., & Hutchinson, C. (2017). Developing teachers' capacities in assessment through career-long professional learning. *Assessment in Education: Principles, Policy & Practice*, *24*(2), 290–307.

Lysaght, Z., & O'Leary, M. (2013). An instrument to audit teachers' use of assessment for learning. *Irish Educational Studies*, *32*(2), 217–232.

Marzano, R. J. (1998). *Models of standards implementation: Implications for the classroom*. Aurora, CO. Retrieved from https://files.eric.ed.gov/fulltext/ED427088.pdf

Marzano, R. J. (2006). *Classroom assessment & grading that work*. Alexandria, VA: ASCD.

Marzano, R. J. (2010). *Formative assessment & standards-based grading*. Bloomington, IN: Solution Tree.

Marzano, R. J., & Haystead, M. W. (2008). *Making standards useful in the classroom*. Alexandria, VA: ASCD.

Marzano, R. J., Heflebower, T., Grift, G., & Warrick, P. (2016). *Collaborative teams that transform schools: The next step in PLCs*. Hawker Brownlow Education.

Marzano, R. J., & Kendall, J. S. (1998). *Implementing standards-based education*. Washington, D.C.: National Education Association.

Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the dimensions of learning model*. Alexandria, VA: ASCD.

McLaughlin, M. W., & Shepard, L. A. (1995). *Improving education through standards-based reform : A report by the National Academy of Education Panel on standards-based education reform*.

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20–32.

McMillan, J. H. (2010). The practical implications of educational aims and contexts for formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 41–58). London: Routledge.

McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, *95*(4), 203–213.

McMillan, J. H., & Turner, A. B. (2014). Understanding student voices about

assessment: Links to learning and motivation. *Paper presented at annual meeting of the American Educational Research Association*. Philidelphia, PA.

McMunn, N., Schenck, P., & McColskey, W. (2003). Standards-based assessment, grading, and reporting in classrooms: Can district training and support change teacher practice? *Paper presented at the annual meeting of the American Educational Research Association*. Chicago, IL.

Mercer, J. (2007). The challenges of insider research in educational institutions: Wielding a double-edged sword and resolving delicate dilemmas. *Oxford Review of Education*, *33*(1), 1–17.

Meshcheryakov, B. G. (2007). Terminology in L.S. Vygotsky's writings. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 155–177). Cambridge, MA: Cambridge University Press.

Miller, J. J. (2013). A better grading system: Standards-based, student-centered assessment. *The English Journal*, *103*(1), 111–118.

Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology*, *22*, 129–152.

Morgan, D. L. (2012). Focus Groups and social interaction. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (pp. 161–176). Thousand Oaks: Sage.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386.

Muñoz, B. M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education. *Phi Delta Kappan*, *96*(7), 64–68.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. *The Elementary School Journal*, *84*(2), 113–130.

National Science Teaching Association. (2020). Retrieved March 8, 2020, from https://www.nextgenscience.org/

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom. *School Field*, *7*(2), 133–144.

Nygren, L., & Oltedal, S. (2015). Constructing a vignette for qualitative comparative family research. *Journal of Comparative Social Work*, *10*(1), 1–14.

O'Connor, K. (1995). Guidelines for grading that support learning and student success. *NASSP Bulletin*, *79*(571), 91–101.

O'Connor, K. (2009). Reforming grading practices in secondary schools. *Principal's Research Review*, *4*(1), 1–7.

O'Connor, K. (2011). *A repair kit for grading: 15 fixes for broken grades* (2nd ed.). Boston, MA: Pearson.

O'Connor, K., & Wormeli, R. (2011). Reporting student learning. *Educational Leadership*, *69*(3), 40–44.

O'Dell, L., Crafter, S., Abreu, G. De, & Cline, T. (2012). The problem of interpretation in vignette methodology in research with young people. *Qualitative Research*, *12*(6), 702–714.

Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools*, *13*(1), 48–63.

Onwuegbuzie, A. J., & Leech, N. L. (2010). Generalization practices in qualitative research: A mixed methods case study. *Quality and Quantity*, *44*(5), 881–892.

Oosterhof, A. (2001). *Classroom applications of educational measurement* (3rd ed.). Upper Saddle River, New Jersey: Prentice-Hall.

Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*(4), 803–830.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, *9*, 129–144.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-

assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133–148.

Payne, G., & Williams, M. (2005). Generalization in qualitative research. *Sociology*, *39*(2), 295–314.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions Psychological and educational considerations* (pp. 47–87). Cambridge, MA: Cambridge University Press.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 85–102.

Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment*, *2*(1), 69–88.

Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies*, *47*(11), 1451–1458.

Ponce, O., & Pagán-Maldonado, N. (2015). Mixed methods research in education: Capturing the complexity of the profession. *International Journal of Educational Excellence*, *1*(1), 111–135.

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, *34*(1), 1–20.

Randall, J., & Engelhard, G. (2009a). Differences between teachers' grading practices in elementary and middle schools. *The Journal of Educational Research*, *102*(3), 175–186.

Randall, J., & Engelhard, G. (2009b). Examining teacher grades using rasch measurement theory. *Journal of Educational Measurement*, *46*(1), 1–18.

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, *26*(7), 1372–1380.

Reeve, J. (2009). Why teachers adopt a controlling motivating style toward

students and how they can become more autonomy supportive. *Educational Psychologist*, *44*(3), 159–175.

Reeves, D. (2004a). *Accountability for learning: How teachers and school leaders can take charge*. Alexandria, VA: ASCD.

Reeves, D. (2004b). The case against the zero. *Phi Delta Kappan*, *86*(4), 324–325.

Reeves, D. (2007). From the bell curve to the mountain: A new vision for achievement, assessment, and equity. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 1–12). Bloomington, IN : Solution Tree .

Reeves, D., Jung, L. A., & O'Connor, K. (2017). What's worth fighting against in grading? *Educational Leadership*, *74*(8), 42–45.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation*, *11*(10), 1–13.

Rowntree, D. (1981). *Statistics without tears: A primer for non-mathematicians*. New York: Charles Scribner's Sons.

Ryan, K. E., Gandha, T., Culbertson, M. J., & Carlson, C. (2014). Focus group evidence: Implications for design and analysis. *American Journal of Evaluation*, *35*(3), 328–345.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, *57*(5), 749–761.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78.

Ryan, R. M., & Deci, E. L. (2009). Promoting self-determined school engagement: Motivation, learning, and well-being. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 171–195). London: Routledge.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory : Basic psychological needs*

in motivation, development, and wellness. New York: Guilford Press.

Ryan, R. M., & Grolnick, W. S. (1986). Origins and pawns in the classroom: Self-report and projective assessments of individual differences in children's perceptions. *Journal of Personality and Social Psychology*, *50*(3), 550–558.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 77–84.

Scarlett, M. H. (2018). "Why did I get a C?": Communicating student performance using standards-based grading. *InSight: A Journal of Scholarly Teaching*, *13*, 59–75.

Schneider, J., & Hutt, E. (2013). Making the grade: A history of the A–F marking scheme. *Journal of Curriculum Studies*, *46*(2), 201–224.

Schoenberg, N. E., & Ravdal, H. (2000). Using vignettes in awareness and attitudinal research. *International Journal of Social Research Methodology*, *3*(1), 63–74.

Scribner, S. (1985). Vygotsky's uses of history. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (pp. 119–145). Cambridge, MA: Cambridge University Press.

Scriffiny, P. L. (2008). Seven reasons for standards-based grading. *Educational Leadership*, *66*(2), 70–74.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shadish, W. R., & Luellen, J. K. (2005). Quasi-experimental Designs. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science.* (pp. 1641–1644). Chichester: John Wiley & Sons.

Sheldon, K. M., Ryan, R. M., Deci, E. L., & Kasser, T. (2004). The independent

effects of goal contents and motives on well-being: It's both what you pursue and why you pursue it. *Personality and Social Psychology Bulletin*, *30*(4), 475–486.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14.

Small, M. L. (2009). `How many cases do I need?': On science and the logic of case selection in field-based research. *Ethnography*, *10*(1), 5–38.

Smithson, J. (2000). Using and analysing focus groups: Limitations and possibilities. *International Journal of Social Research Methodology*, *3*(2), 103–119.

