Policy in Action in Standards-Based Assessment Reform:

An Exploratory Case Study Using Actor Network Theory

by

Nathan Robert Meisner

A Dissertation Presented in Partial Fulfillment
Of the Requirements for Degree
Doctor of Education

Approved November 2021 by the
Graduate Supervisory Committee:

Elisabeth Gee, Chair
Josephine Marsh
Robin Heslip

ARIZONA STATE UNIVERSITY

December 2021

ABSTRACT

A group of educators and administrators in an international school in Thailand collaborated for a year to devise and publish a policy document with aim to reform assessment practices its faculty. The group's beliefs derived from standards-based assessment leaders and its broad aim was to build a more coherent, accurate, and meaningful assessment system. Using Actor Network Theory as its theoretical perspective, this mixed-methods action research study explored the extent that the policy document changed the beliefs and practices of the faculty, the assessment materials within the system itself, and what other factors may also help account for any changes. The first finding is that the policy did lead to observable changes in practices of faculty traced in tests, quizzes, and the gradebooks that record assessments. A second finding is that the impact of the policy as an agent for change depends on the frequency that it is referenced.

ACKNOWLEDGEMENTS

To my wife, Susana Eirene Socorro Cardenas. You sacrificed so much for this. I am forever grateful to you for the sacrifice and for so, so much more. For my children, Mario Franco Meisner, Alessandro Craig Meisner, and whomever else may follow, I pray that you work hard and pursue your own dreams as I have had the privilege of doing through the opportunities gifted to me by my own parents. And may you look up to and inspire one another as my own brothers have me.

My chair, Elisabeth Gee, you have been a tremendous wealth of wisdom.

Thank you to my colleagues who were my think partners along the way, Casey Gaebler and Yujiro Fujiwara.

Thank you to my professional mentors who guide my journey in education, Steve Sibley, Todd Zukewich, Padmini Sankaran, Ocki Fernandez, Amit Khanna, Robin Heslip, Edward Wexler, Michael Emborsky, Derick des Vignes, Amy Greene, Madeleine Bystrom, Sara Kaste, and Jim O'Malley.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER

1 INTRODUCTION 1

National Context.....4

Local Context 11

Personal Context.....19

Research Problem.....24

Research Questions 25

2 THEORETICAL PERSPECTIVES AND RESEARCH GUIDING THE PROJECT.....26

Actor Network Theory 29

Groups 30

Objects as Mediators that Translate 31

Materials as Boundary Objects.....34

The Network: Action, Actors, Agencies, and Actants 36

3 METHODOLOGY 49

Setting.....50

Participants 51

Role of Researcher 51

CHAPTER Page

Innovation 53

Research Design 63

Method of Synthesis 77

Data Synthesis, Inferences, and Conclusions 80

4 RESULTS 81

Introduction 81

Data Analysis 82

Synthesis of Findings 107

5 DISCUSSION AND IMPLICATIONS 111

Discussion 113

Research Question Two 117

Implications 120

Limitations of This Study 122

Final Reflections 123

REFERENCES 126

APPENDIX

A EXCERPTS FROM IST'S HIGH SCHOOL STUDENT HANDBOOK 134

B IST ASSESSMENT BELIEFS AND GUIDING ACTIONS, 2019–2020 137

C SEMI-STRUCTURED INTERVIEW FOR VIDEO-CUED INTERVIEW 143

APPENDIX Page

D SEMI-STRUCTURED INTERVIEW FOR HEAD OF DEPARTMENT....146

E QUESTIONNAIRE QUESTIONS....149

F CONSTRUCT ALIGNMENT FOR HIGH SCHOOL FACULTY SURVEY....153

G OBSERVATIONS OF MATH AND SCIENCE COURSES, TIMESCALES AND TYPES OF ASSIGNMENTS ENTERED IN GRADEBOOKS....156

H EXAMPLE OF GRADEBOOK ENTRIES....170

I OBSERVATIONS AND INFERENCES FROM ASSESSMENT DOCUMENTS IN FS19 AND FS20....172

J INTERVIEW FINDINGS REPORT....178

K IRB APPROVAL....199

LIST OF TABLES

Table Page

1. Assessment Innovations: Old vs New....53

2. HS Grading 2019-2020: Percent, Level, Descriptor....60

3. HS Grading 2020-2021: Level, Descriptor....60

4. Reassessment Policy Guidelines....63

5. Data Collection Instruments and Their Use in Three Phases....64

6. Constructs, Definition, and Variables....65

7. *Video-cued Semi-Structured Interview Question 3*....68

8. Head of Department Semi-Structured Interview Question 7....69

9. Alignment of Research Questions, Constructs, Methods, and Supporting Literature....78

10. Participants in Pre and Post Survey Instrument by Department....84

11. HS Faculty #Beliefs and #Practices....85

12. HS Faculty Beliefs Pre and Post per Department....85

13. HS Faculty Practices Pre and Post per Department....86

14. Minor Assessment Count, Standard Deviation, and Percent Change per Faculty per Department....88

15. Minor Assessment Count, Standard Deviation, and Percent Change per Faculty per Department....89

16. Assessment Documents per Participant....94

17. Research Questions, Codes & Sub-Codes....98

18. Interview Text, Codes, Sub-Codes & Memo with Patterns and Links....100

LIST OF FIGURES

Figure Page

Figure 1. Network of ABGA to Departments 42

Figure 2. Network of ABGA, Assessment Materials, and Departments 44

Figure 3. Network of ABGA, Assessment Materials, T-III-1 and T-III-2 44

Figure 4. Network of ABGA, Assessment Materials, T-III-6 45

Figure 5. School Network and Individual Department Network Without ABGA 46

Figure 6. True-of-You Scale for HS Questionnaire 75

Figure 7. Data Collection, Synthesis, Conclusions 80

CHAPTER 1

INTRODUCTION

The overarching aim of this investigation is to explore how assessment policy change impacts a high school faculty's practices and beliefs. The way that schools measure, apply data from, and report student achievement experienced great changes in the first part of the 21st century. Driving this evolution were two questions: what to teach and how to assess whether it is learned. Near the turn of the 21st century the United States worked in earnest to develop a nationally recognized set of standards to serve as a base for each state's own standards. In this way despite shifts in demographics, work force needs, and political agendas, the question of what to teach is largely being addressed. Detailed, nationally aligned standards that outline what students should know and do by grade level in subjects that range from the arts to the sciences are so prevalent in the American education system that it is easy to take for granted that they exist at all. Adopting common standards, such as the National Governors Association and Council of Chief State School Officers' Common Core standards, has helped teachers share resources, ideas, and activities across school district, county, state, and national borders. It has also helped schools reduce the suitcase curriculum problem, where new teachers come in and departing teacher leave with their own visions of what to teach.

Adopting common standards helps policymakers measure students' progress across the United States and world. One way to do so is through standards-aligned standardized tests. With a consistent set of standards, students' growth can be compared between classrooms and schools around the world. In this area, standards-based reforms have yielded some tangible and controversial results. For example, data from

standardized tests highlight “the persistent test-score gaps in our schools include those between African Americans and Whites, between Latinos and Whites, between students in poverty and wealthier students, between children of parents with little formal education and with greater formal education, and between English learners and native English speaker” (Carter & Welner, 2013, p. 1). Through one lens, the tests lay bare significant, important opportunity gaps in American society from which to build plans to address. However, standardized tests have also led to graft and teaching to tests, contributing to the very problem the tests shed light on. The controversy of standards-aligned standardized tests and their impact on educational outcomes is beyond the scope of this investigation. However, the controversy itself demonstrates the power that standards-based assessment policy reform can have on the priorities of a community with many unintended consequences.

Beyond the debate of standardized tests, the focus of this paper lies in how schools seek to evolve the standards-based assessment practices of their faculty. With the question of what to teach in a high school answered by the adopted course standards, the question of how to assess and report students’ achievement remains. Should a school move away from reporting to letter grades converted from percentages? If the high school diploma is ultimately awarded to those that have shown proficiency at course standards, should students get multiple opportunities to demonstrate that proficiency? If course standards make no mention of any behavioral aspects (i.e., punctuality, raising hands, perseverance, passion, and commitment), do evaluations of students omit that information? The answer to all these questions in my local context is “yes.” A policy document outlining the intended beliefs and guiding practices of its faculty was adopted

in January 2020 with the intention of a full launch in the fall semester in August 2020. I was curious to explore the impact of its adoption. To what extent would it lead to changes in faculty's beliefs and practices?

The impact of policy adoption on faculty is the overarching theme of this investigation. If educational leaders look to change assessment practice, then old systems of measuring and reporting student achievement and the materials that comprise them must align with the new approach. Accordingly, a second theme guiding the research is the role that assessment *materials* (i.e., policy documents, gradebooks, tests, mark schemes, rubrics, homework assignments) play in navigating change. Finally, understanding that social change is multidimensional, there is a question as to what other influences helped navigate the change.

Actor-network theory (ANT), also known as sociology of translation, is a methodological framework that has been used to study a variety of different organizational changes. Within the past several decades, it was taken up by educationalists to investigate impacts of adoption of curriculum, standards, testing, and technological tools. A central tenet of ANT is the role of non-humans in shaping how people behave. Without suggesting objects play a larger role than humans, ANT assumes symmetry between them and that objects have their own agency. It is with this lens that the agency of assessment policy is explored within the assessment network, *infrastructure*, of a high school. More specifically, the research questions guiding this exploration are:

1. How does a new school assessment policy impact educators' assessment beliefs, practices, and infrastructure over one semester?

2. What are teachers' perceptions of factors that led to any changes in their beliefs and practices?

The rest of this chapter will describe the larger, national context of standards-based reforms as they evolved from the 19th through 21st century in the United States. Then, because the local context for this action research takes place in an international school that delivers an American and International Baccalaureate education, context surrounding international schools will be offered. This is followed by a more specific description of the actual local context and my personal role within it.

National Context

The question of consistent and meaningful educational practices goes to the heart of U.S. education history, dating back to the common school movement and carrying through the progressive era, the civil rights era, and the present. This section will explore the broad historical context to help place standardization in U.S. history starting with its roots, then explaining greater federal involvement, and finally exploring two different outcomes of standards-based reforms, high-stakes criterion-referenced tests, and standards-based curriculum, instruction, and assessment. A common theme that emerges throughout this discussion is the importance of measurement of student outcomes, be it letter grades or standardized tests.

The Roots of Standards-Based Reforms

The standards-based reform (SBR) movement has roots that reach as far back as the late 19th century. Thanks to the work of educators and educationalists of the common school movement, the role of education in American culture was elevated by the late 1800s (Dubrow, 2009). However, although the expectation to obtain an education was

normed, there still existed wide variance in purposes and styles. Should schools adopt a model of memorization of knowledge or development of critical thinking skills (National Education Association, 1894)? Are schools meant to track students to academic or vocational endeavors (National Education Association, 1894)? Amidst a rising tide of mass immigration at the turn of the 20th century, the National Education Council (NEC) aimed to address these questions to further direct American education's future (Dubrow, 2009). A group of ten educators chaired nine different conferences attended by ten additional invitees met to discuss these issues to chart a way forward; the group is referred to as The Committee of Ten (National Education Association, 1894). Although other larger concerns were addressed in the conference, the initiating concern was raised by Mr. James Baker, principal of Denver High Schools, who presented a report to NEC on the "uniformity of school programs and in requirements for admissions to college" (National Education Association, 1894, p. 3). Mr. Baker asked for standard, coherent practices in high schools that accurately reported students' achievement to improve admissions procedures. The group was mostly comprised of university professors and sought to identify common subjects that all students would study, regardless of whether or not they intended to attend university. The intent of the Committee of Ten was to offer recommendations for the standardization of education in America (Dubrow, 2009). In actuality, absent coherent standards and benchmarks of learning, it resulted in a very diverse set of practices per state and district informed by perceived expectations of college admissions criteria (Tyler, 1981). Its most immediate impact was to establish a common overarching objective of high schools to be college preparation. This isn't to say there was no dissent. For example, progressive educators seeking to increase graduation

rates during the economic depression of the 1930s embarked on the Eight-Year Study to reframe subject requirements with a vocational lens (Tyler, 1981). Despite their efforts, the common ends of college preparation coupled with disparate understanding of how to do so was largely the norm for the first half of the 20th century. This began to change when the new post–World War II order settled into a state of Cold War and a new player emerged in education.

Growth of Federal Involvement

The launch of Sputnik and ongoing anxieties in the climate of the Cold War era coupled with civil rights pressures and a war on poverty brought an increased presence of federal action in the form of offering financial assistance to schools who met strategic objectives. It was the National Education Defense act of 1958 and Elementary and Secondary Education Act of 1965 (ESEA) that set this new precedent (McGuinn, 2015). Though not initially a significant player, the influence of the federal government created a new dynamic in public education. A key characteristic was that a single vision of education was shared from the top-down to all states, setting the stage for a national standard on content and skills.

Standards-based reforms (SBR) themselves gained more significant ground in the early 1980s. In 1983 the National Commission on Excellence in Education published *A Nation at Risk*, which called course offerings in schools a "curricular smorgasborg" (Klein, 2002). One of the recommendations of the report was that there be regular standardized tests "administered as part of a nationwide (but not Federal) system of State and local standardized tests" (Klein, 2002). States and schools responded by raising graduation requirements and offering more advanced courses, but this did not impact

students' achievement (Hamilton, Stecher, Yuan, & Rand Corporation, 2008, p 17). At the time, the accepted reason was a lack of coherence and clarity as to what content and skills were meant to be taught (Hamilton et al., 2008, p 17). Some organizations responded to these shortcomings. For example, the National Council of Teachers of Mathematics (NCTM) set for itself the aim of writing standard goals for bands of grades, K-4, 5-8, and 9-12, which it accomplished and published in 1989 (Klein, 2002). At the turn of the decade, President George Bush pushed forward grants to organizations to write more standards. Bill Clinton carried this initiative forward into his presidency and the United States saw a rapid increase in the development of standards in a variety of different subjects through grant funds. By 2000 school systems across the United States could access standards and benchmarks for educators to develop targeted lesson plans (Hamilton et al., 2008). Perhaps the best evidence of the prevalence and power of adopted standards lies in a document published in 2007 by the Council of Chief State School Officers (CCSSO). Where there existed a smorgasbord of courses in the 1980s, in 2007 one had access to a catalog of courses of study, academic standards, and curriculum frameworks adopted by each state taken from each state's website (Council of Chief State School Officers, 2007). Take, for example, mathematics: In 2007, standards for mathematics existed in every state, with their formal adoption occurring at some point between 1996 and 2006 (Council of Chief State School Officers, 2007).

From smorgasbord to catalog was an important step toward a more coherent and standardized question of what to teach. In 2008, the CCSS took an even boldest step to change the landscape of education since the creation of the United States (Campbell-Whatley, Dunaway, Hancock, 2016). It published the Common Core State Standards

(CCSS), a set of national standards to serve as a national framework of student and teacher expectations. It was a landmark move aimed at finalizing the move from smorgasbord to catalog to condensed, coherent menu. It also sparked significant controversy. Echoing the times of the anti-federalist papers against ratification of the Constitution itself, some questioned the role of the federal government deciding what every American should learn. Another debate centered on the impact of high stakes standards-aligned, standardized testing on instructional and administrative practices. These unresolved uses led the Department of Education to publicly state 12 years after the CCSS was published that "Common Core is dead" (Greene, 2020). Though the CCSS lasted only 12 years, the process of adopting standards, debating their purpose, and using them in classrooms led to a great deal of collaboration and professional learning. It is from that process that a less socially and politically divisive outcome arose. Some schools and communities who had moved from standards-less to standards-referenced assessment and reporting explored continuing the journey toward standards-based. In a standards-referenced system, a student's achievement is reported relative to the performance standard for each course on the report card (Marzano, 2010, p. 18). A standards-based system is similar, except that a student moves to the next level (i.e., unit or grade level) if and when they have demonstrated proficiency with the standards (Marzano, 2010, *Formative Assessment & Standards-based grading,* p. 18). Implied in the latter is a different timescale and means to assess student achievement.

High-Stakes Criterion-Referenced Tests

A key motivation to adopt nationally recognized standards was the ability to more accurately measure and report students' achievement. The ability to do this would help

drive neoliberal, market-based strategies to achieve greater global competitiveness (Hargreaves & Fullan, 2009). Criterion-referenced tests are one means used to do so. "Criterion-referenced tests and assessments are designed to measure student performance against a fixed set of predetermined criteria or learning standards" (Great Schools Partnership, 2014). Unlike norm-referenced tests, which rank test takers relative to one another using a bell curve, students are assessed on their proficiency to a set of criteria or expected standards (Great Schools Partnership, 2014).

Criterion-referenced testing has increased in importance in the United States since the creation of the National Assessment of Educational Progress (NAEP) in 1964, an assessment created to increase output measures (i.e., student achievement) to educational policy analysis and decision making over input measures (i.e., per pupil expenditures, attendance, number of classrooms, teacher salaries, enrollment) (History and Innovation, 2020). As noted, this comes on the heels of the U.S.S.R.'s successful launch of Sputnik in 1957 and a perception that to win the Cold War the United States needed the most educated population. Also referenced before, a year later Lyndon Johnson pushed forward the ESEA in 1965, offering public funds to schools that qualified. In the 1990s, under the leadership of President Clinton, the Improving America's Schools Act (IASA) in 1994 included in its push toward development of standards the institution of voluntary standardized tests (Hamilton et al., 2008, p 10). In the 21st century, scores on criterion-referenced testing came to be the substantial qualifier for federal and state public funds through President Bush's No Child Left Behind (NCLB) Act of 2002 and President Obama's Every Student Succeeds Act (ESSA) of 2015. The competition for public funds via high stakes tests coupled with other neoliberal reforms led to several controversial

outcomes, including corruption, instructional practices geared solely at test-taking, loss of funding for courses considered ‘specials’ (i.e., art, music, and drama), school closures, and inequitable outcomes. Though beyond the scope of this investigation, several notable outcomes warrant mentioning. First, for better and for worse, the adoption of high-stakes, standardized, criterion-referenced tests are a major outcome of the standards-based reform movement. Second, tying standards to high stakes standardized testing inspired a very emotional and negative response from school communities, casting shade on standards-based reformers’ aims. Third, what remains unanswered is how the benefits of standards-based reforms, those of coherence, accuracy, and equity, can transfer to classroom practices via curriculum, instruction, and *assessment*.

Standards-Based Curriculum, Instruction, and Assessment

Despite differences in motives (i.e., economic global competitiveness or greater equity), at the core of all education reforms is the desire to improve outcomes for students. The movement away from norm-referenced to criterion-referenced assessment was meant to be able to hold all students to the same standardized bar. The coherent set of standards was in and of itself popular among educators (Hargreaves & Fullan, 2009, p. 135). With an eye on what to teach, instructional design linked to standards and benchmarks became a mainstay in schools and teacher education programs. For example, Grant Wiggins and Jay McTighe’s (2011) framework for developing units, Understanding by Design (UbD), begins with educators working backward from the standards and benchmarks they aim to teach, designing assessments and developing big understanding and essential questions that then drive instruction. As a gauge of its popular use, UbD is practiced by thousands of educators, at least 12 books have been

written on the subject, and countless peer-reviewed articles have been published (ASCD, 2020). This is a small demonstration of the power that standards-based reforms have had on reframing unit planning, including *assessment* and instructional practices.

With the evolution of SBR, some have reimagined how learning can be measured and reported. The aims of those who advocate for standards-based assessment reforms (SBAR) include pursuit of greater coherence and clarity (Hamilton et al., 2008, p. 17), equity (Feldman, 2019), accuracy (Schimmer et al., 2018; Feldman, 2019; Marzano, 2010), and focus on learning over achievement. Practices mentioned in relation to these themes include use of standards to improve formative process and to align it to summative outcomes, use of proficiency descriptors to standards as feedback over numerical scores or letter grades, adoption of self-assessment and self-regulation, formal policies on reassessment, separation of achievement from behavior in reporting, reduction of the scale of achievement scores, and discontinuing the practice of averaging marks for final reported scores (Feldman, 2019; Schimmer et al., 2018; Marzano, 2010; Guskey, 2010).

Local Context

International School Context

There may be at least 11,667 international schools operating a $51.7 billion revenue stream employing 564,000 staff and enrolling 5.97 million students world-wide (ISC Research, 2020a). While a mere drop in the bucket of the world's $86 trillion output as of 2018 (The World Bank, 2020), it is a tuition fee revenue stream that rivals the GDP of at least 117 countries in the world (The World Bank, 2017). Equally impressive is the fact that at least one exists in most countries in the world (ISC Research, 2020b).

Considering the geographic and financial scope, it is understandable that there exists a great range of types of international schools that serve different community needs and wants. Some are non-profit and others are not. Some are sponsored by national governments while others are part of multinational corporations. Some have histories that date back to just after World War I, such as the International School of Geneva and Yokohama International School (Hayden, 2006, p. 18), post–World War II and the initial age of oil exploration, such as Escuela Bella Vista in Maracaibo Venezuela (LinkedIn, 2020), and the opening of China to global markets, such as BASIS International School Shenzhen (2020).

The geopolitical and power relations of the history of international schools makes defining them controversial (Bates, 2011). Further, that there is such a variety of schools makes it difficult to precisely define (Phillips, & Schweisfurth, 2014). Additionally confusing the ability to offer a clear definition is the fact that there exists no single, authoritative body that can "grant the right to use the term 'international school' in a school's title" (Hayden, 2006, p. 16). To explore the local context of this investigation, the definition offered by the International Association of School Librarianship's (IASL) at a conference they held in 2009 will be adopted (Marshall, 2019). It is a definition that is comprehensive and cited in several publications referenced in sources common to my local context, such as TIE Online (Nagrath, 2011) and International Schools Search (International Schools Search, 2019). According to IASL, the following set of eight criteria to help define an international school:

1. Transferability of students' education across international schools
2. A moving population (higher than in national public schools)

3. Multinational and multilingual student body
4. An international curriculum (i.e. IB - DP, MYP, PYP)
5. International accreditation (e.g. CIS, IBO, North Eastern ASC, Western Ass. of Schools and colleges, etc.)
6. A transient and multinational teacher population
7. Non-selective student enrollment
8. Usually English or bi-lingual as the language of instruction (Marshall, 2019)

Not all of these criteria must be met for a school to be recognized as being international. Many of these points are generally true of most international schools and are used to understand the type of international school it is. In the section that follows, I will use these criteria to describe the local context of the school at the center of this investigation.

Local School Context as International School

The International School of Thailand (IST) is an international school in all aspects outlined by IASL except its overwhelmingly Thai national student body. IST is a private, non-profit, tuition-based school that offers an education that permits transferability of students to both other international schools as well as to domestic, national schools in the United States. This is because the school offers a program that is accredited by the Western Association of Schools and Colleges (WASC). Specifically, its curriculum is driven by American Education Reaches Out (AERO), “a project supported by the U.S. State Department’s Office of Overseas Schools (A/OPR/OS) and the Overseas Schools Advisory Council to assist schools in developing and implementing standards-based curricula” (American Education Reaches Out, 2020). The high school program

culminates in Advanced Placement (AP) and International Baccalaureate (IB) courses, the latter of which can be used to earn the IB Diploma. Other internationally recognized standardized tests used to help with transferability include the Measures of Academic Proficiency (MAP), Writing Assessment Program (WrAP), and the 6+1 Writing Traits. On average, students outperform national and world averages on all tests in all subjects across all grade levels.

Though my research is centered on the high school division, the whole school offers a preK through 12th grade program that has a capacity of 2000 students but as of 2019 enrolls 1440 students. The high school enrolls 370 students and its student body is mostly Thai (98%), with the remaining 2% comprised of Indian, Chinese, Korean, American, and Canadian students. Because this is a largely local population, the students are not transient. Typically, students who enroll in IST continue on in school to high school graduation, whether they join in kindergarten or ninth grade. The student population speaks at least two languages, English and Thai. English is the key language of instruction. However, Thai law stipulates that all students attend a Thai language course with Thai nationals and must pass a national standardized, Thai language exam. Apart from English and Thai, students are required to take at least two years of another foreign language to graduate—Mandarin, Japanese, Spanish, or French. Enrollment is not highly selective, but there is an entrance exam to place students and to gauge whether the school has the resources to support their needs. This includes both learning support (LS) services and English Language Development (ELD). Finally, the faculty consists of 165 certified members who are almost all expatriates. The majority of the faculty are from the

United States (60%), Thailand (20%), followed by Canada (14%), and completed by Australia, New Zealand, the United Kingdom, Venezuela, and Japan (6%).

School Context with SBAR

It is difficult to pinpoint precisely who, where, or when SBAR made its link to IST. There are at least three opportunities for the connection to be made: accreditation processes, professional organization memberships, and entering faculty and administration. The most direct and observable connection between the international school context and SBAR is made when schools pursue accreditation from governing boards. For IST, this occurs through WASC. A 2015 self-study offers a glimpse into changes that began some time ago. The self-study is organized into the following five topics: 1) Organization for Student Learning, 2) Curriculum, Instruction, and Assessment, 3) Support for Student Personal and Academic Growth, 4) Resource Management and Development, and 5) Thai Language and Culture Standards (SCHOOL, 2015). The second topic required investigation and reflection on what students learn, how they learn, and how assessment is used. The findings on assessment explained that the high school "uses common summative assessments in conjunction with rubrics that were designed with an eye on learning outcomes"(SCHOOL, 2015). The elementary and middle school use standards-based grading while the high school "uses a traditional grading system of percentage and letter grades," and "common assessments and rubrics contribute to a mutual understanding for teachers and students of the expectations for growth and achievement" (SCHOOL, 2015). The self-study was coordinated during the articulation and adoption of division-specific assessment manuals.

Beginning in 2011, each division, elementary, middle, and high school, compiled and/or updated their assessment manuals. Each section was required to clearly articulate its purpose for assessing students, establish transparent assessment policies, articulate the types of assessments used by each section, use of results, and other matters relevant to assessment. Besides providing clarity and consistency to assessment across the sections, the Assessment Manuals addressed grade inflation, set an expectation for frequent and varied formative and summative assessments, and established common language (SCHOOL, 2015).

Areas for growth included "procedures for the collection of student feedback on their learning experiences, reflections on their progress, and their level of engagement" as well as "open lines of communication between teachers and students to provide optimal learning experiences" (SCHOOL, 2015). Clearly IST had been engaged in the collaborative conversations and production of materials that guided the work of assessment. Based on these findings, the high school was already engaged with SBAR in 2015. It is unclear precisely why or when, but since the report five years ago, the assessment manuals in each division cannot be found (Assistant Principal, personal communication, 2020; Principal, personal communication, 2020; Curriculum Coordinator, personal communication, 2020).

To have a more coherent, aligned curriculum across the division, the high school took on the project of articulating and adopting a set of guaranteed standards for each subject and for every grade during the 2016-2017 academic year. As of the beginning of the 2019-2020 academic year, there was no specific document outlining how teachers were to assess, grade, and report student learning. Instead, a set of core beliefs,

definitions, and grade descriptors exist in the high school student handbook. Excerpts are provided in Appendix A. Guidelines are orally communicated to faculty and are articulated by faculty per course syllabi (Assistant Principal, personal communication, November 2019). The known guidelines are described next.

First, assessments are considered major, minor, or final. There is no clearly articulated definition of what constitutes major or minor. The assistant principal explained that examples of major assessments include end-of-unit tests and projects that evaluate students' performance on a large number of guaranteed standards (personal communication, November 2019). Minor assessments are those that occur throughout the unit and cover fewer standards, such as quizzes. Finally, there also exists an end-of-semester final exam, which is meant to cover all or most of the course standards taught during that semester. These three categories, major, minor, and final, make up 45%, 25%, and 30% of the grade, respectively.

Second, teachers are expected to set up and use these weights in their electronic gradebooks. No specific direction is offered teachers in terms of frequency of major or minor assessments in a unit or semester of study. Also, teachers are not required to use the specific titles of "major," "minor," or "final" in the gradebooks. They can also use various other features of the gradebooks, like the option to enter an assignment and label it as "no credit," which means that the score does not get included in the final grade. They may also use a variety of different labels, including but not limited to "homework," "quiz," "pop quiz," "vocab quiz," and "extra credit."

Third, every assignment entered into the gradebook is scored out of a number determined by the teacher, the total marks. The total mark is the total highest possible

score that a student could receive for the combined course assessments. Students' achievement is entered as a number divided by the total marks, which is automatically calculated as a percentage in one column and reported as a corresponding letter grade in the next.

Fourth, there is no requirement for teachers to identify which standard(s) are being assessed in their gradebooks. Some teachers do this some of the time. Instead, to ensure teachers' assessments align with guaranteed standards, they are expected to attach the assessment documents to their online curriculum maps stored in Atlas Rubicon, a digital curriculum mapping platform. According to the curriculum maps, all units designed over the course of the year target the guaranteed standards. Similarly, all assessments are meant to measure students' relative mastery of these standards. As such, grades that are presently reported are standards-referenced.

During the 2017-2018 academic year, the assistant principal, the IB coordinator, several heads of department, and some faculty engaged in discussions about revising the assessment practices of the school to reflect greater coherence, accuracy, and meaning. This group became what was called an Assessment Task Force (ATF) that met informally after school on Thursdays. During the 2018-2019 academic year, the high school division principal became the head of school for all divisions and a new high school principal joined the community. Conversations in the ATF continued in 2018-2019 under the leadership of the new principal with some momentum generated to start the 2019-2020 academic year with significant changes, namely doing away with percentages and adopting a new grading and reporting scale. In April 2019 the principal decided that the students, parents, and faculty were not prepared to take on the change (HS Principal,

personal communication). Further, the principal wanted to make sure that the entire faculty had opportunities to participate in discussions and have a voice in the decision making.

Although I had been in touch with the HS principal since January 2019, I officially joined the community in August 2019 as its High School Dean of Students. The following section will continue to describe the school context through my own lens and actions.

Personal Context

My Background

I have been in education for 14 years, 12 of which were as a social studies classroom teacher. I have both experienced and participated in leading assessment and grading reforms in three different schools. In my first five years as an educator, I served as a social studies teacher in an international school (IS1) in Venezuela from 2007-2012. The school was accredited by the Southern Association of Colleges Schools Council on Accreditation and School Improvement (SACSCASI). Its high school served a primarily Venezuelan-American student body numbering around 100-130 students, almost all of whom spoke English as a second language. Student graduation rates were 100% and almost all students went on to attend university in the United States. In that school I used Understanding by Design to create units, referencing McREL standards and benchmarks. McREL is a non-profit organization that created a compendium of academic standards in the late 1990s through early 2010s until the project lost its funding (McREL International, 2014). Summative assessments in my classes were largely essay-based, projects and presentations, and accounted for 65% of the grade. Homework and

classroom assignments accounted for roughly 15% of the grade. The remaining 20% of the grade was allocated to a semester exam. At the beginning of each year, the social studies department discussed these weights and agreed as a team to keep them. Other departments had different weighting categories. Each of our assignments had a set number of total points. Students' grades were determined by how many points were earned divided by the total number of points. There were twelve levels of achievement: A, A-, B+, B, B-, C+, C, C-, D+, D, D-, and F. Level A, B, C, and D were each divided into ten percentage points 100-90, 89-80, 79-70, and 69-60, respectively. Level F was comprised of 60 points, 59-0.

In August 2012, I moved to Abu Dhabi to continue to serve as a social studies teacher at another international school (IS2). This one was accredited by the Middle States Association of Colleges and Schools (MSA). Its high school serves an international student population of 360 students. Graduation rates were consistently 100% and students went on to attend universities in the United States, Europe, the UK, Asia, and Australia. During my first year at the school, the same assessment criteria and practices existed at the school. Departments determined the weights and categories. The social studies department's weights were the same as the school in Venezuela. In my second year in Abu Dhabi, a new high school principal with extensive academic and professional experience of reforming assessment and grading policies joined IS2. Midway through his first year, he began to speak with faculty about the meaning of assessment and worked with small, interested groups on reforming the assessment policy and practices. By the end of that year, a six-point plan was put into place for the following academic year. First, it adopted a policy of not using homework, formative

assessments, or behavioral approaches to learning (i.e., submission of work on time, participation in class, ability to work independently in class) in the calculation of grades. Only summative assessments would be used in determining the grade for a course. Second, summative assessments were deemed either minor or major with the guiding premise that a minor summative would cover less content and/or not require as complex thinking (i.e., not employ third- or fourth-level questions from Bloom's taxonomy). Third, grades no longer referenced percentage scales, from 0-100. Instead, they were reported solely as letter grades, from A to No Credit, on a scale of 1-12. Fourth, summative assessments would be recorded in the gradebook as minor summative assessments and major summative assessments, but teachers' professional judgment would be used to decide the final grades of students. No mathematical averaging was to formally take place with gradebooks. Fifth, students were not assigned an F for failure; instead they were reported as No Credit. Sixth, reporting progress on formative assessments occurred through four descriptors: limited proficiency, developing proficiency, proficient, and exemplary. These same descriptors were employed in rubrics in all departments and were used to report on summative assessments (i.e., A and A- were exemplary, B+, B, and B- were proficient, etc.).

The process of articulating these changes was more straightforward than putting them into practice. Faculty were very concerned about these changes and frequently asked each other and the principal the following questions: Will this lead to grade inflation? Will students complete their homework and take classroom assignments seriously if they aren't graded? How will I hold students accountable to turning work in on time if I can't deduct achievement levels from their final score? How will effort be

assessed and rewarded? Doesn't this system make the grading process more subjective and leave open the potential for bias? How will students know how well they did? How different will this really be?

It was in this context at IS2 that I began my doctorate program. In initial cycles of research it was clear that teachers had some very different opinions about the grading policies and even more disparate practices. I had not settled on a specific topic for my action research, but the problem of practice emerged. As of August 2018, I was keenly interested in exploring how an organization reconciles teachers' assessment practices with their beliefs, particularly while leadership aims to adopt more standards-based assessment reforms.

Personal Context and SBAR in the International School of Thailand

I joined IST as its high school dean of students for the 2019-2020 academic year. I was in touch with the HS principal since January 2019 and knew about the school's desire to pursue assessment reforms. When I arrived in August, I did some reconnaissance and found that the grading system resembled what I had experienced at IS1 and my first year at IS2, tet at IST, there was far greater variance in practices. Some teachers had adopted their own ways of full standards-based assessment with integrated reassessment practices. Others continued a very traditional practice that included homework and participation in grades. Many new faculty weren't sure exactly how to organize their gradebooks and by association their assessments. I learned about the Assessment Task Force (ATF) early from the assistant principal and found that they were pursuing similar assessment reforms that I had already lived and worked through at IS2. I resolved to join this group and help continue its work.

In the early stages of our work we came to a shared vision of assessment. To do so we used the team's working draft of a document titled "Assessment Beliefs and Guiding Practices" (see Appendix B). Next steps were identified and the group splintered into two. One group worked to prioritize policies to put into place for the 2020-2021 academic year. I joined this group and we settled on the following changes: 1) eliminate the diverse categories of assessments and only report the equivalent of major assessments as "summative assessments;" 2) formative work will be assessed and progress shared, but the evaluation of their achievement in the formative process will not be included in the determination of the final grade; 3) final grades will be determined by professional judgement in referencing the most consistent, recent performance on guaranteed standards *instead* of using any mathematical weighting and averaging algorithm; 4) departments will design and adopt a reassessment policy permitting students to pursue multiple avenues to demonstrate mastery on a particular standard.

The other group invited all faculty to meet during lunch on Mondays to discuss an alternative system of reporting student achievement. The guiding principle was to develop a more accurate, coherent, and meaningful reporting scale. Several iterations were produced, including a familiar 12-point, A-F scale and two other more innovative 7-point scale and 5-point scales. The symbols to be used to report achievement varied as well, including numbers, A-F letters, and emojis. The group originally decided on a 7-point scale for final reporting and a 4-level scale to guide the formative reporting process. This was approved by the administrative council and was to be put into place for the 2020-2021 academic year.

