Students on Assessment: How Students Receive Assessment in a Senior Mathematics Classroom

by

David van Bergeyk

B.A., University of British Columbia, 1997

Thesis Submitted In Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Individual Program
Faculty of Education

SIMON FRASER UNIVERSITY

Spring 2013

Approval

Name: **David Jonathan van Bergeyk**

Degree: **Master of Arts**

Title of Thesis: ***Students on Assessment: How Students Receive Assessment in a Senior Mathematics Classroom***

Examining Committee: **Chair:** David John Pimm
Adjunct Professor

Peter Liljedahl
Senior Supervisor
Associate Professor

Nathalie Sinclair
Committee Member
Associate Professor

Rina Zazkis
External Examiner
Professor
Faculty of Education, SFU

Date Defended/Approved: April 11, 2013

Partial Copyright Licence

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010

Abstract

Over the past several decades, many voices in both the research and professional spheres of education have spoken on assessment, emphasizing its central place in the practice of teaching. But we have heard little of the voices of students on assessment. The purpose of this thesis is to hear what students have to say. In my mathematics classes I have used a variety of progressive assessment techniques for several years, but have sensed a disconnect between my intentions and students' reception of those techniques. In order to describe qualitatively my students' reactions to and reception of my assessment techniques, I surveyed and interviewed members of a senior academic mathematics class. I have concluded that, based on a mix of theoretical and practical factors in their mindsets, the students' reception of assessment does not usually align well with teachers' intentions.

Keywords: assessment; conceptions of assessment; fixed ability vs. growth mindset; didactic contract; norms; decision theory

For my wife and kids, who mean more to me than anything else in the world.

Acknowledgements

I wish to thank Dr. Peter Liljedahl for all that he has contributed to this project and to my professional growth in general.

I also thank Dr. Nathalie Sinclair for her help in refining this thesis.

Table of Contents

1. Introduction

I began my Masters degree knowing that in the end I was going to be addressing some questions regarding assessment in mathematics teaching. In the years immediately preceding the start of my Masters course work, much of my professional energy had gone into changing my own assessment practices. As a new teacher, I had relied for the most part on traditional quiz and test assessment practices, but became increasingly convinced, for various reasons, that those two traditional pillars of mathematics assessment were not enough for me. After several years of teaching senior academic mathematics, I was realizing more and more that scores on tests and quizzes did not always jive with what I felt students should be achieving. It struck me that the mismatch occurred in two directions: some students scored more poorly than both they and I would have guessed they would, and others scored better. In addition to this sense that traditional assessments were not measuring accurately, I began to get the feeling that they also were not measuring what mattered. I had always felt that conceptual understanding mattered, but assumed that traditional assessment was adequate for measuring conceptual understanding, believing as I did that knowing how to do questions procedurally indicated an understanding of the concepts underlying the procedures. I was definitely starting to question this belief.

Based on these developing dissatisfactions with traditional assessment methods, I started experimenting with alternative assessments. It began with a variety of writing assignments in which I asked students to explain in their own words the concepts they were learning. To these strategies, I added alternative ways of demonstrating knowledge, including various projects and open-ended problem-solving assignments. I also tinkered with my traditional approaches, choosing not to count quizzes for marks, and allowing students to retake tests and show corrections for credit. To coordinate my efforts, I tried a few different forms of portfolios as ways for students to collect and demonstrate evidence of their understanding over time. I even ended up dispensing with my previous dependency on a computerized gradebook for collecting and computing

scores and term grades. I adopted a much less rigid, evidence-based evaluation scheme that allows me to include assessment data other than simply numerical scores. These changes did not all occur immediately, nor simultaneously, and have subsequently been adjusted repeatedly as well. Suffice to say, I was, over time, paying close attention to my assessment strategies and trying to solve or address the dissatisfactions I noted above.

These alterations to my practice did not all result from my own original ideas, though. I was influenced in this professional change by several key documents that I came across at the time. This reading was motivated by a professional purpose, to choose and justify strategies that would help me to better measure and promote student learning. These documents, though each in its way is influenced by academic research, are from the realm of professional literature. One of the sources that influenced my newly-formed ideas was in fact the curriculum document my courses were based on (British Columbia Ministry of Education, 2000). Like many teachers, I tended to focus only on the specific learning outcomes in the curriculum document. But as I started to experiment with assessment, I also began to notice the kinds of extra information in the curriculum document that could help to develop a richer assessment program, including the appendix of the document that addresses assessment and evaluation directly (BCED, 2000, Appendix D, pp. D-1 to D-5). Another resource I had at hand was the teacher's guide for the main textbook resource I was using at the time (Davis & Gadanidis, 1999). Here again was a fairly detailed explanation of how and why to assess for understanding in mathematics.

While these resources were already on my shelf at the time, I also started to see and hear (or maybe just pay attention to) references to other influential professional books on assessment. A major influence on my thinking at the time was the aptly named *Rethinking Classroom Assessment with Purpose in Mind* (Western and Northern Canadian Protocol for Collaboration in Education, 2006). This book elaborates on progressive assessment based on the now-familiar framework of Assessment *of*, *for*, and *as* Learning. Other books by people like Rick Stiggins (1997) and Ken O'Connor (2002, 2007), which were all the rage in professional conversations of the day, also came to my attention. The planning approach known as Understanding by Design also influenced my thinking regarding assessment (McTighe & Wiggins, 2001). Similarly, *Smerging Data,* a book put out by the Alberta Assessment Consortium (Mulgrew &

Rawe, 2001), influenced my thinking about grading practices, and allowed me to see that traditional averaging of scores was not the only nor the best way to deal with assessment results. The net result of all this reading was to confirm my conviction that adopting a more progressive assessment approach in my classroom would benefit my students' learning and help me measure and report more accurately on their progress.

I was convinced that I was onto something with my changing assessment practices. References within the documents noted above pointed to academic research that stated a strong case for progressive assessment, and though I did not at the time delve too deeply into the academic literature, I did make some initial forays. Two articles I read at the time, in particular, are worth mentioning. Romagnano (2001) dispels a central myth in mathematics assessment, stating that "objectivity would be wonderful if we could have it, but it does not exist" (p. 31). This article helped to justify my use of increasing, or perhaps I should say merely more obvious, subjectivity in my assessment. The other major article, referenced repeatedly in the documents I was reading, was Black and Wiliam's (1998a) seminal work outlining the effects of formative assessment. This article makes some strong claims about the impact of formative assessment in the classroom, essentially saying it is the single greatest means of improving student achievement. With an endorsement of this sort behind my efforts, and convinced by Romagnano about the myth of objectivity, I was fully motivated to continue with progressive assessment. I expected nothing less than a dramatic transformation in my students' learning, with obvious increases in depth and breadth of understanding, retention and transfer of knowledge, and even motivation to learn.