Smithson, J. (2008). Focus groups. In P. Alasuutari, L. Bickman, & J. Brannen (Eds.), *The SAGE handbook of social research methods* (pp. 357–370). London: Sage.

Soenens, B., Sierens, E., Vansteenkiste, M., Dochy, F., & Goossens, L. (2012). Psychologically controlling teaching: Examining outcomes, antecedents, and mediators. *Journal of Educational Psychology*, *104*(1), 108–120.

Spillman, L. (2014). Mixed methods and the logic of qualitative inference. *Qualitative Sociology*, *37*(2), 189–205.

Stake, R. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 443–466). London: Sage.

Stewart, D., Shamdasani, P., & Rook, D. (2009). Group depth interviews: Focus group research. In L. Bickman & D. J. Rog (Eds.), *The SAGE handbook of applied social research methods* (2nd ed., pp. 589–616). London.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, *87*(4), 324–328.

Stiggins, R. (2010). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 233–250). London: Routledge.

Sun, Y., & Cheng, L. (2014). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, *21*(3), 326–343.

Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *18*(4), 433–449.

Swan, G. M., Guskey, T. R., & Jung, L. A. (2014). Parents' and teachers' perceptions of standards-based and traditional report cards. *Educational Assessment, Evaluation and Accountability*, *26*(3), 289–299.

Taras, M. (2005). Assessment: Summative and formative: Some theortetical reflections. *British Journal of Educational Studies*, *53*(4), 466–478.

Taras, M. (2009). Summative assessment: The missing link for formative assessment. *Journal of Further and Higher Education*, *33*(1), 57–69.

Tashakkori, A., & Teddlie, C. (2009). Integrating qualitative and quantitative approaches to research. In L. Bickman & D. J. Rog (Eds.), *The SAGE handbook of applied social research methods* (2nd ed., pp. 283–317). Thousand Oaks, California: Sage.

Teddlie, C., & Tashakkori, A. (2006). A General typology of research designs featuring mixed methods. *Research in the Schools*, *13*(1), 12–28.

Thomas, S., & Oldfather, P. (1997). Intrinsic motivations, literacy, and assessment practices: "That's my grade. That's me." *Educational Psychologist*, *32*(2), 107–123.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, *31*(3), 25–32.

Thygesen, L. C., & Ersbøll, A. K. (2014). When the entire population is the sample: Strengths and limitations in register-based epidemiology. *European Journal of Epidemiology*, *29*(8), 551–558.

Tocci, C. (2010). An immanent machine: Reconsidering grades, historical and present. *Educational Philosophy and Theory*, *42*(7), 762–778.

Torrance, H. (2012). Formative assessment at the crossroads: Conformative,

deformative and transformative assessment. *Oxford Review of Education*, *38*(3), 323–342.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, *27*(5), 615–631.

Van De Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, *25*(4), 40–48.

van der Veer, R. (2007). Vygotsky in context: 1900-1935. In J. V. Wertsch, H. Daniels, & M. Cole (Eds.), *The Cambridge companion to Vygotsky* (pp. 21–49). Cambridge, MA: Cambridge University Press.

Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist*, *41*(1), 19–31.

Vansteenkiste, M., Niemiec, C. P., & Soenens, B. (2010). The development of the five mini-theories of self-determination theory: An historical overview, emerging trends, and future directions. In T. Urdan & S. Karabenick (Eds.), *Advances in motivation and achievement, vol. 16: The decade ahead* (pp. 105–166). Emerald Publishing.

Vansteenkiste, M., Ryan, R. M., & Deci, E. L. (2008). Self-determination theory and the explanatory role of psychological needs in human well-being. In L. Bruni, F. Comim, & M. Pugno (Eds.), *Capabilities and happiness* (pp. 187–223). Oxford: Oxford University Press.

Venkatesh, V., Brown, S. A., & Sullivan, Y. W. (2016). Guidelines for conducting mixed-methods research: An extension and illustration. *Journal of the Association of Information Systems*, *17*(7), 435–495.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard university press.

Vygotsky, L. (1986). *Thought and language*. (A. Kozulin, Ed.). Cambridge, MA: The MIT Press.