However, with the advent of COVID-19, remote learning, and limitations on large gatherings, the principal opted to adopt the 12-point, A-F scale. The reasoning was that in order to adopt the less familiar scale, parents and students would need to be educated about the change. COVID-19 measures made that process too difficult from February through June. Further, communication about the changes in expectations was made through a foreign, remote platform. Although two years later the team is more fluent with coordinating faculty meetings and professional development online, at the time the effectiveness of the programs was unknown and stalled. The most lasting, enduring change was the adoption of the Assessment Belief and Guiding Actions (ABGA) as the new official assessment policy.

Research Problem

The school was only verbally informed of policies in the past. It has now formally adopted and published a policy that outlines the beliefs and guiding practices. It has also identified a prioritized set of expectations that are meant to more greatly align the faculty's assessment beliefs and guiding practice with SBAR. The particulars of these pieces are outlined in more detail in Chapter 3, in the description of the innovation of this action research. Based on my experience with the process at IS2, the shift of beliefs and practices takes deliberate, purposeful action. In particular, teachers tend to reuse materials that have proven successful in years past. These materials include homework, practice problems, instructional strategies, tests, projects, rubrics, assessment calendars, and even frameworks for gradebooks. Together, these materials create an assessment infrastructure that requires restructuring to match the new beliefs outlined in the adopted policy.

Research Questions

It is in the explained larger national, international, local, and personal context that I explore the impact of the policy on teachers' assessment beliefs and practices through changes in its assessment infrastructure using the following questions:

1. How does a new school assessment policy impact educators' assessment beliefs, practices, and infrastructure over one semester?
2. What are teachers' perceptions of factors that led to any changes in their beliefs and practices?

CHAPTER 2

THEORETICAL PERSPECTIVES AND RESEARCH GUIDING THE PROJECT

Consider a typical Monday morning in a high school. Educators walk through the campus gates with satchels, bags, and backpacks slung over shoulders. They walk along the sidewalk raising hands and smiling to bid good morning to one another as they walk into the school building to check in at the fingerprint reader and promptly sanitize their hands with the automatic sanitizer-sprayer just to the right of it. Onward to the elevator they get out at their respective floors, enter their classrooms, settle into their desks, and open their laptops. The Google Document(s) with the lesson plans are pulled up from a web browser. So is PowerSchool gradebook, Google Classroom, and FlipGrid to review students' submitted work and taking note of those that have yet to do so. After a quick check through e-mails and a glance at the news, the lesson plans with guiding questions are written on the whiteboard and in the hallways gather teenage voices. A bell rings and the conversations continue into the classroom. Javier strides in and sits at the table by the window; most recently he has achieved "exemplary proficiency" despite starting the year as "emerging." The teacher smiles. Susana joins Javier at the table; she has also shown the same growth. Jake walks in with Robert to sit near the front. Jake has been at :approaching proficiency" across all standards all year. Robert is similarly consistent but at "proficient." The essays that were submitted last Thursday are returned to the students with green ink comments on the papers' margins and in the text. There is a rubric with notes on it attached as well. Javier and Robert meet on the carpet soon after the papers are returned to share their feedback with one other while Susana and Jake do the same sitting on yoga balls by a coffee table. There is a space on their rubric for students to write one

personal strength and one opportunity to grow. The students then take a picture of their rubrics with their phones, laptops, or tablets and upload it to their portfolio on Google Classroom—just in time before the bell rings.

Truthfully, although that sequence may be a typical Monday morning for the economics teacher I shadowed for a week, there is hardly anything very typical more broadly. Other schools may not have sidewalks, gates, fingerprint readers, laptops, or bells. They could be a single-story building with no elevators. The classrooms may not have yoga balls and coffee tables. They also may not have teachers' desks. The differences I've experienced in schools and classrooms has me wonder how a school's architecture enables specific types of social interaction and influences beliefs and practices about learning. More relevant to this investigation, schools may not use Google Docs, Google Classroom, or FlipGrid. The gradebook, assignments, and activities may be managed through other platforms. Teachers may use red ink, no rubrics, no exemplars, and score them on percentages that translate to letter grades—A to F, "excellent" to "failure," instead of "exemplary" to "emerging proficiency." Perhaps students do not have space to reflect on the rubric and do not keep portfolios of their work. I wonder how different assessment infrastructures enable specific types of social interaction and influences beliefs and practices about learning.

Schools are highly complex, social institutions that Weick (1976) likens to an "unconventional soccer match," one where there are multiple goals on a sloped round field without any specific identified team and each individual able to name their own goals at any point in time to their advantage. This complexity of schools makes the inevitable problems that arise "wicked" (Rittel and Webber, 1973). Rittel and Webber

(1973) note that unlike the problems in the natural sciences, where there may be a clear identifiable solution derived from replicable verifiable data based on multiple experiments, the problems in society derive from people. There is no clear-cut if-then in the human experience, no series of "outputs of actions," making it difficult to identify any "center" or beginning of a problem (Rittel & Webber, 1973, p. 159). In this context of complexity and wickedness, we lean on theoretical frameworks and models to help understand schools and how to improve them.

Actor Network Theory (ANT) offers theoretical tools that help cut through some of the wickedness by mapping where, how, and to what extent beliefs are related to practice. Although not a strictly defined recipe or procedure nor a "totalizing theory of the world and its problems," it offers a set of *ideas* that can be put into practice to understand the world (Fenwick & Edwards, 2012, p. i). ANT's definition is intentionally vague because it is meant to remain somewhat flexible; its progenitors do not wish that ANT become "reified as an immutable research strategy… with methodological baggages" (Fenwick & Edwards, 2010, p. 166). It could be referred to as a set of tools, but Law (2009) prefers that it be named a "framework" or "sensibility," again pushing against the idea that it is a strict methodology. For the purposes of this investigation, I will henceforth refer to it as a framework because it is through ANT that I aim to construct a material semiotic means to explore the problem of practice in my workplace.

The ANT framework has been applied to a variety of contexts, including technology development, construction of scientific knowledge, organizational management, and more recently to the field of education. In this section I present the history and development of ANT, define it, outline the aspects most relevant to my own

research, and summarize how it has been applied to research in the field of education to date so as to best apply those methods to my own research interests. I conclude by offering a model of the network developed with the ANT framework and the policy document that reifies the intent of the SBAR at IST.

Actor Network Theory

> *"the historical name is 'actor-network-theory', a name that is so awkward, so confusing, so meaningless that it deserves to be kept."*
>
> (Latour, 2005)

The aim of this section is to cut through some of the confusion behind the name and assemble enough meaning to use the ANT toolkit as the base of my methodology, analysis, and inferences. ANT emerged in the 1980s from a social-scientific tradition developed in the Center de Sociologie de l'Innovation (CSI) in Paris, France, in the 1970s and 1980s (Muniesa, 2015, p. 81). In the context of growing numbers of personal computers and the realities of emergent information networks, Callon, an engineer and sociologist, and Latour, a philosopher and sociologist, led and worked with others—Akrich, Hennion, Law, Star, and Bowker to name a few—to develop an interdisciplinary approach to articulate a "new social theory adjusted to science and technology studies" (Latour, 2005, Loc 266). The group drew from their diverse experiences to explore how knowledge is discovered and constructed in scientific communities. For example, Callon's works in the 1980s centered on domestication of scallops and fishermen (1984) and electric vehicles (1986). Latour's (1988) work explored the codification of the person Pasteur into *pasteur*ization, the process. Law's (1987) work examined how maritime and navigational technologies of Portugal enabled its colonization. Although diverse, their

work coalesced into a framework to examine the nature of *groups*, the power of *materials* to mediate and translate meaning between them, and the importance of *action* in the real, observable reality. ANT uses the ideas of groups, materials, and actions "to render the social world as *flat* as possible" to view the links between them (Latour, 2005, Loc 401). As Fenwick (2012), a spokesperson for the use of ANT in education, puts it, ANT "looks closely at the translation process, tracing how an entity, human or nonhuman, becomes selected, enticed, persuaded and partially or fully changed in ways that mobilize it to join the network's movements" (p. 99).

Despite there not existing any clear blueprint for how to employ ANT (indeed its progenitors aim to actively resist such blueprints), below I offer an outline of the most relevant ANT ideas to be used as tools that are most practical to my research: groups, objects as mediators, materials as boundary objects, and the network.

Groups

Organizations, like all larger societies, are comprised of groups of individuals. To understand how groups can be influenced as well as how they influence others, a clear picture of what a group is and why they form is presented from an ANT perspective. The first assumption that ANT makes about groups is that they are defined by "spokespersons which 'speak for' the group existence…. justifying the group's existence, invoking rules and precedents and … measuring up one definition against all the others" (Latour, 2005, p. 31). Without a group maker, no group exists. Schools themselves are group makers, the brand and building "speak" into existence a student population, faculty, and group. Within the school faculty there are many formal and informal groups, such as teaching partners, departments, divisions, lunch mates, professional learning communities, and

coaches. Second, the ties that make up one group exist in juxtaposition to alternative "anti-groups" (Latour, 2005, p. 31). Although there may be multi-membership to different groups (i.e., a teacher can be employed to teach a Spanish and Economics course), the groups that individuals are members of remain defined by the anti-groups. Third, when groups form and reform the spokesperson is always looking for ways to define them (Latour, 2005, p. 32). A new principal to a school may engage in new norm-setting protocols and introduce new meeting agenda policies. Fourth, an ANT researcher must count themselves among the spokespersons that make the group possible. Whether as a member or an outsider looking in, the researcher's words and actions help define the group relative to others.

Objects as Mediators that Translate

Society is comprised of humans and their social interactions. Latour points out the etymology of "social is *seq-*, *sequi* and the first meaning is "to follow." The Latin *socius* denotes "companion, an associate" (2005, p. 5). In that light, understanding society, the word "social" can be viewed "as a very peculiar *movement* of re-*association* [italics added for emphasis] and reassembling" (Latour, 2005, p. 6). ANT is wary of the sociology of the social's dependence on examining humans as the sole agents of forming society. "All social scientists agree that groups have to be made and remade anew through some other non-social means" (Latour, 2005, p. 36). A defining feature of ANT's view of how groups make and remake themselves is through objects' mediation. School-wide faculty meetings *refresh* social ties; branded T-shirts, mascots, and cheers *enliven* the passions of school solidarity; and traditions, like the selection of the valedictorian of a graduating class based on grade point averages, are invented to *celebrate* individuals of

the group. Non-ANT social scientists would claim that these objects and events simply act as intermediaries, transporting "meaning or force without transformation: defining inputs is enough to define its outputs" (Latour, 2005, p. 39). Among the faculty, staff, and students there exists a reservoir of social forces that the means, events, objects, and policies tap into but cannot change (Latour, 2005, p. 38).

This view is denied by ANT. The events, objects, and policies by which a group achieves the social will mediate; they "transform, translate, distort, and modify the meaning of the elements they are supposed to carry" (Latour, 2005, p. 38). Latour takes the role of materials further, suggesting that without it, a group would not exist. As Durkheim, a founder of modern sociology, wrote, "a clan is essentially a reunion of individuals who bear the same name and rally around the same sign. Take away the name and the sign which materializes it and the clan is no longer representable" (Latour, 2005, p. 38).

The role that objects play in groups is the most controversial point of ANT. In a theoretical attempt to assemble actors in groups to develop a durable whole from which to understand observable actions, objects are seen "as being full-blown actors…. They can sometimes 'express' power relations, 'symbolize' social hierarches, 'reinforce' social inequalities, and 'reify' gender relations… things might authorize, allow, afford, encourage, permit, suggest, influence, block, render possible, forbid, and so on" (Latour, 2005, p. 72). This is not to be interpreted as meaning objects and humans are equivalent in the actor network, (Latour, 2005, p. 75) yet this does have important implications. With the perspective that objects are actors that have agency in the complex network of groups, the meeting space and organization of desks at a division-wide faculty meeting

can help understand the type of learning community envisioned by its leaders. A cheer at a sport event that derides the opposing team instead of solely celebrating the home team may impact those students' behavior with the opposing team after the game and even years after graduation—as an example, consider university sports rivalries such as Duke versus the University of North Carolina at Chapel Hill. The exclusionary and highly competitive policy of selecting valedictorians based on highest average grade point averages may motivate students to be aggressive in seeking higher grade point averages.

Or not. Latour notes that not all objects are always mediators. In fact, even those that mediate may also become intermediaries. Latour offers the example of a wall. "Once built, the wall of bricks does not utter a word—even though the group of workmen goes on talking and graffiti may proliferate on its surface" (Latour, 2005, p. 79). Extending the example, consider the Berlin Wall. Its construction had the agency of reshaping families, economies, and politics for decades. And still, those that walked along or near it on a typical Monday morning may not have been impacted in a most immediate sense, if only because the wall and its initial agency had been normed. To this phenomenon Latour notes that "objects appear associable with one another and with social ties only momentarily" (2005, Location1482). So, in another moment, when the wall came down, its agency was revived clearly and most obviously not only in Berlin but all around the world, foreshadowing and enabling the near end of the Cold War.

There are also walls all over a high school campus. Take classrooms, for example. They are usually a series of four walled rooms, occupied by one teacher who teaches a class in a specified subject or field (i.e., Mathematics). They group students and teachers in uninterrupted space for 45-60 minute periods of time during the day. The classrooms

shape friendships of students based on who is assigned to what subject and teacher. They block faculty from the view of fellow faculty and administrators, privatizing their practice. Just like the Berlin Wall, though, the wall becomes silent.

ANT researchers develop strategies to offer the voice to walls to learn of their agency and with what other actors they are associable. The first is to study objects in their workshops, where they "live a clearly multiple and complex life through meetings, plans, sketches, regulations, and trials" (Latour, 2005, p. 80). The second is to approach the object with a sense of ignorance to reproduce the novelty of the object (Latour, 2005, p. 80). A third is to observe them in the situation of an accident or break down, when "all of a sudden, completely silent intermediaries become full-blown mediators" (Latour, 2005, p. 80). A fourth way is to review the historical documents to produce like a historian the moment that the object was created (Latour, 2005, p. 81).

Materials as Boundary Objects

In studying how groups coordinate work with materials across space and time, Star and Griesemer (1989) propose the concept of the boundary object, one that has been used in other theories, including Wenger's (1998) communities of practice. Specifically, Star and Greisemer present a model for understanding how heterogeneous groups overcome "central tensions" due to divergent viewpoints that arise in scientific endeavors when trying to cooperate to generalize findings (p. 387). Their historical case study of the Museum of Vertebrate Zoology at the University of California, Berkley, examined how in restructuring a museum to become a research museum, Grinnell, its director, standardized the methods of collecting and describing artifacts. Standardization was pursued to limit "potential difference in ***beliefs*** [emphasis added] about evolution or

higher-order questions" (p. 407). Difference in beliefs stemmed from competing visions of various stakeholders' perspectives and their social worlds, including the "amateur naturalists, professional biologists, the general public, philanthropists, conservationists, university administrators, preparators and taxidermists" (p. 396). When the actors work independently, the problem of translation and integrity of generalizations do not arise. It is when they overlap that they do. Grinnell successfully navigated the boundary tensions that existed between the social worlds by carefully constructing protocols of collecting and labeling the artifacts.

From their case study, Star and Griesemer (1989) identify four boundary objects that they are careful to explain are "only analytic distinctions, in the sense that [they were] really dealing here with systems of boundary object which are themselves heterogenous" (p. 410). First, is the repository, such as a library or museum, that "[is] built to deal with problems of heterogeneity caused by difference in unit of analysis" (p. 410). Second, is the ideal type, "such as a diagram, atlas or other description which in fact does not accurately describe the details of any one locality or thing" (p. 410). It is abstracted and vague but is a "'good enough' road map for all parties" (p. 410). Third, the coincident boundaries are "common objects which have the same boundaries but different internal contents" and arise when there exists "different means of aggregating data and when work is distributed over a large-scale geographic area" (p. 410). The fourth and final one is the standardized forms, which are "devised as methods of common communication across dispersed work groups" (p. 411). Like what Latour called "immutable mobiles," these forms are "objects which can be transported over a long distance and convey unchanging information" (p. 411).

A final concept that that Star and Grisemer (1989) reference is that of marginality, "a critical concept for understanding the ways in which the boundaries of social worlds are constructed, and the kinds of navigation and articulation performed by those with multiple memberships" (p. 411). This idea of marginality is of particular import because it is within these individuals and/or groups that the boundary tensions arise the most because it is along the margin that the greatest divergence in views are expressed. "People resolve the problems of marginality in a variety of ways: by passing on one side or another, denying one side, oscillating between worlds, or by forming a new social world composed of others like themselves" (p. 411). In managing change, the development of boundary objects must account for the concerns that arise from marginality. In closing, Star and Griesemer (1989) offer an important cautionary note at the end: "But the protocols are not simply the imposition of one world's vision on the rest; if they are, they are sure to fail. Rather, boundary objects act as anchors or bridges, however temporary" (p. 414).

The Network: Action, Actors, Agencies, and Actants

The world is not observable in any single instance, individual, group, or object independently. The world is in perpetual motion, where individuals, groups, and objects act and are acted upon. Further, according to ANT, this action is not a planned, sequenced set of events; "action is not done under the full control of consciousness; action should rather be felt as a node, a knot, and a conglomerate of many surprising sets of agencies that have to be slowly disentangled" (Latour, 2005, p. 44). What is observable are the various layers of actions that can be untied to better understand what, who, and why something occurs.

ANT offers a way to map out actors, agencies, and actants. First, between the nodes, within the network, there are actors and actants. Actors are "participants in the course of action" (Latour, 2005, p. 71). They are "not the source of an action but the moving target of a vast array of entities swarming toward it" (p. 46). Agency is the term used to describe the effect that one actor produces on the other. Agency does something, makes some difference, and transforms (p. 52). Second, borrowing from the literary figuration, ANT posits that agency is figured. Returning to the examples of the school-wide faculty meeting, the school spirit, and the valedictorian, we naturally adopt and accept the figuration of their agencies to refresh, to enliven, to nominate, and to celebrate. An actant is the actor that acts, that has agency and mediates.

The aim of ANT is to offer an account of how and what occurred. A "good account [is] one that traces a network" (Latour, 2005, p. 128). The term calls to mind the technical networks of electricity, trains, and the internet as well as the informal way of associating human agents in sociology of organization (Latour, 2005, p. 129). The network is meant to be "a concept, not a thing…. a tool to help describe something, not what is being described" (Latour, 2005, p. 131). So although an actor network is meant to observe actions as agencies like the movements of trains from station to station along tracks to stations, it is less of an input-to-output transaction and more a mapping of the flow of translations between actors, human and non-human, in their concatenations.

Networks can last a very short time and influence a very small geography or they can "become so durable and apparently powerful in education, exerting influence across far-flung geographic spaces and time period" (Fenwick & Edwards, 2012, Location 356). The connections between actors "come together and connect, changing one another to

form links" (Location 323). The way that they become so strong is through the number of allies and connections that they make (Fenwick & Edwards, 2012, Location 369).

ANT in Educational Research

ANT has experienced a "lively trajectory" in the social sciences since the 1980s (Fenwick & Edwards, 2012, Location 237). Although its main application has been to science and technology studies, it is finding greater appeal in the fields of health and organizational studies (Vickers & Fox as cited in Hamilton, 2012, p. 40). There are increasingly examples of its use in education to study policy reform strategy (Hamilton, 2012; Gorur, 2012), emergence and adoption of tools (Pierce, 2015; Nespor, 2012; Fox, 2005; Parker, 2017), and the impact and various translations of curriculum in different contexts (Edwards, 2012; Zhang & Hayden, 2016; Mulcahy, 2012). This section will summarize these works with an emphasis on the diversity of data and methods used by these authors.

Policy Reforms

Hamilton (2012) uses observational records including field notes, interviews with participants, and collections of documents and artifacts to follow trajectories of actants, people, projects, discourses, and objects. She uses that data to create three ethnographies of how policy "can impose order on the disorderly flow of social life, creating knowledge that eventually becomes unquestioned truth" (p. 52). Comparing the three ethnographies, she concludes that, though a controversial ontology, ANT's "dynamic view of social life" fits the policy process and analysis well. ANT is a helpful tool to make sense of the "complicated networks of human agents to carry a policy from the stage of formation to implementation" (Hamilton, 2012, p. 42). Gorur's (2012) work was part of her doctoral

research and was built on two interviews of high-ranking Programme for International Student Assessment (PISA) officials. Gorur takes a classical ANT approach, conceptualizing the PISA as a laboratory and the ways that the assessment and its results are translated as scientific facts (Timmons, 2018, p. 58). In this way, she extends the original purpose of ANT, the study of scientific knowledge, into a metaphor for her own context.

Adoption of Tools

Pierce (2015), Fox (2005), Nespor (2012), and Parker (2017) develop and apply the tool of mapping through historical record, explain how ANT can turn technologies viewed as mundane into true actors, and elaborate on the significant role they play in the passage points in networks. Pierce (2015) uses ANT to explore the boundary between nature and culture when the first genetically engineered animal food product, AquaBounty Technologies' AquAdvantage salmon, was brought to market. To do this, he builds a "map of the life of salmon and demonstrates its pedagogical uses for developing an alternative scientific literacy in both science education and social studies contexts" (p. 87). Nespor (2012) explores how two devices, a computer-assisted interactive video module and an assistive communication device, mediate change. Nespor's methods include interviews of administrators and professors, documentary analysis, and classroom observations over a span of sixteen years. Nespor (2012) finds that the tools are successfully integrated into the work of groups when they become what Callon and Law call "obligatory passage points" or "points of stabilizations" within their network (p. 10). Parker (2017) also suggests how materials become important passage points in processes of education. Using vignettes offered from observations and audio-recordings of Personal

Education Plan (PEP) meetings, Parker finds that the PEP document is a key actor in the meetings, offering entry points of discussion and a framework to move the meeting along (p. 159).

Translations of Curriculum

Edwards (2012), Zhang and Hayden (2016), and Mulcahy (2012) find that the greater the material diversity of the setting, the more diverse the translations and outcomes. Edwards (2012) uses ANT to empirically study how a prescribed, standardized curriculum is translated into a multiplicity of practices. Edwards narrates two case studies as vignettes to "illuminate curriculum-making as network effects" in diverse settings, a hospitality-specific vocational college setting and a traditional, academic school setting (p. 31). He concludes that what curriculum standards are materially "enrolled and translated into it [make] a big difference in terms of both practice and what is learnt" (Edwards, 2012, p. 37). Zhang and Hayden (2016) used an exploratory, ethnographic case study approach with document analysis, semi-structured interviews, and classroom observations in a Chinese school to interpret the implications of delivering an Ontario secondary school curriculum concurrent with a Chinese national high school curriculum. They found significant dynamism, instability, and interchangeability among the actors as the two systems' practices intermingled. Finally, Mulcahy (2012) explored how professional teaching standards and "technologies (tools) in the service of broader social and cultural agendas" (p. 79) can do more than "describe pre-existing realities such as accomplished teaching practice or accomplished teachers; they actively produce them" (p. 80). Her research was centered on data from video-recorded classrooms and video-

stimulated interviews with students and teachers used to develop case studies in eight different schools across three major Australian states.

An Actor-Network Theory of Assessment Policy Reform

Like Hamilton (2012) and Gorur (2012), the following model aims at exploring policy reform. As Gorur has done, it will extend the scientific knowledge metaphor at the root of ANT's origins to treat the school, its various groups, and how they interpret the policy reform as a laboratory and how it is translated into various materials and practices as the scientific facts. Like Pierce (2015), the model I offer is meant to assemble an "epistemological model … capable of capturing the complex series of actors and institutions that are involved in the practice" of assessment reform (2015, p. 85). The model is a network meant to serve as a map of the life of the policy reform. The network will be comprised of various passage points, some of which may emerge as the obligatory passage points that Nespor (2012) and Parker (2017) discovered.

There is an anticipated diversity of translations of the policy reform that Edwards (2012), Zhang and Hayden (2016), and Mulcahy (2012) find in their research. Some of this may be because of the different materials teachers engage with (i.e., science department versus language arts). Other differences may be based on forms of resistance or misunderstandings. This is the root of the problem of practice I aim to explore. In the same way that the professional standards of Mulcahy's (2012) work "actively produced" accomplished teachers, this action research explores how a policy reform document that reifies a small group of change agents' assessment beliefs can actively produce practices amongst the high school faculty that reflect those beliefs.

In schools there are some traditional groupings that exist within the faculty. Some of these are obligatory and others are not. Departments exist by subject level. At IST there exist six different departments: language arts, foreign languages, social sciences, sciences, mathematics, and the arts. There are department heads who act as the formal speakers of the group. Several materials help define them, such as e-mail groups, shared Google Drives, and running agendas. There are two buildings on campus and multiple floors for each. Department faculty are intentionally moved together in the same building and most often on the same floor, geographically coordinating their work. In division-wide faculty meetings, these groups are frequently asked to sit with one another, further defining them within the larger groups.

Figure 1 illustrates an idealized network of different high school departments connected by the Assessment Beliefs and Guiding Actions (ABGA) document.

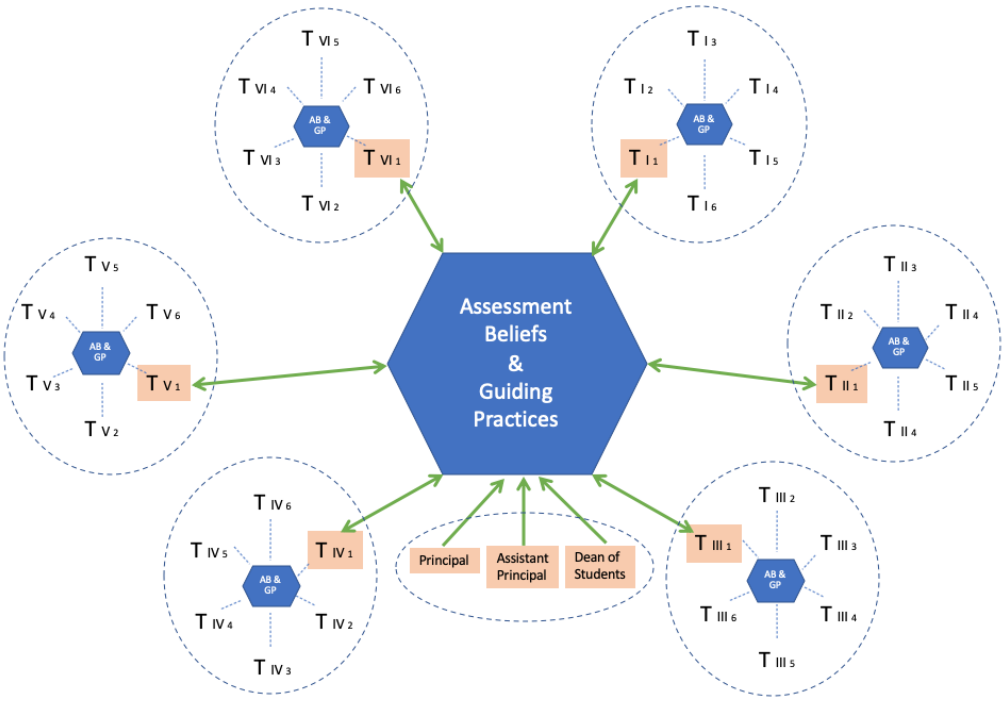


Figure 1. Network of ABGA to Departments

The hexagon at the center of the figure is the ABGA policy document. The ABGA is ultimately the articulation and reification of the work of the ATF and the collaboration of department heads. This is indicated by the green lines connecting to the pink rectangles. Inside those rectangles are the key stakeholders that created the document. The T-1 entries indicate a department head. For example, T-III-1 is the department head for social studies. The other pink rectangles are the members of the administrative team, the principal, assistant principal, and dean of students. Most green lines have arrows on either side, indicating that the individual influenced the ABGA and that the ABGA influences that individual's practice. The principal, assistant principal, and dean of students have only a single directional arrow because these individuals do not use the document to assess students the way teachers do; the document does not directly impact their practice. There are six groups of six teachers encircled by a dashed line around the ABGA. Each teacher is denoted with a T-entry. The T stands for teacher. The Roman numeral refers to a different department. And any number that is not 1 is a member of the department that is not the department head. The Arabic number does not denote hierarchy. In the center of their work is the ABGA represented by a smaller hexagon with dashed lines extending from it to the teachers. This is meant to indicate how the policy is translated by each teacher in their practice. An alternative way to draw this would be to have dashed lines extending from the larger, original SB&AP. However, I contend that the SB&GP will be translated in a particular way within the department, by its team. This group's, the departments, own translations are captured in Figure 2. It is an idealized network of the relevant ABGA document with the assessment materials that comprise the assessment infrastructure they work within. Because the red and green font

is too small to read, I offer Figure 3, the network between T-III-1 and T-III-2, and Figure 4, the network of T-III-6, respectively.

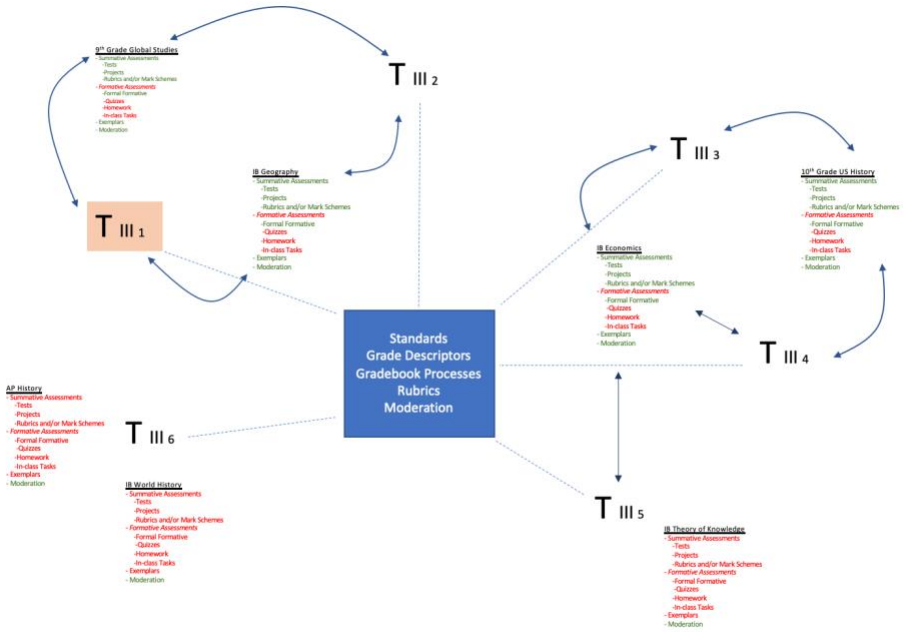


Figure 2. Network of ABGA, Assessment Materials, and Departments

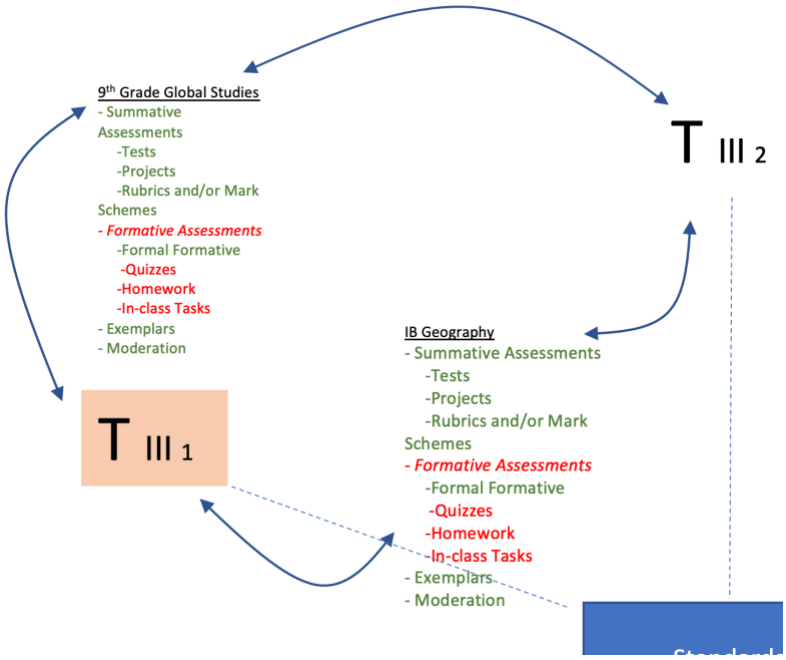


Figure 3. Network of ABGA, Assessment Materials, T-III-1 and T-III-2

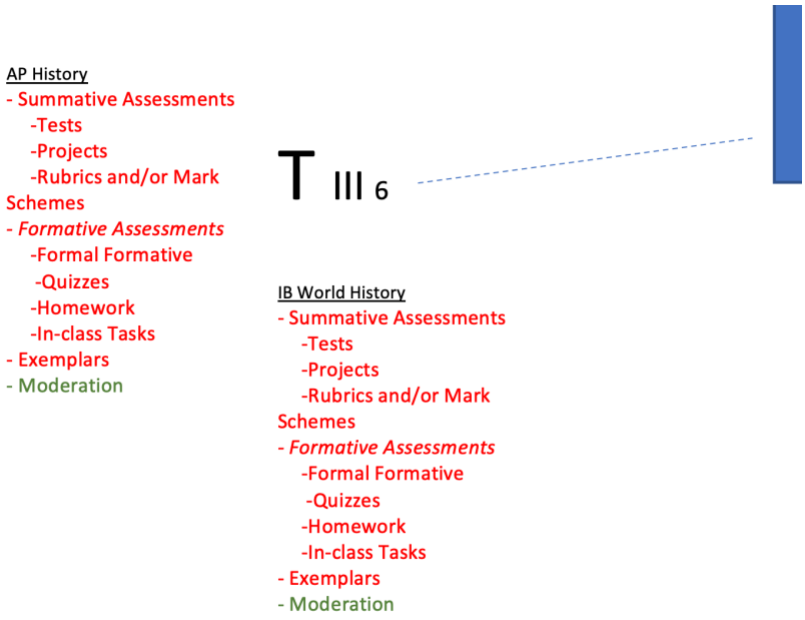


Figure 4. Network of ABGA, Assessment Materials, T-III-6

At the center are the points from the SB&GP that are interpreted as a group: Standards, Grade Descriptors, Gradebook Processes, Rubrics, and Moderation of student work. Time in meetings together is allocated specifically to develop shared understanding of these assessment materials and practices. It is a priority of the administration, the principal and assistant principal, to do so because shared understandings reduce diverse translations and these materials will inform the creation of other materials. Teachers are still denoted T entries. The pink T-III-1 is the department head. This department is the social studies department. Teachers T-III-1 and T-III-2 in the top left-hand corner of the diagram are connected through the courses they teach, ninth grade Global Studies and International Baccalaureate (IB) Geography. Underneath each course title are materials that comprise the assessment infrastructure. Those listed in green are materials that the two work on together, summative assessments that include tests and projects, and their

rubrics and mark schemes, exemplars, and the practice of moderation. Those listed in red are materials that the two do not necessarily work on together but may be shared with one another, the formative assessments like quizzes, homework assignments, and other in-class tasks. Notably, there is one, formal formative assessment in green under the formative process. Teachers T-III-3 and T-III-4 share a U.S. History class. They also share with T-III-5 an IB Economics class. Teachers T-III-5 and T-III-6 each have a class that they do not share with anyone else. Because there is no connection with any other teacher, all materials in those courses are listed as red, save for the practice of moderation because it is by definition a shared endeavor.

It is instructive to also consider a network without the ABGA. Figure 5 shows the school network and an individual department network side by side.