But this is not what actually happened. My efforts did not fail outright, by any means, for I did indeed see some increases in those grand goals. But the transformation was far less dramatic than I had been hoping for. Though I did see that to some extent students were turning their focus to concept development and understanding, some remained decidedly resistant to and put out by my alternative practices. While some students commented that they appreciated a chance to demonstrate understanding in different ways, and liked having a less rigid timeline on their learning, others seemed exasperated by having to do more than just get questions right on tests. Basically, the ideals that were taking shape in my mind as a teacher were not completely jiving with the realities I faced in my classroom.

While living this tension as a teacher, I began my Masters program, and have subsequently decided to focus my research project on exactly this tension. I have explored some of the factors in my classroom that contribute to this tension, and will attempt to describe in part why the idealization of progressive assessment for teachers does not always accord with student perceptions of same. In particular, this research project focuses on the factors in students' mindsets that influence their reaction to progressive assessment. In the next chapter, I will review the literature that shows how teachers idealize progressive assessment as well as some relevant literature about student mindsets. Chapter 2 concludes with my research question, which arises from my sense of professional discontent and fits in the intersection of the literature on assessment ideals and the literature on student conceptions. In chapter 3 I describe the methodology I used to address the question, which involves using the features of student mindsets as a lens to interpret my qualitative data. Chapters 4 and 5 present the results of my research, first explaining each student's unique reactions, and then collecting the common themes that connect some of the students' reactions together. In chapter 6 I discuss some theoretical considerations from the fields of economics and psychology that allow me to extend the analysis of the common themes. Finally, in chapter 7 I give some conclusions and answer my research question.

2. Assessment: Influences and perceptions

The field of assessment in mathematics education is full of tension. This metaphorical tension is felt of course by the classroom teacher, though perhaps only subconsciously, as she is pulled in different directions by competing influences. But the reader of the research literature also perceives a field in tension when he sees a decades-old discussion with as yet unresolved problems in defining terms and explaining phenomena (Frey & Schmitt, 2007; Taras, 2008). As part of the personal journey I described above regarding my own assessment practices, I went looking in the literature for a coherent vision of and justification for progressive assessment, but did not find it. Instead I found that assessment practices are influenced by competing philosophies, and that any single assessment decision is often based on a sometimes inconsistent mix of the various influences. I have sorted the influences into three broad categories that pull in different directions on a teacher as he or she makes assessment decisions. Each broad category has one central force at its center, with some related sub-categories around it that exert influence in similar ways. The three central influences are tradition, theories of knowledge, and the didactical situation. Note that this mapping I am imposing on the field is far from perfect, and is not meant to imply that any single influence is primarily responsible for one assessment decision or another. Rather, I am attempting to sort the influences to reveal which ones coincide more closely and which ones pull in competing directions. But the idea is of overlapping influences rather than sharp delineations. My attempt here is to map the tensions teachers feel and to depict as well as possible what the literature portrays as an ideal of progressive assessment within this map of competing influences.

2.1. Major Influence: Tradition

The first major influence to examine is perhaps one of the strongest: tradition. Tradition influences assessment decisions in at least two distinct ways. Every individual

has a personal tradition of assessment experiences, and the system as a whole also has a defined historical tradition: put simply, this tradition is the quiz and test approach (Buhagiar, 2007; Romagnano, 2001). For individuals making assessment decisions, their own prior experiences as subjects of assessment make up the largest pool of example assessment strategies for them to draw on. In many cases, too, educators happen to be people who achieved at least moderate success in their past assessment experiences, which contributes to a strong intuitive sense of the rightness of these approaches. Given that all individuals who end up as policy makers, parents, administrators, and other stakeholders in education have similarly experienced a personal tradition of assessment, it is unsurprising that the system as a whole is also influenced by a sense of tradition (Gipps, 1994). In fact, the influence of tradition is so strong and insidious that, as Shepard (2005) points out, sometimes assessment methods that are being implemented as new and innovative are in fact merely thinly-disguised traditional approaches.

In any practical field, there will exist conservative influences that resist change, and the field of assessment, at a practical level, is no different. Note that my description here of this influence is making no reference to any of the particular ideologies that created the 'traditional' practices in the first place. Those are other influences to be described later. Here I am just considering the influence of traditional practices for the sake of tradition itself; in other words, I am talking about the 'because that's the way we've always done it' mindset.

Of course, tradition for its own sake is very rarely a person's conscious choice of influence. Usually, practitioners will refer to some other reason for using traditional practices when they justify their choices. These justifications point to other influences that contribute to the tensions in the field. Though they may relate strongly to tradition, they are nonetheless distinct, and may in fact influence a person who might otherwise reject tradition on its own terms.

2.1.1. Efficiency

The first such related influence, then, is efficiency. In practice, many assessment choices are based on what methods are most easily implemented given classroom or

research limitations (Buhagiar, 2007). In the classroom, for example, a teacher may choose multiple-choice tests (a traditional approach) for the reason of efficiency. But it should be noted that he or she may also be influenced by efficiency in designing an alternative assessment device – choosing particular constraints on a project, for example, so that the final products are easier to grade. Ginsburg (2009), in establishing the need to support an alternative assessment approach (in the case of his article, an interview) by extensive professional development, is tacitly acknowledging the teacher's need to feel efficient. The influence of efficiency, though, is related to tradition's influence because even though teachers may not be consciously looking for traditional techniques, once they give in to the efficiency pull, they are more easily drawn to traditional techniques as well. A parallel tendency occurs at the level of system-wide assessment approaches, as noted by both Shepard (2005) and Wiliam (2007). Districts and other education authorities sometimes attempt to implement "formative assessment" programs using commercially available benchmark or interim tests. While well-intentioned, these programs end up being "'early-warning summative' assessments rather than … true formative assessments" (Shepard, 2005, p. 2).

2.1.2. Ranking and Sorting

A second influence, also often responsible for system-wide interventions, that is related to tradition is the ranking and sorting influence. In fact ranking and sorting is even more closely linked to tradition, and is in fact responsible for the very presence of the quiz and test tradition itself (Buhagiar, 2007). While the tradition may have risen out of the pressure or need to sort people for jobs and such, I contend that the tradition influence for its own sake has somewhat overshadowed this ranking impulse, and that many teachers would not consciously admit to feeling the ranking influence. But assigning any sort of single statistic to a student is essentially a ranking act. The term "grade" is used in common language to refer to levels of quality (think Grade A beef or eggs). Any time a teacher consciously or subconsciously applies bell-curvish thinking to their assessment approach, they are under the influence of the ranking impulse. This happens often with mathematics teachers whenever there is a competitive grading impulse in their classroom, whenever the focus of marking papers becomes looking for errors, whenever, basically, teachers become students' opponents (Strickland &

Strickland, 1998). The power of the sorting influence is again often unacknowledged, but amounts to producing "definitions of who people are" (Strickland & Strickland, 1998, p. 131), such that any assessment becomes high stakes inadvertently whenever ranking and sorting is at play (Gipps, 1994). Again, there is an insidious element to the pull of sorting, even influencing the setting of criteria or the development of exemplars in alternative assessment, for, as William Angoff states, "lurking behind the criterion-referenced evaluation, perhaps even responsible for it, is the norm-referenced evaluation" (quoted in Wiliam, 2010, p. 254). I take this to mean that even when we presume to assess students against some external standard and not directly against each other, we are somehow defining the standard based on tacit ranking and sorting.