Wasserstein, R. (2016). American statistical association releases statement on statistical significance and p-values: Provides principles to improve the conduct and interpretation of quantitative science. *American Statistical Association News*, 1–3.

Watts, H. K. (1996). Bridges freeze before roads. In T. R. Guskey (Ed.), *Communicating student learning: 1996 ASCD yearbook* (pp. 6–12). Alexandria, VA: ASCD.

Welsh, M. E., D'Agostino, J. V., & Kaniskan, B. (2013). Grading as a reform effort: Do standards-based grades converge with test scores? *Educational Measurement: Issues and Practice*, *32*(2), 26–36.

Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.

Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press.

Wertsch, J. V. (1994). The primacy of mediated action in sociocultural studies. *Mind, Culture, and Activity*, *1*(4), 202–208.

Wertsch, J. V. (1995). The need for action in sociocultural research. In J. V. Wertsch, P. del Rio, & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 56–74). Cambridge, MA: Cambridge University Press.

Wertsch, J. V. (1998). *Mind as action*. New York, NY: Oxford University Press.

Wertsch, J. V. (2007). Mediation. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 178–192). Cambridge: Cambridge University Press.

Wertsch, J. V., del Rio, P., & Alvarez, A. (1995). Sociocultural studies: History, action, and mediation. In J. V. Wertsch, P. del Rio, & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 1–34). Cambridge, MA: Cambridge

University Press.

Wertsch, J. V., & Rupert, L. J. (1993). The authority of cultural tools in a sociocultural approach to mediated agency. *Cognition and Instruction*, *11*(3–4), 227–239.

Wertsch, J. V., & Tulviste, P. (1996). L. S. Vygotsky and contemporary developmental psychology. In H. Daniels (Ed.), *An introduction to Vygotsky* (pp. 53–74). London: Routledge.

West, P. (1982). *Reproducing naturally occurring stories: Vignettes in survey research* Aberdeen: MRC Medical Sociology Unit.

White, P. (2009). *Developing research questions : A guide for social scientists*. New York: Palgrave Macmillan.

Wibeck, V., Dahlgren, M. A., & Öberg, G. (2007). Learning in focus groups. *Qualitative Research*, *7*(2), 249–267.

Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Alexandria, VA: ASCD.

Wiggins, G., & McTighe, J. (2011). *Understanding by design guide to creating high-quality units*. Alexandria, VA: ASCD.

Wiggins, G., & McTighe, J. (2012). *The understanding by design guide to advanced concepts in creating and reviewing units*. Alexandria, VA: ASCD.

Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18–40). London: Routledge.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Wiliam, D. (2017). Learning and assessment: A long and winding road?

Assessment in Education: Principles, Policy & Practice, *24*(3), 309–316.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–610.

Wilmot, S., & Ratcliffe, J. (2002). Principles of distributive justice used by members of the general public in the allocation of donor liver grafts for transplantation: A qualitative study. *Health Expectations*, *5*(3), 199–209.

Wolff, B., Knodel, J., & Sittitrai, W. (1993). Focus groups and surveys as complementary research methods: A case example. In D. L. Morgan (Ed.), *Successful Focus Groups: Advancing the State of the Art* (pp. 118–136). Thousand Oaks: Sage.

Wood, M. (2005). Bootstrapped confidence intervals as an approach to statistical inference. *Organizational Research Methods*, *8*(4), 454–470.

Wormeli, R. (2006). *Fair isn't always equal : Assessing and grading in the differentiated classroom*. Portland, Maine: Stenhouse.

Wright, D. B. (2003). Making friends with your data: Improving how statistics are conducted and reported. *British Journal of Educational Psychology*, *73*(1), 123–136.

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 140–160.

Yin, R. K. (2013). Validity and generalization in future case study evaluations. *Evaluation*, *19*(3), 321–332.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Thousand Oaks, California: Sage.

Zinchenko, V. P. (1985). Vygotsky's ideas about units for the analysis of mind. In J. V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives* (pp. 94–118). Cambridge: Cambridge University Press.

Zinchenko, V. P. (2007). Thought and word: The approaches of L.S. Vygotsky

and G.G. Shpet. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 212–245). Cambridge: Cambridge University Press.