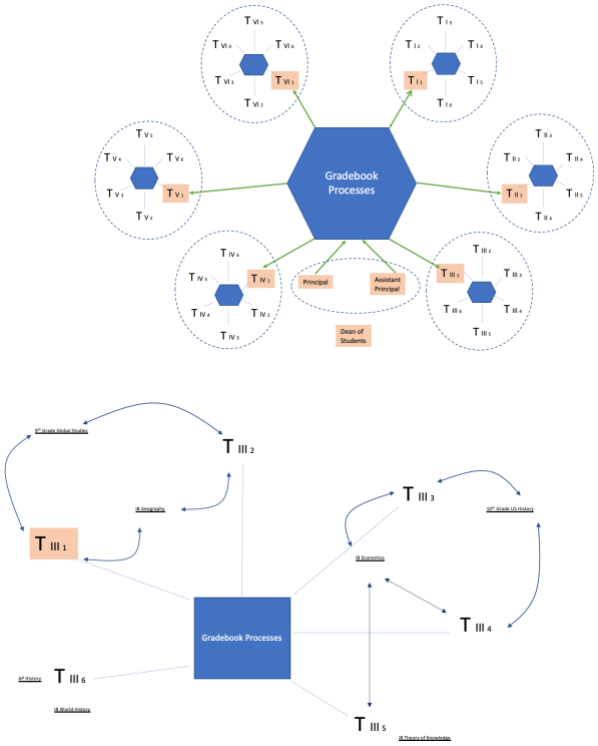


Figure 5. School Network and Individual Department Network Without ABGA

Instead, only a set of Gradebook Processes exist. Arrows that go into developing the processes come straight from administration. In this case, the dean of students, who has no academic responsibilities, lies outside of the network. Arrows from the Gradebook Processes go out to the Department Head, indicating that the processes are then shared with the team members. However, those processes are translated in many different forms by teachers. Teaching teams still teach the same classes together, but there are no materials that explicitly link their practice, such as the summative assessments that include tests and projects, their rubrics and mark schemes, exemplars, the practice of moderation, formal formatives, homework, etc.

ANT proposes that moving from Gradebook Practices to clear Assessment Beliefs and Guiding Practices will reorient groups, their assessment materials, and their practices. New materials will be created. Old ones will be discarded, often reluctantly. Some materials within the assessment infrastructure may need to stay because they act as boundary object, serving another network. For example, the gradebook software is one of these materials. First, because gradebooks record so many heterogeneous types of assessments, they are a repository of information, like Star and Greisemer's (1989) study of the museum. Second, it is the place where achievement is reported abstractly and vaguely while still being good enough for the many other groups such as teachers, students, parents, student support services, administration, and program evaluators to use to perform their work and responsibilities. Third, the gradebooks "have the same boundaries but different internal contents," those contents being at minimum an explanation of the standards being assessed. Fourth, as alluded to in the second point, different groups receive this common communication and use this information for their

work. Finally, when there is a policy change, there may be significant marginality as teachers' practices oscillate between what they did before in their gradebook and what they do now. What remains to be seen is whether gradebook software and/or other materials act as anchors or bridges for the school's intended change.

CHAPTER 3

METHODOLOGY

> *Follow the actors in their weaving through things they have added to social skills so as to render more durable the constantly shifting interactions.* (Latour, 2005, p. 68)

A new assessment and reporting ***policy*** was developed at the International School of Thailand (IST) with administration and faculty to improve assessment accuracy, coherence, and meaning. It was adopted and launched in August 2020 with a series of formal professional development opportunities, including sessions with an assessment expert during orientation, work in departments during orientation, division work during student release days, and faculty meetings after school. The purpose of this action research case study was to explore how change in school policy influenced teacher beliefs and practices. As alluded to in Chapter 1 and explained in Chapter 2, how I aimed to explore organizational change amidst the reform was to study its materials, specifically the assessment infrastructure, the way different assessment materials were coordinated and constructed. Latour (2005) suggests a metaphor of a supermarket. "We would call 'social' not any specific shelf or aisle, but the multiple modifications made throughout the whole place in the organization of all the goods …. [The] shifts reveal to the observer which new combinations are explored and which paths will be taken (… defined as a 'network')" (p. 65). With this lens, the exploration also included how the policy influenced the assessment materials of the school, its infrastructure.

Chapter 2 explained relevant tools from Latour's description of ANT, shared relevant educational research that used ANT as its framework, and concluded with an

application of it to construct of a map, an idealized network of a school and a department that practices beliefs aligned to the ABGA. It listed a host of different materials that were anticipated to emerge and/or evolve under the new system.

The purpose of this chapter is to describe how I studied the modification and reorganization of assessment infrastructure at IST using ANT tools in a multi-stage convergent parallel mixed design action research study. The chapter describes the setting, the participants, the role of the researcher, and the innovation. Then it details the research design by introducing the constructs and sub-constructs that guide the instruments. Each instrument is then explained with examples of prompts for participants, showing how they were linked to the constructs and sub-constructs. For qualitative instruments, the coding procedure that was employed to interpret the findings is identified and explained. Anchoring all this work are the following research questions:

1. How does a new school assessment policy impact educators' assessment beliefs, practices, and infrastructure over one semester?
2. What are teachers' perceptions of factors that led to any changes in their beliefs and practices?

Setting

This case study took place in a non-profit, private, college preparatory international high school in Bangkok, Thailand. At the time of the exploration, the school was accredited by the Western Association of Schools and Colleges, enabling it to offer students the equivalent of an American high school diploma. The school's curriculum included a robust selection of courses and included both Advanced Placement (AP) courses and International Baccalaureate (IB) courses which may lead to the IB Diploma.

At the time of the investigation there were 430 students in the high school. The families' income profile was of high-middle to high socioeconomic status. The student population of the high school was largely Thai nationals. The school used Measures of Academic Progress (MAP Growth), Writing Assessment Program (WrAP), Scholastic Aptitude Test (SAT), and Advanced Placement (AP) and International Baccalaureate (IB) scores to measure student achievement relative to peer organizations and the world. On average, students outperformed national and world averages on all tests in all subjects.

Participants

The participants in this innovation included the entire high school faculty. There were 66 faculty members on contract for the 2020-2021 academic year. Because this was a school-wide policy change, all faculty were participants in the innovation. Depending on the specific tool, questionnaire or interview, different numbers of participants partook in specific aspects of the research.

Role of Researcher

My changing role within the school was an important part of my ability to gather rich data. I was both an insider and outsider, a participant and observer. As my professional role with the institution began as the high school dean of students, it afforded me the opportunity to work closely with the instructional leadership team, administration, and interested faculty members on developing policy documents that guided the faculty. Some of these interactions were formal and formed part of early cycles of research. Other interactions were informal and still guided my thinking. The nature of my professional status made me an insider, observing through very personal,

social interactions such as hallway conversation, lunch meetings, faculty meetings, and semi-structured interviews (Ivankova, 2015, p. 57).

My role changed after six months. I served as interim elementary school assistant principal. My office moved to one that was at most 100 meters away. On the one hand, this meant I participated less in shaping professional development opportunities for faculty to take on the innovation. Further, I missed out on many organic social interactions that occurred in hallways or in the cafeteria. On the other hand, because I worked closely with the principal and assistant principal as dean of students, there may have existed some trust concerns among the participants. Having moved to the elementary school, I expected that the more formal instruments, the interviews and questionnaires, would yield more open responses. I also remained a participant in the process in that I shared my learnings and inferences from them with those that remained in control of the change. Finally, I had an opinion of assessment and grading practices. It was not radical, but it was progressive. I did not believe the policy changes went far enough to align with my own beliefs, but they were steps in a direction I supported.

Lastly, my positionality influenced my perspective on the school's decision to pursue the assessment reforms. Prior to my arrival here, I worked at a school that underwent these changes. I heard from others and felt personally the anxiety and excitement about the changes. I engaged in countless formal and informational conversations with faculty, students, parents, and administration on the topic. As a social studies teacher, for five years I was immersed in the changes and tried various forms of standards-based assessment. These experiences helped me work with the team to direct the change. However, it also placed me on the opposite end of the assessment innovation

bridge. Faculty and administrators were embarking on the change that I felt I had already crossed. This disadvantaged me because I needed to temper projecting my own experiences on the experiences of this community. Undoubtedly, comparisons will arise, but I needed to remain alert for contrasts. This was a different context and different version of the innovation.

Innovation

The innovation was a school-wide adoption of a new assessment policy that outlined a new means of measuring and reporting student achievement. Its aims were to make assessment more accurate, meaningful, and coherent. Accurate was defined as the truest measurement of a student's achievement. Meaningful was defined as a clearer description of a student's achievement toward course goals. Coherent was defined as logical and consistent reporting both across and within grade levels and courses. Table 1 outlines practices addressed in the policy that existed for the 2019-2020 and 2020-2021 academic years. The practices were divided into four themes: 1) determination of grade, 2) reduction in graded materials, 3) grading scale, and 4) reassessment policy.

Table 1.

Assessment Innovations: Old vs New

	2019-2020	2020-2021	Intended Outcome

I. Determination of Final Mark	• Final grades are determined by the averaging of scores to "major assessments," which account for 42% of final score, and "minor assessments," which account for 28% of the final score, plus a final exam, which accounts for 30% of the final score. • No set number of major or minor assessments are mandated per category. • No set number of points are mandated per major or minor assessment.	• **No averaging of grades for the final mark.** Professional judgement of faculty to identify the most consistent, recent measure of a student's achievement to the standards set out by the curriculum of the course will *determine grade*.	• #Accuracy • #Coherence
II. Reduction in graded material	• Homework, quizzes, notebook checks, and other formative assessments are permitted to be included as "minor assessments."	• **Only "major" summative assessments** are used in determination of the grade.	• #Accuracy • #Coherence
III. Grading Scale	• Grades are reported as percentages that translate into letter grades +A to F, where any final score below 60% is an F.	• **Grades are reported to a 12-level scale.** No percentages on any assessment are used. All scores are reported as letter grades A, A-, B+, B, B-, C+, C, C-, D+, D, D-, F.	• #Accurate • #Meaning
IV. Reassessment	• No school-wide reassessment policy is in place. Some faculty and departments have a system of extra credit or dropping of one assessment in place. Others have a more complete reassessment policy.	• **A school-wide reassessment policy** guides teacher practice. Only students that have completed the formative process (i.e. quizzes, homework, notebook, checks, and other formative assessments) are permitted to reassess. They have two weeks after the summative assessment is taken to complete that work and/or meet with the teacher to show evidence of working to improve in order to actually sit for the reassessment.	• #Accurate • #Coherent

Determination of Grade

The change in determination of the final mark was meant to make teachers' practice more coherent and accurate. Prior to the policy, teachers were given the

following weighting parameters for three categories in their gradebook: major (42%), minor (28%), and final exam (30%). The logic was that assessments taken throughout the semester would count for most of the final grade - 70%. Major assessments taken throughout the semester accounted for 60% of the 70% allotment (42%) and minor assessments count for 40% of it (28%). The remaining 30% of the final grade was accounted for in the final exam. Although the numbers created the appearance of logic and consistency, the practice of teachers within this system told a different story. Teachers across and within grade levels and courses used a wide variety of different points allocated to assignments, types of assignments, and frequency of assignments entered into their gradebooks. This made for illogical and inconsistent outcomes. For example, one unintended consequence related to frequency of minor assessments relative to major assessments is how an individual minor assessment can be worth as much as or even more than major assessments in the final calculation of students' grades. For example, some gradebooks included up to eight major assessments and only two minor assessments per category, making each major assessment worth less than the minor assessments in the final grade, 5.25% and 14% respectively. Other teachers may have had up to 40 different minor assessments, making each one worth 1.4% of the final grade. Adding to the mathematical complexity in calculation and valuation of assessments was that some student scores were reported as greater than the total possible points (i.e., 75/70).

There also existed incoherence in practice when determining the difference between a minor and major assessment. One might have asked what the difference was between a minor and a major assessment and likely heard very logical and consistent

explanations. A minor assessment might not require the same depth of knowledge (i.e., recall instead of synthesis and application). It could be more knowledge-based instead of skills based (i.e., vocabulary quizzes versus document-based essays). These justifications created the appearance of sound assessment strategies. Yet, as demonstrated through the aforementioned mathematical examples, in practice the outcome might undermine the intent that the minor assessment be worth the same as or more than the major. In other words, a vocabulary quiz could be worth more than an essay despite the English teacher's logical and consistent justification for the need to have the two categories.

Eliminating the weights and categories aimed to allow faculty to use professional judgement less encumbered by the mathematics. The very same justifications centered on depth of knowledge and knowledge versus skills could become more relevant. For example, a teacher could simply not use vocabulary quizzes and reference students' performance on document-based essays to evaluate students' ability to comprehend a set of texts and communicate an argument effectively.

The change was also meant to make reporting student achievement more accurate, so as to report a student's true academic achievement. Prior to the policy document, the gradebook used the mathematical mean as a measurement of central tendency to report a final grade. This greatly undermined accuracy because the mean is so susceptible to extreme scores (Schimmer et al., 2018, Redos, Retakes, and Reassessment in Action section, para. 15). This is particularly true when schools adopt a top-heavy percent-to-letter grade conversion, where 100-60% is a passing mark and 59-00% is failure. Consider the case of a student who earns 40% on a math assessment at the beginning of the year. The student would need to earn 100% on the next two assessments to raise the

grade to 80%, a B-. The move to faculty's use of professional judgement offers the teacher two other ways to evaluate student achievement. First, the teacher could review the assessments, the standards that they were meant to measure, and decide if the student had achieved them. Returning to the example and assuming the student did earn 100% on the two last assessments, the standards on the first assessment, solving quadratic equations, may have been incorporated to some degree in the second and/or third assessments, quadratic word problems and graphing quadratic equations. Alternatively, if this wasn't the case, the teacher could still use other ways of calculating an average, the median or mode, both of which return a value of 100%. This could be a more accurate report of the student's achievement.

Reduction in Graded Material

The use of only "major" assessments was in line with the previous change and was meant to both clarify the practice of the determination of the grade as well ensure a more coherent and accurate report of student achievement. Minor assessments took the form of pop quizzes, vocab quizzes, homework assignments, and lecture notes checks, to name a few examples. Teachers within and across grade levels and courses were more likely to have a consistent definition and justification of what a minor assessment is than to have consistent practice of using them. As noted earlier, there were different frequencies and types of minor assessments across and within grade levels and courses. Eliminating this category aimed to help clarify that practice. Further, it aimed to create greater consistency in terms of the kinds of assessments that count toward a final grade. Pop quizzes, vocab quizzes, homework assignments, and lecture notes would not count for any class, eliminating that inconsistency.

As to the accuracy, the true report of student achievement, the motivating philosophy behind this change was that assessments that are more skills-based and require more complex depth of knowledge are preferable to the recall-based simpler assessments. There was an assumption that the assessments more like the former include assessment of the latter, making assessing that work redundant. Because final marks would require some type of measurement of central tendency, to include separate, shallower depth of knowledge assessments as data to determine said marks would skew the final results. For example, a student may be able to recall definitions of English words for a vocab quiz and not be able to understand or recognize them in a short story nor use them effectively in a written report. Does the performance on the quiz more accurately demonstrate students' command of the English language than the latter? The school's assessment philosophy was that it does not.

One could argue that some simpler tasks could be used as evidence of achievement of learning. However, if a minor assessment only targets part of a standard that is later evaluated in a major assessment, then the minor assessment is more aptly used as a formative assessment to direct further instruction. Teachers could use information from a student's performance on a vocab quiz to improve students' understanding of the language to prepare them for the short story. The students' achievement on the culminating short story would then be the most accurate representation of learning.

Grading Scale and Descriptors

Reducing the scale was meant to make grades more meaningful and accurate. Reported "scales provide learners with a natural progression of quality that runs from

simplest to the most sophisticated" (Schimmer, 2018, Proficiency Scales and Rubrics in Action section, para. 10). The scope of the scale implies clear delineation of student achievement from one point to the next. The ability to delineate requires clear justification of student performance from one point to the next. The greater the scope, the more delineations and justification required. Reducing the scale implies the opposite. Consider for example the difficulty in offering a meaningful distinction between the following percentages: 89%, 88%, and 87%. It would be difficult to describe the difference between the works that earned different marks. If this was a calculation of number correct over total number of questions on a test, then the reason would be missing one or two points on the test. This is not meaningful feedback. In fact, it is not likely that distinction would be necessary, because in traditional grading schemes all of the scores are the equivalent of a B+. On the one hand, one could ask, "then why question the system at all?" On the other hand, how can one ignore a system that creates unnecessarily large scales of reporting? If one were to include in the example a score of 90%, the conversation changes. The distinction between a 90%, an A-, and an 89%, a B+, may more easily be articulated. This is one reason to reduce the grading scale.

The other reason was to improve accuracy of teachers' assessment. As early as 1912, educators challenged the growing use of percentage scoring to consistently assess student performance. Starch and Elliot showed that both high school English teachers (1912) and geometry teachers (1913) demonstrated great variation in assigning grades on the same assignment. In English the scores ranged from 50 to 98 and in geometry they ranged from 28 to 95 (Guskey, 2015, p. 25). Scales eventually shrunk to as low as three and more commonly five by the 1920s, until a modern resurgence of the use of

percentages in the 1980s (Guskey, 2015, p. 26). In 2011, Hunter Brimi replicated Starch and Eilliot's work to examine relative accuracy in teachers' assessment and found "among the seventy-three teachers who responded, scores ranged from 50-96" (in Guskey, 2015, p. 26). The intent of the change was to reduce the variance to make for a more accurate account of student achievement. Table 2 illustrates the system of reporting grades in the 2019-2020 academic year.

Table 2.

HS Grading 2019-2020: Percent, Level, Descriptor

Percent	Level	Descriptor
100-94	A	a thorough understanding of knowledge and skills, with the ability to apply them accurately in complex situations
93-90	A-	
89-87	B+	a strong understanding of knowledge and skills, with the ability to apply them accurately in a variety of situations
86-83	B	
82-80	B-	
79-77	C+	a sound understanding of knowledge and skills, with the ability to apply them accurately in predictable situations
76-73	C	
72-70	C-	
69-67	D+	an adequate understanding of knowledge and skills, with a limited ability to apply them in routine situations
66-63	D	
62-60	D-	
59-0	F	a limited understanding of knowledge and skills, and unable to apply them

The 2019-2020 school year had a 101 percentage point scale (including zero) that translated into five different levels, A-F, which in turn have three further gradations, plus, neutral, and minus (except for the top level, A).

Table 3.

HS Grading 2020-2021: Level, Descriptor

Level	Descriptor

A A-	a thorough understanding of knowledge and skills, with the ability to apply them accurately in complex situations
B+ B B-	a strong understanding of knowledge and skills, with the ability to apply them accurately in a variety of situations
C+ C C-	a sound understanding of knowledge and skills, with the ability to apply them accurately in predictable situations
D+ D D-	an adequate understanding of knowledge and skills, with a limited ability to apply them in routine situations
F	a limited understanding of knowledge and skills, and unable to apply them

At face value, the grading scale for 2020-2021 presented in Table 3 looks the same as the years prior, simply without the percentage scale. This is true and this simple move does much. It reduced the grading scale substantially, from 101 different marks to only 12 and made reporting student achievement more meaningful, because it is a clearer description of a student's achievement related to a course's goals. For example, differences in student achievement between scores of 90, 89, and 88 are theoretically possible. In practice, though, the difference between 90 and 89 is likely a conversation on the difference of the different levels, the difference between an A and a B. The difference between an 89 and an 88 becomes less significant because they both translate to a B+. For a grade to be meaningful to students, teachers "would have to make transparent the specific differences between [them]" (Schimmer et. al, 2018, Proficiency Scales and Rubrics in Action section, para. 10). If one cannot explain a difference, then why have one?

Reassessment

The motivation behind reassessment policy is to ensure all teachers across and within grade levels and courses consistently and accurately report student achievement. Prior to the policy adoption, reassessment strategies were framed by the positive intent of helping students. Some teachers framed it as an incentive to work harder and as an opportunity to show mastery of the standard. Although the latter reason was aligned with the school's adopted beliefs and was another motivation for this particular part of the policy, the fact that there were a variety of different practices makes for inconsistency. For example, some departments adopted a practice of dropping the lowest summative assessment. Some teachers offered extra credit opportunities. Others offered alternative means of assessment after the class attempt, like an in-person interview of the content.

One reason teachers were in favor of reassessment is that they believed there were both internal and external variables that influenced the outcome of a first attempt. Internally, the student may not perform well because s/he just doesn't understand the content yet. Reasons for this could be students' processing time, level of effort put forth in the formative work, and/or the teacher may not have noticed the misunderstanding to offer feedback in time. It does not mean that the student will not understand the content, only that the student has not understood it *yet*. Basing an end of semester mark on a static point of verification would be an inaccurate report of student understanding. Also, externally, social dynamics of family and friendships as well as a student's health can impact the ability to perform on an assessment. In this case, the assessment will not accurately measure that student's performance. Without an opportunity to reassess, it will be used in determining the student's achievement. The following table outlines the

guiding principles of the reassessment policy, which was aimed at ensuring all students have a consistent opportunity to demonstrate their true academic achievement.

Table 4.

Reassessment Policy Guidelines

1.	All students are eligible for second chance learning opportunities for summative assessments regardless of original grades (encourage mastery for guaranteed standards).
2.	Second chance learning opportunities will be given only after an authentic/reasonable attempt was made on the previous assessment.
3.	Students provide evidence that they have completed a learning plan/reflection sheet (submit with original attempt) presented to the teacher prior to a reassessment opportunity.
4.	Second chance learning opportunities must be completed within a reasonable time frame of the original assessment's posting.
5.	No second chance assessments will be offered within the last five days of the marking period. (except with extraordinary circumstances)
6.	Reassessment should be highly focused and demonstrate learning of the standard—rather than on replicating the modality of the assessment. When possible, teachers should consider different modalities when deciding on how to reassess.
7.	The highest mark will be recorded as the grade
8.	Students can have reassessment opportunities during Flex block and other times as determined by the teacher.
9.	Practices need to be consistent among PLTs (particularly with common courses)

Research Design

This action research case study employed a multi-stage convergent parallel mixed methods design (Creswell, 2014, p. 219) to explore the impact of an assessment policy change on both a faculty's beliefs and practices as well as the assessment infrastructure of the school. A case study is one in which a researcher collects in-depth information employing a variety of methods over a sustained period of time (Creswell, 2014, p. 19).

As a convergent design, this study collected "both quantitative and qualitative data, analyzes them separately, and then compares the results to see if the findings confirm or disconfirm each other" (Creswell, 2014, p. 219). This study was meant "to evaluate the impact of a program or project" to "help address a larger program objective" and as such, a multistage evaluation design was used (Creswell & Guetterman, 2019, p. 559). There were several predetermined stages that built (or informed) " each other throughout the study" (Creswell & Guetterman, 2019, p. 559). For that reason, there was some planned fluidity, much like what Creswell and Guetterman (2019) suggest, "in light of the formative" evaluation of a multistage design (p. 559). Table 5 outlines the data collection instruments, the type of data collected (qualitative and/or quantitative), and in which phase they were employed.

Table 5.

Data Collection Instruments and Their Use in Three Phases

Method	Qual/Quan	Pre	Mid	Post
Video-cued Interview	Qualitative	✓		✓
Faculty Questionnaire	Qualitative Quantitative	✓		✓
Document Analysis (gradebooks, rubrics, tests, progress report comments)	Qualitative	✓	✓*	✓
Heads of Department Interview	Qualitative		✓	✓

**in this phase progress reports will be analyzed in lieu of exams*

True to the concurrent, parallel design, in the pre and post phases, there were both qualitative and quantitative measurements. All instruments used specific constructs and sub-constructs in their creation and analysis. The table is followed by a clearer

description of the instruments, beginning with the constructs and sub-constructs that they aimed to explore.

Constructs

The constructs for this investigation were: #Beliefs, #Practice, #Policy Alignment, and #Material. The constructs related to the overarching themes of this investigation. It was an exploration of the extent and ways that a faculty's beliefs and practices aligned with those that were intended to be adopted by the school's policy reform. It did this through ANT's framework, looking closely at the materials that were created, used, and left behind in the course of reorienting the network/infrastructure of assessment. Table 6 identifies and defines the constructs used in developing the tools for data collection and formed a basis for my analysis.

Table 6.

Constructs, Definition, and Variables

Construct	Definition	Sub-constructs
#Beliefs	The purpose(s) of assessment in schools.	Direct Indirect
#Practice	The way teachers interact with assessment.	Summative Assessments Formative Assessments Reassessment Semester Grades
#Policy Alignment	The extent that the beliefs and practices of the faculty resemble the stated beliefs and practices outlined in the Assessment Beliefs and Guiding Practices policy.	Determination of Final Mark Number of graded materials Grading Scale Reassessment
#Material	The resources that teachers use and create to measure student learning.	Tests Rubrics Formal Formative Homework In-class tasks Exemplars Gradebook Curriculum Maps

Beliefs. Beliefs about assessments were centered on their purposes. The two sub-constructs explored under beliefs were direct purposes, such as the measurement of student learning, and indirect purposes, such as motivation of students to perform work.

Practice. Practice referred to different interactions that teachers had with assessment. The four sub-constructs related specifically to practices targeted by the innovation. Summative assessments were expected to be designed collaboratively with clear links to standards. Students' marks were meant to be moderated in teaching teams or departments. Formative assessments were also meant to relate specifically to the course standards and were not to be included in determination of final grades. Reassessment opportunities were the expectation for all summative assessments. Finally, the semester grades were determined from the most consistent, relevant evidence of student achievement – not from calculation of statistical means.

Policy Alignment. Policy alignment was meant to measure the extent to which the beliefs and practices were aligned with the intent of the ABGA to make a more coherent, accurate, and meaningful assessment. The beliefs about these intentions and the ways they were practiced were explored through the following four sub-constructs: the determination of final mark, number of graded materials, grading scale, and reassessment.

Materials. This study assumed that the resources that teachers used and created to measure student learning would help explain teachers' beliefs as well as transform them. Tests, rubrics, homework, in-class tasks, gradebooks, and curriculum maps were all part of the infrastructure.

Data Collection Instruments

True to the mixed-methods action research design, the instruments used in this investigation were both qualitative and quantitative. Each tool was designed with the constructs in mind so as to align the data with the research questions. Of the qualitative tools, two types of interviews and document analysis were used, each one twice. There was one quantitative instrument, a full, high school faculty questionnaire, which was also distributed twice. A description of each instrument, the sampling procedures used in each, and the means by which I aimed to analyze and interpret the data are provided next.

Interviews

As Nespor (2012), Gorur (2012), Mulcahy (2012), and Zhang and Hayden (2016) did in their own ANT education research, interviews formed a key component in understanding how individuals navigate the assessment changes both in their beliefs as well as in their practice.

Video-cued interview (See Appendix C). Video-cued multivocal ethnography was pioneered by Tobin (1991) in his study of preschool in three cultures. Tobin used video recordings of preschools in three different cultures as a cue in to get "concrete and evocative than verbally asking about the same events" (Tobin, 2019, p. 258). The strategy of cuing an interview was adopted by Mulcahy (2012) in her ANT-rooted exploration of professional teaching standards in Australia. I aimed to use the method to measure and understand faculty members' beliefs and practices before and after one semester of engagement with the new assessment and grading practices. The video recording made was of a teacher from a different school describing what assessments had been included in her gradebook, when and why, and how a student's final grade was determined using

professional judgement. The video was shared with two members in the local context, each randomly sampled from the mathematics and another from the science department. After faculty viewed the video, a semi-structured interview was conducted with each member separately to solicit their reactions with probing questions that aimed at faculty's beliefs and practices; see Appendix C. The interview was conducted twice, once before the innovation and once at the end of the first semester, after grades were published.

Questions and Constructs. The questions were constructed to elicit reactions to the video that relate specifically to IST's reforms. Table 7 offers an example of a question and how I aimed to explore specific constructs and variables from the participants' response.

Table 7.

Video-cued Semi-Structured Interview Question 3

Question	Code derived from construct
3) Considering the formative assessment process, what is your take on the difference between informal and formal formative assessments, how they are used, and how they are reported? ***Probe for*** *take on the descriptors, number of assessments, importance of including task-oriented work reported as incomplete and/or complete*	#NumberOfAssessments #TypeOfAssessment #Policy Alignment #Beliefs

The question was constructed to ask about a specific practice articulated in the video. The teacher described how there were both informal and formal formative assessments. The latter were meant to resemble the style and format of summative assessments to a large extent and be recorded in a gradebook. This was not a common practice at IST. Depending on the response of the participant, I probed for information

related to the number and type of assessments. I anticipated being able to identify information related to a faculty's beliefs about assessment and the extent that they aligned with the intent of the ABGA.

Head of Department Interview (See Appendix D). Using a semi-structured interview protocol, the heads of department were interviewed twice, once mid-way through the semester and once at the end. The reason for conducting interviews mid-way through was to allow time for practices to occur after the policy had been adopted. By that time, summative assessments were administered and progress reports written, meaning that sufficient engagement with the various facets of the assessment infrastructure had taken place. A final interview was conducted at the end, after a complete cycle of engagement with the policy reform. The objective of the interview was to explore the practices, events, and materials that had developed since the assessment reforms.

***Questions and Constructs*.** Table 8 offers an example of a question and how I aimed to explore specific constructs and variables from the participants' response.

Table 8.

Head of Department Semi-Structured Interview Question 7

Question	Code derived from construct
7) Working with others, how do you ensure #coherence of assessment practices?	#NumberOfAssessment #TypeOfAssessment #tests #assignments #quizzes #projects

This question asked about a particular practice expected in the policy reform, coherence in assessments. Using a positive presupposition to anchor the participant's thinking, that they *do* work with others, I asked for ways that they ensured coherent practices. Anticipating the question of what coherence means, I used the anticipated codes to describe and probe.

Sampling

Both sets of interviews were conducted in two departments, mathematics and science. Ideally, research would have been conducted across all departments. However, due to the intention of gathering a large, diverse amount of data to provide rich detail and limited time, I selected two departments to study closely. The mathematics and science departments were selected because these two fields may have offered the most similar types of assessments; mathematics and equations are often used in the scientific disciplines. This was a nonprobability sampling technique and it biased any inferences that were extrapolated for the school community.

Video-cued interview sampling. I employed a purposive sampling strategy, "to (a) select a purposive sample that represents a broader group of cases as closely as possible [and] (b) set up comparisons among different types of cases" (Teddlie & Yu, 2007). The mathematics and science departments were selected for reasons already outlined. The faculty per department who were asked to participate in the video-cued interview were randomly selected. The names of the heads of department were omitted from the sample because I asked them specifically to participate in a separate interview (see following section). Then, names were be drawn from a hat.

Head of department interview sampling. Again, purposive sampling was used. The heads of the mathematics and science departments were purposefully sampled because of their dual roles. As leaders in their departments, they had insight into three domains and two groups. Their department and the instructional leadership team (comprised of all department heads, the principal, and the assistant principal) were two groups and domains. As classroom teachers themselves, they had firsthand experience in that domain as well.

Interview data analysis. The interviews were recorded, transcribed, and coded using a sub-coding strategy, when one "assigns a second-order tag after a primary code to detail or enrich the entry" (Saldaña, 2013, p. 69). In the pre-coding phase I "[circled, highlighted, bolded, underlined, or colored] rich or significant participant quotes or passages that [struck me] – those "codable moments" worthy of attention" (Boyatzis, 1998, as cited in Saldaña, 2013, p. 19). Then, in the first cycle I re-read and coded the transcript with the constructs and sub-construsts that coded "moments" (Saldaña, 2013, p. 22). The constructs served as "parent codes" and the sub-constructs as "children codes" (Gibbs, 2007, as cited in Saldaña, 2013, p. 77). These codes were how I planned to "perceive and interpret what is happening in the data," (Saldaña, 2013, p. 7). Specifically, I created a spreadsheet in Microsoft Excel with the transcribed text in the far left column and both parent and children codes as headers. I made a mark and comment for each line of the text that related to the codes. In the second cycle, I considered how and to what extent the original constructs and sub-constructs matched the actual coding filters. Finally, analytic memos were written to connect the transcripts and codes on a personal level as it relates to my research questions, my code choices, emergent patterns, and

potential links between other participants' responses (Saldaña, 2013, pp. 43-45). The "outcome, categorization, or analytical reflection" derived from all interviews were narrated in a descriptive analysis, organized by the constructs, per each department (Saldaña, 2013); see Appendix J. Finally, inferences were drawn, recorded, and reported as case studies per department.

Documents

Documentary analysis has been used in case studies as a precursor to other methods, to offer context throughout an investigation in conjunction with other methods, or as the major method used (Simmons, 2009, p. 64). Simmons (2009) notes that "formal document analysis tends to be used less than interview and observation in case study research and its potential for adding depth to a case has not perhaps been fully exploited" (p. 63). In ANT research, as the documents are the materials in the nodes of the network which act as agents in it, they can play a significant role. Just as Edwards (2012), Hamilton (2012), Parker (2017), and Zhang and Hayden (2016) use documentary analysis in their ANT research to collaborate with interviews and observations, this investigation used the strategy in conjunction with interviews and a questionnaire to specifically investigate how both the practices and the physical materials of the faculty change. Specifically, the materials that the faculty used within the assessment infrastructure of the school, gradebook copies, progress report comments, rubrics for projects, and final exams were those that formed the nodes, or what Callon refers to as 'moments' (as cited by Hamilton, 2012, p. 43), that were the key documents investigated. This technique is supported by Simmons (2009) specifically for gauging whether a school's practices are in line with its policies (p. 63).

Sampling. In line with the participant sampling procedures outlined for interviews, the same departments, mathematics and science, were targeted. This helped with triangulation of data as well as assisted in offering rich detail for analysis for the narrations. In gathering the documents, each individual that was interviewed in both the mathematics and science departments were asked to share final exams, rubrics for projects, and a unit test from the Fall 2019 semester and Fall 2020 semester. The decision to ask the same participants from the video-cued interviews and the heads of was made so as to have data to triangulate between interview data and document analysis. The decision not to include all members was made out of practicality in terms of amount of data and my capacity to code and interpret it.

Gradebooks. The gradebooks that teachers use are a record of assessment information, both formative and summative, that the teacher has chosen to record. For this type of document analysis, all faculty members in both the mathematics and science departments were referenced. The total number of both minor and major summative assessments was counted, the number of minor assessments before major summative assessments was counted, and the different types of additional categories used were noted. These were compared across teaching teams and within the department. These observations were made for the fall semester 2019-2020 (F19) and fall semester 2020-2021 (F20) for comparison and contrast. Those semesters were used with the assumption that the same material was most likely taught and assessed for each respective course and would make for more meaningful interpretations and findings.

Gathering the documents. In my professional position at the school I had access to all teachers' gradebooks. I also had permission from the school to use this information in

my investigation. I printed out copies of the documents and ensured student names were deleted and teachers' names were appropriately translated into participant codes (i.e., Mr. John from Science became Participant #1S).

Rubrics. Rubrics defined here are documents used by teachers to articulate in writing a set of criteria for scoring projects, papers, or tests. Criteria may relate directly to course standards or they may not. According to the ABGA, it was ideal that they do. The descriptors of the criteria offer a continuum of performance, using words such as exemplary, proficient, satisfactory, and unsatisfactory. These descriptors can reflect the assessment descriptors adopted by the school for each letter grade or not. Again, according to the ABGA, it was ideal that they do. To analyze the documents, the consistency of criteria in the rubrics between teachers and alignment of criteria to course standards was observed. Similarly, the consistency between teaching teams of the descriptors of the continuum of performance as well as alignment to the school's descriptors was noted. These observations were made for the fall semester 2019-2020 and fall semester 2020-2021 for comparison and contrast.

Gathering the documents. I reached out to each of the participants and asked to see a rubric for a project that they completed in the fall semester of 2019-2020 and the fall semester of 2020-2021.

Final Exams. At the end of each semester, final exams are conducted for each core class. The guidance for designing the exams was that they cover a significant number of course standards, that they should be rigorous enough so that most students complete them within 90 minutes, and that all students can complete it within 120 minutes. The exams were observed for consistency between teachers who taught the

same course, types of questions asked, and way that they were scored. For consistency, I aimed to look at the actual questions and format of the test. I observed the questions and compared them to the adopted course standards to view the degree to which they were linked. Further, I viewed how they were scored and reported out. These observations were made for the fall semester 2019-2020 and fall semester 2020-2021 for comparison and contrast.

Gathering the documents. I reached out to each interview participant for copies of their final exams from both fall semester 2019-2020 and fall semester 2020-2021.

High School Faculty Questionnaire

The final instrument used was the only quantitative instrument (see Appendix E). It was a questionnaire comprised of 40 questions. To my knowledge, the use of a questionnaire is not typical of ANT research. Further, though used to gather demographic data at the end of an interview or include it in an attempt to "boost the generalizability of case" (Simmons, 2009, p. 164), it is not an essential part of a case study. However, in the context of this action research, the data was expected to 1) help with triangulation of the qualitative analysis and 2) understand the perspectives of the members of the department who were participants in neither the video-cued nor the head of department interviews.