While ranking of students is an influence on assessment in any discipline, there is, I believe, a unique way in which ranking influences assessment in mathematics. At district, state, and national levels, scores on mathematics tests in particular are taken as indicators of the general well-being of an education system (Gipps, 1994; Pegg, 2003), and indeed as indicators of broader economic well-being. As such, enormous political and social pressures bear on mathematics assessment decisions, because not only are individuals being ranked and sorted by the data, but so too are schools, districts, and even entire societies.

2.1.3. Precision

Intimately connected to the impulse to rank and sort is the allure of objective-seeming data. I call this the precision influence. This is the impulse to have assessments that provide results in the form of percentages, preferably with several decimal places. This is also the impulse that encourages people to gather many numbers for their gradebooks, like feeding a voracious electronic beast. Obviously the connection to ranking is strong, because the more precise the data, the more hair-splittingly ranked students can be. As I mentioned in the introduction, though, Romagnano (2001) explains that the assumption of objectivity in such data is a myth: "Objectivity, like the mythical pot of gold at the end of the rainbow, would be wonderful if we could have it, but it does not exist" (p. 31). Even without the desire for objectivity, per se, the ideal of precision still influences practice. In different conversations with several colleagues about rubric design, I have witnessed firsthand their desire to turn a device used to give open-ended

but useful feedback to students into a device used to generate useful scores for their gradebooks. Here the rubric is an acknowledged subjective tool, but there is still a strong draw towards 'the number.'

2.2. Major Influence: Theories of Knowledge

The precision influence is also intimately connected with another force in the field, one that is almost entirely at odds with the ranking and sorting impulse. This is the central influence that I will call theories of knowledge. This impulse relates to the central concept that assessment is measuring *learning*. It sounds ridiculous to suggest that assessment is ever anything but measuring learning, I know. And yet we have already seen that at least three other influences affect our assessment decisions without necessarily being about the measurement of the goals of education (i.e., learning). Perhaps this suggests that theories of knowledge serve as another main category distanced somewhat from tradition and its related influences.

So, when a practitioner turns his attention toward measuring learning, he is feeling the pull of the theories of knowledge. As Bodin (1993, p.116) says, "any assessment is underlaid by a particular perception of the nature of knowledge," and consequently assessment decisions are sensitive to changes in one's perceptions of knowledge (Strickland & Strickland, 1998). The influence of theories of knowledge often manifests itself as a pull away from tradition because practitioners become disillusioned with the picture of knowledge that underlies the test and quiz tradition. In general, non-traditional assessment approaches (i.e., alternative assessments) are uniformly based on a broader perspective of what constitutes important knowledge (Pegg, 2003). The impetus for broader perspectives on knowledge is of course nothing new, going back at least as far as the seminal work by Bloom (1956) in which he delineates his taxonomy. An interesting question to which we will return is why these old ideas have not gained more widespread approval in the decades since Bloom wrote. More recently, however, researchers in assessment are calling for not just a *broader* definition of what constitutes the knowledge to be measured, but also *better* definitions. To this end, Wiliam (2010) highlights the need to move beyond analyzing assessments for validity at a technical level towards analyzing the validity of the constructs they were set up to measure in the

first place. That is, one may establish a particular assessment as technically valid, but still question whether one can infer with confidence about the knowledge in the population assessed. Within the formative assessment research literature, the idea of better (and more broadly) defining the learning to be acquired is central (Popham, 2008; Sadler, 1989).

2.2.1. Mathematical Knowledge

The knowledge influence as discussed above entails a general consideration of knowledge in the abstract, not any particular subject matter. Within the field of assessment in mathematics, close to the general "theories of knowledge" influence, there is also an influence of how mathematics itself is perceived. That is, what we think and believe about mathematics itself exerts its influence on assessment decisions (Raymond, 1996). Central to any discussion of this influence is the debate over the relative importance of procedural fluency as opposed to conceptual understanding (Hiebert & Lefevre, 1987; Skemp, 1976). As is the case with a broadening view of knowledge generally, the past several decades of research in mathematics education have emphasized a broadening view of what it means to do mathematics (Bahr, 2007; Boaler, 1998; Galbraith, 1993). Moreover, a clearer picture is forming of the psychological and social mechanisms by which mathematical concepts develop (Ginsburg, 2009; Schiralli & Sinclair, 2003). Here again the mathematics influence, if it pulls the practitioner towards it, often pulls away from tradition. As Buhagiar (2007) has outlined, the quiz and test tradition was established based on the economic need to rank and sort people for advancement or further education, well before many of these ideas matured about what school mathematics is or should be.

It should be noted however, that both the knowledge and mathematics forces can still exert influence on someone who is otherwise quite attracted to tradition. There is nothing to say that a teacher would not try to improve in general the tests and quizzes he administers to be more representative of various types of knowledge, nor that a teacher would not fuss over particular quiz questions in order to better represent what she believes to be accurate mathematical concepts. Similarly, there may be interplay between the precision influence and these two forces, in that the desire to measure

better might well be reflected in a desire to get more measurements and more numerically exact measurements.

A somewhat more interesting consequence of the pull of the knowledge and mathematics influences is what happens when the practitioner realizes an accurate description of mathematics, and indeed any knowledge, is an impossible ideal. Romagnano (2001) makes this clear with respect to mathematics learning when he notes that a "student's mathematical understanding…is a 'mental concept' and as such can be observed only indirectly" (p.31). Bodin (1993) goes further in saying that "assessing demands the making of faithful reports about an unknown entity" (p. 116). Elaborating on this idea, he implies that assessment has its own version of Heisenberg's uncertainty principle, whereby the more precisely you pin down exactly where a student's knowledge currently is, the less you know about the direction or rate or quality of her learning progress. Even with respect to the clearer picture of developmental psychology that I mentioned earlier, there is some question as to its value, as Bodin (1993) and Mellin-Olsen (1993) both imply that all learners' develop idiosyncratically anyway. Remember too that in general constructivist learning theories have come to dominate educational theory (Boaler, 1998; Gipps, 1994), so that it is pretty widely accepted that students create their own understanding rather than receiving transmissions of knowledge. For Mellin-Olsen (1993), this belief causes him to ask "How can the educator assess the quality of students' knowledge when they acquire knowledge in their own ways, as long as the educator does not know about those ways?" (p. 147)

So what does happen when this realization dawns? Possibly, the knowledge and mathematics influences are weakened by this realization, which allows some of the other forces to exert more influence. In the cynical scenario, efficiency or tradition might take over here, and the teacher will go back to what is typical and what is easiest because he feels anything 'better' is a pipe-dream anyway. Or alternatively, and more positively, the teacher may be pulled instead towards increased focus on the actual and idiosyncratic didactical situations in her classroom. Bodin (1993) explains that the development within mathematics education research of the "didactical approach" did exactly this for the research field, allowing researchers to lay aside the notion of measurement, given that the question of what is to be measured cannot be answered. In my mapping of the

influences on the field, this focus on the actual classroom represents a third set of forces, a pair that I will call the pedagogy influence and the learner influence.