Questions and Constructs. Participants were asked to indicate how "true of you" a set of statements were for their beliefs (questions 1-16) and practices (questions 17-39). The degree of "true of you" is a seven-point ordinal (Figure 6).

0	1 2	3 4 5	6 7
Irrelevant	Not true of me now	Somewhat true of me now	Very true of me now

Figure 6. True-of-You Scale for HS Questionnaire

I included ‘0 – Irrelevant’ as an option because in testing the questionnaire, some teachers expressed that prompts relating to collaboration were not relevant to their practice, because they were the only teacher of their subject. In reality there were three categories, “not true of me now,” “somewhat true of me now,” and “very true of me now.” I chose to break down the three into seven because “people respond to the ordinal position of categories as well as to the descriptors,” which makes comparing answers across groups and samples over time more accurate (Fowler, 2014, p. 90).

Each question from 1-40 was based on language derived from the ABGA document. For example, the second belief under the category assessment design offers the following practice: “All assessments are aligned to specific course standards” (see Appendix E). The first question of the questionnaire relating to faculty beliefs is: “All summative assessments should align to the adopted course standards.” Each question also relates to a specific construct (See Appendix F). For example, the same question relates to the construct #Beliefs and #Policy Alignment.

Sampling and Validity. The high school faculty questionnaire could have had up to 66 responses if the entire faculty responded. I aimed for 44 responses, assuming a 60% response rate. Cronbach’s alpha was used to test the instrument’s internal consistency. Specifically, scale reliabilities of the instrument were conducted using overall reliability and inter-item reliability statistics. Descriptive statistics of each were used to help examine “how variable [the scales and associations] are and what shape the distribution value takes” (Shi & McLarty, 2009, p. 9). Upon determining the reliability of the

instrument, the means and standard deviations for each construct were calculated per department.

Method of Synthesis

Having described the entire process, I present Table 9 to illustrate how the entire methodology came together, from research questions to constructs, data collection instruments, data analysis, and supporting ANT education research.

Table 9.

Alignment of Research Questions, Constructs, Methods, and Supporting Literature

Research Question	**Constructs**	**Methods**			**Supporting Literature**
		Data Collection Instrument	Corresponding Question	Data Analysis	
1) How does a new school assessment policy impact educators' assessment beliefs, practices, and infrastructure over one semester?					
1 a →How does an institution's adoption of a ***new policy*** of assessment and grading impact a faculty's ***beliefs*** about assessment over one semester?	#Beliefs	Video-cued Interview	Q3, Q4, Q5, Q6, Q8	Subcoding	Mulcahy, 2012
	#Beliefs #Material	Faculty Questionnaire	Q1-Q16	Means and Standard Deviation	none
1 b →How does an institution's adoption of a ***new policy*** of assessment and grading impact a faculty's assessment ***practices*** over one semester?	#Practice	Video cued Interview	Q7	Process Coding	Mulcahy, 2012
	#Practice #Material	Head of Department Interview	Q2, Q3, Q4, Q5, Q6, Q8	Process Descriptive Coding	Nespor, 2012 Gorur, 2012 Mulcahy, 2012 Zhang & Hayden 2016
	#Practice #Material	Faculty Questionnaire	Q17-39	Means and Standard Deviation	none
	#Practice #Material	Documents	Gradebook Tests Rubrics Progress Report	Documentary Analysis	Hamilton, 2012 Zhang & Hayden 2016

1 c → How does an institution's adoption of a ***new policy*** of assessment and grading impact a school's assessment ***infrastructure*** impact assessment reform?	#Belief #Policy Alignment #Material	Video cued Interview	Q9	Value Coding	Nespor, 2012 Gorur, 2012 Mulcahy, 2012 Zhang & Hayden 2016
	#Practice #Policy Alignment #Material	Head of Department Interview	Q2, Q3, Q4, Q5, Q6, Q8	Process Coding	
	#Practice #Policy Alignment #Material	Faculty Questionnaire	Q33-Q39	Means and Standard Deviation	none
2) What are teachers' perceptions of factors that led to any changes in their beliefs and practices?	#Policy Alignment	Head of Department Interview	Q9, Q10	Subcoding	Nespor, 2012 Gorur, 2012 Mulcahy, 2012 Zhang & Hayden 2016
	#Policy Alignment	Video cued Interview	Q10	Subcoding	Mulcahy, 2012

Data Synthesis, Inferences, and Conclusions

After data were gathered and analyzed through the various coding measures, it was synthesized into two descriptive narratives, one for each department, organized by the constructs #Beliefs, #Practices, and #Materials; see Appendix J. The purpose of the narrative is to summarize and report the findings per the data sets per department. That is, they offer a description of the beliefs and practices of each department over the course of the semester, leaning on pre, mid, and post data. Then, those narratives were compared and contrasted with direct reference to the research questions and reported in the concluding section of Chapter 4. Figure 7 offers a final illustration of the entire process.

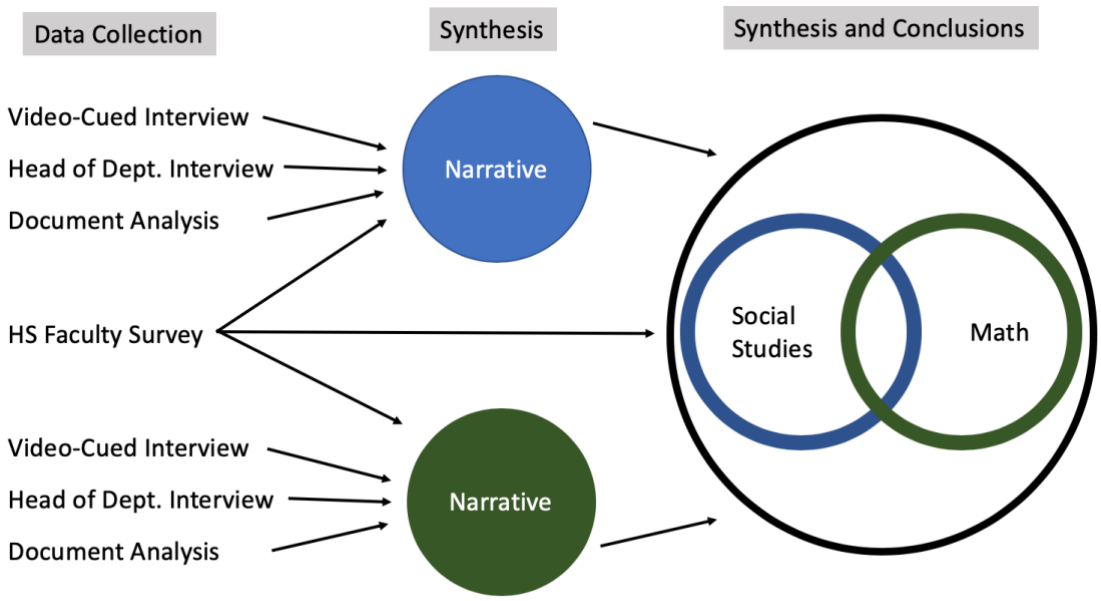


Figure 7. Data Collection, Synthesis, Conclusions

CHAPTER 4

RESULTS

Introduction

This action research project intended to fulfill two purposes. From a pragmatic perspective, it intended to explore the impact of the adoption of an assessment policy in a secondary school on its faculty. The results of the exploration were to be used to continue improving assessment practices via targeted interventions. The policy was written to improve accuracy, coherence, and meaning of assessments. It was adopted in the fall semester of the 2020-2021 academic year. A survey, video-cued interviews, semi-structured interviews, and document analysis of gradebooks and assessments were used to observe any measurable change. This chapter presents the findings of the study. It will analyze the data from the survey, gradebooks, tests, and interviews. Then it will synthesize those findings in a discussion of the results. The second objective was to use actor-network theory as a lens to design the study and understand what happened to determine what can happen next. That discussion is undertaken in Chapter 5. Both the data analysis and discussion sections are organized by the research questions:

1. How does a new school assessment policy impact educators' assessment beliefs, practices, and infrastructure over one semester?
2. What are teachers' perceptions of factors that led to any changes in their beliefs and practices?

Data Analysis

Survey

I distributed the survey instrument at the beginning and end of the fall 2020 semester with the purpose of quantitatively exploring changes in assessment beliefs and practices. The constructs it aimed to explore were educator beliefs about the purpose of assessments and the faculty's practices employed on summative assessments, formative assessments, reassessments, and the determination of grades. These were used to determine the extent to which those beliefs and practices align with what is articulated in the Assessment Beliefs and Guiding Actions (ABGA) policy document. There were a total of 40 items, 15 specifically relating to beliefs and 25 to practices (see Appendix F). Each item was designed using language directly from the ABGA to explore the faculty's relative alignment of beliefs and practices with ABGA over one semester. In other words, the items on the survey were statements taken from the ABGA, to which the participants responded by indicating the extent the statement was true for them on a seven-item Likert scale:

0	**1** **2**	**3** **4** **5**	**6** **7**
Irrelevant	Not true of me now	Somewhat true of me now	Very true of me now

Survey Measurement Methods and Hypotheses

To measure the relative alignment of beliefs and practice to the policy and changes, the mean scores of each item were calculated. Because each item was created based on the ABGA, a self-reported score 6 or 7, "true of me now," indicates alignment with the ABGA. Any score between 3 and 5 self-reports "somewhat true," which is translated into

weaker alignment. A score of 1 or 2, “not true of me,” is unaligned. Some items were phrased such that a response of “very true” meant that their beliefs and/or practices went against the policy (i.e., Q28, I assign some work in class that does not align to the adopted course standards). For those items, the scores were reversed in SPSS (i.e., 7=1, 6=2, 5=4, 4=4, 3=5, 2=1).

Participants

The total number of respondents the first questionnaire was 23 (*n = 23*). Out of a total 57 faculty members, this results in a 40% response rate. The second distribution yielded 16 responses, with a response rate of 28%. Identifying information of the respondents was not taken, eliminating the possibility of measuring change via paired sampled t-tests. This step was done to ensure anonymity of the participants. At the time of distributing the survey this was important due to my positionality as an elementary school interim assistant principal (AP) in the school. Although I did not have an evaluative role in the high school, the school was in a period of transition with the high school AP departing and the faculty’s knowledge of my interest in that role. To ensure that the participants’ responses did not impact our professional relationships, I opted not to collect identifying information. As a researcher, I regret not having done so. Without that information I was unable to run a paired t-test to measure changes of individual respondents. Further, response rates were very low, rendering independent t-tests unhelpful. Therefore, any comparisons made are at best speculative. Table 10 shows the known respondents’ department membership for both iterations of the survey, pre and post.

The survey is somewhat representative of the full faculty in the sense that at least one faculty member from each department responded to each survey both times. The most representative group is the English language development (ELD) department, followed by the arts and then the sciences. The science department has the most consistent participation; half the department participated both times. The other departments all had a participation rate of 33% or less both times, some as little as only 1 member (or 10%) of the department.

Table 10.

Participants in Pre and Post Survey Instrument by Department

Department	Pre		Post	
	N	% of Dept	N	% of Dept
English Language Development	4	100%	2	50%
Arts	4	80%	2	40%
Sciences	4	50%	4	50%
Language Arts	2	29%	1	10%
Mathematics	2	29%	2	29%
World Languages	2	20%	1	10%
Social Studies	1	11%	3	33%
Unknown*	4	*	2	*
Total	23	40%**	16	28%**

**The responses left blank are treated as 'unknown' and omit the % of department membership.*
**These percentages are calculated from the total number of faculty of 57.*

Internal Consistency

The survey instrument and constructs were evaluated for internal consistency using IBM SPSS© software to calculate Cronbach's alpha on its first administration, the pretest. The primary constructs are #Beliefs and #Practices. The instrument proved internally consistent overall. The entire 40-item instrument yielded a score of 0.882. The alpha for beliefs is lower than practices, 0.586 and 0.855 respectively.

Change in Beliefs

Because there were no matched samples, changes in beliefs were measured by calculating the mean and standard deviation of each construct on the pretest and posttest. These calculations were conducted overall as well as per department.

Beliefs and practices overall. Table 11 shows the mean and standard deviations results for all respondents at the beginning (Pre) and end (Post) of the semester.

Table 11.

HS Faculty #Beliefs and #Practices

Constructs	N	Pretest M(SD)	N	Posttest M(SD)
#Beliefs	23	5.11(.68)	16	4.79(.58)
#Practices	0	4.41(.75)	16	4.52(.70)

Note. the number of valid responses (N) for #Practice construct in the pretest was lower than the construct #Beliefs because 3 participants did not respond to all items associated with that construct.

The notable trend from the results is that the faculty self-report being consistently somewhat aligned in beliefs and practices at the beginning and end of the semester.

Beliefs per department. Table 12 disaggregates respondents' reported scores on the construct #Beliefs at the beginning and the end of the semester by department. It includes the number (N) respondents per department, the mean score, and its standard deviation.

Faculty reported beliefs that are somewhat aligned across all departments in both the pre- and posttests. Inferences about changes in alignment cannot be reliably made due to differences in participation in the pre- and posttests.

Table 12.

HS Faculty Beliefs Pre and Post per Department

Department	Beliefs			
	N	Pretest M(SD)	N	Posttest M(SD)
English Language Development	4	5.37(.95)	2	5.56(.71)
Arts	4	4.95(.36)	2	4.69(.35)
Sciences	4	4.89(.54)	4	4.75(.63)
Language Arts	2	5.81(.35)	1	5.56(.00)
Mathematics	2	3.90(.22)	2	4.28(.40)
World Languages	2	4.78(.04)	1	4.19(.00)
Social Studies	1	5.31(.00)	3	4.78(.40)
Unknown	4	5.61(.36)	1	4.56(.00)
Total	23	5.11(.68)	15	4.79(.58)

Note: Some respondents did not report which department they belonged to and were treated as Unknown.

Practices per department. Table 13 disaggregates respondents reported scores on the construct #Practices at the beginning and the end of the semester by department. It includes the number (N) respondents per department, the mean score, and its standard deviation.

Table 13.

HS Faculty Practices Pre and Post per Department

Department	Practices			
	N	Pretest—M(SD)	N	Posttest—M(SD)
English Language Development	4	4.69(.95)	2	4.84(.72)
Arts	4	4.47(.32)	2	4.24(.52)
Sciences	4	4.89(1.35)	4	4.38(.80)
Language Arts	2	4.9(.85)	1	4.63(.00)
Mathematics	2	3.83(.51)	2	3.69(.11)
World Languages	2	4.27(.50)	1	5.82(.00)
Social Studies	1	4.30(.00)	3	4.80(.52)
Unknown	1	5.26(.00)	1	4.46(.00)
Total	20	4.41(.75)	16	4.53(.70)

Note: Some respondents did not report which department they belonged to and were treated as Unknown.

Reported practices are somewhat aligned across all departments in both the pre- and posttests. Inferences about changes in alignment cannot be reliably made due to differences in participation in the pre- and posttests.

Alignment of Beliefs and Practices to Policy

Across both constructs the statements related to both beliefs and practices are "somewhat true" for faculty both at the beginning and at the end of the semester, indicating that faculty who responded began the semester and ended it somewhat aligned with the school policy. These findings cannot be generalized to the entire faculty because there might be biases in who responded (i.e., faculty whose beliefs and practices are less aligned may have been less likely to participate). For another perspective on possible changes in practice, I turn to teacher gradebooks.

Gradebooks

The gradebooks of teachers were analyzed to observe changes in practice. The ABGA stipulated specific items be eliminated from reporting, namely the use of bonus points, extra credit, and rote homework assignments (i.e., practice problems, worksheets, vocabulary lists, guided questions, memorization). To explore the impact this made on assessment practices, the gradebooks of every course in the math and science departments were analyzed for number, timescale, and type of assessments recorded in two semesters, fall semester 2019-2020 (FS19) and fall semester 2020-2021 (FS20). See Appendix G for data related to each course, teacher, timescale, and type of assessment for each department.

Participants

A total of 24 science and 25 math courses were offered each semester for a total of 49 courses. There was a total of six science and six math teachers. The screenshots of the digital gradebooks from PowerSchool were taken for every course at the end of two academic semesters, FS19 and FS20. See Appendix H to view a sample of the student-view gradebook used. The total number of gradebooks collected for FS19 and FS20 is 90. There are more gradebooks collected than courses because some courses are taught by more than one teacher at a time.

Number of Minor and Major Assessments

Table 14 reports the mean number of minor assessments recorded in gradebooks per teacher by department in two semesters, FS19 and FS20. The mean is calculated by total number of minor assessments recorded in all of an individual teacher's gradebooks divided by number of courses that individual taught. Percent change was calculated to compare the changes in teachers' and departments practices. A negative percent change indicates a decrease in the number of minor assessments.

Table 14.

Minor Assessment Count, Standard Deviation, and Percent Change per Faculty per Department

Subject	Faculty N	Minor M(SD)		% Change
		FS19	FS20	
Math	1*	37.33(7.02)	0(0)	-100
	2**	5.25(2.36)	4.25(2.5)	-19
	3	2(1.41)	2.4(.55)	20
	4	19.33(2.89)	20.4(4.34)	6
	5	27(7.21)	5(1.63)	-81

	6	13(4.24)	0.33(.58)	-97
	Combined	**16.11(13.5)**	**6.33(7.87)**	**-61**
Science	1*	6.5(2.87)	5.25(1.26)	-19
	2**	15(6.68)	6(2.65)	-60
	3	18.25(3.4)	2.25(2.63)	-88
	4	8(4.00)	8.6(8.71)	8
	5	7.25(2.5)	4.4(1.52)	-39
	6	4.5(.58)	3.33(.58)	-26
	Combined	**10(5.42)**	**5.13(2.22)**	**-49**
Science & Math		**12.76(6.14)**	**5.73(10.48)**	**-55**

Note. Faculty numbers 1 and 2 participated in the head of department and video-cued interviews respectively.

Table 15 reports the mean number of major assessments recorded in gradebooks per teacher by department in two semesters, FS19 and FS20. The mean is calculated by total number of major assessments recorded in all of an individual teacher's gradebooks divided by number of courses that individual taught. Like the minor assessments, percent change was calculated to compare teachers' and departments practices.

Table 15.

Minor Assessment Count, Standard Deviation, and Percent Change per Faculty per Department

Subject	Faculty N	Major M(SD)		% Change
		FS19	FS20	
Math	1*	4.33(.58)	8(1.00)	85
	2**	3.5(1.29)	2(.82)	-43
	3	4.5(1.29)	3(.71)	-33
	4	4.33(0.58)	3.8(1.64)	-12
	5	4(2.00)	4.25(.96)	6
	6	9(7.07)	8.33(5.78)	-7
	Combined	**4.63(2.52)**	**4.5(2.99)**	**-3**
Science	1*	5.75(2.06)	4(1.15)	-30

	2**	5.75(3.59)	5.33(2.51)	-7
	3	5.5(2.08)	6(3.16)	9
	4	3.67(1.53)	2.4(1.14)	-35
	5	3.5(1.29)	3.6(.55)	3
	6	3.5(1.00)	4(.00)	14
	Combined	**4.65(2.17)**	**4.08(1.95)**	**-12**
Science & Math		**4.64**	**5.73(10.48)**	**-8**

Note. Faculty numbers 1 and 2 participated in the head of department and video-cued interviews respectively.

Math Department. There is observable change in practice in regard to number and types of assessments. The average number of minor assessments in math courses fell by 61%, from 16 to 6 between the FS19 and FS20. There was also an increase in coherence of practice, with the standard deviation of average number of minor assessments reported falling from 13.4 to 7.62. Using the average number of minor assessments reported in FS20 as a lens to compare the change of the faculty from one semester to the other, two of the six faculty members already reported less than 6 graded minor assessments in FS19. These two may have had practices already aligned with the intent of the policy reform. Of the remaining four faculty, three reduced the number of minor assessments by as much as 80% on average. One of the three reduced the number reported from 37 minor assessments to omitting them entirely. Notably, one faculty member did not change his practice; the average number of minor assessments he offered in his courses remained between 19 and 21 in FS19 and FS20.

As to major assessments, excluding the end of semester exam, the average number recorded remained constant at around 4 to 5 per semester. There was no discernible change, with a decrease of 3% from FS19 to FS20. Notably, there is an exception. As noted earlier, faculty 1 stopped including minor assessments in his

gradebook and on average increased the number of major assessments by 85%, from 4 to 8. Faculty 1 is also the math head of department (MHOD) and participated in the semi-structured interview; I will provide more insight into his practices in my discussion of his interview.

Further, the types of assignments and assessment range widely across subjects in the FS19 semester. Those labels include quizzes, continuous assessment, chapter titles, practice, investigation, group work, notebook checks, test corrections, class work, work check, spiral, workbooks, notebook check, and homework (see Appendix G). In the FS20 semester, labels consisted almost exclusively of quiz, test, and project (see Appendix G). The exception was for the courses offered by teacher 4, who continued to use the homework, test corrections, quizzes, assignments, and workbook.

Finally, as can be expected from decreasing the number of total minor assessments, the timescale of those assessments changed across the department (see Appendix G). Notably, there is usually at least one minor assessment offered before a major assessment. Major assessments occur in 4- to 8-week timescales.

Science Department. Like the math department, there are observable changes in practice regarding the number and types of assessments. The average number of minor assessments fell by 50%, from 10 to 6. There was also an increase in coherence of practice, with the standard deviation of average number of minor assessments falling from 5.42 to 2.22. One of the six faculty members submitted less than six graded minor assessments in FS19. One other faculty member, the science head of department (SHOD), submitted on average 6.5. This would indicate that these two may have had practices already mostly aligned with the intent of the policy reform. Of the remaining

four faculty, two of them reported as many as 15 and 18 minor assessments in FS19 and experienced the greatest percent change, 60% and 88% respectively, bringing it closer to the groups average of 5 in FS20.

As to major assessments, excluding the end of semester exam, the average number of major assessments remains constant at 4 to 5 per semester, excluding the end of semester exam. The types of assessments were not as varied as the math department but still included activities and assignments the FS19 semester. Labels included labs, quizzes, presentations, tests, activity, packets, activities, chapter titles. In FS20 the labels included quizzes, tests, and labs. There was one teacher, teacher 5, who included worksheets and presentations in the gradebook in FS20.

Like the math department, with the decreasing the number of total minor assessments, the timescale of those assessments changed across the department. Also like the math department, there is usually at least one minor assessment offered before a major assessment and major assessments occur every 4 to 8 weeks (see Appendix G).

All Math vs Science. In FS19 there was a clear difference between the assessment practices per department. By FS20, those differences were hardly discernible. On average, there were more minor assessments recorded in math than science courses in FS19, 16 to 10. The difference between the two is less obvious in FS20, 6.33 and 5.13. In fact, excluding data from math teacher 4 (who did not observably change practices at all), the average number of minor assessments reported in math is lower than that of science. Also, the math department demonstrates reduced variance in practices in FS19 to FS20 as measured by standard deviation, at 13.4 to 5.42 (see Table 14). Of all twelve faculty members, the three who submitted the most minor assessments in FS19 were all in the

math department (see Table 14). On the other hand, so were the faculty members who submitted the least and third-least number of minor assessments. In FS20, there is more consistency in both departments. The average number of minor assessments offered in both departments is between 5 and 6. The average number of major assessments offered is also between 4 and 5. Although the standard deviation in math is higher than science in FS20 at 7.62 to 2.22, discounting data from the outlier in the math aligns both departments' variance at 2.2.

Thus, the overall gradebook data indicates greater coherence of practice after one semester with the new guidelines. I now look more closely at the assessments themselves.

Test Data

I requested copies of rubrics, assessments, and final exams from a total of 4 participants, two teachers from the Math department and two from the Science department from two different semesters, FS19 and FS20. These participants were the same teachers to take part in the interviews. A total of 16 documents, 4 per participant, were shared with me. Table 16 lists the assessments per participant in each department.

The ABGA explains that assessments are designed to offer varied means to demonstrate mastery of the standard and should have clearly communicated purposes and methods of evaluation. To observe that belief in practice and any changes in practice, the command terms, scoring system, and organization of the assessments' content were observed (see Appendix I).

Table 16.

Assessment Documents per Participant

	Science		Math	
	SHOD	SVC	MHOD	MVC
Assessment Documents	• HLY2_P1 Semester Final 2020 • HLY2 P2 Semester Exam 2020 • AP Chemistry • General Chemistry Final Assessment	• Static Electricity Quiz • Bohr Model Flipgrid Formative Assessment • Introduction to Chemistry and Physics Fall Semester Exam 2019 • Introduction to Chemistry and Physics Fall Semester Exam 2020	• 10th Grade Paced Math FS29 • 10th Grade Paced Math FS20 • IB MAI HL Semester Exam December 2019 • IB MAI HL Semester Exam December 2020	• Dec Exam IB MAI SL 2019 • Math 9 Reg Dec Exam 2019 • IB MAI SL Y1 Unit 6 Test FS20 • Math 9 Unit 2 Test

Math Department. The command terms used were more diverse for students in the higher level classes, specifically the IB course Math Applications and Interpretations (MAI) at the standard and high levels. In the year 9 and year 10 courses, the most common command terms in FS19 and FS20 were “find” and “solve.” These were less common in the IB courses. Regarding changes of practice from year to year, there was no noticeable change in types of command terms within each course. In fact, the year 10 paced math course used identical problems. However, a key difference is the total number of questions on that test— the FS19 version included 25 and the FS20 only had 15.

What is consistent in the math department is the use of marks to assess student work. A mark is a number of potential points a student receives per each question. There appears to be a progressive increase in how much each question is worth on each test. The questions that are worth more marks appear to require more steps to answer them

and/or are actually two questions in one (i.e., solve for AND graph). These typically appear at the end of the test. Also, the value of the marks increases in the upper years' classes. In the second year of the IB MAI course, only one question was worth two marks while the majority were worth at least five. The 9th and 10th grade tests did not include word problems with applications beyond two-dimensional geometry. This may relate to the fact that there are not many questions worth more than two marks on 9th and 10th grade tests.

Science Department. The science department appears to be making changes to the form of their assessments and considering ways to make changes to their substance. A different rate of change per teacher has led to a lack of coherence within the department. The quizzes used by the Intro to Chemistry class in FS19 and FS20 are similar in that they are assessed by rubrics on a scale of A to D and not by total marks. However, the rubrics themselves are different. The rubric used in the FS19 is completion-based (i.e., has visuals, uses specific vocabulary) to evaluate student performance. The rubric used in FS20 does not include a completion or frequency of correctness component, using terms "minimal, some, and complete" instead. Further, the rubric's different bands of achievement are described using adjectives that correspond more clearly to grade descriptors. Specifically, the highest letter grade, A, requires "extension" of knowledge, B requires demonstration of sound knowledge, C requires satisfactory knowledge, and D requires a fundamental amount of knowledge.

The final exam for Intro to Chemistry 9 is also somewhat different. Though it is more in form than substance, there is a structural change to the test. Specifically, there are at least 16 questions repeated verbatim from FS19 to FS20. There is also a combination

of multiple choice and short-answer questions in both. In this way, the substance of the test has not changed significantly. However, the questions themselves are organized differently. The FS19 version included two sections, one for multiple choice and one for short-answer. The FS20 version breaks the exam into 8 sections, each corresponding to a different unit of the test. In each section, there is a range of 3 to 5 multiple choice questions and 2 to 3 short-answer questions. In both iterations, the grade is evaluated out of a total number of marks and converted into a percentage.

The IB and AP semester exams look very much like the actual, external IB and AP exams. Interestingly, the IB examinations look very similar to the FS19 exam, with a set of multiple choice questions followed by short-answer questions.

The General Chemistry final assessment for FS20 is in some ways the most standards-based aligned as well as grading-reporting misaligned in the sample of both math and science assessments. First, it is not a test but a project. Students are meant to answer the question: How does technology use electromagnetic radiation? In some ways it more closely resembles the FS20 quiz for Intro to Chemistry in that there is a rubric that resembles the grade descriptors. It is more detailed in that it lists every standard covered in the course. It also is completed on a scale of 7: A, A-, B, B-, C, D, F. This does not translate equally to the school-wide grading scale of 12: A, A-, B+, B, B-, C+, C, C-, D+, D, D-, F. In fact, the scale of seven is further translated into a scale of three, A for exceeding, B for mastering, and C-F for concerning.

Math and Science Synthesis. The math and science departments use consistent practices in the externally assessed courses of the IB and AP. The tests resemble the final

IB and AP tests that students will take in both form and substance. These practices did not change from FS19 to FS20.

In the courses that are not IB or AP, there is less consistency between the math and science departments. First, the math department has not changed its practices, whereas it appears that the science department has made some changes. In the Intro to Chemistry course, the rubric used for quizzes changed slightly. The final exam for the course appears to be restructured by unit. This may mean there is a more intentional focus on being able to assess and report discrete skills. The Intro to Chemistry course has shown the most innovation in assessment style. It is very clearly standards-aligned and has moved away from a percentage system of evaluation. At face value, it also appears to have moved away from the school-wide reporting scale. For more insight into these changes, I now move to the interview findings.

Interview Findings

I conducted two sets of interviews with four participants in the math and science departments. The first set of interviews was a video-cued interview with a randomly selected member of the department who was not the department head. These were conducted at the beginning and end of the FS20 semester. The second set of interviews was semi-structured and took place with the heads of the departments in the middle and the end of the FS20 semester. Questions from the interviews were written to explore the constructs of #Beliefs and #Practices through the lens of changing alignment to the ABGA or to examine how two other constructs, #Materials and #Others, played in the change. The interview data were analyzed using a deductive coding strategy with first-order parent codes and second-order children codes. The parent codes are the original

constructs, #Beliefs, #Practices, #Materials and #Others. The children codes were derived from standards-based assessment themes, Summative, Formative, Reassessment, Determination of Final Mark, Tests, Rubrics, Formative, Quiz, and Gradebook. Table 17 demonstrates how each research question was tagged by each code and the corresponding sub-codes that related.

Table 17.

Research Questions, Codes & Sub-Codes

Question	Code	Sub-Code
1. How does a new school assessment policy impact educators' assessment beliefs, practices, and infrastructure over one semester?	#Beliefs #Practices #Material	#Summative, #Formative, #Reassessment, #Determination of Final Mark, #Tests, #Rubrics, #Formative, #Quiz, #Gradebook #ABGA
2. What are teachers' perceptions of factors that led to any changes in their beliefs and practices?	#Other	#Leadership, #Administration

After the interviews were transcribed and organized by questions, an initial cycle of coding was implemented to gauge the extent to which the codes were effective and to identify "'codable moments' worthy of attention" (Boyatzis, 1998, as cited in Saldaña, 2013, p. 19). A spreadsheet was generated in Microsoft Excel with the transcript from the interview listed in the first column and the codes listed across the header of the table. The text was reviewed and cross referenced with the codes, identifying what code applied to what section of the text. Descriptive memos were also kept as personal reflections as inferences and insights emerged. One outcome of this cycle was the determination of the effectiveness of the codes. Most of codes were determined to be effective. Other sub-

codes were introduced to account for how the materials and other influences shaped the change: Collaboration, Standardization, Leadership, Administration. This is why the list of codes in Table 17 differs from that which was anticipated and reported in Chapter 3. Some sub-codes were deemed either redundant or not a good fit while other new codes emerged from this first cycle.

In the second cycle, a descriptive narrative the interviews was created that synthesized the most frequently coded excerpts in the first cycle. The report was organized by department and parent codes, #Beliefs, #Practices, #Materials, and #Others. See Appendix J for the full report.

In the third cycle, excerpts were re-examined through the lens of the constructs and additional memos to identify and report patterns and links from the descriptive narrative report. Table 18 shows examples of how excerpts of an interviewee were coded, sub-coded, and interpreted. These were then synthesized as department case studies and are shared following Table 18.

Table 18.

Interview Text, Codes, Sub-Codes & Memo with Patterns and Links

Interview	Interview Text	Code	Sub-code	Memo
SHOD	I helped write it [the ABGA]. So I'm really familiar with it and I really take it to heart.	#Materials #Beliefs	#ABGA	The SHOD is aligned. His participation in writing the document helped.
SHOD	they're [ABGA] in line with some things I've done in the past at other schools …like reading an old friend.	#Material #Practices	#ABGA	The SHOD is aligned. His past experience in another school helped shape align it.
SHOD	he was "shoehorning [IST] assessment policy into what I was already doing…." He felt he had to "to take the way I was asking questions and transform that into a percentile style grading assessment	#Material #Practices	#Determination of final mark	The assessment infrastructure at IST made it difficult for him to practice the standards-based assessment strategies he was already familiar with.
SVC	"formative assessments before the summative assessment" can help students "actually fix those gaps."…. "it also helped me to reflect on my teaching.	#Materials #Beliefs	#Formative #Summative	Very aligned to the spirit of standards-based assessments philosophy. The second part of this references the purpose of formative assessment, *for* learning of student and teacher.
SVC	In the first interview she felt that the weighting of the assessments should be an individual "teacher decision," because the person needs "the knowledge of… the assessment and then what standards or what learning object is, what skill was assessed. So I think the teacher has to decide the maximum mark." In the second interview she stated explicitly and passionately on the point of collaboration, "that's very important. And even I do that. So we have a [professional learning team (PLT)] and we work together … to create the assignment and then we have rubrics and plan by the whole team and then we have marking scheme or sometimes rubrics."	#Beliefs #Practice #Materials	#Detemination of final mark	The teacher expresses a change in beliefs and practices. Originally, determination of the mark is an individual decision. In the second interview she works in the PLT to develop the materials to inform those decisions.

SVC	Assessment “is based upon previously determined rubrics, criteria, mark schemes, and other evaluative tools”	#Materials #Practice	#Formative #Summative #Rubric #MarkSchemes	She is using the standards-based assessment materials in her work.
SVC	she will “consider only the recent [assessment]” if it is “more extensive and it includes the previous assessment.”	#Practices	#Summative #Determination of final mark	This aligns with the ABGA’s point that assessment “allows for multiple opportunities for learners to demonstrate their knowledge, skills, and understanding” (Appendix A).
SVC	She will let students slide on submitted work if “the performance in the summative assessment is reasonably good.” But, if it is not, then those students will be called in to do the work during the flex period	#Practices	#Determination of final mark #Reassessment	This is very much in the spirit of standards-based assessment, where students are pushed to demonstrate the learning. There is some tension here, because it means she won’t hold others as accountable if they already do well on the summative.
SVC	She said, “in terms of the work ethics, I feel that it's not that strong.”	#Beliefs		This is a concern that was brought out.
SVC	quizzes that she calls “minor, which is actually graded. It’s formal formative assessment. We actually grade it and it happens once in three weeks. Sometimes, once in two weeks. There's no such timeline. Depends on the difficulty level	#Beliefs #Practices #Materials	#Quizzes	Here there is tension with the standards-based philosophy because quizzes are graded and count towards a final mark, but she uses the term “formative,” which is meant to be *for* learning not *of* learning.
SVC	“make sure that [students] understand the basics before we move on to the complex.”	#Beliefs #Practices #Materials	#Quizzes #Determination of Final Grade	This carries over from the previous comment. So, the quiz is a type of benchmark of learning. The question is, what to do if they don’t get it? She indicates that she ‘fills the gaps’ and reflects on her own practice. However, if it is still in the gradebook, then it will count against them.
SVC	“a very small portion of that unit.”	#Beliefs #Practices	#Determination of final grade	Here she justifies entering the mark by stating it does not count as much.

Science case study. The ABGA document was initially translated at the beginning of the semester by the entire science department. The team went point by point through the document to ensure agreement and understanding. The SHOD was relieved to adopt the new policy. He worked in a school prior that had already made many of the changes and expressed some frustration for having to return to traditional practices. Specifically, he noted that "they're in line with some things I've done in the past at other schools," with different, SBG architecture. Therefore, the assessment reforms were "like reading an old friend." Indeed, SHOD felt so strongly about assessment practices on campus that he enlisted early to play a major role in articulating the ABGA document. His science colleague who participated in the video-cued interview (SVC) was also well primed for its adoption. She spent the year before meeting with her department members discussing assessment practices and potential forthcoming changes.