2.2.2. Curriculum

Before I elaborate on those two influences, however, I should first add another influence to the category of theories of knowledge. In close proximity to the knowledge and mathematics influences is the curriculum. The curriculum is exerting its influence whenever a practitioner makes assessment decisions based on his or her direct interpretation of published standards or learning outcomes. This may be as simple as setting up a traditional quiz to 'cover' particular outcomes, or be as complex as realigning a gradebook to reflect outcomes-based assessment as opposed to task type categories. Obviously the established curriculum is related to and even defined by widely held ideas about what constitutes knowledge in general and mathematical knowledge in particular (Pegg, 2003), but it is not the same thing as those two and their separate influences on the practitioner's view of assessment. In fact, it is quite possible that a practitioner may be quite oblivious to or even dismissive of theories of knowledge, for example, and yet will base assessment decisions on mandated curriculum. "Today, as the mathematics curriculum projects a new philosophy of a dynamic mathematics curriculum, assessment of the learning of that curriculum has also become more dynamic" (Raymond, 1996, p.3). As Ridgway (1998) points out, curriculum ambitions can influence assessment practices and bring them more in line with ideas about what constitutes important learning and significant mathematics. Conversely, of course, if the pull of tradition or other influences is too strong, assessment practices may instead represent a barrier to those ideals (Ridgway, 1998). Thus, rather than curriculum defining assessment, assessment as defined by other influences ends up defining "the actual curriculum" (Ramsden, quoted in Pegg, 2003, p. 228).

2.3. Major Influence: The Didactical Situation

Much like what can happen when an educator realizes that true measurement of knowledge is impossible, sometimes assessment tensions also pull (or repel) a teacher away from the influence of curriculum. When the real focus shifts to the actual didactical

situation in the classroom, the intended or explicit curriculum becomes less important than the received curriculum. I can speak from personal experience of the influence of what I will call a pedagogy influence on my own assessment practice, which is the force that causes changes in specific teaching practices as a result of assessment information. It is a widely stated belief in the assessment literature that, as Strickland and Strickland (1998) put it, "assessment drives instruction" (p. 27). This is one of the fundamental facets of the formative assessment movement: that assessment information should lead to adjustments in teaching (Black & Wiliam, 1998a, 1998b, 2009; Frey & Schmitt, 2007; Taras, 2008; Wiliam, 2007). While this facet may be fundamental and widely held, it contributes to the tension felt by classroom teachers because they may not always possess the ability to appropriately adapt teaching practices to meet the assessed needs (Heritage, Kim, Vendlinski, & Herman, 2009). Moreover, the relationship goes both ways, because choosing certain teaching practices also forces one to adopt certain assessment strategies (Black & Wiliam, 2009; Strickland & Strickland, 1998). Adopting problem-based learning, for example, or more technology-rich activities, forces a teacher to consider how better to assess the kinds of competencies required of and created by such learning tasks (as opposed to traditional testing which typically measures recall of demonstrated procedures). Obviously there exists some interplay, then, between the pedagogy influence and theories of knowledge and mathematics.

2.3.1. The Individual Learner

Whereas the pedagogy influence represents the teacher side of the didactical situation, the final influence I will describe represents the other half: the learner. The learner's influence on assessment is the pull towards the consideration of the actual, individual learner and the unique accommodations he or she makes in light of the learning activities. I use accommodation here in the Piagetian (1970) sense of adjustments to a changing view of reality. Note that, while obviously the individual learner and theories of knowledge and mathematics are related and sometimes exert aligned forces, the learner influence here represents something different from the teacher's increasing awareness of general developmental theories. Rather, this force is the one that influences assessment decisions because of what Kyle or Kim are doing or

needing in class *today*. Hewitt (2004, 2010), in his research on forms of immediate feedback that teachers employ during teaching, points out that teachers need to choose a form of feedback that is effective for the particular pedagogical situation, and that one form may be nearly useless in one situation but most effective in another. According to Hewitt (2010), experienced and effective teachers employ a student-centric and context-specific "kitbag" (p. 263) of feedback strategies that allows them to respond to student actions in productive ways, and which is informed by a rich personal pedagogy. A nice example of what this might look like is given by Harkness (2008), who describes the learner-centered focus as a "believing game" (p. 243) played by a teacher who is looking to make sense of the individual learner's ideas in a positive manner. In general, of course, several individual students' needs sometimes coincide or overlap in significant ways, which allows teachers to do one thing for the benefit of many (which reflects again the pull of efficiency). But the main effect of the learner influence is to focus a teacher's attention on the individual, idiosyncratic learner.

The learner influence shows up in the literature as the other key aspect of formative assessment, which is to say the ownership of learning by the students (Black & Wiliam, 1998a, 2009; Wiliam, 2007). The students themselves need to become users of assessment information to adjust their learning strategies, set goals, and monitor progress (Frey & Schmitt, 2007). This is elsewhere in the literature described as assessment *as* learning, which centers on the concept of metacognition, students' awareness of their own understanding (WNCP, 2006). Assessment as learning is sometimes taken to be a poor cousin to the other two modes of assessment in the theoretical framework of assessment of, for, and as learning, but I contend that it may well be the most important. Certainly, for those teachers who come to realize the impossibility of exactly measuring some entity called "knowledge," and who decide that intended curricula are also ultimately not the main point, while nevertheless acknowledging that traditional approaches leave something to be desired, assessment practices that center on the learner as agent become paramount. Such teachers put aside the need to assess all students in the same ways, and instead allow for idiosyncratic demonstrations of understanding:

> The illusion of justice that consists of treating in the same way all the year-long students where knowledge is not the same, is an important obstacle to a real formative assessment (Bodin, 1993, p. 139)

This is not to say that the teacher throws away all other influences and applies assessment haphazardly and unthinkingly to different students in different ways. Rather, the teacher, recognizing the difficulties inherent in the act of assessing, nevertheless chooses to employ a variety of what Ginsburg, Jacobs, and Lopez (1993) have called "thinking-oriented approaches to assessment [which] allow the teacher, at all grade levels, to create a rich and practical *theory of the individual student*" (p. 158, emphasis added). This ideal, of a rich and practical theory of each student, is at the center of the learner influence.

2.4. Magnetic Field Metaphor

All this talk of influences and forces and impulses, with references to overlapping or pulling in different directions, may be a confusing approach to describing the field of assessment. The reader may be finding it difficult to see this as much of a mapping at all. Certainly I myself have questioned whether I am in fact failing the principle of parsimony with this organizing approach, multiplying entities unnecessarily. But in fact I feel that the complexity of the description is part of its appeal, for it captures what I have come to realize is the complexity of the field itself. Occam's razor (Spade & Panaccio, 2011) and the principle of parsimony do not say that simple theories account for everything, but merely that we should tend to prefer simpler explanations if they account for all observable phenomena. Or, as Einstein reportedly has said, "Make everything as simple as possible, but no simpler." Perhaps a metaphor might help at this point to make sense of the tensions in the field of assessment. It strikes me that a useful analogy (though imperfect, as all analogies are) is to compare the field of assessment to a literal magnetic field, with multiple magnets. The reader or practitioner in this field is like a pendulum swinging above the magnets. The behaviour of an actual pendulum in this situation, as a simple example of a chaotic dynamical system, proves tremendously hard to predict, and is extremely sensitive to initial conditions. By analogy, any individual navigating the field of assessment in mathematics education is influenced uniquely by

the various ‘magnets’ that exert their pull on him or her. Of course, this analogy is oversimplified, for in reality, we must acknowledge that the magnets in assessment are moving around over time, and growing or shrinking in strength, and also acknowledge that in such a system, which includes people acting willfully, the pendulum itself would exert an influence. As I mentioned at the outset, accounting for the observed influences on teachers in assessment is a difficult job, and I like the magnetic field image for how it captures the sense of competing or aligning forces, the sense of motion and change over time, and the sense of chaotic unpredictability.

While I have tried to present each magnet somewhat objectively to account for the various forces that cause the tension in the field, the reader probably nevertheless has discerned some emphasis on a particular path through the magnetic field. The literature implies (or outright asserts) that some of the influences, especially those centered on tradition, are negative influences overall, and a preference is given to assessment techniques that are further from those magnets. There is indeed a trending away from the tradition magnet, with its emphasis on measurement, precision, and ranking, towards a focus on learning, and in the end on the unique classroom situation (Gipps, 1994; Strickland & Strickland, 1998). I am not sure that all researchers would agree the final ideal is some version of Ginsburg’s ‘rich and practical theory of the individual learner,’ (Ginsburg, et al., 1993, p. 158) especially those researchers who are heavily influenced by the standards movement in the United States, but I contend that the learner magnet does represent a logical endpoint of much of the theorizing around assessment.

2.4.1. Paradigm Shift

This sense of a trend in the literature, a shift of focus, suggests another organizing metaphor to explain the tensions in the field: a paradigm shift. The phrase ‘progressive assessment,’ with which I began this investigation, implies to me movement, or progress, from one particular set of ideas to another. Ideological and philosophical positions can be called the paradigm governing a particular view. Thus the purpose of this literature review has been in a way to map a paradigm shift.

Paradigm shift is a phrase that has entered the popular consciousness, and is somewhat taken for granted. It is useful to examine the central ideas of Thomas Kuhn's (1970) description of this phrase and to discuss my mapping of assessment theory in light of Kuhn's ideas. There are some compelling features that help to explain the development of ideas in assessment, but there may be some ways in which Kuhn's concept, about science and its rational progress, do not fully apply. In the humanities there are often coexisting central ideas from which people may choose, none of which supersedes the others, but in science, paradigm shifts represent a decided move towards a better vision. To what extent is assessment in education scientific in nature?

Paradigms are the central commitments held by members of a scientific community. They are the ideas that shape worldviews and structure the work of individuals in a discipline. The process of shifting paradigms involves human agency and social tensions and not merely a battle of pure logic. The process begins when dissatisfaction with the current ideas arises, and some members of the group locate intelligible and plausible alternatives that promise fruitful lines of inquiry. It is not that the new ideas arrive fully formed and push older, otherwise healthy ideas out, but rather that older, sickly ideas start to show their age and inability to keep up with the times, and thus give some young upstart ideas a chance to prove themselves. After a paradigm shift, 20/20 hindsight often makes the inevitability of the new ideas seem obvious, but during a paradigm shift, which can take many years, there are no foregone conclusions. Some of these features, particularly the way in which the process can take time, help to explain the slow pace of change in educational assessment. Kuhn's (1970) description of paradigm shift allows us to see that the ideas that, though old, are central to "progressive" assessment are part of the long, slow process of conceptual change.

Another complicating factor to consider is how ideas in education reflect political ideologies. It strikes me that the ideas in assessment and in education in general often divide along lines that reflect conservative and liberal divisions. This goes back to the question of how scientific is the field of assessment. Science as a discipline tries to be as apolitical as possible (though obviously politics impacts scientific progress: think cloning and genetically modified food). But education is not just touched by politics; it is central to politics. Thus, the tradition magnet in my metaphor retains strength due to

conservative political elements despite the general sense among researchers that it ought to be weakened or outright removed from the field.

2.5. The Teacher's Ideal of Assessment

This notion of changing the location or strength of various magnets is how the paradigm shift metaphor connects with the magnetic field metaphor. If we truly are in the process of shifting paradigms, then we should see the magnets that define the traditional influences either weakening over time, or being removed from the field altogether. Another way of looking at it is to imagine the pendula of individual practitioners and researchers entering the field with initial conditions that determine a path less influenced by the tradition magnet. If my sense is true, that there is a paradigm shift underway, then whether the magnet of tradition is weakening, or the pendulum of the practitioner enters the field under new conditions, the effect is the same: more pendula reach an end state in the neighbourhood of the learner magnet. This is my sense of the ideal of progressive assessment as revealed in the literature. Certainly my own path through the magnetic field of assessment has brought me to a position in which I idealize assessment as a subjective process focused on promoting and maximizing the learning of meaningful mathematics. I see assessment being ideally unique to each individual learner, such that each student should receive from me just the information needed at just the right time to help him or her understand better or more deeply. Of all the magnets I have personally reacted most strongly to, the ranking and sorting impulse is in my view most negative, as my ideal is that assessment of one student should have no bearing on assessment of others, and learning is cheapened and undermined when it is made to serve external, economic needs rather than be an end in itself. This is the ideal of assessment I take from my review of the literature, and is the context within which I seek to understand my students' reactions to my progressive assessment efforts.

2.6. Student Conceptions

But this is not enough. In order to better understand my students' reactions, we also need to have some sense of what the literature says about student mindsets in

general, and, as much as possible, about student conceptions of assessment in particular. The review of literature thus far has focused exclusively on how teachers and researchers conceive of assessment and its purposes, forms, benefits, and so on. My principal concern in this project is with the intersection of these idealistic teacher conceptions (namely, my own) and the reality of how students are responding to the actual assessment program in the classroom. What research says about student attitudes and beliefs, therefore, bears directly on my investigation. I will summarize some of the relevant research on student mindsets by grouping theories into four categories of student conceptions: conceptions of self, conceptions of mathematics, conceptions of school, and finally, conceptions of assessment itself. Obviously these categories overlap and interact with each other, but they nevertheless serve as a useful organizing structure for this review of the literature. Also, the fourth category, conceptions of assessment, clearly is a subset of conceptions of school, but I pull it out separately because it is my main concern in this investigation and I wish to situate my research within the context of this focused body of work.