Once the ABGA was adopted and in combination with the school's recent adoption of Next Generation Science Standards (NGSS), the SHOD's creativity in his own classes as well as leadership with others thrived. In adopting NGSS, there was both a need and an opportunity to rewrite curriculum and redesign assessments. As he put it, it was "a perfect storm kind of thing. It all came together at once." First, there were what he called the "mechanical" pieces. He worked hard to prepare rubrics that appropriately reflect NGSS standards and IST grade descriptors with his team. In that work, there was an intent to move away from traditional assessments, like multiple choice tests, to encourage students to articulate their thinking. In his general chemistry class he did not give a paper test all semester, having students do projects, create posters, and make

videos instead. Further, to ensure that his assessments are in line with the school's philosophy, he takes out the ABGA and measures the test against its points.

Aside from rubric writing, other mechanics include the creation of and reporting out of students' achievement on assessments. A key part of assessment writing is generating strong questions that lead students to engage with the material meaningfully. Although the SHOD did not specifically encourage his team to move away from paper tests, he worked with them to write the tests, moderate students' work on them, and reflect on the effectiveness of the tests and rubrics. The concerted move to rubrics and away from percent-style reporting meant that the policy of entering only letter grades into gradebooks went smoothly for the team in 9th and 10th grade general science courses. The SVC noted that they "work as a team to create the assignment and then we have rubrics and plan by the whole team and then we have marking scheme or sometimes rubrics." Also, as to formative assignments that once comprised part of the grade, they were now all collected as a portfolio and used as a requirement for students to qualify for reassessment. In other words, the team agreed that if a student wished to seek reassessment opportunities, s/he would need to demonstrate that s/he did the work of the course; this was evidence of trying to learn the material.

There remain two points of tension. The first is the relative freedom that the policy still gives its faculty. Coherence of practices within departments and even across classes of the same teacher has not been achieved. The SHOD interpreted the policy to have creative license to go beyond the intent of administration. At least one of his rubrics showed a conversion from achievement to letter grade that was both innovative and unconventional. It hinted at his admission that, after thinking and working through his

assessment practices, he would prefer to do away with grades altogether. On the other hand, both the SHOD and SVC still retain a number of traditional elements of the former system in IB and AP courses. For example, quizzes remain a staple in their assessment systems. They also continue to recycle old IB and AP tests as their summative assessments.

Another issue lies in the reassessment policy's impact on the teacher-learner relationship. Both SHOD and SVC are very aligned in the belief that the most recent evidence of student achievement is the more accurate, especially if it demonstrates growth. As the SVC noted, the team has "started disregarding the previous assessment if [they] are actually testing the same skill." They also both work hard to meet with students during content support time to help them learn their mistakes and do better. For those that don't take the opportunity, the SHOD mandates students who receive a letter grade of D or F to come in for help. However, the SVC finds that "in terms of the work ethics, I feel that it's not that strong." With the ever-present opportunity to reassess, she isn't confident students put their best effort forward on the first attempt. Whereas she felt she spent more time with students prior to the test to ensure they understood the content, now she spends more time with them after the test.

Math case study. In the math department, the ABGA was mobilized and translated by the head of department with his team at the beginning of the semester. The group went point by point connecting the beliefs and guiding practices to their work. The MHOD was part of the writing and adoption process so was well versed in the document. There were points that he and his team did not completely understand nor agree with, but the MHOD himself was very familiar with the document all the same. The others in his

department were not. The meeting at the beginning of the year was the only time that the document was formally translated by the team. Otherwise, it was referred to when a question arose from practice, particularly regarding reassessment policies.

As to the practices themselves, the MHOD spearheaded changes in his own classroom. For example, he no longer enters any quizzes into gradebooks as an assessment that counts toward the final determination of the grade; they were previously entered as minor summative assessments and now he only enters major summative assessments. This opened his practice to more students' self-assessments and group work. Specifically, he noted "suddenly it's like, well, if it's not going in the gradebook, why don't I give you the solution key as well and you can mark it on your own and ask questions." He noticed that this created more engagement and opportunities to learn from the assessments themselves. It inspired him to try to leave the final 10 to 15 minutes after a test to allow students to compare their responses. Again, he noted how the students thrived from the immediate feedback as well as saw how students attacked the different problems.

Another change the MHOD has made is to the tests themselves. In fact, department-wide "tests aren't going to look the way they used to." First, as previously noted, he aims to budget his time differently for tests, giving students time to review the tests at the end of the block. He also aims to spiral the content on his tests. Spiraling refers to addressing standards from previous units on current units. This was common practice for homework assignments but not tests themselves. He reasoned that the practice builds reassessment into his assessment systems without recreating old tests.

Finally, the tests are constructed very differently. Initially, tests were constructed and evaluated in a norm-referenced system. The MHOD noticed his team designed a test for an "ideal 'A' student," by manipulating the number of questions and degree of difficulty based on their students. His math colleague who participated in the video-cued interviews (MVC) explained that initially, upon evaluating students' work on a percentage scale, he would curve students' grades up to ensure a normal distribution of A, B, C, etc. After adopting the policy, tests are more descriptor-based and less percent-based. As the MHOD put it, they don't report "[students'] raw points that [they] got, but what level of understanding were [they] able to show?" The MHOD creates rubrics for his tests before writing the test. He also aims to ensure that there are sufficient opportunities to demonstrate both proficiency and mastery of the content. No longer leaning on percent, if students can answer one of three of the most difficult questions, they have proven mastery and will receive a letter-grade A.

Two changes continue to cause tension in the department. The first has to do with reassessment. Philosophically, the math teachers are in favor of enabling students to demonstrate their true achievement of the standards. They believe students grow over the course of a semester and wish to accurately report student achievement. On the practical level, it is a lot of work for teachers to recreate tests. Although the MHOD experimented with a spiraling assessment system, he struggled to keep up with it over the course of the semester. Other teachers find it is very time-consuming to make entirely new tests. This is compounded by students missing the scheduled reassessment. For the sake of test integrity, teachers already feel pressure to make two tests in case students miss the first test session. Missing a second opportunity means making yet another test.

The use of quizzes in the assessment system is the second node of tension. The math department feels pressured by the standards-based culture to do away with them. The argument against quizzes is that they do not offer students a true ability to master content – a core tenet of assessment design articulated in the ABGA. The MVC refuted the argument, stating that students in his class have ample opportunities to learn new, difficult material before a quiz and that his quizzes offered students an opportunity to demonstrate mastery of what was learned. Further, he insisted that students were not intrinsically motivated. Specifically, he noted how "we would like that students are intrinsically motivated to learn and do well, but obviously that's not the case." He continued the argument that quizzes motivate students to remain focused over the course of a four- to six-week unit.

Finally, a key piece to the department's ability to change practices has been the time they spend working on tests together. The MHOD takes more of a hands-off approach to leading his team. Although he will show what he does and model new ideas, he also defers to his team members who have an average of at least 15 years of teaching experience. The greatest opportunities for change lie in when teaching teams write tests.

Synthesis of Findings

Beliefs

The survey and interview results confirm that the beliefs of the faculty are somewhat to mostly aligned. As reported earlier, the questionnaire recorded a mean of 5.11 and 4.68 before and after the semester, respectively. The standard deviations of 0.68 and 0.58, respectively, indicate they were already somewhat to mostly aligned. This was the case in the science and math departments, as well. The heads of departments both

expressed enthusiasm and readiness for as well as experience with the new changes. The other members were also ready for the adoption of the policy and shared some reluctance to take on the full change, specifically as it pertains to the formative assessments.

Practices

The practices of the faculty were somewhat aligned before the policy adoption and mostly aligned after a semester of its adoption. The survey results indicate that the faculty were somewhat aligned before and after the semester, reporting lower means than the beliefs, 4.41 and 4.52, respectively. Yet, the results of from gradebook analysis, assessment analysis, and the interviews indicate that practices are more closely aligned than faculty self-reported. The gradebooks saw a steep reduction in the use of minor assessments in the science and math departments, 49% and 61%, respectively. Although the reduction of minor assessments alone does not confirm an alignment, it does indicate a shift in focus of faculty's use of minor assessments to measure learning. The reduction in the types of categories that faculty report also indicates a shift in focus. Further, there were changes to the tests and quizzes of both math and science classes that show intent to align those assessments with the ABGP. Comments made in the interview substantiate the claim that the way tests are made has undergone changes in both the science and math departments. Both the tests and the interview data show that rubrics and grade descriptors play a larger role in the assessment design. Again, results from the interviews indicate that the heads of departments were more innovative than the other members of the department. Finally, both before and after the policy was adopted, teachers created assessments that were driven by standards. The science department adopted new standards, which may have helped facilitate the change in the 9th and 10th grade courses.

Materials

The intended purpose of assessments has generally not changed; teachers use assessments to evaluate student achievement, report it out at the end of the semester, and record it on a student's transcript to be used for university applications. However, what has changed are the parts that make up the overarching grade for the student. There are fewer minor assessments and assignments recorded in the gradebook. Having said that, the raw data may be somewhat deceiving. Interviews confirmed that although many assignments are no longer reported as minor assessments in the grade book, those same assignments still exist. In both the math and science departments, for students to earn reassessment opportunities, they must demonstrate that they have done the assignments. The science department includes a portfolio assignment in their gradebook, which is a catchall for the assignments. Still, the fact that they are largely omitted from the gradebook shows some change and greater alignment. Further, the change has begun a movement toward experimenting with what had been previously used, by accepting them as portfolio assignments and using results on quizzes for peer and self-assessment. In the same vein of viewing the first semester as the beginning of a change, the substance of the assessments has started to shift. Both interview and document analysis show that teachers are rethinking the format of the tests, decreasing the number of simple questions in favor of more complex ones, and assigning more intentional weights aligned with degree of complexity.

Other

Leadership style, opportunities to collaborate, adoption of new standards, and professional development were all mentioned as other important factors that facilitated

adoption of standards-based assessment practices. Administrative leadership was mentioned as a reason for stalling the change. The science department's struggles with the gradebook's percentage system derived from inaction and wavering of decision-making. The MHOD also expressed frustration with "backtracking" which opened the "barn door," permitting teachers to revert to their usual practices. The SHOD took a more hands-on approach working with his team than the MHOD. The SHOD shared that one reason he was able to do so was that his department adopted new standards, NGSS, for the 9th and 10th grade classes. This factor was very important for science because it necessitated greater collaboration between teams and more coaching from him. The MHOD shared that most collaboration in his department occurred in teaching teams and specifically over assessment construction. The importance of those moments was highlighted when he shared a concern for the next academic year, when fewer teaching teams would exist in his department.

CHAPTER 5

DISCUSSION AND IMPLICATIONS

A funny thing happened when I got into my first classroom. Yes, I was innovative in some ways, using online discussion forums and interactive fishbowls for Socratic seminars. Although this may at first seem fairly tame in 2021, this was 2007, the year of the first iPhone and MySpace was still a thing. My creativity wasn't just in the digital space. My classes got out of the classroom to sit in the courtyards overdressed for Venezuelan tropical heat to enact French Revolution–styled salons. We even published public service announcements for the Salvation Army orphanage on a new platform called "YouTube," which was two years old at the time. These were mostly very good, engaging activities. I am happy that I took risks to engage my students. Still, I never had a very creative thought about assessment practice.

Curiously, many of the *things* in my classroom echoed my own experiences as a student. In high school I had textbooks, three-ring binders, overhead projectors, pencils, pens, and paper. As a teacher of high school students my classrooms still had those things, plus laptops, smartphones, Microsoft Word, Google, and so many more tools and resources on the internet. And still teaching was fundamentally the same. I substituted away from physical toward digital resources, yes. Students had access to immense resources online, yes. They could publish their own content for the world to see, yes. Yet I am unsure there was any real transformation. Especially in terms of assessment infrastructure, nothing tangible changed. Homework check – check. Journal entries – check. Pop quiz – check. Participation – check. Even extra credit – check. Sure, I had rubrics and linked them to Mid-continent Research for Education and Learning (McREL)

History standards the best I could, but we didn't have all-encompassing Common Core standards. Instead, I used IB exams as models for tasks and tests. This seemed logical and practical because students would be better prepared to take IB courses and its exams. I did not experience any alternative grading system from weighted percentages on teacher-discretion categories for at least my first seven years in the profession. When I was introduced to them, there was friction —some friction in my beliefs but mostly in my practice. Now in my fourteenth year in education I have learned that I was not alone. I embarked on this exploration wondering why. I conclude the investigation with an answer and many new questions.

Chapter 1 detailed the larger context of how long it took for standards-based grading to emerge as a major theme. Since at least 1894, there has been movement to have "uniformity of school programs" (National Education Association, 1894, p. 3). Only in the more recent decades has greater uniformity been achieved on the back of an expanding role of federal government. One consequence of adoption of standards has been a rethinking of how to assess and report student achievement. Chapter 1 also detailed both my own and my present school's journey with standards-based assessment practices. The difficulty of innovating a school's beliefs and practices about assessment was at the center of this dissertation. Rittel and Webber's wicked problems (1973) and Weick's metaphor for school as an "unconventional soccer match" helped me put words to the difficulty.

Chapter 2 outlined what Actor-Network Theory (ANT) was and how it could be used to cut through the wickedness of assessment reforms by breaking down its intricacies through the lens of groups, materials, and actions. Chapter 3 used the ANT

tools to outline the methods of this multi-stage convergent parallel mixed methods case study. Chapter 4 reported the findings through the lens of action research's pragmatism to detail the impact of the adoption of the assessment policy on the school. In this final chapter I will synthesize the pragmatic aspect of the investigation with the theoretical ANT lens and discuss the findings and implications of my research questions:

1. How does a new school assessment policy impact educators' assessment beliefs, practices, and infrastructure over one semester?
2. What are teachers' perceptions of factors that led to any changes in their beliefs and practices

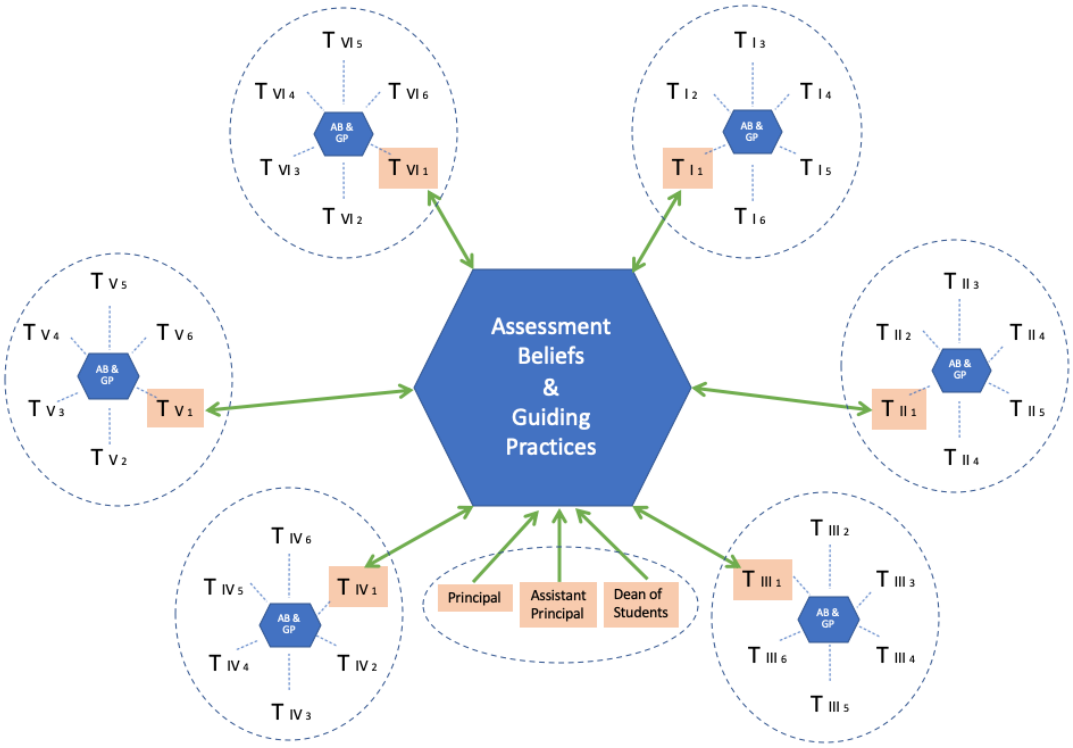
Discussion

Research Question One

> *In addition to 'determining' and serving as a 'backdrop for human action', things might authorize, allow, afford, encourage, permit, suggest, influence, block, render possible, forbid, and so on.* (Latouor, 2005, p. 71)

The first question relates to the impact the policy may have on assessment beliefs, practices, and infrastructure. The original thought was that the policy document may serve as an important material to help *determine* the school's assessment culture. ANT theory suggests that groups make and remake themselves through an object's mediation. If the Assessment Beliefs and Guiding Actions (ABGA) policy document successfully reified the beliefs of a significant number of key stakeholders, then both *the policy* and its creators could *influence* others by transferring those beliefs to them and guiding their practices, rendering change. This exploration found this to be a possible explanation for

the faculty's behavior over one semester. Figure 1 in Chapter 3 is shown again here to demonstrate how the policy was believed to potentially direct the efforts of the heads of department as well as the department's other members in their work on writing summative and formative assessments.

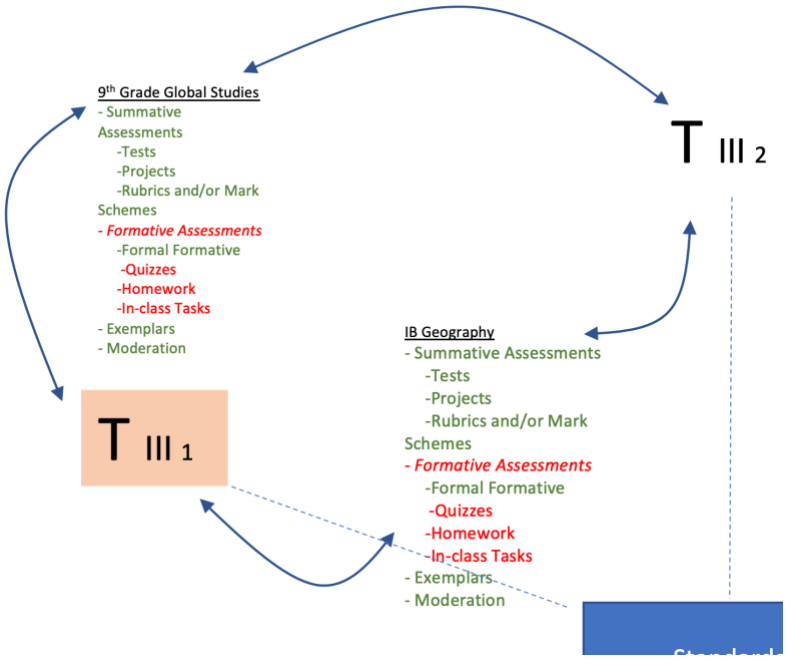


The ABGA is the centerpiece of the infrastructure. Having been worked on by the instructional leadership team, comprised of administration and heads of department, the ideas it reified are commonly shared. Findings from Chapter 4 show that the heads of departments proved to be very aligned and the most progressive in their practice. Once the ABGA was published, the model presumed that those already familiar with the assessment strategies would feel free to pursue them. Indeed, the science head of department (SHOD) specifically said that he felt he was given "quote unquote

permission" to do so. That he also took out the document as a reference for building assessments means that it greatly influenced his assessment processes.

For those whom the assessment strategies are uncharted territory the model posits that the ABGA would serve as a road map in collaboration with the heads of department and teaching partners. Again, the SHOD stated he took out his hard copy of the document and worked with his teams on this. The ABGA did not have the same impact on the math head of department (MHOD) and this is explored more when I discuss the implications of the research.

Figure 3 from Chapter 2 is reproduced here to demonstrate how other materials of the assessment infrastructure traced and translated actions of the faculty, TIII1 and TIII2.



At the center of the department's network are the key components of ABGA: standards alignment, use of grade descriptors over numbers, the gradebook procedures, rubrics, and moderation of assessments. The creation and use of these objects are the

assessment practices and they are informed by beliefs. When teachers collaborate, they do with these documents and through dialogue that *encourages* alignment of the department's beliefs and actions. It is possible that the more that the materials were worked on in collaboration with other actors, the more *influence* they may have. Beliefs are translated while working with and on the objects. It stands to reason that the more one is in contact with the object, the more the beliefs may be engaged and influenced. Based on both the MHOD and SHOD's experiences, this is true for tests and rubrics, where the ideas of the ABGA were translated to the tests. The actions of building the tests, speaking out the ideas to one another, and consolidating any differences in opinions may have helped *align* beliefs and actions between the people and the materials. The document analysis showed that the tests are more aligned to the ABGA. Further, as noted in the interview analysis of the video-cued math faculty member (VMC)'s practice, the practice of norm-referencing students has transformed to a standards-based and grade descriptor approach because of the way the tests are generated. As additional proof of the power of collaboration with materials, the MHOD expressed concern for the reforms if fewer teaching teams were possible due to budget cuts.

The model proposed in Chapter 2 proved to be incorrect in at least one way. In the aforementioned Figure 3, "grade book processes" were included as part of the ABGA. In fact, the ABGA did not specifically outline how to use the grade book. Despite this, the number and types of minor assessments became more coherent in both the science and math departments. However, the original intent of eliminating minor assessments and use of professional judgement in determining grades was not achieved. Teachers still rely heavily if not solely on an algorithm of specific weights given to certain types of

assessments that are automatically averaged and published as a final score. This idea of different categories and weights for different assessments was not explicitly stated in the ABGA. That is at least one reason why teachers use their gradebook is not consistent. This is most evident in the tension surrounding the purpose of having the category "minor assessments." There is a belief that they motivate students to do work to learn the material. In this way, it *blocked* further change in practices.

In conclusion, the ABGA was as effective as it was *touched*. Using the terminology of the Likert scale developed for the survey instrument, the beliefs and practices of the faculty can be best described as "somewhat aligned" to the policy both before the semester and observably more aligned after it. Although difficult to measure change solely from the survey, results from the document analyses and interviews show that there was observable improvement in accuracy, meaning, and coherence. This is evident in the way teachers create assessments, mark them, and enter them into their gradebooks. It is also evident in how they speak about the materials that shape their practices. The network that was created using ANT proved useful in exploring and understanding the impact of the ABGA's adoption.

Research Question Two

> *This, of course, does not mean that these participants [materials] 'determine' the action, that baskets 'cause' the fetching of provisions or that hammers 'impose' the hitting of the nail.* (Latouor, YEAR, p. 71)

The question of the social emerges when the ties in which on is entangled begin to unravel; the social is further detected through surprising movement from one association to the next. (Latour, YEAR, p. 247)

Latour preambles his introduction to ANT with a harsh critique of those that assign agency to a vague social force. The "social cannot be construed as a kind of material or domain" and cannot be used to provide an explanation (Latour, 2005, p.1). ANT uses the materials to best understand how humans organize and reorganize themselves. Yet, he also is careful to note at the end of *Reassembling the Social* that it isn't so much that the "laws of the social world" do not exist (p. 245). Rather, it is that they are not forces in and of themselves, "they are not behind the scene, above our heads and before the action, but after the action, below the participants and smack in the foreground" (p. 245). Latour calls this background a kind of *plasma*, "not yet formatted, not yet measured, not yet socialized, not yet engaged in metrological chains, and not yet covered, surveyed, mobilized, or subjectified" (p. 243). In that vein, there are at least three social forces not directly tied to the ABGA material that were unraveled from the exploration as plausible additional social factors that impacted the reform: leadership style, opportunities to collaborate, and professional development.

As to leadership, the two heads of department had different approaches to how they led their teams. The SHOD was more hands-on and directive. He worked closely with his team to consider questions, build rubrics, write tests, and design alternative assessments for his department in the 9th and 10th grade classes. It is in constructing, revising, and dialoguing *with these materials* that his style is visible. Further, the

adoption of new standards in the science department necessitated a greater role on his part. He received the most training in the new curriculum, making him the resident expert and voice for the new courses. So in some ways, the standards themselves empowered his leadership. The MHOD took a different approach. He modeled ideas, shared his own innovative practices, and allowed his team to take his suggestions. Again, his ideas took *shape* in his spiraling-assessments, use of quizzes, and development of rubrics. He was less hands- on, but he built a quiet trust, one that led them to come to him with questions. These two teacher leaders also served two departments that are comprised of very different demographics and histories. People with social capital within the group may have impacted the heads of department's capacity to lead.

Further to the point of leadership, both department heads expressed a desire to have stronger, clearer leadership from administration. The ABGA was fully adopted and published with talk of how it would specifically change some grading and reporting procedures. Those procedures were rolled back in the first few weeks of the academic year and left the heads of department wondering how to answer specific assessment questions for their team.

Linked to leadership is the opportunity to collaborate. The heads of departments were innovative in their practices and more innovative in the classes that they did not share a teaching partner. In the classes that they shared, they spent a great deal of time working on assessments, rubrics, and moderating work. According to the MHOD, some of those meetings could be contentious. It was in those moments that their potential to lead was greatest.

Professional development was also an important factor. The science department met frequently with the curriculum coordinator in professional learning teams to discuss assessment practices. An expert was brought to campus to discuss the philosophy behind standards-based assessments. Feedback on that session was largely positive and helped motivate the faculty to move forward. Referencing the administration's retreat from adopting specific gradebook procedures again, the momentum from those sessions early in the semester was largely lost.

Implications

> *The figured world is mediated and materialized through cultural artefacts and can thus lever change in social practices* (Hamilton, 2012, p. Location 1999)

If materials do have agency, it is when they are *touched* by human actors. The materials that are touched most often have the most potential agency, the most leverage to reconfigure the world. Assessment practices use specific materials, rubrics, quizzes, tests, and gradebooks. All of these materials are nodes where ideas are communicated. In my practice, the administrative team and I have already begun to look more carefully at these nodes as opportunities for the 2022-2023 academic year.

First, as to the ABGA document, there is a movement to adopt a user-friendly version of it. The policy was originally published as a four-page outline with beliefs written in one color and actions written in another. The dean of students, curriculum coordinator and I will simplify that document so that it resembles more of a checklist. Heath & Heath (2010) explain the power of checklists in air travel, intensive care units,

and business acquisitions at Cisco to build better habits. “You may not save a life, but you’ll sure avoid a painful blind spot” (p. 259). The idea is that if the school wants its tests to be adopted with fidelity to the policy, then it can design the document to be *touched* more often if it is practical to its faculty.

The second move is to be explicit with the gradebook procedures. For example, the school will not include as a potential category “minor assessments.” Instead, it will include “formative” as a category that carries zero weight and “summative” that carries all of it. The school’s gradebook software offers a few measures of central tendency to be used to calculate final grades, mean, mode, and median. The mean is currently used. It will be proposed to consider the use of the median to generate a *drafted* grade for faculty to then employ their *professional judgement* in determining a final grade. This type of shift will require significant dialogue, which is the third move. Seeing as how the most aligned individuals to ABGA were those who had been embedded in the process, the administration will engage as many people as possible in the construction and adoption of the checklist and gradebook procedures.

New iterations of rubrics, tests, and other summative assessments will emerge. Some standardization of practice must be determined. This kind of work requires strong leadership from the heads of department and administrators. Leadership style proved to be important part of *how* and *how often* materials were touched, so a common set of expectations for heads of departments can be articulated. More than just a clear set of expectations is a need to develop the capacity of the leaders. The skills of facilitating department meetings, moderation of work, and writing of rubrics and tests will need development.

ANT was very useful in seeing the way that individuals in the school interacted with one another. It helped identify specific moments where important action takes place. Knowing where the action takes place helps one influence it. It can be used for a number of processes on campus outside of assessment. The finance team has a process for budgets. The facilities team has a process for maintenance. The office of professional learning and curriculum has its own software for teachers to document the school's curriculum and processes for suggesting or approving professional development. Flattening these systems by carefully considering the materials used will help find the nodes that action takes place and help influence it.

For future research, I am interested in exploring the power dynamics of grades. The theme of grades as motivators was an original curiosity of mine in the earliest cycles of research. Specifically, it persistently came up in interviews regarding the use of quizzes in units to motivate students to work and learn. Similarly, the power of gradebook software is a new interest of mine.

Limitations of This Study

One of the greatest limitations to my exploration was the survey. The survey itself proved to be valid enough. However, there was a very low response rate on both the pre- and posttest. Further, I chose not to collect any identifying information on it so as to solicit the most genuine responses. I reasoned that my position as an administrator, someone who evaluates the faculty, may turn people away. This meant that I could not use a two-tailed t-test to measure individuals' change. No concrete conclusions could be inferred from these results.

The rest of my data collection was conducted on two departments out of a total of six. This limits any conclusions to those departments. My original intent was to examine every department. Yet, as I built the data instruments and methods, it was clear that it would be untenable. Still, I resisted only doing one department because it was important that I be able to compare the experiences of two departments.

The most comprehensive instrument I used was the document analysis of the gradebooks, where I looked at every gradebook for all faculty in the two departments. Conclusions on the impact of numbers and types of assessments are therefore most representative. However, the other instruments, interviews and tests, were not. I selected two members from each department. Again, this decision was made out of pragmatism. However, I resisted interviewing only one member. It was important that I understand the perspective of both the head of department as well as another member of that department to juxtapose those views. Because there were still four other members of each department from whom I did not collect data, any conclusions on the department overall and comparisons between HODs and other members must be treated as tentative.

Another limitation is the timescale of my research. One semester is only five months. Human and organizational change is a process, not a moment. Assessment reforms are no exception.

Final Reflections

In the spirit of ANT, my final reflection will be on what agency this document, my dissertation, has had on me. I have spent a significant amount of time on this work. I have *touched* this document for at least two years. In turn, it has *touched* me. For example, every citation traces an action, a document read and annotated, a summary

written in my notes and tagged with a main idea, to finally be pulled together and *translated* by me into a new reified idea – communicated, here. Noteworthy is the timescale of the citations, with some reaching at least as far back as the conference of the Committee of Ten; their work continues to impact the work of aspiring leader-scholars like me. Practically all the citations that inform this work were drawn from ASU's One Search library research page, a boundary object styled like Star and Griesemer's (1989) subject of study, the museum, where the work of scientists are stored, catalogued, and sorted for aspiring educators, doctors, engineers, and researchers.

Arizona State University's (ASU) Doctor of Education (EdD) brought a group into being. Within my 2021 ASU EdD cohort, a smaller group formed for peer-review and moral support. Our dissertations and degree requirements brought us together as a network. We shared our thoughts on research methodologies and theoretical perspectives as well as drafts of chapters. One by one, we defended the dissertations and completed the requirements, loosening the ties of the network, rendering the group's meaning less practical – it may dissolve altogether. For now, the nodes that continue to connect us are tweets and a hashtag, #ASUEdD.

Perhaps the most lasting personal impact lies in how each chapter traces a particular skill that now informs my professional practice. Chapter 1 is a pause to examine from a bird's eye view the forest for the trees before diving down to detail the terrain up close and personally. This I continue to do on school policy related to our program of study, attendance, advisory programs, and student leadership programs. Chapter 2 identifies connections between the problem of practice to theory to help anchor both the methods of undertaking an actionable solution as well as how to evaluate it. My

mind now consistently moves between problems and theory. Our new advisory program sits in community theory. My counselor's push for earlier information for students and their parents on colleges is centered in navigational capital. And my entire career in international education rests in a strange space of critical theory as I ponder what role they and I play in persisting power dynamics of the global north and south. Chapter 3 sets up the work to be done. I used the skill of forward planning the evidence and methods of evaluation in my administrative council's action plan that was submitted to our accreditation board. I ensured that the action plan document included evidence of accomplishing a task and means of assessing it. Further, our teacher performance evaluation system (TPES) is quasi-action research. Its framework includes a goal, evidence that informs the goal, action to be taken to reach the goal, and means for assessing it. Combining my skills from Chapter 2, I'd like to become fluent enough in multiple theories to be able to center their work in theoretical perspectives. My point, though, is that chapter 3 has enabled me to feel fluent in thinking through ways that I can evaluate others' and my own work. Chapter 4 presented in this dissertation traces far less than I actually gathered. A peak at the breadth can be seen in the appendices. That is the skill I have taken with me, the ability to distill information to its most salient points. Admittedly, it may also be the space where I have the most learning still do to. And here we are in Chapter 5. At first glance it bookends this journey. What I am trying to express in these final reflections is that it is simply a marker. In *closing* this final paragraph, what is really happening is that this dissertation is *rendering possible* another set of opportunities that I am excited to pursue.

REFERENCES

Abu Dhabi Education Council. (2017). Irtiqaa Inspection Report for American Community School. Abu Dhabi, UAE. Retrieved from https://www.adec.ac.ae/en/education/keyinitiatives/pages/irtiqaa-reports.aspx

Altheide, D. (2008). Ethnographic content analysis. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 288-288). Thousand Oaks, CA: Sage. doi: 10.4135/9781412963909.n149

Amantino de Andrade, J. (2017). Analysis of public action instrumentation from the perspective of the actor-network theory: social technology and rural education in Rondônia. *Revista de Administração Pública*, *51*(3), 407–430. Retrieved from http://search.proquest.com/docview/1908620089/

American Education Reaches Out. (2020). *AERO*. Retrieved from http://www.projectaero.org/

AP at a Glance. (2019). *College Board: AP Central.* Retrieved from https://apcentral.collegeboard.org/about-ap/ap-a-glance

ASCD. (2020). *Understanding by Design Retrieved*. from http://www.ascd.org/research-a-topic/understanding-by-design-resources.aspx

Bandura, A. (2005). The evolution of social theory. In K. G. Smith & M. A. Hitt (Eds.), *Great minds in management* (pp. 9-35). Oxford: Oxford University Press.

Bandura, Davidson, Davidson, Hannah Frances, Davidson, John, & Alexander Street Press. (2003). *Bandura's social cognitive theory an introduction*. San Luis Obispo, CA: Davidson Films.

Butchart, R., & McEwan, B. (1997). *Classroom Discipline in American Schools*. Albany: State University of New York Press.

BASIS International School Schenzen (2020). *BASIS Curriculum Schools Story*. Retrieved from http://www.basisinternationalsz.com/about-basis-shenzhen/the-basis-story.php

Bassey, M. (1999). *Case study research in educational settings*. Open University Press.

Bates, R. (2011). *Schooling internationally globalisation, internationalisation, and the future for international schools* (1st ed.). Routledge.

Benaquisto, L. (2008). Axial coding. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 52-52). Thousand Oaks, CA: Sage. doi: 10.4135/9781412963909.n30

Biberman-Shalev, Liat, Sabbagh, Clara, Resh, Nura, & Kramarski, Bracha. (2011). Grading Styles and Disciplinary Expertise: The Mediating Role of the Teacher's Perception of the Subject Matter. *Teaching and Teacher Education: An International Journal of Research and Studies, 27*(5), 831-840.

Butchart, R., & McEwan, B. (1997). *Classroom Discipline in American Schools*. Albany: State University of New York Press.

Callon, M. (1984). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *The Sociological Review, 32*(1_suppl), 196–233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

Callon, M. (1986). The Sociology of an Actor Network: The Case of the Electric Vehicle. In Callon, M., Law, J., & Rip, A. (Eds.). *Mapping the dynamics of science and technology : sociology of science in the real world.* Macmillan.

Carter, P. L., Welner, K. G. (2013). *Closing the Opportunity Gap: What America must do to give every child an even chance.* Oxford: Oxford University Press.

Charmaz, K. & Bryant, A. (2008). Grounded theory. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 375-376). Thousand Oaks, CA: Sage. doi: 10.4135/9781412963909.n189

Cicciola, E., Foschi, R., & Lombardo, G. (2014). Making up intelligence scales: De Sanctis's and Binet's tests, 1905 and after. *History of Psychology, 17*(3), 223-236.

Council of Chief State School Officers. (2007, May). *State Content Standards*. Retrieved from http://programs.ccsso.org/content/pdfs/2006-07%20Content%20Standards%20FINAL.pdf

Crawford, A. (1936). A History of Examinations : An Historical Study of Examinations and Grading Systems in Early American Universities. *Journal of Higher Education, 7*(5), 283-284.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles: Sage.

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Boston, MA: Pearson. (with MyEducationLab access) ISBN: 9780134519364.

Dubrow, G. (2009). Committee of ten. In E. F. Provenzo & A. B. Provenzo (Eds.), *Encyclopedia of the social and cultural foundations of education* (pp. 150-150). Thousand Oaks, CA: Sage. doi: 10.4135/9781412963992.n74

Durm, M. (1993). An A is not an A is not an A: A History of Grading. *The Educational Forum* 57(3), 294-297. doi: 10.1080/00131729309335429

Edwards, R. (2012). Translating the Prescribed into the Enacted Curriculum in College and School. In Fenwick, T., & Edwards, R. (Eds.). *Researching education through actor-network theory* (pp. 23-39). Malden, MA: John Wiley & Sons.

Erdley, C., & Asher, S. (1994). *Social-Cognitive Processes of Behaviorally Withdrawn Children*. S.l.]: Distributed by ERIC Clearinghouse.), aggressive (Kennedy, E., & Perry, D. (1993). *Social-Cognitive Correlates of Aggression and Victimization*. S.l.]: Distributed by ERIC Clearinghouse.

Fenwick, T. (2011). Reading Educational Reform with Actor Network Theory: Fluid spaces, otherings, and ambivalences. *Educational Philosophy and Theory: Actor-Network Theory in Education: A Focus on Educational Change*, *43*(Suppl), 114–134. https://doi.org/10.1111/j.1469-5812.2009.00609.x

Fenwick, Tara, and Richard Edwards. Actor-Network Theory in Education, Taylor & Francis Group, 2010. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/asulib-ebooks/detail.action?docID=544021.

Fjeldsoe, B., Miller, Y., Marshall, A., & Kazak, Anne E. (2013). Social Cognitive Mediators of the Effect of the MobileMums Intervention on Physical Activity. *Health Psychology*, *32*(7), 729-738.

Fowler, F. (2014) *Survey Research Methods* (Fifth Edition). Thousand Oaks, CA: Sage.

Fox, S. (2005). An actor-network critique of community in higher education: implications for networked learning. *Studies in Higher Education*, *30*(1), 95–110. https://doi.org/10.1080/0307507052000307821

French, M., Homer, J., Popovici, I., and Robins, P. (2014). What You Do in High School Matters: High School GPA, Educational Attainment, and Labor Market Earnings as a Young Adult. *Eastern Economic Journal*, *41*(3), 370-386.

Fullan, M. (2015) *The New Meaning of Educational Change*. Teachers College Press.

Gemus, J. (2010). College Achievement and Earnings. *IDEAS Working Paper Series from RePEc*, IDEAS Working Paper Series from RePEc, 2010.

Great Schools Partnership. (2014, April). *Criterion-referenced Test*. The Glossary of Education Reform. Retrieved from https://www.edglossary.org/criterion-referenced-test/

Greene, P. (2020, January 20). Common Core is dead. *Forbes*. Retrieved from https://www.forbes.com/sites/petergreene/2020/01/30/common-core-is-dead-long-live-common-core/#162df9165e65

Guskey. T. (2014) *On your mark.* Solution Tree.

Guskey, T. (2002). Professional Development and Teacher Change. *Teachers and Teaching : Theory and Practice, 8*(3), 381-391.

Hamilton, L. S., Stecher, B. M., Yuan, K., & Rand Coroportation. (2008). *Standards-Based Reform in the United States: History, Research, and Future Directions*. Washington, DC: Rand.

Hamilton, M. (2012). Unruly Practices: What a sociology of translations can offer to educational policy analysis. In Fenwick, T., & Edwards, R. (Eds.). *Researching education through actor-network theory* (pp. 23-39). Malden, MA: John Wiley & Sons.

Hargreaves, A., & Fullan, M. (Eds.). (2009). Change wars. Bloomington, IN: Solution Tree. ISBN: 9781934009314

Hayden, M. (2006). *Introduction to International Education: International Schools and their Communities*. Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781446213292

Heath, C., & Heath, D. (2010). *Switch: How to change things when change is hard.* New York, NY: Broadway Books. ISBN: 9780385528757

History and Innovation. (2020). *National Assessment of Educational Progress*. Retrieved from https://nces.ed.gov/nationsreportcard/about/timeline.aspx

History of the first SAT. (n.d.) *Frontline.* Retrieved from https://www.pbs.org/wgbh/pages/frontline/shows/sats/where/timeline.html

How is Grade Point Average Calculated? (2011, April 7). Retrieved from https://nces.ed.gov/nationsreportcard/hsts/howgpa.aspx

International School Search (2019, October 1). International Schools – What's the difference. *International School Search.* Retrieved from https://www.internationalschoolsearch.com/news/international-schools-whats-the-difference#:~:text=In%202009%2C%20the%20IASL%20(International,referenced)%20international%20school%20criteria%20checklist%3A&text=A%20moving%20population%20(higher%20than,%E2%80%93%20DP%2C%20MYP%20etc.)

ISC Research. (2020a, June). *Data and intelligence on the world's K-12 international schools market*. Retrieved from https://www.iscresearch.com/

ISC Research. (2020b, July). *Data and Intel*. Retrieved from https://www.iscresearch.com/data

Klein, D. (2002). *A Brief History of American K-12 Mathematics Education in the 20th Century*. Distributed by ERIC Clearinghouse.

Kuznets, S. (1934) National Income, 1929-1932. Senate Document No. 124, 73rd Congress, 2nd Session.

Latour, B., & Latour, B. (1988). *The pasteurization of France*. Harvard University Press.

Latour, B. (1984). The Powers of Association. *The Sociological Review*, *32*(1_suppl), 264–280. https://doi.org/10.1111/j.1467-954X.1984.tb00115.x

Law, J. (1987). On the Social Explanation of Technical Change: The Case of the Portuguese Maritime Expansion. *Technology and Culture*, *28*(2), 227–252. https://doi.org/10.2307/3105566

Law, J. (2008). Actor network theory and material semiotics. In B. S. Turner (Ed.), *Blackwell companions to sociology: The new Blackwell companion to social theory*. Blackwell Publishers. Credo Reference: https://login.ezproxy1.lib.asu.edu/login?url=https://search.credoreference.com/content/entry/bksoctheory/actor_network_theory_and_material_semiotics/0?institutionId=2261

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meaning in ecosocial systems. *Mind, Culture, and Activity*, *7*(4), 273-290.

LinkedIn (2020). *Escuela Bella Vista*. Retrieved from https://www.linkedin.com/company/escuela-bella-vista#:~:text=Escuela%20Bella%20Vista%20was%20founded,pre%2Dkindergarten%20program%20were%20added.

Marshall, J. (2019). *Introduction to Comparative and International Education* (Second Edition). Thousand Oaks, CA: Sage.

Martin Moreno, C., Rivera, D., Hekler, E., Peet, M., & Tsakalis, K. (2016). *A System Identification and Control Engineering Approach for Optimizing mHealth Behavioral Interventions Based on Social Cognitive Theory* [ProQuest Dissertations Publishing]. http://search.proquest.com/docview/1825302917/

McGuinn, P. (2015). Schooling the State: ESEA and the Evolution of the U.S. Department of Education. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 1(3), 77-94. Russell Sage Foundation. Retrieved February 5, 2019, from Project MUSE database.

McREL International. (2014) Retrieved from http://www2.mcrel.org/compendium/kSkillsIntro.asp

McTighe, J., & Wiggins, G. (2004). *The Understanding by Design professional development workbook.* Alexandria, VA: Association for Supervision and Curriculum Development.

Meshkani Z. (2004). High school achievement as a predictor for university performance. *Journal of Medical Education, 5*(1), 11-12.

Muniesa, F. (2015). Actor-Network Theory. In *International Encyclopedia of the Social & Behavioral Sciences* (Second ed., pp. 80-84). Elsevier.

Nagrath, C. (2011). What Makes a School International? *TIE Online.* Retrieved from https://www.tieonline.com/article/87/what-makes-a-school-international

National Education Association of the United States, Committee of Ten on Secondary School Studies. *Report of the Committee of Ten on Secondary School Studies: With the Reports of the Conferences Arranged by the Committee*. Published for the National Educational Association by the American Book, 1894. Web.

Ojeda, Lizette, Flores, Lisa Y., & Navarro, Rachel L. (2011). Social Cognitive Predictors of Mexican American College Students' Academic and Life Satisfaction. *Journal of Counseling Psychology,58*(1), 61-71.Pezzi, F., Donelli, T., & Marin, A. (2016). School Failure in the Perception of Adolescents, Parents and Teachers. *Psico-USF, 21*(2), 319-330.

Parker, E. (2017). An actor-network theory reading of change for children in public care. *British Educational Research Journal*, *43*(1), 151–167. https://doi.org/10.1002/berj.3257

Phillips, D., & Schweisfurth, M. (2014). *Comparative and international education : an introduction to theory, method, and practice* (Second edition.). Bloomsbury Academic.

Pierce, C. (2015). Learning about a fish from an ANT: actor network theory and science education in the postgenomic era. *Cultural Studies of Science Education*, *10*(1), 83–107. https://doi.org/10.1007/s11422-013-9498-3

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155-169.

Saito, H. (2010). Actor-network theory of cosmopolitan education. *Journal of Curriculum Studies*, *42*(3), 333–351. https://doi.org/10.1080/00220270903494261

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Thousand Oaks, CA: Sage.

Schimmer, T., Hillman, G., Stalets, M. (2018). *Standards-based learning in action.* Bloomington, IN: Solution Tree Press.

SCHOOL (2015). *SCHOOL Self-Study Report Presented to WASC*. (report).

Shippee, N., & Owens, T. (2011). GPA, Depression, and Drinking: A Longitudinal Comparison of High School Boys and Girls. *Sociological Perspectives, 54*(3), 351-376.

Shi, R., & McLarty, J. W. (2009). Descriptive statistics. *Annals of Allergy, Asthma & Immunology, 103*(4), 9-14. doi:10.1016/s1081-1206(10)60815-0

Shippee, N., & Owens, T. (2011). GPA, Depression, and Drinking: A Longitudinal Comparison of High School Boys and Girls. *Sociological Perspectives, 54*(3), 351-376.

Shulman, L. S. (2002). Making differences: A table of learning. *Change: The Magazine of Higher Learning*, 34:6, 36-44, DOI: 10.1080/00091380209605567

Simons, H. (2009). *Case study research in practice*. Retrieved from https://ebookcentral-proquest-com.ezproxy1.lib.asu.edu

Smith & Glass (1987) section on Multiple Regression in Smith, M. L., & Glass, G. V. (1987). Correlational studies. In M. L. Smith and G. V Glass, *Research and Evaluation in Education and the Social Sciences* (pp. 198-224), Needham Heights, MA: Allyn and Bacon.

Standards for Accreditation and Requirements for Affiliation (2018). Retrieved from https://www.msche.org/standards/

Star, S. L. (2010). This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values*, *35*(5), 601–617. https://doi.org/10.1177/0162243910377624

Star, S., & Griesemer, J. (1989). Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*(3), 387-420.

Starch, D., & Elliott, E. (1912). Reliability of the Grading of High-School Work in English. *The School Review*, *20*(7), 442–457. https://doi.org/10.1086/435971

Starch, D., & Elliott, E. (1913). Reliability of the Grading of High-School Work in Mathematics. *The School Review*, *21*(4), 254-259. http://www.jstor.com/stable/1076246

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*(77), 77-100. doi:10.1177/2345678906292430

The story of ACT. (2019). Retrieved from https://www.act.org/content/act/en/about-act.html

The history of the IB. (2017). Retrieved from AP at a Glance. (2019). *College Board: AP Central.* Retrieved from https://apcentral.collegeboard.org/about-ap/ap-a-glance

The World Bank. (2020). *GDP(Current US$).* Retrieved from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

The World Bank. (2017). *Development Indicators: Size of the economy*. Retrieved from http://wdi.worldbank.org/table/WV.1

Timescales. 2011. In *Merriam-Webster.com*. Retrieved May 8, 2011, from https://www.merriam-webster.com/dictionary/timescale

Timmons, J. (2018). *Learning Architectures in Higher Education: Beyond Communities of Practice*. Bloomsbury Academic.

Tobia, E. (April 2007). The Professional Teaching and Learning Cycle: Implementing a Standards-Based Approach to Professional Development. Retrieved from http://www.sedl.org/pubs/sedl-letter/v19n01/professional-teaching-and-learning-cycle.html

Tobin, J. (1991). *Preschool in Three Cultures: Japan, China and the United States*. Princeton, NJ: Yale University Press.

Tobin, J. (2019). The Origins of the Video-Cued Multivocal Ethnographic Method. *Anthropology & Education Quarterly, 50*(3), 255-269.

Tobin, J., Hsueh, Y., & Karasawa, M. (2009). Preschool in three cultures revisited : china, japan, and the united states. Retrieved from: https://ebookcentral-proquest-com.ezproxy1.lib.asu.edu/lib/asulib-ebooks/detail.action?docID=471788

Tyler, R. (1981). Curriculum development since 1900. *Educational Leadership, 38*(8), 598-601.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative ScienceQquarterly*, 1-19.

Wiggins, G., & McTighe, J. (2011). *The Understanding by Design guide to creating high-quality units.* Alexandria, VA: ASCD.

Wayne, U. (2011). Leaders in the Historical Study of American Education Clifferd, G. J. (2011). Serendipity, Or So It Would Seem. Leaders in the Historical Study of American Education

Zhang, Z., & Heydon, R. (2016). The changing landscape of literacy curriculum in a Sino-Canada transnational education programme: An actor-network theory informed case study. *Journal of Curriculum Studies*, *48*(4), 547–564. https://doi.org/10.1080/00220272.2015.1090626

APPENDIX A

EXCERPTS FROM IST'S HIGH SCHOOL STUDENT HANDBOOK

Assessment, Grading and Reporting

IST Assessment Pre-K–12: Core Beliefs

1. The main purpose of assessment is to communicate to parents and students the level of student achievement.

2. Assessment must be fair and equitable.

3. Teachers should use a variety of assessment techniques to inform instruction, check for understanding, and provide valuable feedback.

4. Comprehensive assessment combines formative and summative techniques.

5. Domains such as behavior, attitude, and effort should be assessed and reported separately from actual achievement grades.

6. Assessment results should be made accessible to students and parents as soon as they are available.

7. Students should know and understand measurement criteria before assessment is carried out.

8. Assessment should be used as neither reward nor punishment.

9. Assessment should be clearly aligned with the school's POP's

Definitions

- Assessment is the continual process of gathering and analyzing evidence of student progress for the purposes of evaluation and the provision of meaningful feedback on student achievement, with the ultimate goal of supporting student learning.

- Standards-based Assessment derives from clearly identified criteria upon which learning is measured.

- Evaluation is the assigning of a mark or grade, based on professional judgment of student progress against objective criteria, to represent a level of academic achievement.

- Formative Assessment is conducted throughout instruction to provide preliminary evaluation and promote learning through regular feedback.

- Summative Assessment is conducted at the end of a specified period of instruction to allow students to synthesize and demonstrate knowledge and skills acquired and refined over time, and to provide comprehensive evaluation of student learning.

- Balanced Assessment integrates formative and summative assessment from a variety of sources and through a variety of methods.

- Diagnostic Assessment is primarily used prior to enrolment to make appropriate decisions regarding admission, placement, and support. Teachers may also design and administer diagnostic assessments at the beginning of a course or unit to determine prior knowledge and skills.

- Reporting is the formal method of informing students, parents, and appropriate outside institutions of student progress through Report Cards, Progress Letters and Transcripts.

- Integrity is adherence to moral and ethical principles, soundness of moral character, honesty.

High School

The IST High School grading system follows the National Association of Secondary School Principals (NASSP) standards and guidelines for assigning grades.

On school transcripts, grades shown are as follows (*only IB HL and AP courses are weighted):

Percentage Value	*Letter Grade*	*Regular Grade Point*	*Weighted Grade Point**
93 - 100	***A***	***4.00***	***4.50***
90 – 92	***A-***	***3.67***	***4.17***
87 – 89	***B+***	***3.33***	***3.83***
83 – 86	***B***	***3.00***	***3.50***
80 – 82	***B-***	***2.67***	***3.17***
77 – 79	***C+***	***2.33***	***2.83***
73 – 76	***C***	***2.00***	***2.50***
70 – 72	***C-***	***1.67***	***2.17***
67 – 69	***D+***	***1.33***	***1.33***
63 – 66	***D***	***1.00***	***1.00***
60 – 62	***D-***	***0.67***	***0.67***
59 or below	***F***	***0.00***	***0.00***

- A “C” grade represents satisfactory completion of all assignments at a minimum level.
- A “B” grade represents work that is very good and shows signs of high achievement.
- An “A” grade is reserved for students whose work achieves excellence and is clearly exceptional

APPENDIX B

IST ASSESSMENT BELIEFS AND GUIDING ACTIONS, 2019–2020

We, the members of Ruamrudee International School, believe assessment informs and shapes our learning community, cultivating student growth through purposeful **design, evaluation, feedback, reflection,** and **reporting**.

From the Purpose Statement above, the Assessment Committee identified five overarching, interrelated areas: **Design, Evaluation, Feedback, Reflection,** and **Reporting**. For each, the group identified belief statements, which are in blue, that stem from our schoolwide Collective Commitments (see below). The committee then determined inseparable guiding action statements, or policies, which are in green, that reflect those beliefs. The committee hopes that this clearly communicates the beliefs and priorities of our community, as well as how these beliefs are portrayed in action. Please see the Appendix for the Common Assessment Guide, Meta Rubric, Moderation Process, Strategies *(need to revise)*, Assessment Design: Some Key Questions, Assessment Strategies *(definitions)*, and a glossary of terms.

We believe that assessments are **designed** to:

a. Reflect learning experiences related to our guaranteed, viable curriculum
   a. All questions, prompts, and tasks relate to learning experiences.
   b. It is acceptable to ask students to reasonably extend knowledge that has been taught, e.g., to reach for an "E" (ES/MS) or an "A" (HS), so long as the criteria is related to the curriculum.

b. Be aligned to documented course standards
   a. We use backward design when planning.
   b. All assessments are aligned to specific course standards.

c. Allow students with varied abilities to demonstrate their level of mastery in multiple ways
   a. Multiple types of assessments or modes of assessment are used or offered to all students without limitations of time, etc.
   b. Accommodations/modifications listed in a Student Support Plan must be followed
      i. Modifications must be pre-approved by the Student Support Team and reflected during the reporting process
      ii. Teachers should consult the Student Support Plan and case manager to determine any necessary modifications
      iii. Modifications should be planned and implemented before an assessment

d. Provide students with opportunities to demonstrate mastery of the standard
   a. Assessment design allows students to show proficiency at the highest level for each reporting standard by the end of the reporting period. Therefore, there needs to be a way for a student to be "Exemplary" or to "make complex or new connections" for each Guaranteed Standard on one or more assessment(s) within a unit.
      i. Examples of assessments that might not allow demonstration of exceeding the standard include exit tickets, quick checks for understanding, etc.

e. Have a clearly communicated purpose and method

a. Assessment methods, purposes, and/or expectations are communicated clearly to students at the beginning of each unit of study, prior to any actual assessment activity, so students know what they need to do.
b. Rubrics or other helpful, guiding documents, such as mark schemes, are to be shared with students toward the beginning of the unit.
c. Exemplars are encouraged, when appropriate.

f. Be created collaboratively

a. PLTs work together to create and/or plan all common assessments (see the common assessment document)
b. Rubrics or mark schemes are created collaboratively, so common expectations are understood prior to teaching a unit
c. Ideally offer moderation of exemplars before teaching *(Possibly link the Meta-rubric)*

g. Engage students

a. Many aspects should be considered when choosing and creating assessment tasks.
b. Engagement includes student interest, relevance, authenticity, autonomy, collaboration, higher-order thinking skills, and self-assessment.

h. Be developmentally appropriate

a. Use professional judgment and assess at the appropriate grade level, especially with multi-grade band standards.

We believe that the **evaluation** of student performance:

i. Is based upon previously determined rubrics, criteria, mark schemes, and other evaluative tools

j. Allows for multiple opportunities for learners to demonstrate their knowledge, skills, and understanding

k. Communicates student proficiency through agreed-upon systems that use descriptors or commonly defined levels of learning, such as EMPB, ABCD, 1–7 (IB), or 1–5 (AP).

l. Assesses behaviors separately from academic standards

a. Evaluations are not used punitively to address non-academic issues.
b. Feedback can be given on both academic performance/evidence as well as behavior but in clearly separated grades or comments.
c. Academic dishonesty will result in behavioral consequences per each section-specific handbook. These actions will not impact a student's academic grade and are instead reflected in his or her Habits of Mind grade.

m. Is a result of professional collaboration and moderation to ensure consistency, reliability, and fidelity in the assessment of student learning

a. Evaluation is consistent and standardized among course sections and across school years.

We believe that effective **feedback** should:

n. **Be timely, continual, and actionable**
o. **Provide a description of what students need to do in order to meet a learning goal in relation to their current level of achievement**
p. **Be based on skills or knowledge required to achieve standards and/or Habits of Mind**
q. **Constitute a system of consistent communication to students and parents**
    a. Examples of feedback may include conferencing, written feedback, oral feedback, feedback using digital tools, and peer feedback.

r. We believe that **student reflection** should:

s. **Be developed by explicitly teaching reflective skills that are essential for students to think deeply about their performance**
    a. Skills are explicitly taught and developed in the context of guaranteed or additional standards.
    b. Reflection best happens both individually and collaboratively.
t. **Provide ongoing opportunities to act on feedback from assessments to strengthen learning**
    a. Students are asked to reflect on their performance on standards both formally and informally
        i. Informal examples include intentional pauses to allow for processing, quick verbal conferences, short bulleted lists, journal entries, self-assessing on a rubric, or emoticons to note how they feel or how hard they worked.
        ii. Formal examples include:
            1. Students tracking their progress toward meeting the Guaranteed Standards and Habits of Mind
            2. Portfolios or other collections of evidence of work
            3. Linking certain questions to given standards on written assessments
            4. Rubrics or mark schemes with full descriptors or criteria
            5. Checklists
    b. Reflection is most useful when it is timely and continual.
        i. Reflection occurs at all stages in the learning process.
        ii. Examples include reflections during tasks, on complete units, or semester-long and can be about a task or a process and can include Habits of Mind.
u. **Help students know what skills they have mastered, allow them to plan their progress, and set realistic and timely goals**
v. **Be student-centered and habitual**
w. **Develop students' goal-setting skills and build ownership of their own learning**

We believe that **teacher reflection** should:

x. **Be used to plan subsequent instruction that is based on evidence of student learning**
y. **Be an ongoing and vital part of the learning process**
    - **a.** Teachers need to reflect on the validity of assessments, as well as what the evidence is saying about future learning needs.
    - **b.** Teachers should reflect on the results of all student's work in order to identify areas of growth and challenges (in comparison to past demonstrated knowledge of that individual student).
    - **c.** Teachers should <u>moderate</u> common assessments.
    - **d.** Reflection needs to be timely to be useful.
z. **Happen both individually and collaboratively**
    - **a.** Teachers look at how the assessment allowed students to demonstrate mastery and where the assessment can be improved.
    - **b.** Teachers reflect collaboratively, looking at each student's progress on all of the Guaranteed Standards.
    - **c.** One of the goals of our PLTs is for teacher reflection to happen individually, during PLTs, and in other collaborative groups.

We believe that **<u>reporting</u>** consists of:

aa. **Teachers using their professional judgment to select which assessments constitute the body of evidence (to report on)**
    a. Teachers should consider factors such as the student's ability to demonstrate going beyond mastering the standard and the validity of the assessment itself when determining how. important an assessment should be in relation to the other assessments given in the course.
    b. Teachers should keep in mind that not all assessments give the same depth of insight into comprehension—a lab report and an exit ticket give different levels of insight.
    c. Commonly taught courses report a similar body of evidence that represents student learning toward the guaranteed course Standards.
    d. Reassessment of Guaranteed Standards is offered to all students.
    e. Commonly taught courses must offer similar reassessment opportunities.
    f. Appropriate assessment practices eliminate the need for bonus points and extra credit.
    g. Rote homework assignments are not incorporated into the final academic grade for a course. Instead, homework completion informs the Habits of Mind.
        i. "Rote" assignments include practice problems, worksheets, vocab lists, guided questions, memorization, etc.
        ii. Non-"rote" assignments include essays, lab reports, and creative long-term assignments/projects.
    h. Teachers determine the relative value of individual assessments when concluding the student's end-of-semester grade.

i. If there is not enough evidence to assess a certain guaranteed standard, a grade of No Credit (HS) or No Evidence (ES/MS) needs to be reported and noted in comments.

bb. Curating the "best" evidence that is current, relevant, accurate, and supportive of learning

a. A simple recall question, for example, gives less insight into mastery than a question requiring the application of skills or knowledge.
b. Priority is given to assessments completed during class as opposed to work taken home.
c. Assessing a student's individual mastery as part of a group must be based solely on that individual's work.
d. When determining a student's final grade, achievement over time, as well as more recent performances, are important to consider. Therefore it is necessary to look beyond the most recent one or two grades.
e. When the evidence of learning could justify two differing grades, consider the momentum of the student's progress.

cc. Clear and consistent communication to all stakeholders

a. Sections will work to define the frequency of reporting.

APPENDIX C

SEMI-STRUCTURED INTERVIEW FOR VIDEO-CUED INTERVIEW

This document outlines the process and set of questions to be used in the video-cued interview.

Process

Stage 1. Make sure that the participant has brought or given in advance the signed consent form. If the participant does not have it, then begin by having the participant sign a copy of the form that you brought with you to the interview.

Stage 2. Then, say the following to the participant:

> "*Thank you for taking the time to participate in this video-cued interview. The video is a 20-minute conversation I had with another colleague in a different school. Using their gradebook as the point of departure, we discussed the faculty's assessment practices in her context. I am supplying you with a piece of paper and pencil for you to jot down ideas and notes with as you like. After the video, I will ask you a few questions.*
>
> *Do you have any questions before we get started?"*

Wait and answer questions as needed.

Set up the video, distribute the paper and pencil, and leave the room.

Stage 3. Return to the room, begin recording the audio and use the semi-structured interview protocol below. The aim is for this conversation to flow freely, allowing the participant to share beliefs uninhibited and in *their* direction. Having said that, there are some key points to probe on that are linked to the theoretical perspectives, aims of the research, and aims of the innovation.

Questions	Anticipated Codes
Introduction	
1) What did you think of the video?	
2) Do you have any questions about what was described?	
Formal Formative	
3) Considering the formative assessment process, what is your take on the difference between informal and formal formative assessments, how they are used, and how they are reported?	#NumberOfAssessments #TypeOfAssessment #Accuracy #Meaning
Probe for *take on the descriptors, number of assessments, importance of including task-oriented work reported as incomplete and/or complete*	
4) There is a policy of one formal formative assessment before a summative assessment. What is your reaction to	#NumberOfAssessments #PolicyDocument

that idea? ***Probe for*** *take on number of assessments and top-down directive as a 'policy.'*	
Determination of Grade	
5) The teacher spent some time talking through how she and her team determine the end of year grade. What was your impression of that? ***Probe for*** *the themes of collaborative process, lack of averages, standards and objectives aligned, conversion to letter grade*	#ScaleReduced #StandardsLinked #TypeOfAssessment #Coherence #Accuracy #Meaning
6) Work that is not submitted is recorded as Not Turned In. That theme came up a few times in the recording. What are some of your thoughts on how that works? ***Probe for*** *the idea of not penalizing students on their work.*	#Reassessment #Accuracy #Meaning
Concluding	
7) In considering your own assessment *practice*, how similar or different is this? ***Probe for*** *ideas of coherence, accuracy, and meaning of grades.*	#Practices #Coherence #Accuracy #Meaning
8) Reflecting on your beliefs about the purpose of assessment, to what extent do the practices from this video agree with your own *beliefs*? ***Probe for*** *ideas of coherence, accuracy, and meaning of grades.*	#Beliefs #Coherence #Accuracy #Meaning
9) What is your impression of the gradebook platform relative to the one you use, here?	#Inhibit #Neutral #Support #Materials
10) Do you have any final comments or ideas you'd like to share?	

Stage 4. Thank the participant for participating in the interview. Turn off the recording. Then ask one other question: *"How was the interview?"* Listen closely in an informal way. Upon leaving the room, sit down to write (or voice record) a brief reflection of the final question and the entire interview.

APPENDIX D

SEMI-STRUCTURED INTERVIEW FOR HEAD OF DEPARTMENT

This document outlines the process and set of questions to be used in the interview of the head of department.

Process

Stage 1. Make sure that the participant has brought or given in advance the signed consent form. If the participant does not have it, then begin by having the participant sign a copy of the form that you brought with you to the interview.

Stage 2. Use semi-structured interview protocol below. To the right of the questions are the anticipated codes that are linked to the theoretical perspectives, aims of the research, and aims of the innovation.

Questions	Anticipated Codes
Introduction	
1) How are the assessment reforms going for *you*? For your *team*? For the *division*?	#Groups
Levels of Use (modified from Hall and Hord)	
2) What kinds of #changes are you making in your assessment #practices?	#Practices
3) How are you coordinating your assessment practices with others?	#Group #Network
4) Are you planning, exploring, or imagining making any future modifications to your practices?	#Beliefs
Purpose of Assessment Reform	
5) When thinking about #accuracy of assessment of learning, what do you and your team do?	#rubrics, #StandardsLinked #ScaleReduced
6) As you consider the #meaning of assessment for learning, what do you and your team discuss?	#meaning #rubrics #StandardsLinked #ScaleReduced
7) Working with others, how do you ensure #coherence of assessment practices?	#NumberOfAssessment #TypeOfAssessment #tests #assignments #quizzes #projects
Actor Network Specific Questions (maybe best questionnaire)	

8) How do you use the following documents in your work? … is this different from last year?

a. Policy Document
b. Handbook
c. Gradebook
d. IB Criteria
e. Grade Descriptors

#Practices
#PolicyDocument
#Handbook
#Gradebook
#GradeDescriptors
#tests #assignments
#quizzes #projects

In closing

9) What other *opportunities* have you faced in your #meetings with your team?

#Rhetoric

10) How about *challenges*? What other challenges have you faced in your #meetings with your team?

11) Do you have any final #beliefs about or #practices with the assessment reform that you care to share?

APPENDIX E

QUESTIONNAIRE QUESTIONS

The following are the questions used to generate the electronic questionnaire. The final format is shared electronically with Qualtrics Software.

0	1 2	3 4 5	6 7
Irrelevant	Not true of me now	Somewhat true of me now	Very true of me now

	All summative assessments should align to the adopted course standards.	0 1 2 3 4 5 6 7
	All projects should align to adopted course standards.	0 1 2 3 4 5 6 7
	All tests should align to adopted course standards.	0 1 2 3 4 5 6 7
	The most accurate way to measure student achievement is by designing summative assessments collaboratively.	0 1 2 3 4 5 6 7
	The need to differentiate learning experiences for students makes collaboration on summative assessments less of a priority.	0 1 2 3 4 5 6 7
	One of the most accurate way to evaluate students is using predetermined criteria, such as rubrics or mark schemes, that align to course standards.	0 1 2 3 4 5 6 7
	The most accurate way to evaluate student achievement is through professional collaboration and moderation.	0 1 2 3 4 5 6 7
	There are some learning experiences in my class that do not align specifically to course standards.	0 1 2 3 4 5 6 7
	All formative assessments should align to the adopted course standards.	0 1 2 3 4 5 6 7
0	All in-class activities should align to adopted course standards.	0 1 2 3 4 5 6 7
1	Students' proficiency towards all course standards grows over time.	0 1 2 3 4 5 6 7
2	The most accurate way to report student learning is to use the mathematical mean to average students' grades.	0 1 2 3 4 5 6 7
3	Only students' academic achievement should be included in the end of semester reported grade.	0 1 2 3 4 5 6 7
4	The habits of mind criteria is a sufficient to reporting students' behavior.	0 1 2 3 4 5 6 7
5	The larger the scale to report student achievement, the more accurate the evaluation.	0 1 2 3 4 5 6 7

6	I know the degree that my students have achieved proficiency of the course standards with or without the mathematical mean.	0 1 2 3 4 5 6 7
7	I ensure that questions for tests align clearly to adopted course standards.	0 1 2 3 4 5 6 7
8	I work with my teaching team to develop all summative assessments.	0 1 2 3 4 5 6 7
9	All rubrics for my course are developed my professional learning team.	0 1 2 3 4 5 6 7
0	All rubrics for my course are developed my teaching team partner.	0 1 2 3 4 5 6 7
1	The rubrics and/or other evaluative tools used in my classes only include reference to the adopted course standards.	0 1 2 3 4 5 6 7
2	Homework are included as evidence of students' level of proficiency for the semester grade.	0 1 2 3 4 5 6 7
3	Pop quizzes are included as evidence of students' level of proficiency for the semester grade.	0 1 2 3 4 5 6 7
4	Practice questions are included to determine students' level of proficiency for the semester grade.	0 1 2 3 4 5 6 7
5	I assign some work in class that does not align to the adopted course standards.	0 1 2 3 4 5 6 7
6	I create learning experiences derived specifically from the adopted course standards.	0 1 2 3 4 5 6 7
7	I work with my teaching team to create learning experiences derived from course standards.	0 1 2 3 4 5 6 7
8	I offer opportunities to students to earn additional, 'bonus' points on tests or quizzes.	0 1 2 3 4 5 6 7
9	Students are given multiple opportunities to demonstrate that they have achieved proficiency of course standards.	0 1 2 3 4 5 6 7
0	I use students' achievement on most current, relevant, and accurate assessments to determine students' end of semester grade,	0 1 2 3 4 5 6 7
1	I collaborate with my teaching team to determine students' end of semester grades.	0 1 2 3 4 5 6 7
2	I collaborate with my professional learning team to determine students' end of semester grades.	0 1 2 3 4 5 6 7
3	I use the school adopted Beliefs and Guiding Actions document to develop my assessment plan for my classes.	0 1 2 3 4 5 6 7
4	I use my gradebook (Powerschool) to help me ensure summative assessments align to adopted course standards.	0 1 2 3 4 5 6 7
5	I use my gradebook (Powerschool) to help me ensure formative assessments align to adopted course standards.	0 1 2 3 4 5 6 7
6	I use Atlas to help ensure summative assessments align to adopted course standards.	0 1 2 3 4 5 6 7
7	I use Atlas to help ensure formative assessments align to adopted course standards	0 1 2 3 4 5 6 7

8	I have my own system, aside from Powerschool and Atlas, to help ensure summative assessments align to adopted course standards.	0 1 2 3 4 5 6 7
9	I have my own system, aside from Powerschool and Atlas, to help ensure formatuve assessments align to adopted course standards.	0 1 2 3 4 5 6 7

APPENDIX F

CONSTRUCT ALIGNMENT FOR HIGH SCHOOL FACULTY SURVEY

Question Number	Belief (B) or Practice (P)	Standards Linked	Accurate	Coherent	Summative	Formative	Semester Grade	Reassessment	Materials
1	B	X			X				
2	B	X			X				
3	B	X			X				
4	B			X	X				
5	B			X	X				
6	B	X	X		X				X
7	B		X	X	X				
8	B	X				X			
9	B	X				X			
10	B	X				X			
11	B						X		
12	B		X				X		
13	B	X					X		
14	B						X		
15	B				X		X		
16	B						X		
17	P	X			X				
18	P			X	X				
19	P			X	X				X
20	P			X	X				X
21	P				X				X
22	P					X	X		X
23	P					X	X		X
24	P					X	X		
25	P	X				X			

26	P	X			X			
27	P		X		X			
28	P	X				X		
29	P	X				X		
30	P					X		
31	P		X			X		
32	P		X			X		
33	P							X
34	P	X		X				X
35	P	X		X				X
36	P							X
37	P							X
38	P	X						X
39	P	X						X