2.6.1. Self

A student's attitudes and beliefs about herself colour her reactions to all circumstances in the classroom, but will particularly impact how assessment events are received. Given that assessment either feeds back information to allow students to improve or provides an evaluation of achievement, how students react to assessment will be affected by how they conceive of themselves as learners. While a number of perspectives on student psychology, including interrelated concepts such as confidence, motivation, self-efficacy, agency, and identity, could be applied here, the main theory that has influenced my thinking is Carole Dweck's (1999, 2008) work on fixed ability versus growth mindsets. Dweck's (1999) theory suggests students can have either an entity theory of intelligence or an incremental theory, meaning that they conceive of their intelligence or ability as innate and unchanging, or as capable of improvement, particularly with effort. According to this theory, students with a fixed ability mindset will tend not to put in extra effort on things for which they receive negative feedback, because they perceive the negative feedback as indication of low ability, which cannot be improved. The ideal of assessment as described above in the literature on teacher

conceptions quite clearly presupposes the students hold an improvement conception, seeing as how it emphasizes the use of assessment information to deepen levels of understanding.

2.6.2. Mathematics

Speaking of understanding, students will also hold conceptions of the content they are learning that impact how they receive assessment. What it means to know and do mathematics is not just a factor, as discussed above, in teachers' decisions regarding assessment. Students, too, will have particular "epistemological conceptions" of the subject, as Star and Hoffman (2005) measured using a survey called the Conceptions of Mathematics Inventory. This survey device indicates student conceptions on seven scales: Composition of Mathematical Knowledge, Structure of Mathematical Knowledge, Status of Mathematical Knowledge, Doing Mathematics, Validating Ideas in Mathematics, Learning Mathematics, and Usefulness of Mathematics. Each scale has a traditional end and a more reform-oriented end, and individual student conceptions vary along the spectrum, with a student who mostly agrees with the Likert-type survey items having conceptions of mathematics that are "consistent with the aims of recent reform documents" (Star & Hoffman, 2005, p. 28). In several longitudinal studies in the UK and the US, Boaler (1998, 2002b; Boaler & Staples, 2008) has shown that in markedly different school contexts, it is possible for students to develop markedly different conceptions of mathematics as a subject. Moreover, Boaler's (2002a) work reveals some interplay between conceptions of subject and conceptions of self, as students who learned in reform-oriented contexts developed a greater sense of agency in keeping with their view of mathematics knowledge as more flexible. Star's (2002; Smith et al., 2000) research on student conceptions similarly shows that the conceptions students hold can be different based on whether they receive a reform curriculum or a traditional curriculum, and shows also that what conceptions students hold will influence how they respond to a change from one environment to another. Given that assessment communicates to students information about what aspects of the subject are valued, these student conceptions of mathematics will influence how they receive progressive assessment as idealized above.

2.6.3. School

The student's conception of a particular subject, i.e., mathematics, fits within the student's broader conceptions of school itself. The key organizing theory in this category for my research is Brousseau's *didactiques* (1997), a theory that describes learning situations as interactions between three elements in a triangle: teacher, student, content. A central feature of this theory is that the agents in the triangle take on roles governed by expectations, or norms. Thus, students in general, but also each student in particular, will bring to the learning situation a set of attitudes and beliefs, often subconscious, about what is expected of them as students, of the teacher as teacher, and of the content as material to be learned. The category of conceptions of mathematics could be described within this theory as the student's expectations concerning content, or what they consider 'normal' when learning mathematics specifically. The entire set of implicit norms governing the didactic triangle is referred to as the didactic contract, to which individuals give tacit approval when they undertake their role. When the expectations are not met in the learning situation, it is called a "breach" of the didactic contract (Herbst & Chazan, 2012, p. 607). In the context of my investigation, I am most interested in how students respond when the teacher behaviour or the content breaks the norms held by the students' conceptions of how school, and particularly the mathematics classroom, is supposed to operate. Assessment events, as enacted by teachers, are governed by norms, and assessment information or feedback forms part of the nature of the content in the didactic triangle.

2.6.4. Assessment

Part of how students view school, then, and how they view mathematics in school, is how they conceive of assessment. As mentioned, it is relevant to review some studies that address student conceptions of assessment specifically, because assessment is the aspect of the learning situation I am investigating. There is not as much research in this specific domain, however, as there is in student conceptions more generally, but two recent studies about students' assessment beliefs and attitudes will serve as a useful starting point for my investigation. Bagley (2010) borrows from cultural anthropology a "Relational Models Theory" as an organizing structure to investigate how the nature of the relationship between teacher and student affects students' reception of

assessment. The four models in the theory, Authority Ranking, Market Pricing, Communal Sharing, and Equality Matching, represent different types of power distribution. Bagley uses the generic model and specifies a typology to describe assessment practices: Authority Ranking means teachers hold absolute authority and hand down assessment information, Market Pricing describes assessment as a negotiated interaction where students receive results in exchange for meeting clearly stated expectations, Communal Sharing is a model of balanced power in which students and teachers identify and ascribe to common goals for student success and assessment information is arrived at by consensus rather than delivered authoritatively or as an exchange, and Equality Matching refers to equivocal assessment of teachers by students, students by teachers, and students by students. Though this theory could be applied by an outside observer who categorizes observed relationships, Bagley applies the theory in her study to which of the four models *students* perceive in their reception of assessment. Bagley reports that, though Authority Ranking is dominant, students who perceive the Communal Sharing and Market Pricing models experience, understandably, a greater sense of empowerment in the assessment interaction, and hence in their learning situation.

The other work on student conceptions of assessment that I feel is worth mentioning is Brown's summary of various studies that employed a survey device called the Student Conceptions of Assessment inventory (2011). This paper is perhaps even more pertinent to my research than Bagley. Brown's summary of the survey results shows that certain student conceptions of assessment correlate to what he calls adaptive learning behaviours, while other conceptions correspond to maladaptive behaviours. The survey instrument "summarizes student conceptions of assessment as four inter-related constructs" which Brown labels Improvement, External, Affect, Irrelevance (2011, p.734). The Improvement construct refers to the degree to which the student conceives of assessment as central to improving teaching and learning. The External construct indicates whether students feel little or no personal control in or over assessment. The Affect construct refers to how much students respond emotionally to assessment events and assessment feedback. And, finally, the Irrelevance construct indicates the degree to which students ignore or disregard assessment. Perhaps unsurprisingly, the Improvement construct alone corresponds to adaptive learning

behaviours, while the other three all indicate maladaptive orientations. Brown describes the adaptive orientation as a growth orientation, echoing Dweck's (1999, 2008) use of that term re: mindsets, and calls the maladaptive orientation a well-being orientation, indicating that a person so disposed will tend to emphasize affective, emotional, low-effort well-being over the sometimes hard work required to grow and learn. Clearly, given my sense that not all students in my classroom are responding positively nor as I intend to my progressive assessment program, Brown's and Bagley's theories about student conceptions of assessment are germane to this investigation. The two theories together help to reveal that an important element of student reactions to assessment is the locus of control.

This research project, then, is situated against the background of the teacher ideal of assessment as described earlier and within the theories about student mindsets laid out above. The teacher ideal, which emphasizes individualized assessment for the purposes of deepening and enriching a conceptual understanding of content, with little or no comparative ranking and sorting at play, tacitly assumes a certain frame of mind in students. The research on student mindsets shows, though, that student conceptions vary, which may help to explain the variety in responses to assessment.