APPENDIX G

OBSERVATIONS OF MATH AND SCIENCE COURSES, TIMESCALES AND TYPES OF ASSIGNMENTS ENTERED IN GRADEBOOKS

Math					Science				
Semester	Course	Teacher	Timescale	Types of Assignments	Semester	Course	Teacher	Timescale	Types of Assignments
FS19	Accelerated Math 10	3	The time between the first and second was two months, the. Second and third was one month.	Minors are quizes. There is one "continuous assessment". Tests covered "chapters".	FS19	AP Biology	4	Two minors before a major is standard.	Majors are tests, minors are mostly labs, with one homework assignment.
FS20	Accelerated Math 10	2	3 early minors then an early major, next majors two months later	Majors are all tests and out of 50 (consistently), Minors include quizzes and practice sets and are not included in the grade	FS20	AP Biology	4	Majors happen roughly once per month. Minors happen at least once a week	Minors include quizzes and what may be activities.
FS20	Accelerated Math 10	3	There is one minor before each major. Minors are at least one month between each other and three weeks before a major.	Minors are quizzes and Majors are tests.	FS19	AP Chemistry	2	At least one minor per week. Majors happened 1-2 weeks apart.	Minor assessments were labs, quizzes, investigations (like several parts to this), and activities. A "reflection" was a category, though not included in final grade. Major assessments were labs and tests
FS19	Accelerated Math 9	6	There is roughly one minor per week, unless there are majors in that week. Majors occur roughly 4 to 5 weeks apart.	Minors are mostly quizzes, there is one activity, a work check, a pretest, and a packet. Quizzes are out of 10. Others out of 5. Majors are tests and are out of 10.	FS20	AP Chemistry	2	Minors about every 2 weeks. Majors 2.5 to 4 weeks	Minors are labs and quizzes and Majors are unit tests

FS20	Accelerated Math 9	6	There is one major per unit. Each unit is 4-5 weeks apart.	Majors are titled as unit titles, which could mean anything.	FS19	AP Physics	6	There is an assessment once every 1-3 weeks.	Minors are quizzes and majors are tests.
FS19	AP Calculus AB	5	There is just under 1 minor per week. Sometimes 2 a week and sometimes none. The first major assessment is a few weeks into the year. The second and third are a month apart. The fourth is 2 months.	Minors are either "practice" or "quiz", ranging from 5 to 15, mostly 5. Quizzes are out of 100. It appears there no chapter 1 test. Majors are chapter tests and are out of 100.	FS20	AP Physics	6	There is an assessment once every 1-3 weeks.	Minors are quizzes and majors are tests.
FS20	AP Calculus AB	5	There appears to be a minor before each major. Majors are roughly 4-5 weeks apart.	Minors are quizzes and are out of 100. Majors are tests and out of 100.	FS19	AP Physics C	6	There is an assessment once every 1-3 weeks.	Minors are quizzes and majors are tests.
FS19	AP Calculus BC	6	There is at least one minor before each major. Majors appear in pairs, one with and one without a calculator. This means there may be more like 7 major assessment "events"	Minors are mostly quizzes, there is one activity and another packet. Quizzes are out of 10. Others out of 5. Majors are tests with or without calculators and are all out of 10.	FS20	AP Physics C	6	There is an assessment once every 1-3 weeks.	Minors are quizzes and majors are tests.
FS20	AP Calculus BC	6	There are no minors. The	Majors were titled as unit titles, which could	FS19	Biology	5	No real pattern to Minors with a chunk	Minors are mostly labs, worksheets,

			majors can be 1 to 4 weeks apart. There were 2-3 majors per unit, as they were titled by unit number.	mean anything (text, quiz, project). The final exam attaches a syllabus rubric.				in the beginning then none for a while. Majors happened in the 10th week and then a month later, followed by the final exam	quizzes, and activities.
	FS20	Design Technology	2	Midway through there was a minor and then a major.	Minor was excluded. One major was the assessment. This is a semester long class	FS20	Biology	4	There is no distinguishable pattern to the minors. The major assessment occurs in the middle of the semester, 10 weeks into it.	Minors are mostly quizzes and activities
	FS20	Financial Math	1	about a month (3-5 weeks) apart on tests	Three types of majors, Tests, projects, and a habits of mind piece that is not graded but entered at the end	FS20	Biology	5	No real pattern. There is a break in the middle of the semester and more majors at the end, with none in the beginning.	Minors are quizzes, assignments, and presentations. Majors are tests.
	FS20	IB Math AA HL Y1	3	There are two weeks between the first three majors. Then there are two minors 6 weeks later spaced two weeks apart. And then a final exam.	Majors are tests and minors are quizzes. Majors are scored out of 100. Minors are scored out of 40	FS19	Chemistry	2	Minors about 1 per week. Majors occur 4-6 weeks.	Minors are activity-related works and Majors are tests.
	FS19	IB Math AA HL Y1	3	Two months pass between minor summatives. The first three majors are within 1-2	Majors are tests, Minors are quizzes, they are all scored out of 10o. There is an unknown "investigations/group	FS20	Chemistry	2	Minors about 2-4 weeks, depending on flow of majors. Majors 3-6 weeks	Minors don't indicate quiz or activity. Majors are unit tests

			weeks of each other, then 2 months apart followed by 1 month.	work" that isn't clear what kind it is. Fs were present at 45/100. Not sure what that means…					
	FS20	IB Math AA HL Y2	3	Majors are about a month apart from one another. A minor is had at least two weeks before. The semester opens with a major.	Minors are out of 40 and majors out of 100. One minor is "excluded" for some reason.	FS20	Environ-mental Science	5	There is an assessment once every 1-3 weeks. There is not always a minor before a major.	Minors are quizzes, assignments, and presentations. Majors are tests.
	FS20	IB Math AA SL Y1	4	There is usually at least one minor a week. Majors occurred once per each of the first two months.	Minors are mostly "Tasks" with a few "quizzes" and one "Test corrections". Some tasks are marked only as completed with a check mark. Other tasks are out of 10. Quizzes are out of 40 or 50. One test was out of 50 and the other out of 100.	FS19	Health Science	4	Mostly project based, and some tests, minors are quizzes and diaries	
	FS19	IB Math AA SL Y1	3	There is greater variety in minor and major assessments. Almost all Majors are preceded by a minor. There is a two week difference between the first two majors and then a two month	Minors are quizzes and Majors are tests.	FS19	IB Biology HL Y1	5	Roughly every two weeks there are several assessments turned in or completed at once.	Lab, assignments, and quizzes (hard to tell because of titles of the assignment... could be quizzes of the topic)

			difference between the next one.						
FS20	IB Math AA SL Y2	3	There are two majors at the beginning before any minors. The majors are about a month and a half apart. The first minor happens three months into the semester and is followed by a second one a month later.	The Majors are all out of 100. The minors are out of 40 and 60. One of the majors is a project, the math investigation.	FS20	IB Biology HL Y1	4	Minors happen around once a month, the majors don't happen until the last two months.	Minors are quizzes and majors are tests with one lab
FS20	IB Math AI HL Y1	2	Minors are about 2 weeks apart. First major was two months in and the next one was two months later	Minors are excluded from grade and are scored mostly out of 20. Majors are out of 100. The final exam was out of 10 and student scored an 8.1. That feels odd.	FS20	IB Biology HL Y1	5	No clear pattern. Aside from first month, when there is no assessment, there is one assessment once every 1-3 weeks.	There is a "formative" entry at the beginning.
FS19	IB Math AI HL Y1	2	No consistency. One early and then three at end. Could be entered in late? Could be offered reassessment?	Majors are all tests, Minors are quizzes and a project, most assessments are out of 30. Two minors are out of 5 and one major is out of 60. the mode is closer to 30	FS19	IB Biology HL Y2	4	Minor are around once every other week. There are two major tests 10 weeks apart.	Minors include quizzes, worksheets, labs, and homework
FS20	IB Math AI HL Y2	2	Minors are about 2 weeks apart. First major was two months in and the next one was	Minors are quizzes and are out of 20, and include 2 "mock exams" which were actually out of 57/59. The Major	FS19	IB Biology HL Y2	5	Once every other week or so, but then there is a jump in October where there is no minor. There	Lab, assignments, and quizzes (hard to tell because of titles of the assignment... could

				two months later	assessments are tests which are out of 100.				are more major assessments in the final part of the semester	be quizzes of the topic)
	FS20	IB Math AI SL Y1	1	About a month (3-5 weeks) apart on tests	Three types of majors, quizzes out of 10 and tests out of 40, and a habits of mind piece that is not graded but entered at the end	FS20	IB Biology HL Y2	5	No clear pattern. Many assessments in the first month, then nothing for a while, and then the last month	Minors are quizzes, assignments, and presentations. Majors are tests.
	FS19	IB Math AI SL Y1	1	about a month (3-5 weeks) apart on tests	Majors are all tests. Minors include notebook checks per unit, and what appear to be titles of topics, and a mid-unit quiz worth 5x more than others. All notebook checks and titles are full marks, where quiz and tests student doesn't earn. The comment from teacher notes the discrepancy and suggests test anxiety and/or skills lacking	FS20	IB Biology SL Y1	4	One a month for the most part. One major after 10 weeks and one at the end.	Minors are mostly quizzes and activities
	FS19	IB Math AI SL Y1	4	There is usually at least one minor a week. Majors occur once a month with about a month in between.	Minors are "tasks" "quizzes" "test corrections" "class work" and part of an "investigation". Majors are tests. Quizzes are out of about 30, other minors are out of 10, and the tests are between 50-65. No consistency there.	FS19	IB Biology SL Y2	5	No real pattern. There is a break in the middle of the semester and more majors at the end, with none in the beginning.	Lab, assignments, and quizzes (hard to tell because of titles of the assignment... could be quizzes of the topic)

FS20	IB Math AI SL Y2	1	about a month (3-5 weeks) apart on tests	Four types of majors, quizzes out of 10 and tests out of 40, and a habits of mind piece that is not graded but entered at the end, and an math investigation IA	FS19	IB Chemistry HL Y1	2	Minors about 1 per week. Majors 3-6 weeks apart.	Minors are quizzes, investigations, and labs. Majors include mostly "topic tests" and one lab.
FS20	IB Math AI SL Y2	4	There is usually at least one minor per week. There are three early majors, one is a mock, and then the other two are about a month apart.	Minors are mostly tasks, assignments, and quizzes, but also include drafts, investigations, and test corrections. Most minors are out of 10. Quizzes were worth up to 20. One Major was an early quiz (out of 10). Otherwise they were out of 40 to 100 points	FS19	IB Chemistry HL Y1	3	Mostly 2-5 weeks apart. They happen in spurts almost	Test (topics) lab, activity, quiz
FS19	IB Math HL Y2	3	First three tests occur within a week of one another. Then a month between the other three.	Only majors and all of them tests. The last one is called "test option" which I am not sure what means.	FS20	IB Chemistry HL Y1	3	Roughly one a month	there are 11 "check mark" ones. I am not sure what this means. there is an interesting "formative average quiz" entered into the document as a grade and a cumulative quiz, and a semester portfolio
FS19	IB Math Studies Y2	2	No consistency. One early and then three at end. Could be entered in late? Could be offered	Majors are tests and "scaffolds for the IA", mostly in the range of 40-60. Minors include a "work check," quizzes, and a "spiral," which I	FS19	IB Chemistry HL Y2	3	About 5 weeks apart with a few project oriented assessments at near end of semester	task, lab, scales, some are omitted but entered so students see the conversion to IB scales

			reassessment? There are a few that look like scaffolded assessment for the IA.	don't know what is					
	FS20	Intro to Computer Programming	6	There is one minor. Then the rest are Majors. There is no real pattern. Sometimes there are several in a row, perhaps part of larger project.	Minor was a basic "commands" quiz. The Majors quizzes and completion..	FS20	IB Chemistry HL Y2	2	Minors don't have any clear consistency. Majors don't either, between 1, 2, and 6 weeks. There is typically an assessment every week, except in November.	Minors are mostly quizzes and one IA work assignment. Majors are topic tests
	FS20	Math 10	3	Minors are about 2 months apart, they happen 2 weeks before the major.	Minors are quizzes and out of 40. The Majors are out of 100.	FS20	IB Chemistry HL Y2	3	Roughly one a month	There are a few check marks for "Drafts" of works. Some lab work, quizzes, and tests
	FS20	Math 10	4	There is a little less than one minor per week. Majors are about 6 to 8 weeks apart.	Minors are tasks, quizzes, assignments, workbook, and test corrections. Tasks and quizzed being most frequent. They are out of 10, though one quiz is out of 20 and another out of 39. A written assignment is out of 100. The majors are all out of 100 and are tests.	FS19	IB Chemistry SL Y1	3	About 5 weeks apart with a few project oriented assessments at near end of semester	task, lab, scales, some are omitted but entered so students see the conversion to IB scales
	FS19	Math 10	4	There is usually at least one minor a week. Majors occur once a	Minors are "assignments", "workbooks" "quizzes" "test corrections". They	FS20	IB Chemistry SL Y2	3	There is one assessment per week, either minor or major.	Mixture of lab, investigations, and tests

			month with about a month in between.	typically out of 10 or 30. Tests are mostly out of around 50.					
FS19	Math 11	1	about a month (3-5 weeks) apart on tests	Majors are all tests. Minors include notebook checks per unit, and what appear to be titles of topics, and a mid-unit quiz worth 5x more than others. All notebook checks and titles are full marks, where quiz and tests student doesn't earn. The comment from teacher notes the discrepancy and suggests test anxiety and/or skills lacking	FS19	IB Chemistry SL Y2	2	Minors about 1 per month. Majors about once per month. HOWEVER, there are other IA-sort and reflections. There is about one assessment/entry per week.	Minors are quizzes, labs, and investigations.
FS19	Math 12	2	2 early minors, one mid major and one end major	Majors and minors were all termed "SAT" review or test. The final major was a project. Majors were out of 100 and the minors were 15 then 18	FS20	IB ESS SL Y1	5	There is an assessment once every 1-3 weeks. There is not always a minor before a major.	Minors are presentations, quizzes, and worksheets and questions. Majors are tests
FS20	Math 9	4	There is usually at least one minor per week. There is a major every month.	Minors are homework, assignments, quizzes, and test corrections after each major. Minors are mostly 10-20, though a quiz is usually worth at least 20. One is worth 30. Test corrections are worth 5 to 5.5 (odd?). Tests are worth 64, 70, and 100.	FS20	IB ESS SL Y2	4	One major a month, at least once no minor between majors.	Minors are labs, tests, and quizzes

FS20	Math 9	5	There is a minor before every major. It appears 1-3 weeks before the Major, depending on how long the time between majors is. Majors are about 4-5 weeks apart, except the first two.	Minors are quizzes and are our of 100. Majors are tests and out of 100.	FS19	IB Physics HL Y1	1	There is one major per month. Minor occur about every week, except when there is a Major	Majors are mostly "class tests" except the lab. They are usually out of around 30 points, but no consistency. One is out of 100. Minors are quizzes or labs.
FS19	Math 9	1	about a month (3-5 weeks) apart on tests	Majors are all tests. Minors include notebook checks per unit, and what appear to be titles of topics, and a mid-unit quiz worth 5x more than others. All notebook checks and titles are full marks, where quiz and tests student doesn't earn. The comment from teacher notes the discrepancy and suggests test anxiety and/or skills lacking	FS19	IB Physics HL Y2	6	There is an assessment once every 1-3 weeks.	Minors are quizzes and majors are tests.
FS19	Math 9	4	There is usually at least one minor a week. Majors occur once a month with about a month in between.	Minors are "HW" "quizzes" "test corrections". They typically out of 5 or 10. Quizzes are out of higher marks, ranging from 18 to 56. Tests are out of 54 to 76. No real consistency there.	FS20	IB Physics HL Y2	1	Major assessments are 5 weeks apart. Some major assessments are in two parts, one on each different day. Minor assessments are about 1-2 weeks apart.	Major assessments are tests. Minor assessments include quizzes, a "portfolio of assignments" and scaffolded project work for Ias.

FS19	Math 9	5	There is at least 1 minor per week. There is one major per month, roughly 4-5 weeks between one another.	Minors are either practice or quizzes. Practice appears to reference specific page numbers in a book and are out of 5, sometimes reporting 4.4 and 4.2. Quizzes are out of 100. Majors are tests and are out of 100, too.	FS19	IB Physics SL Y1	1	There is a major every 4-5 weeks. There is a minor every 1-2 weeks, unless there is a major.	Minors are labs, quizzes, or presentations. Majors are all class tests. Two of the lowest scores are marked as "not counted towards grade"
FS19	Paced Math 10	2	3 early minors then majors were not consistently spaced, one was a month apart and the other two months apart	Two early minors were not scored/entered, Minors and majors were out of around 30, minors were quizzes and majors were tests	FS20	IB Physics SL Y1	6	There is an assessment once every 1-3 weeks.	Minors are quizzes and majors are tests.
FS20	Paced Math 9	5	There are at least two minors before each major, depending on the length of time between majors. The last major is the longest apart, so there are three minors. Majors are about 5 to 7 weeks apart.	Minors are quizzes and are out of 100. Majors are tests and out of 100.	FS20	IB Physics SL Y2	1	Major assessments are 5 weeks apart. Some major assessments are in two parts, one on each different day. Minor assessments are about 1-2 weeks apart.	Major assessments are tests. Minor assessments include quizzes, a "portfolio of assignments" and scaffolded project work for IAs.
FS19	Paced Math 9	5	There is about 2 minor per week. Majors occur one month after the start of school and then two months later.	Minors are "assignments", "workbooks" "quizzes" "test corrections". They typically out of 10 or 30. Tests are mostly out of around 50.	FS19	IB Physics SL Y2	1	Seems to be a string of assessments in September, then a break. It moves to two a month. Only minors were in the early times	Minors did not count towards grade. The first three majors did not count either. Assessments were mostly IA related

										or class tests.
	FS20	Pre-Calculus	5	There is usually one minor before each major, except the last major assessment. Major assessments are around 4 weeks apart. The last assessment is 3 weeks apart.	Minors are quizzes and are our of 100. Majors are tests and out of 100.	FS19	Intro to Chemistry and Physics	1	The first month's minors and majors were all geared around lab work. Then the minors were review and worksheet related followed by class tests.	
	FS20	Statistics	4	There is about one minor per week. The major assessments happen about 30 days and 60 days apart. The caveat is that there are two majors per unit, one is a test and the other is a project.	Minors are all tasks out of 10. Majors are tests and project. Tests are either out of 50 or 100. The projects are 25 and 65.	FS19	Intro to Chemistry and Physics	3	About 8 weeks between summative assessments	Packets, Activities, Titles,
						FS20	Intro to Chemistry and Physics	1	The first 7 assessments were minor, followed by 3 majors. Minors were 1-3 weeks apart. Major assessments were 2 weeks apart at the end.	Minor assessments were pre-assessments, reassessments, lab work related items, and a "portfolio of assignments." Major assessments were unit tests.
						FS20	Intro to Chemistry	3	All minor until the last 6 weeks, then	This was an interesting way of

	and Physics		Major	setting it up... minors and then majors toward the end. There are a few "re-assessments" in the gradebook.
FS19	Physics	6	There is an assessment once every 1-3 weeks.	Minors are quizzes and majors are tests.
FS20	Physics	1	Majors were about 4-5 weeks apart. There were 4 minors before the first major, then a single minor between the 2nd and 3rd. Minors were about 1 week apart or halfway between the majors.	Minor assessments were task oriented and a "portfolio of assignments". Majors were unit tests. The portfolio is directly linked to the approaches to learning or "practices"

APPENDIX H

EXAMPLE OF GRADEBOOK ENTRIES

Math 9 Exp. P1(A-C) -

Final Letter Grade[1]:	D-
Teacher Comments:	
Section Description:	In Semester 2 Regular Math 9 will cover solutions of systems of linear equations, factoring, solving and drawing quadratics, as well as the properties of triangles and quadrilaterals.

PowerTeacher Pro Assignments

Due Date	Category	Assignment	Flags	Score	Grade	
12/19/2019	Final Exam	December Exam		69.5/140	F	
12/09/2019	Minor	Unit 5 Test corrections		10/10	A	
12/05/2019	Major	UNIT 5 TEST		44/74	F	View
11/18/2019	Minor	Unit 4 Test corrections		10/10	A	View
11/13/2019	Major	Unit 4 Test		29/45	D	
11/12/2019	Minor	Unit 4 Hwk 2		8/10	B-	
11/04/2019	Minor	Unit 4 Quiz		26.5/37	C-	View
11/04/2019	Minor	Unit 4 Homework 1		10/10	A	
10/18/2019	Minor	Unit 3 Test corrections		10/10	A	
10/11/2019	Major	Unit 3 Test		31/73	F	
10/09/2019	Minor	Hmk U3		10/10	A	
09/27/2019	Minor	Unit 3 Quiz		17.5/56	F	
09/17/2019	Minor	Unit 3 H/W		4.05/5	B-	
09/16/2019	Minor	Unit 2 Test corrections		10/10	A	
09/10/2019	Major	Unit 2 Test A		48/76	D	
09/10/2019	Minor	Bonus		--/10		
09/05/2019	Minor	Unit 2 H/W		4.55/5	A-	
09/02/2019	Minor	Unit 2 Quiz 1		35/54	D	
09/02/2019	Minor	Unit 1 Test Corrections		5/5	A	
08/23/2019	Major	Unit 1 Test		38.4/62	D-	
08/23/2019	Minor	Unit 1 H/W		5/5	A	
08/19/2019	Minor	Chapter 1 Quiz		13/18	C-	

Grade stored on: 01/13/2020

APPENDIX I

OBSERVATIONS AND INFERENCES FROM ASSESSMENT DOCUMENTS IN FS19 AND FS20

Teacher	Term	Assessment Title	Observations	Inferences and Considerations
SVC	2020-2021	Static Electricity Quiz	There are no points or marks associated with this quiz. Four operative command terms: Show, Show, Explain, Explain Scoring Rubric is descriptor based and on scale of four, ABCD. Minimal evidence is D and C. Some evidence with some errors is B. Complete evidence to justify is A. Fundamental conceptual knowledge is D. Satisfactory knowledge is C. Sound knowledge is B. Broad based, relevant, and extension of knowledge is A.	The rubric being very descriptor-based may lead to thinking along descriptors for feedback and evaluation. The small scale, ABCD, may lead to more accurate results per student, though not across students.
SVC	2020-2021	Bohr Model Flipgrid Formative Assessment	Rubric is along task completion, clarity, use of visuals, and appropriate vocabulary. Not using the same rubric as Static Electricity Quiz. No percentile, but there is a number entered associated with it. There is a frequency piece to this… all, most, some	The rubric being task oriented and frequency-based nature of this may lead to more completion-style marks. Not sure what the reason is for not using the descriptor-based quiz.
SVC	2019-2020	Introduction to Chemistry and Physics Fall Semester Exam 2019	Composed of two parts. The first section has 21 multiple choice questions, each worth one mark. Offers some practical formulas that students aren't meant to memorize. Does not break down section by marks. Just the first part, multiple choice. Almost all questions are identifying some knowledge from recall. They lead with "which (11x)" "when" "what" and a complete the sentence. There are three table/graph interpretation questions. Part B is short answer, is out of 66 marks, and uses the following prompts: Classify, state, explain, use of a model, table, completing the model, equation completions, interpretation of model, distinguish, plotting and planning. Roughly 1/3 recall in multiple choice questions. The other 2/3 is more higher order thinking and application. There is some planning, skills based pieces.	This seems like a balance of recall and higher order thinking. I am not sure what the reason is for the split, 1/3 to 2/3.
SVC	2020-2021	Introduction to Chemistry and Physics Fall	Composed of 8 parts, each has on average 10 points, but range from 7 to 13 marks. Each part refers to a section of the syllabus, i.e., Safety, Atomic structure, Periodic table, Chemical Bonding.	Very different style test. Instead of combining a bunch of recall and then separately some application, it

		Semester Exam 2020	Almost every part has a combination of multiple choice (27 total) AND short answers. One section has only short answer. There is a note in the teacher's copy that one question was used for reassessment in the past. The grade is still marked as a percent. There are actually more multiple choice questions on this version of the test. I can see some repeat questions across the two tests, 16 total. Same style questions more or less.	is divided into sections. It may be that there is an equal number of questions for both, but the division is clearly "Standards-based." Still, because it is still percent-based and the questions are mostly the same, it feels more of an adaptive than transformative change. In the end, this makes it more clear for students and teachers to see and analyze results.
SHOD	2020-2021	HLY2_P1_Semester_Final_2020 HLY2_P2_Semester_Exam_2020	IB Style Exams with 23 multiple choice and then short answer 57 marks. Interesting that this is the same style as the 9th grade exam. The key difference is how the second FS20 9th grade exam is organized, but in essence is very similar.	It could be a coincidence that the style of this exam is the same as the 9th grade one. I do not believe this is the case.
SHOD	2019-2020	AP Chemistry	Multiple choice and free response combination. This is modeled off of an actual AP exam.	The exam may be similar because the teacher is preparing the student for the style of exam. It may also be the same because the exam already assesses students on the skills that have been taught, making it appear redundant to make a new one.
SHOD	2020-2021	General Chemistry Final Assessment	Project to answer the following question: How does technology use Electromagnetic Radiation? Graded using a rubric along scale of three for all 15 standards. The scale is concerns, mastering, and exceeding. Detail is offered for all "mastering," which is just a statement of the standard. 2/15 of the standards offer some detail for what "concerns" and "exceed" can look like. A final mark is offered out of 7, A, A-, B, B-, C, D, F.	The SHOD admitted in his interview that he was unafraid of trying new things. He has reworked the final grading scale into one out of seven and set up his own way of reaching that scale using a scale of three. The style is very descriptor and standards based. It also goes beyond the boundaries of the ABGA.

			Exceeding is A, Mastering is B, and concerns is C-F To reach highest marks is it application with extensive evidence. The instructions in the exam are highly correlated to the standards of the course. They are shared verbatim with students as points to address.	
MVC	2019_@020	Dec Exam IB MAI SL 2019	Command terms include: Round, find (prevalent), show, suggest, explain, classify, construct, simplify, justify. Questions are listed and not categorized by any type of unit. However, the questions are grouped by a theme (i.e. statistics, geometry) without explicitly stating so. n Marks are offered per question. The highest marks are for command terms justify and construct and for "find" and "show," but for when there are clearly several steps to finding and showing based on the information offered. Pixilation indicates screenshots taken from other sources.	This reminds me of an IB test. This makes sense because it is an IB class. It doesn't perfectly mirror one because it is the first semester of a two-year long course.
MVC	2019_2020	Math 9 Reg Dec Exam 2019	On the cover of the test one can see three teachers' names. This implies that the three teachers teach the course and use the same assessment. Command terms include: solve, determine, find, graph, write an equation for, identify, how long, interpret, name, state. No explicit thematic organization. There is some increase in difficulty in the questions. The majority relate to understanding types of equations and how they relate to one another geometrically (i.e. on planes, circles, intersections). No clear explanation of points *on the test*. Marks range from 1-4. No clear correspondence of command term to marks, except "find," which is most common, will not be higher than 2.	This reminds me of the type of test I took 24 years ago. There is no authentic word problem contextualizing the equations in any real world application. There is no prose writing in this test, only numeric expressions and words.
MVC	2020_2021	IB MAI SL Y1 Unit 6 Test FS20	Command terms include: match, provide interpretation, plot, find, calculate, comment, how much, predict, evaluate, graph	Command terms in the IB style tests seem deeper than the grade 9

			Questions are divided further into sub-questions. Some questions have increasingly more marks associated with each sub-question. The greatest number is 4, for calculating the equation that best fits the one drawn by hand.	math. Could be a consequence of lower math level. There are authentic word problems contextualizing the statistics and equations.
MVC	2019_2020	Math 9 Unit 2 Test	Explain, write, solve, Answer Points increase in number over course of the test, the most being at the end, at 5	The way the points increase appears to be very intentional, requiring students to complete two parts to a question answer/solve AND graph. There is no authentic word problem contextualizing the equations in any real world application. There is a section to explain in prose.
MHOD	2019_2020	10th Grade Paced Math FS29	Total of 55 Marks. Definitions of some vocabulary offered, including three command terms, solve, find, state. Three formulas offered, as well. Command terms include: evaluate, find, for what values is, provide, state, sketch, how high, is (yes or no), graph, circle best fit, expand expression, simplify, identify, complete table, solve. The entire test deals with functions, beginning with simple knowledge questions about its components, then progressing to ability to sketch and interpret the meaning of the visual representation of the function. The functions themselves get progressively complex and are mapped across the x-y coordinate plane. The marks for each question get progressively higher. There is one question worth a total of 8 marks. It is an application to an authentic experience, building a tray.	This reminds me of a test I would take 24 years ago. There is a progression of difficulty and marks. There is one authentic word problem contextualizing the equations in any real world application. There I no section requiring explanation in prose.
MHOD	2020_2021	10th Grade Paced Math FS20	The total number of marks is not listed, but it tallies to 61. Command terms include: evaluate, find, for what values is, provide, state,	I can tell right away that this is practically identical to the final exam from before. The same

			sketch, how high, is (yes or no), graph, expand expression, simplify, identify, complete table, factor, solve. The entire test deals with functions, beginning with simple knowledge questions about its components, then progressing to ability to sketch and interpret the meaning of the visual representation of the function. The functions themselves get progressively complex and are mapped across the x-y coordinate plane. The marks for each question get progressively higher. There is no word problem question.	structure and marking scheme. The same command terms. The values have changed, but hardly anything else. One significant change is that there is no authentic word problem contextualizing the equations in any real world application. There I no section requiring explanation in prose.
MHOD	2019_2020	IB MAI HL Semester Exam December 2019	The exam opens with a table that lists the units, skills related to them, and the number of marks associated with each one. There is a financial application and a few extensions in geometry and vectors at the end, too. Each of these count for significantly more marks. Command terms include: fill in, model, how many, explain why or why not, write down, draw, explain the meaning, find, describe, for what values does, name, calculate, estimate, determine. Some of the questions are complex in nature, offering a scenario and then asking a series of questions about the scenario. Many of the questions are word problems or application based. These questions sometimes have more simple parts appear to scaffold some of the thinking for students.	This exam is very well organized and appears clear on what is being assessed, where, and to what extent it is valued in the overall marking. The command terms are diverse. There are many questions that offer the authentic application. None require prose writing.
MHOD	2020_2021	IB MAI HL Semester Exam December 2020	Command terms include: show, find, determine, state, write, at what time, shade the region, how many, provide, estimate, comment on, construct, calculate, integrate, what is, indicate. No question is worth less than 2 marks. The majority are worth at least 4.	This is a very different kind of question for the same course. It looks more like a common test from the IB. Perhaps these are screenshots.

APPENDIX J

INTERVIEW FINDINGS REPORT

Science Department

Beliefs. The science department faculty's beliefs are largely aligned with what is articulated in the policy document. The Science Head of Department (SHOD) stated outright that the policy's points "are right in line with my philosophy." This may be because he "helped write it. So I'm really familiar with it and I really take it to heart." It also is because "they're in line with some things I've done in the past at other schools," with different, SBG architecture. Therefore, the assessment reforms were "like reading an old friend." In fact, he contrasted his experiences of the new policy with what he did first arrived at IST, that he was "shoehorning [IST] assessment policy into what I was already doing…." He felt he had to "to take the way I was asking questions and transform that into a percentile style grading assessment." This translation of his beliefs as to what assessment should look like into what IST wanted from him took additional effort.

His science team member who took part in the video cued interview (SVC) was not directly involved in the crafting of the policy. However, she shared a few specific beliefs about assessment that also demonstrated alignment. First, she noted that assessment information is not only a way to report student achievement, but that multiple "formative assessments before the summative assessment" can help students "actually fix those gaps." This comment aligns with the following point on assessment from the ABGA (Appendix A): assessment is meant to "provide ongoing opportunities to act on feedback from assessments to strengthen learning." Second, she notes that the information from assessment is for students, "but it also helped me to reflect on my teaching." The ABGA Policy Document states that assessment should be used for reflection to "be used to plan subsequent instruction that is based on evidence of student

learning." By the end of the semester, she became more aligned on ABGA's note that assessment be "a result of professional collaboration and moderation to ensure consistency, reliability, and fidelity in the assessment of student learning" (Appendix A). In the first interview she felt that the weighting of the assessments should be an individual "teacher decision," because the person needs "the knowledge of… the assessment and then what standards or what learning object is, what skill was assessed. So I think the teacher has to decide the maximum mark." In the second interview she stated explicitly and passionately on the point of collaboration, "that's very important. And even I do that. So we have a [professional learning team (PLT)] and we work together … to create the assignment and then we have rubrics and plan by the whole team and then we have marking scheme or sometimes rubrics." Further, her remarks on rubrics and mark schemes align with the ABGA documents direction that assessment "is based upon previously determined rubrics, criteria, mark schemes, and other evaluative tools" (Appendix A). Finally, she remarks that she will "consider only the recent [assessment]" if it is "more extensive and it includes the previous assessment." This aligns with the ABGA's point that assessment "allows for multiple opportunities for learners to demonstrate their knowledge, skills, and understanding" (Appendix A). This note in the ABGA is meant to be the mindset adopted for reassessment of student work.

There were a few notes of relative tension between her beliefs and that of the policy by the end of the semester. First, she agrees that one should not penalize students who do not submit work, which aligns with the ABGA's point that teachers assess "behaviors separately from academic standards." (Appendix A). On this point, however, there is some tension. She will let students slide on submitted work if "the performance in

the summative assessment is reasonably good." But, if it is not, then those students will be called in to do the work during the flex period. She also is concerned that the policy breeds an unintended consequence of lowering the work ethic of students. She said, "in terms of the work ethics, I feel that it's not that strong." These are opinions she holds that do not specifically impact her grades of students, but she believes are consequences of the policy.

Lastly, it is on the discussion of quizzes that there is most misalignment. She commented that she and her team administer quizzes that she calls "minor, which is actually graded. It's formal formative assessment. We actually grade it and it happens once in three weeks. Sometimes, once in two weeks. There's no such timeline. Depends on the difficulty level." Their purpose is to "make sure that [students] understand the basics before we move on to the complex." The belief behind the practice has students' interest at heart, offering a kind of dipstick for students and teachers covering complex curriculum. Yet, the AGPA states that all assessments should "provide students with opportunities to demonstrate mastery of the standard." A quiz that checks low-level content before moving on to the complex is a valuable learning opportunity, but should not count towards the overall, even if it is only counts for "a very small portion of that unit."

Practices. The team made changes over the course of the semester. According to the SHOD, some of those changes were "mechanical issues," such as "how to use a rubric to grade, how to ask the questions, how to prepare the kids to answer questions rather than using a multiple-choice test," and how to "ask things in different ways to get that mastery rather than just the rote" on tests. Other changes are more significant but

appear to depend on teacher comfort levels. For example, the SHOD explained with pride in his tenor that he "hasn't given a paper test in Gen Chem all year. It's been through projects and posters and videos." It was the assessment policy that "really gave me, I think, quote-unquote permission to do it." Further, he notes that he has made strides "in terms of reassessment, which is something that I wasn't really doing a whole lot of before this." This practice continues to evolve. His latest change in reassessment is how to reach the unmotivated students who appear to accept the grades of "an F or D." He and his team noted that students who were earning Cs and Bs were more likely to seek opportunities to reassess while students who earned lower than that did not. Instead of letting the grade stay in the gradebook for the students, he contacts them to set up a time for them to come in for additional instruction and finally reassessment in some way.

As far as "the assessments themselves, I have seen a big switch in most of my department teachers moving away from rote style tests to more application focus." Again, a central part of this change is in rethinking the questions being posed, to "get the students to explain rather than just state something and to get them to build those cause-effect relationships and to analyze the patters in a meaningful way." Further the team writes "tests together." In that process "we put a lot of thought into what are we really trying to get at with these questions." Then they moderate them by exchanging "exemplars and grading those. And we make sure we check each other when we're breaking the rubrics, make sure we're grading, assessing the same way, the same thing."

The SVC may be less personally invested in assessment innovation but was willingly engaged in adopting the changes at the school-wide expected pace. She notes that she no longer enters assignments and the day-to-day formative assessments into the

gradebook. These are collected as a portfolio. She has changed her practices from entering all in-class and homework assignments, quizzes, and summative assessments in the gradebook to one that only includes minor and major assessments. The SHOD has done the same thing, calling his gradebook really "clean," now. SVC used to take a more privatized approach to assessment, stating at the beginning of the semester that "the teacher has to decide the maximum mark" on assessments. By the end of the semester her response was very different. She said she thinks collaboration is "very important. And even [she does that]."