2.7. Research Question

Broadly speaking, my research examines the intersection of *ideal* with *real* in mathematics assessment. As is always the case when ideals intersect with reality, there is potential for disconnect and discord. The personal tension I have felt in my own practice stems primarily from the fact that not all students react positively to my efforts to improve assessment. Specifically, then, my question is **how do students receive assessment in my classroom**?

3. Methodology

To explore the research question from the previous chapter, the researcher collected qualitative data via a survey and interviews within the context of a senior mathematics class. In this chapter, I will describe the salient features of the classroom context and the participants, as well as the data collection and analysis methods.

3.1. Setting: Problem-based Learning to Promote Understanding in a Principles of Mathematics 12 Classroom

The research setting was a Principles of Mathematics 12 class taught by the researcher in the first semester of the academic year. The Principles of Mathematics 12 course is an elective academic mathematics course intended to prepare university-bound students for post-secondary programs involving the study of calculus. As such, the type of student who takes the course is an academically-motivated student who has experienced at least moderate success in school in general and who has aspirations to enter university, usually immediately after graduating high school. As a teacher, I have taught this particular course in all but one of the past eight years. Given the status of the course as a challenging academic capstone which needs to prepare students for success in further schooling, I have put significant energy over the years into refining my teaching of Principles of Mathematics 12.

In my first few years of delivering the course, the bulk of my energy went into trying to create a course that anticipated what I believed students were likely to encounter in a university-level mathematics course. More recently, however, my focus has shifted towards improving more generally my teaching of conceptual understanding, because I believe that whether what I do looks like a university course or not is less relevant than whether my students are learning deeply and well. If they know their stuff, they should be ready for university level courses, as I see it, whether or not the style of

delivery is familiar. It is worth noting, then, that in addition to progressive assessment strategies, which I will detail more fully below, my research participants were also encountering a mathematics classroom in which the focus is deliberately placed on conceptual understanding and problem-solving.

I begin my course each semester with a focused week of open-ended problem-solving, not necessarily curricular in nature. Students work in randomized groups on several problems each day and I try to highlight at this time the kinds of traits that make for successful problem-solving. The intent is to establish a classroom culture of collaboration and mathematical sense-making. My teaching style throughout the course now includes problem-based learning strategies in which students work in random groups, sometimes standing at whiteboards or writing on windows. I do not usually deliver traditional notes, which students copy verbatim from the board, but rather provide time in class after we work on problems and discuss solutions for students to summarize their understandings in their own words. I have not completely eschewed direct instruction, but my students did encounter more alternative instruction techniques in academic mathematics than they were used to.

3.2. Setting: Assessment Strategies

As part of my general efforts to focus teaching and learning on robust conceptual understanding, the Principles of Mathematics 12 course is one of the courses in which I have attempted the most thorough implementation of progressive assessment. It was in Principles of Mathematics 12 that I first developed the dissatisfactions with traditional assessment I noted in the introduction, and it was in Principles of Mathematics 12 that I first began experimenting with writing assignments and other alternative assessments. I felt, therefore, that it would be students in that course whose reactions to progressive assessment would be most relevant to measure.

One of the key features of my assessment approach in Principles of Mathematics 12 is unit writing assignments. I adopted these assessments some years ago in an effort to measure more directly how students understand concepts as opposed to procedures. Towards the end of each of six units of study, I ask students to explain in their own

words some aspect of the central concepts in that unit. The prompt may be quite simply stated, as in "Write to explain what is meant by a logarithm," or more complex, as in "With reference to transformations on the basic features of the graphs of $y = \sin x$ and $y = \cos x$, write to explain how the parameters a, b, c, and d in $y = a\sin b(x - c) + d$ reflect the features of a real-world periodic situation." When I give the assignment, I usually conduct a brief discussion to make sure students understand what the question is asking, and I give a broad brushstrokes sketch of what an adequate response will need to contain. Students usually have about a week to complete these assignments, mostly on their own time, but they are free to seek help from each other or get feedback from me on their writing before they hand it in. Most of the time students end up writing about a page or so to complete these assignments, although I never state a desired length. These assignments then form a part of the major summative evaluation of the unit, along with a traditional unit test and a unit problem to solve.

Each of the unit problem assignments is a multi-step problem that represents the main type of situation to which the mathematics of a particular unit can be applied. For example, in the exponential functions and logarithms unit, the unit problem involves modelling a population's growth, and answering predictive questions based on the model. The problems are not strictly open-ended, because specific answers are usually warranted, but they do allow for multiple solution strategies, so they are less procedural. I also require that students submit the unit problems with fully-worked solutions, and tell them I am looking for them to demonstrate the conceptual understanding that underlies the procedures they use to solve the problems. These unit problems are usually assigned at the same time as the writing assignments, with a similar timeline for completion, and again students can seek help if desired. The students must submit their own individual assignment, but they may choose, and often do, to work together.

Though I give due dates for these assignments, I accept late assignments for full credit (but note the tardiness for purposes of evaluating work habits). I always tell students that I give due dates to help them organize their busy student lives, but that there are no deadlines, per se. This policy is part of my general strategy to relax some of the typical demands of a classroom in order to allow students to demonstrate their best performance and to remove anxieties that might hinder that performance. Another way I

attempt to remove anxieties is to allow and encourage rewrites of tests. In order to rewrite a test, though, students must demonstrate improved understanding by correcting, with explanations, all their mistakes on the first test, and completing a test reflection sheet.

The test reflection sheet is one of the newer strategies I have employed. It includes a chart that matches up the test items to the learning outcomes for each unit. Next to each item, students are to record whether they got that item right or wrong, and if wrong, whether it was a simple error or some misunderstanding. After going through the test in this way, they reflect on what they seemed to know well and what areas they still did not understand fully. Then they self-evaluate the degree to which they understand each learning outcome. Students then re-write a different version of the test, or simply a section if the test reflection revealed a deficiency in only one area.

The test reflection sheet is an example of a strategy I am using to try to blur the lines between summative and formative assessment. My relaxed due dates and opportunities for rewrites make the summative evaluations at the end of each unit a chance to improve learning, which makes them also formative assessments. I also, though, use other formative assessment strategies within the progression of learning. I give regular (almost daily) short quizzes that do not strictly count for marks. They are intended to show students what they should be capable of based on the previous day's learning, and for me to measure whether on the whole the class grasped yesterday's concept(s). If the daily quiz does not go well for a majority, I spend more time addressing the concept, while if some few struggle with it, I try to help them in a small group at some convenient point in the subsequent class period. I say the quizzes do not strictly count for marks because I do not record and count them towards an average grade like I did in the past, but they can contribute to summative evaluation because I ask students to keep in their portfolios quizzes on which they score over eighty percent. These collected quizzes can then help influence the unit or term letter grade we decide on during interviews later. (More on that process below).