Both the SHOD and SVC have started entering letter grades into the gradebook instead of percentages. However, the SHOD explained that the software automatically reports the letter into a percentage, undermining the intent. The SVC also admitted that while she enters the letter grades into the gradebook, she sometimes changes it because she feels the gradebook's default conversion to percent is inequitable. It automatically converts the letter grade into the lowest percentage value. At the end of the semester, it uses all the minor and major assessment to tabulate a final score based on those numbers, not the letters. For example, an A is entered in as a 94% and a B is an 83%. The average between those two numbers is 88% which is reported as a B+. However, if the student scored an actual 98% and an 86% on the tests, those scores would average to a 92% which is reported as an A-. In this scenario, regardless of the actual scores, based on the ABGA policy, the test results would be entered as 94% and 83%, creating the discrepancy. SVC admits that she reports the more accurate numbers (98% and 86%) in the gradebook when she feels it is necessary as a workaround.

Materials. The policy document is something that SHOD references often when writing tests and developing assessment plans. He stated that “it’s in a folder over there. I’m using it all the time.” He also states that it is also always in the back of his mind because he “helped write it. So [he’s] really familiar with it and [he] really [takes] it to heart.” As further evidence that remains a part of his practice, he explains that he used part of it verbatim in sharing his reassessment policy students. In the interview, he talked about how at the beginning of the semester many faculty were tired of talking about it and were ready for the policy document. “I know at the beginning of this year I was hearing a lot of teachers say, you know, I'm tired of talking about this, let's just do it. They're just ready for a policy or they were done talking about it.” He couldn’t speak to how much the document was referenced by other faculty, but he said it was referenced when he collaborated with others on writing assessments. The SVC made no mention of it directly.

The grade book is used differently by both, which expresses where they lie along the standards-based grading continuum. The SHOD noted that the policy document had him reflect on what was included in the grade book and he notes this year that “there is a lot of stuff [that] just didn’t need to be there. It wasn’t affecting anything. So this year, [he] really cut way back on what’s in there?” The formative pieces he is substituting another tool for, Google Classroom. So his gradebook is, as he put it, “really clean.” The SCV stated that the gradebook is not used for the common assignments anymore. Instead, she uses a portfolio for that work. She seems satisfied with this move, overall, but she expressed perceived value in having more assignments for parent information. Both the SHOD and SVC noted that the software misinterpreted their intent to some extent. The

practice of entering only letter grades is something they both use, but the SHOD knows "it calculates numbers and students are going to look at that and [he] can't really do anything about that," implying that he would rather students not concentrate on the number as feedback. He assured me, though, that "it is all letters for [him]."

The SVC also only enters letters in the gradebook, but sometimes changes the number because she believes its means of calculating the grade is inequitable. A full explanation for why is included in the previous section. Despite the inequality of the way the gradebook calculates the grade with letter grade conversions to percent, she expressed a strong opinion that the gradebook be the mechanism that determines the grade at the end. "The [gradebook] should automatically enter the score. For [her] to judge - the judgment can vary for different teachers. So there can be no inconsistency … for different subjects." The calculation appears to take away some of what she calls "judgement." In fact, she recommends that there "should be some professional development for teachers to tell how you need to decide the maximum score based on the skill and the content, the depth of the skill and the content." In other words, she suggests the fix is not to do away with the calculations/algorithm of the gradebook, but to develop the skills of teachers to use it better.

Standards appear to play a strong role in shaping assessments and lessons. For courses that are externally assess, such as the AP and IB courses, the role appears more top-down. The SHOD uses the IB criteria "all the time. Whenever we're developing a test or a quiz or a lab or anything, a lesson, even. That's always foremost in our mind." This is in part because they are drawing questions from previous IB tests instead of writing their own. For NGSS courses they write their own assessments from scratch. Still, though

the SHOD explains that he references them “all the time. Every Day. [He’s] looking at just the high standards. When [he has] those projects and [is] building rubrics, [he’s] looking at [grade descriptors] mostly…. It’s like this lesson is all about exceeding. So let’s really look at what we’re doing here.” The key difference between the NGSS and the IB/AP courses is that he and his team develop the questions themselves because tests and their questions aren’t already written. The combination of the adoption of the new NGSS standards and the assessment reforms was “a perfect storm kind of thing. It all came together at once.” The reason is that “NGSS is very application focused. It's very explain yourself, you know, cite your evidence and what your justification focus.” This helped shift the team from rote assessments to a new kind that is more aligned with the policy.

For the SVC the standards play a stronger role in its power to align her on the reassessment policy. When it comes to reassessment, she notes that the team “started disregarding the previous assessment if [they] are actually testing the same skill.” If standards are constructed in a way that they identify discrete content and skills, then it makes it impractical to not consider all tests as evidence of student achievement. Also, the standards influence a weight of the assessment and grade, “to consider the standards, you know, the skills, what they're assessing.”

Rubrics are clearly a place where teachers collaborate. For both the SHOD and SVC IT came down to a matter of accuracy. The SHOD referenced rubrics specifically when asked about how the team made sure assessments were accurate: “[they] co-grade and swap exemplars.” Here, again, he mentioned the importance of referencing the rubrics, too, to “make sure we’re grading, assessing the same way, the same thing.” The SVC corroborates this practice: “We work as a team to create the assignment and then we

have rubrics and plan by the whole team and then we have marking scheme or sometimes rubrics."

Only the SHOD had anything to say about grade descriptors. The Grade descriptors "are foremost in our minds when building instruments and those are the ways I use that language with students, when I'm discussing with them."

Other. Leadership was an important theme that came up in both interviews. The SHOD explained that he worked very hard with his teaching team partners to develop the assessments. He insisted on the team working on developing the tests and rubrics together. The SVC also referenced "multiple meetings with [the curriculum coordinator]" and "professional development session" with an expert at orientation that helped shape her practices and beliefs. Another important aspect was the adoption of Next Generation Science Standards (NGSS) standards for 9th and 10th grade. According to the SHOD it was a "perfect storm kind of thing."

Math Department

Beliefs. The math department's beliefs were moving towards alignment before the policy document was published and adopted. Perhaps because he was an important player in writing the ABGA document, the department head was very aligned, evidenced by his unprompted, enthusiastic use of metaphors to explain how to reassess and move students towards mastery of a standard - "when you do the high jump and you clear five foot, they don't put the bar back at five foot" - and baseball player's batting average to explain his shift away from percentage-based grading – "a good baseball player doesn't hit, you know, a thousand percent… and, he doesn't every time he swings, he doesn't hit a homerun." He explained that he used the policy document with his team at the beginning

of the year to create a shared understanding and set of expectations. There were some items that the team was already doing and others that they did not agree with.

By the end of the semester, there is greater consensus in the math department that SBG is a means to achieving more accurate assessment. When the MHOD was asked in the middle of the semester how they ensure assessments are accurate, he explained that while he “does not know… because that's up to each individual teacher… I can say that it's better than what it was before, which was just they scored in this percentile.” The impetus for the change is largely due to the work that they are doing on writing different tests. Also, he feels math class assessments generally tend to be more accurate than others because they “probably match the business we do during our class periods…. I don’t know how to get any closer to what we do in class versus what's on the test and what’s done in math class, right?” At the interview at the end of the semester, he expressed with great confidence that “we do a great job with that [accurate assessments].” He did not elaborate.

The other math department member interviewed, MVC, was initially very concerned that moving away from an average that is automatically generated to teacher’s professional judgment would lead to inaccuracies by allowing an “element of emotion and subjectivity” to creep into decisions about grades for students. He also had practical concerns about the ability of a reassessment policy to accurately reflect the course’s intended learning targets. “Trying to [incorporate questions] on a final exam for every topic [covered over a semester] while trying to get an accurate picture that could potentially replace their original attempt - there's this gap that I feel is there.” In his first interview, when invited to offer any additional comments, he explained how confident he

was in his ability to make accurate assessments and that his point system on tests was already "quasi-descriptor based." On the other hand, he reflected that "this standard based system or whatever you want to call it system, could be better for just creating a consistent system across the whole school." This comment indicates that there is a perception that there is a lack of coherence across classrooms and departments which the policy may help with.

There is less consensus between the MHOD and MVC on what role assessment plays in the process of learning. As to the informal feedback, MVC noted in both the first and second interview that informal assessment "are just assignments and just walking around every day and looking at their work and giving feedback like that." The MHOD beliefs strongly in the role of ongoing feedback, as well. The role that quizzes play is one that is almost contentious. Quizzes are formal assessments that do not cover a full unit of standards and are usually completed bi-weekly or mid-unit. Both the MHOD and MVC used similar logic in justifying quizzes being used as minor summative assessments that count towards students' final grades. The MHOD explained that the team feels students have "had many opportunities to show their understanding [through]… the in class work, … all the homeworks, [and] a practice quiz…" For that reason, because they have "had opportunities for feedback," that the quiz is no longer a formative assessment *for* learning but a summative assessment *of* learning. The MVC justified that "the students would have gotten enough material and lessons to achieve the highest rating in the topics that are covered on that quiz" which is why they are formally reported as part of the grade. By the end of the semester there is a change in their narratives and evidence that the team has reflected on this point. The MVC calls the practice "debatable," acknowledging that there

is argument against its practice and that there is room for it. He maintains that he still believes quizzes "should be graded and they should have some effect on the overall grade." Yet, his justification includes the weight of the quizzes and impact on overall grade. "It impacts the grade, but it won't pull it all the way down or all the way up. The majority of the grade is still determined by the major assessments." Finally, the MHOD explained in the second interview that minor assessments are "one thing that [they] may change."

There was agreement by the MHOD and MVC that assessment information needs to be shared with other stakeholders, not only students and teachers. Namely, it should be offered to parents and counselors. The MHOD cautioned that excluding class assignments, homework, and quiz information in grade books "might be confusing for some parents that don't realize that certain things are excluded." Yet, he is still "fine with [excluding them.]" The MVC also expressed that while he no longer enters as much into his gradebook, some system for "alerting parents and counselors if [students] were building up tasks" is helpful to ensure that students "were actually attempting to learn the stuff." The belief here is that student learning is not a two-way street, between teacher and student, but that it is a network of people helping the student – a network that needs information on students' progress.

There is a sense that assessments also serve as a motivator for students to do the work to learn. The MHOD noted that the decision to either include or exclude assessments is a "slippery slope in terms of motivation" for both students and teachers. For students, if a piece of work does not count towards a grade, they may not put their full effort in. For teachers, he explained that he may not spend as much time offering

teacher-directed feedback to students. The MVC was more direct, expressing in the beginning of the semester that “we would like that students are intrinsically motivated to learn and do well, but obviously that's not the case.” In fact, it is for that reason that he starts to call into question validity of quizzes as markers of student achievement if they are not graded. “If you want them to be a measure of progress and they're not graded, you can't guarantee that the students will be motivated to try their best on that formative.” While there is still not clear consensus as to how they motivate, there is an agreed belief that they do in some way.

Practices. The creation of tests has been a significant change in practice. The MHOD explained at the beginning of the semester that “tests aren’t going to look the way they used to” and that “most people were on board with the concept of - if we want to really test for higher level skills, you can’t have one hard question on the test.” At the end of the semester, he explains “that's probably the biggest change department-wide.” The key is that they are “trying to say, not what was [students’] raw points that [they] got, but what level of understanding were [they] able to show?” The MVC also took time to detail how quizzes and tests have changed. During the interview he took one out to share and stated, “this is a quiz, not a test, but it's the same idea.” Questions are labeled by three ‘levels,’ level three being “the more difficult “applications kind of questions.” The MHOD and MVC do things a little differently in that the MHOD creates the rubrics for the test before and/or during the creation of the test. The MVC tallies points while he grades and “that's how I determine these grade boundaries” using the levels as “the descriptors that I work off.” So, while the practice is still to record a percent, it could be

that “80% [of questions correct] reflects more towards” an A than having to get at least 94% of the questions correct.

The MHOD explains how this has been a significant change in practice because before they “were reverse engineering their tests with how they felt their students were going to do. So as they're building a test, it's like, should I add another hard question and another hard question?... it's like this weird game that they're trying to play with, with their ideal 'A' student in mind, just making sure that that ideal ‘A’ student still gets an A." This norm-referencing style of grading was also referenced by the MVC. He explained that “I would grade the assessment and determine the raw score. Then, considering the average of the entire class I would define a curve such that the average of the class is around a B or B-, but never curving more than 6 or 7%.” The intent was to accommodate “unforeseen circumstances,” like a question that was too difficult or a topic that wasn’t covered but was accidentally included on the test.

The MHOD is more innovative in his practices. Using the high jump metaphor, he attempted to spiral assessments, whereby he would include questions from the previous unit on the current unit to reassess students. This proved to be difficult to sustain, because he does not appear to practice it by the end of the semester. As to quizzes, he does not always grade it and give feedback. Instead, because he does not grade it, “suddenly it's like, well, if it's not going in the gradebook, why don't I give you the solution key as well and you can mark it on your own and ask questions.” The MVC has changed his testing practice and no longer includes assignments and informal formative assessments in the gradebook. He also offers reassessment according to the policy. Where he used to drop

the lowest grade, he now offers students the opportunity to reassess any test so long as they show that they completed all the in-class assignments and homework beforehand.

Materials. The policy document is not something the MHOD and the math department references often. Instead, the team "went over that a lot when that process was going down and basically we approach the like. Is there any of this that we don't like? And we would go over it bullet point by bullet point and see like what disagreed with our current practices." Aside from some "nit picking over the vagueness of it… we don't go through it a lot only because there weren't any issues." Aside from the beginning of the semester, it would be referenced like a "rule book."

There was a point on policy and its impact on practice in the gradebook. Specifically, the MHOD noted that the policy should "have been firmer with the shape of how things would have looked." Instead, it did not go far enough with clarity of practice, which could have empowered and enabled change. The administration had started with some clarity and then

backtracked." This led to at least some personal frustration because it left it open to interpretation, breeding less coherence. The fact that some change was asked and then revoked early had the power of stalling changes in his department. For the team, he notes that the school "had a more narrow defined assessment plan and then things got opened up right after the beginning of the year." There was some "backtracking" by administration on how things were reported. He notes that normally he likes to lead a group by "consensus…, but this seems like a strange instance where they just wanted to be told what exactly to do and then the not knowing was more stressful than being told what to do."

Aside from potentially shaping the practice, the gradebook is a communicator. Inputting grades sends messages to parents, students and counselors. When reporting progress, the SVC noted a desire to have consistent language for the formative and summative process to ensure parents understand. Further, when students are not submitting work, it is important for "alerting parents and counselors if they were building up tasks." He went on to state that to help with consistency, all parties could be "trained on [the gradebook platform] just to understand what [it is telling [different stakeholders] and the data that is relevant versus is not relevant." With the new practice of entering less assignments, the MHOD pointed out, "it may take a month or five weeks before you've built up enough" documentation in the gradebook, which " might be confusing for some parents that don't realize that certain things are excluded."

A third role that the gradebook may play is motivational. The MVC explained that the power in the "previous system, let's call it using percents and ABCDs and grading formative assessments and inputting those grades, putting in a zero for an assignment usually prompted the student to see that zero and come to me or whatever teacher the next day and say, how can I fix this is very immediate kind of response." The MHOD pointed out that it changed his own motivation, as well. When an assignment/assessment wasn't graded, he decided he would not personally give as much feedback on the work. It changed his practice by instead using the opportunity for students to collaborate, share, and ask him questions with the answer key. He's "less strict on the way the formative plays out now. Now that I've done away with that in the grade book, I've been really flexible with that."

Lastly, the gradebook can be designed to offer some perceived separation of the teacher from the process of determining a student's grade. The MVC believes that the point-based, percent-styled gradebook offers a level of separation and objectivity in the evaluation of student performance, which is ultimately more accurate. He expressed strong discomfort with the practice of teacher's using professional discretion to look at the body of evidence for recent and consistent marks to determine a grade. Professional discretion allows an "element of emotion and subjectivity to creep into those decisions for students." Whereas "using the standard grading system… I have never really gotten to the end of a year and looked at a student's grade and thought … that grade wasn't accurate." Except, he made a comment that sometimes there were "students who are on the boundary" and using "a little subjectivity on my part, giving them a little bump." The difference lays in the degree of professional discretion offered between the systems. In the point-based, percent-styled gradebook professional discretion is not stated outright. Points are accumulated and translated into percentages and letter grades. The discretion lies in adding and subtracting points to change outcomes. In a system that explicitly states that teachers look at the body for consistency and recency, discretion is the advertised practice.

Standards play a key role in much of the work that the teachers do. The MHOD explains that he uses these frequently on a personal level. "I just print out the book for myself and then when I teach it with the group I highlight it, and I check over it often and make little notes if I need to cycle back here. So this little note for myself to know that we didn't do it enough here… And then by April, everything's going to be highlighted and we'll have two or three weeks to review." As a team, they meet to unpack and discuss

the standards to build horizontal alignment of students' classes, charting clear pathways and opportunities for students and for their standards to spiral in those classes. In fact, more than assessment practices this semester, his team looked through the standards to determine which course will teach what content at what grade level and class. He shared an impressive bulletin board in the math meeting room with each standard printed and laminated and charted up with tacks and lines drawn between them.

As to assessment reform, the way standards are written and the number of standards that are required impact the way that teachers think about and practice assessment. In the MS and/or in other subjects, there are less standards and/or they are more skill-oriented standards. When written this way, it may be easier to shift practices beyond averaging. The MVC stated that with the Common Core "strands" or "dimensions," one can be "more generally assessed multiple times and then standards that are more specific that aren't assessed once or twice." But in high school "the standards are more specific" and so are "assessed once or twice." In the MS … "they are building up their relationship with math and these generalized skills are important." There is less control in the HS because it appears that there are more standards. The MVC is clear that "if there were a set of generalizable standards that were assessed multiple times then I would lean towards the process [of using the most consistent and recent evidence to determine a final grade]." However, "if those standards and summative assessments are more specific then this process of evaluating the most recent and most common scores, I think is not best process for the situation too meet the specific standards that are assessed once or twice."

Externally assessed courses, the IB and AP, have their own standards and style of assessment. This both drives and limits the team's work. The MHOD went as far as to say teaching the IB courses, "most teachers still feel a little boxed in with what that final assessment, that we know is going to look like the IB exam." This leads them to, "teaching to the test in those junior and senior level courses… in both content and type of question." The MVC independently noted the influence of the externally assessed coursework beyond just the junior and senior years. "Once you get to high school it starts to become a lot more about preparing for AP courses and SATs that have exams that are courses that focus on specific skills… especially IB courses."

The creation of rubrics helped the MHOD to move away from "raw points" to a more "descriptor-based" assessment of learning. And an unexpected outcome of the switch for him was "the actual marking … is a lot more enjoyable because you're just making comments to students at that point and not really counting points." This movement to comments appears to enable more meaningful feedback to students. Instead of the feedback being a number indicating a level of correctness, outlying and patterned mistakes are observed by the teacher and shared with the student for improvement.

Other. The connection of the importance of team teaching was made early in the interview when asked about how reforms were going for him. There was an interesting contradiction because he claimed to be "lucky" to be the only one teaching one course. Yet, while this gave him freedom to try new things in his own classes, where teachers are more traditional they have freedom to continue old practices. On the other hand, reforms are a "slower process" with team teaching due to difficulty of getting members to agree. In fact, he confessed that there is a "planning for confrontation… because you're

assuming it could be a contentious conversation." The contention arises largely around the making of tests.

Another point that was brought up was both the MHOD's leadership style and that of administration. To leadership style, the MHOD took a more hands off approach. He explained, "I don't often tell them…. I just use a model of... I will I will show you what I'm doing. I would explain why I think this is the right way to do it, but it's optional if you want to." His role is more of a facilitator and coach. But he does note that there are some things that "have to be a certain way because they're coming from the administration." The MHOD felt like there should have been more clear direction and less wavering. One directive about reporting grades was rescinded a week into the semester and "that was just, I think that was just a failure in leadership in the way that they got rolled out. I think I think most of the people were ready to go with it." In fact, the move may have worsened the situation because "once you make that headway, I would not advise backtracking, because then then it's like let the barn door open."

APPENDIX K

IRB APPROVAL

INSTRUCTIONS
Complete each section of the application. Based on the nature of the research being proposed some sections may not apply. Those sections can be marked as N/A. Remember that the IRB is concerned with risks and benefits to the research participant and your responses should clearly reflect these issues. You (the PI) need to retain the most recent protocol document for future revisions. Questions can be addressed to research.integrity@asu.edu. PIs are strongly encouraged to complete this application with words and terms used to describe the protocol is geared towards someone not specialized in the PI's area of expertise.

IRB: 1. Protocol Title:

IRB: 2. Background and Objectives

2.1 List the specific aims or research questions in 300 words or less.
2.2 Refer to findings relevant to the risks and benefits to participants in the proposed research.
2.3 Identify any past studies by ID number that are related to this study. If the work was done elsewhere, indicate the location.

TIPS for streamlining the review time:

- ✓ Two paragraphs or less is recommended.
- ✓ Do not submit sections of funded grants or similar. The IRB will request additional information, if needed.

Response: I aim to explore how a high school faculty adopts a significant assessment policy change into their practice and beliefs over a single academic semester, August 2020 to January 2021. The research questions are:

1. How does an institution's adoption of a new policy of assessment and grading impact a faculty's beliefs about assessment over one semester?
2. How does an institution's adoption of a new policy of assessment and grading impact a faculty's assessment practices over one semester?
3. How similar or different are changes in beliefs and practices among teachers?
4. What are teachers' perceptions of factors that led to any changes in their beliefs and practices?

The benefits are mainly a moment to purposefully reflect on practices related to a work-related reform/innovation. There are no foreseeable risks to engaging in the research.

IRB: 3. Data Use – What are the intended uses of the data generated from this project?
Examples include: Dissertation, thesis, undergraduate project, publication/journal article, conferences/presentations, results released to agency, organization, employer, or school. If other, then describe.

Response: This work is primarily part of my dissertation for my Education Doctorate with ASU. I do aim to share the findings via journals and conference papers, as well. Further, I will share my work with the team of educators and leaders with whom I have undertaken the study.

tIRB: 4. Inclusion and Exclusion Criteria

4.1 List criteria that define who will be included or excluded in your final sample.

Indicate if each of the following special (vulnerable/protected) populations is included or excluded:

- Minors (under 18)
- Adults who are unable to consent (impaired decision-making capacity)
- Prisoners
- Economically or educationally disadvantaged individuals

4.2 If not obvious, what is the rationale for the exclusion of special populations?

4.3 What procedures will be used to determine inclusion/exclusion of special populations?

TIPS for streamlining the review time.

- ✓ Research involving only data analyses should only describe variables included in the dataset that will be used.
- ✓ For any research which includes or may likely include children/minors or adults unable to consent, review content [here]
- ✓ For research targeting Native Americans or populations with a high Native American demographic, or on or near tribal lands, review content [here]

For research involving minors on campus, review content [here]

Response: I will share my survey questionnaire with the entire high school faculty and hope for full participation – these are included. I aim to conduct video-cued interviews with six faculty members, one from each department: English Language Arts, Foreign Languages, Social Studies, Science, Mathematics, and the Arts. I do this because I want all curricula to be represented and I want to be able to compare different groups' experiences. The names of each faculty member per department will be randomly selected. I will reach out to them. If one declines, then another will be randomly selected. Mid-way through the investigation I aim to interview the heads of departments, the Principal, and the Assistant Principal. Minors, adults who are unable to consent (impaired decision-making capacity), prisoners, and/or economically or educationally disadvantaged individuals are all excluded from this research because their input is not specifically part of the investigation.

IRB: 5. Number of Participants

Indicate the total number of individuals you expect to recruit and enroll. For secondary data analyses, the response should reflect the number of cases in the dataset.

Response: 66 faculty for the survey questionnaire, 6 individuals for the video-cued interview, 6 heads of department faculty for the check-in interviews, and two administrators, for a total of up to 80 faculty. However, though I do not anticipate a need, I request some flexibility from the IRB to include more *faculty* in interviews as needed.

IRB: 6. Recruitment Methods

6.1 Identify who will be doing the recruitment and consenting of participants.

6.2 Identify when, where, and how potential participants will be identified, recruited, and consented.

6.3 Name materials that will be used (e.g., recruitment materials such as emails, flyers, advertisements, etc.) Please upload each recruitment material as a separate document, Name the document: recruitment_methods_email/flyer/advertisement_dd-mm-yyyy

6.4 Describe the procedures relevant to using materials (e.g., consent form).

✓

Response:

6.1 I will be doing the recruitment. The Head of School has given me consent to recruit.

6.2 The participants will be recruited at the beginning of August at the earliest. I will reach out to them in an email and, as needed and/or requested, follow up in person to be clear. I know that I need all faculty for the survey questionnaires, pre and post. I mentioned earlier that I would like six faculty from each of the six departments. I would also like to reach out to the heads of departments, the assistant principal, and the principal. They have been 'identified' because I aim to piece together a set of experiences surrounding the network of assessment and grading on campus.

6.3 For consent, I have prepared three documents. One is for the video-cued interview protocol. Another is for the heads of departments (HOD), and another is for the survey. The first two require signatures. I will provide the document to them electronically as well as have copies ready for them to sign at the interview. I have consent from the Head of School to conduct the interviews, questionnaire, and access documents for analysis. The consent letters and that from the Head of School are attached at the end of this IRB Proposal.

6.4 I will send out the email with the consent form. I will accept their digital signatures on the document, file them for safekeeping, and scan them to keep digital copies. Also, I will have hard copies for their signature at the interview if they prefer that option. In any and all interviews, I will ensure to ask and have on recording that they consent to being recorded.

IRB: 7. Study Procedures

7.1 List research procedure step by step (e.g., interventions, surveys, focus groups, observations, lab procedures, secondary data collection, accessing student or other records for research purposes, and follow-ups). Upload one attachment, dated, with all the materials relevant to this section. Name the document: supporting documents dd-mm-yyyy

7.2 For each procedure listed, describe who will be conducting it, where it will be performed, how long is participation in each procedure, and how/what data will be collected in each procedure.

7.3 Report the total period and span of time for the procedures (if applicable the timeline for follow ups).

7.4 For secondary data analyses, identify if it is a public dataset (please include a weblink where the data will be accessed from, if applicable). If not, describe the contents of the dataset, how it will be accessed, and attach data use agreement(s) if relevant.

TIPS for streamlining the review time.

- ✓ Ensure that research materials and procedures are explicitly connected to the articulated aims or research questions (from section 2 above).
- ✓ In some cases, a table enumerating the name of the measures, corresponding citation (if any), number of items, sources of data, time/wave if a repeated measures design can help the IRB streamline the review time.

Response:

1. The new policy has been adopted by the school. Faculty will begin their work with it in the coming year indefinitely.
2. I will conduct and record six video-cued interviews during orientation week and/or the week afterwards in the faculty's classrooms or some other appropriate place on campus.
3. I will distribute the HS Faculty Survey Questionnaire during orientation and/or the week afterwards in an email and give the faculty one week to complete it.
4. I will conduct and record interviews with the heads of the department after progress reports with quarter one comments are completed in the faculty's classrooms or some other appropriate place on campus.
5. I will analyze the quarter one comments of the same six faculty who participated in the video-cued interviews on my computer by accessing the electronic gradebook.
6. I will distribute the HS Faculty Survey Questionnaire during orientation and/or the week afterwards in an email and give the faculty one week to complete it.
7. I will conduct and record six video-cued interviews during orientation week and/or the week afterwards in the faculty's classrooms or some other appropriate place on campus.
8. I will collect final exams and copies of tests from the same six faculty members who participated in the video-cued interviews after final grades are entered for the semester in January 2021.
9. I will conduct and record interviews with the heads of the department after progress reports with quarter one comments are completed in the faculty's classrooms or some other appropriate place on campus after final grades are entered for the semester in January 2021.
10. I will distribute the HS Faculty Survey Questionnaire after final grades are entered for the semester in January 2021.

IRB: 8. Compensation

8.1 Report the amount and timing of any compensation or credit to participants.
8.2 Identify the source of the funds to compensate participants.
8.3 Justify that the compensation to participants to indicate it is reasonable and/or how the compensation amount was determined.
8.4 Describe the procedures for distributing the compensation or assigning the credit to participants.

TIPS for streamlining the review time.

- ✓ If partial compensation or credit will be given or if completion of all elements is required, explain the rationale or a plan to avoid coercion
- ✓ For extra or course credit guidance, see "Research on educational programs or in classrooms" on the following page: https://researchintegrity.asu.edu/human-subjects/special-considerations.
- ✓ For compensation over $100.00, review "Research Subject Compensation" at: https://researchintegrity.asu.edu/human-subjects/special-considerations for more information.

Response: I will offer a coffee voucher to participants in each of the high school survey questionnaires. The funds for this will be from my own wallet. It is really only for an additional incentive to get survey results from the entire faculty. There are several cafes on campus. A single cup of coffee for each US$ 1-2. With 66 faculty, that is potentially US$132. Upon completing the survey, faculty will be directed to a page with a "secret word." They will email me with the secret word and I will be sure to pass on the voucher to them in their mailbox or in person.

IRB: 9. Risk to Participants

List the reasonably foreseeable risks, discomforts, or inconveniences related to participation in the research.

TIPS for streamlining the review time.

- ✓ Consider the broad definition of "minimal risk" as the probability and magnitude of harm or discomfort anticipated in the research that are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.
- ✓ Consider physical, psychological, social, legal, and economic risks.
- ✓ If there are risks, clearly describe the plan for mitigating the identified risks.

Response: In discussing their beliefs and practices, if they do not already naturally align with the school's new policy, then the faculty may feel threatened. I am an assistant principal of the elementary division, which means I have no evaluative power of them, but do work in circles of other administrators. This may be perceived as a threat.

It may be difficult to get data early in the year from faculty. There is a lot going on at the beginning of a school year. Especially with pandemic-sized precautions in place, participating in this study may not be their topmost concern.

IRB: 10. Potential Direct Benefits to Participants

List the potential direct benefits to research participants. If there are risks noted in 9 (above), articulated benefits should outweigh such risks. These benefits are not to society or others not considered participants in the proposed research. Indicate if there is no direct benefit. A direct benefit comes as a direct result of the subject's participation in the research. An indirect benefit may be incidental to the subject's participation. Do not include compensation as a benefit.

Response: There are no foreseeable risks.

IRB: **11. Privacy and Confidentiality**

Indicate the steps that will be taken to protect the participant's privacy.

11.1 Identify who will have **access to the data**.
11.2 Identify where, how, and how long data will be **stored** (e.g. ASU secure server, ASU cloud storage, filing cabinets).
11.3 Describe the procedures for **sharing, managing and destroying data**.
11.4 Describe any special measures to **protect** any extremely sensitive data (e.g. password protection, encryption, certificates of confidentiality, separation of identifiers and data, secured storage, etc.).
11.5 Describe how any **audio or video recordings** will be managed, secured, and/or de-identified.
11.6 Describe how will any signed consent, assent, and/or parental permission forms be secured and how long they will be maintained. These forms should separate from the rest of the study data.
11.7 Describe how any data will be **de-identified**, linked or tracked (e.g. master-list, contact list, reproducible participant ID, randomized ID, etc.). Outline the specific procedures and processes that will be followed.
11.8 Describe any and all identifying or contact information that will be collected for any reason during the course of the study and how it will be secured or protected. This includes contact information collected for follow-up, compensation, linking data, or recruitment.
11.9 For studies accessing existing data sets, clearly describe whether or not the data requires a Data Use Agreement or any other contracts/agreements to access it for research purposes.
11.10 For any data that may be covered under FERPA (student grades, etc.) additional information and requirements is available at https://researchintegrity.asu.edu/human-subjects/special-considerations.

Response: Only I will have access to the data. The data will be stored on my ASU Google Drive for a maximum of 24 months. The data will be deleted from the drive. The ASU Google Drive is password protected. Audio recordings will be saved in the same ASU Google Drive. The names of each participant will be done in order of the interview (i.e. Faculty One, Administrator Two, Head of Department 1). I will keep a separate document that identifies the person and their role to ensure I can keep them straight. This document will be held in my own personal Google Drive, separate from the ASU Google Drive. Consent forms will also be stored on the ASU Google Drive and kept for 24 months. Only names and department affiliations will be stored. I will also ask and keep some basic professional, demographic information like number of years as an educator, highest degree, etc. I will not be accessing existing data sets. Nor will I be using and analyzing students' data.

IRB: 12. Consent
Describe the procedures that will be used to obtain consent or assent (and/or parental permission).

12.1 Who will be responsible for consenting participants?
12.2 Where will the consent process take place?
12.3 How will the consent be obtained (e.g., verbal, digital signature)?

TIPS for streamlining the review time.

- ✓ If participants who do not speak English will be enrolled, describe the process to ensure that the oral and/or written information provided to those participants will be in their preferred language. Indicate the language that will be used by those obtaining consent. For translation requirements, see Translating documents and materials under https://researchintegrity.asu.edu/human-subjects/protocol-submission
- ✓ Translated consent forms should be submitted after the English is version of all relevant materials are approved. Alternatively, submit translation certification letter.
- ✓ **If a waiver for the informed consent process is requested, justify the waiver in terms of each of the following: (a) The research involves no more than minimal risk to the subjects; (b) The waiver or alteration will not adversely affect the rights and welfare of the subjects; (c) The research could not practicably be carried out without the waiver or alteration; and (d) Whenever appropriate, the subjects will be provided with additional pertinent information after participation.** Studies involving confidential, one time, or anonymous data need not justify a waiver. A verbal consent or implied consent after reading a cover letter is sufficient.
- ✓ ASU consent templates are [here].
- ✓ Consents and related materials need to be congruent with the content of the application.

Response:

I am responsible for consenting participants. The process will take place on campus. A digital or physical signature will be taken on the consent form.

IRB: 13. Site(s) or locations where research will be conducted.
List the sites or locations where interactions with participants will occur-

- Identify where research procedures will be performed.
- For research conducted outside of the ASU describe:
  - Site-specific regulations or customs affecting the research.
  - Local scientific and ethical review structures in place.
- For research conducted outside of the United States/United States Territories describe:
  - Safeguards to ensure participants are protected.
- For information on international research, review the content [here].

For research conducted with secondary data (archived data):

- List what data will be collected and from where.
- Describe whether or not the data requires a Data Use Agreement or any other contracts/agreements to access it for research purposes.
- For any data that may be covered under FERPA (student grades, etc.) additional information and requirements is available [here].
- For any data that may be covered under FERPA (student grades, homework assignments, student ID numbers etc.), additional information and requirements is available [here].

Response: Ruamrudee International School, Min Buri, Bangkok, Thailand.

IRB: **14. Human Subjects Certification from Training.**

Provide the names of the members of the research team.

ASU affiliated individuals do not need attach Certificates. Non-ASU investigators and research team members anticipated to manage data and/or interact
with participants, need to provide the most recent CITI training for human participants available at www.citiprogram.org. Certificates are valid for 4 years.

TIPS for streamlining the review time.

- ✓ If any of the study team members have not completed training through ASU's CITI training (i.e. they completed training at another university), copies of their completion reports will need to be uploaded when you submit.
- ✓ For any team members who are affiliated with another institution, please see "Collaborating with other institutions" [here]
- ✓ The IRB will verify that team members have completed IRB training. Details on how to complete IRB CITI training through ASU are [here]

Response: Nathan Robert Meisner, Dr. Elisabeth Gee

General Tips:

- Have all members of the research team complete IRB training before submitting.
- Ensure that all your instruments, recruitment materials, study instruments, and consent forms are submitted via ERA when you submit your protocol document. Templates are [here]
- Submit a complete protocol. Don't ask questions in the protocol – submit with your best option and, if not appropriate, revisions will be requested.
- If your study has undeveloped phases, clearly indicate in the protocol document that the details and materials for those phases will be submitted via a modification when ready.
- Review all materials for consistency. Ensure that the procedures, lengths of participation, dates, etc., are consistent across all the materials you submit for review.
- Only ASU faculty, full time staff may serve as the PI. Students may prepare the submission by listing the faculty member as the PI. The submit button will only be visible to the PI.
- Information on how and what to submit with your study in ERA is [here]. Note that if you are a student, you will need to have your Principal Investigator submit.
- For details on how to submit this document as part of a study for review and approval by the ASU IRB, visit https://researchintegrity.asu.edu/human-subjects/protocol-submission.