I also regularly use individual whiteboards in my class as a response system for me to gauge the level of understanding as a lesson progresses. I can pose various types of questions as we go and ask students to hold up their responses on the whiteboards,

which allows me to very quickly check whether the whole class is getting the idea and which students are not there yet. The individual whiteboards work fairly well for simple conceptual questions, where I might ask students to make a prediction or choose from among some options (e.g., in which quadrant, I through IV, will the terminal arm of $\frac{7\pi}{6}$ lie?). Students use them for longer procedural questions, though, too, because I generally give them time to try to solve a problem first before I work through it as an example. The whiteboards seem to help more students attempt the problem in an engaged manner than would happen if they were just copying down an example during verbatim notes.

In addition to these formative strategies and the alternative summative assessments described above, there are some progressive elements to the ways in which I evaluate performance and generate grades and percentages to report. On both the writing assignments and unit problems, I do not assign a score per se, but rather put written comments and an evaluative code using what I call my check-plus system. This system of codes I devised some years ago in an effort to remove traditional scores and letter grades from my evaluations so that students could focus more closely on the quality of their work and whether it meets the requirements of assignments. I felt at the time that scores and letter grades quickly switch to percentages in students' minds, and percentages become the goal in and of themselves, rather than learning the concepts that the percentages are supposed to be a measure of. The check-plus system is really just a five-point scale with a plus (+) at the top, a minus (—) at the bottom, and a check (√) in the middle. Check-plus (√+) indicates an evaluation in between check and plus, while check-minus (√—) indicates an evaluation in between check and minus. I try to explain this scale to students without referencing traditional letter grades and percentages, but I do use language like that used to describe the letter grades in the Ministry of Education (2009) guidelines on reporting. So I refer to a plus, for example, as reflecting outstanding or excellent work, something that exceeds what I would expect in general from students in the course. I say that a check-plus suggests very good performance, and even a check indicates that the assignment is good (that is partly why I chose the symbols I did, because a check does convey "good"). Thus there is, ironically perhaps, a correlation in my mind between the symbols and traditional letter grades,

which may well reflect some of my own tensions regarding assessment and evaluation that I referenced in the literature review above. But again, the use of the symbols with written, descriptive feedback is intended to help students focus on how well they have hit the targets intended for the assignment. As such the feedback is summative, but, given that rewrites are possible, the feedback is also meant to help students see where things could be improved, and is hence formative assessment.

When I get to the end of a term, therefore, I do not have exclusively numerical data to contend with as I determine a letter grade and percentage to put on the report cards. I should note, too, that in general I do not report letter grades and percentages at more frequent intervals than the required two report cards per term. I do not keep students' performance a secret from them, however, but rather seek to be as clear as possible about how they are doing, just not in an aggregated score kind of way. If students want or need to know "where they are at," during a term, I will discuss with them what evidence we have collected so far, ask what they feel that demonstrates, and give them a ball-park suggestion of what level (i.e., letter-grade) they are performing at.

In fact, such a discussion occurs more formally between each student and myself at the end of each term, when I am obligated to pin down a particular letter grade and accompanying percentage. The students bring their portfolio to an interview with me, and we look at their unit test scores, symbols given on unit problems and writing assignments, and any other evidence (e.g., the quizzes I mentioned above) that they want to present or which I may have noted. From this evidence, we decide first upon an appropriate letter grade, and then on an appropriate percentage within that letter grade's corresponding range. It is usually quite straightforward to choose the letter grade, as the set of scores and codes usually has a "mode" value. The percentage is sometimes a bit trickier, but the student and I always reach a consensus on a score that we both feel reflects his or her performance.

In terms of Bagley's (2010) relational models, these assessment and grading practices, at least as intended, reflect my move from an Authority Ranking structure towards a Communal Sharing structure, with some Market Pricing included as well. Bagley describes a student who perceived the Communal Sharing model as "able to interpret the feedback he received from his teachers … as relevant and meaningful to

his own success, not just as an arbitrary and random whim of their authority" (p.95). It is similarly my goal for students to develop this awareness of assessment results as something they own and can use to adjust learning behaviours. My attempts to clarify expectations with test reflection sheets and rubrics would be examples of Market Pricing relational models, because I am trying to make it possible for students to earn certain results by meeting clearly defined criteria. While these are my intended models, though, my students may perceive something else, which again raises the question of how in reality my students are receiving assessment.

3.3. Participants

The participants in this research project, then, were a class of Principles of Mathematics 12 students experiencing the teaching and assessing strategies as described above. The survey portion of the data collection was conducted about three quarters of the way through the course, and the interviews were done after the course was over. The students had plenty of time to react to the teaching and assessing strategies by the time I collected data. It is the students' reactions to the assessment strategies in particular that I am seeking to understand more fully. The class of twenty-one students and their parents were informed of the nature of my research project and given an opportunity to decline to participate (with of course no adverse effect on any evaluation or status in the course). All but three students consented to participate. These eighteen students completed a survey as described below. From this pool of eighteen students, based largely on the nature of their responses to the survey, six were invited to participate in the second part of the investigation, the follow-up interviews. All six of these students consented to participate in the interviews as described below.

3.4. Survey

As mentioned, I surveyed all eighteen students who consented to participate at about the three-quarters mark of the course. In order to generate the actual survey items used for data collection, I undertook some informal discussions with students in several of my classes early in the term, in an attempt to identify productive directions for inquiry.

I tried an open response survey and some small group questioning in which I asked students to comment on how they respond to assessments in mathematics classes. Based on these discussions, I developed the survey items below. The survey took place during the latter portion of an eighty-minute class period, so the students had about forty-five minutes in which to complete it. (The non-participating students worked on their regular assignment at the time.) The survey consisted of five open response questions related to preferred types of feedback, to students' perceptions of the usefulness of different types of feedback, and to how they respond to results of assessments. Students had space on the survey to write several sentences in response to each open-ended question. I focused on these responses because I assumed that students are most aware of assessment in the episodes where information is fed back to them in the form of results, and therefore those elements of the assessment program would be good places to look for how they received assessment. These items also focused on students' perceptions of *usefulness* because I felt that inquiring about usefulness was a relatively value-neutral way to measure the nature of student reactions. Use implies purpose, so determining what uses students have for assessment information helps to identify the purposes for their learning behaviours in the classroom. Given that the idealization of assessment suggests assessment results ought to be useful for promoting learning, determining whether students do or do not use such information, and their reasons for doing so, should help to identify the factors influencing student perceptions of progressive assessment.

It should be noted here that my use of the term feedback in the survey items is not the same as the way Hewitt (2004, 2010) uses it to describe the kind of responses to student behaviours that teachers make during the learning situation. The term feedback in the following survey items refers more to the kind of information provided to students to indicate the quality of their work, whether it be traditionally evaluative like letter grades or scores, or more descriptive in nature, such as rubric information or comments.

The following are the survey items used:

1. What is your preferred feedback for assignments submitted to your teacher (this can be in-class assignments or homework and can have been completed individually or in groups)? Why? What do you do with this feedback?

2. What is your preferred feedback for quizzes? Why? What do you do with this feedback?
3. What is your preferred feedback for end of unit tests? Why? What do you do with this feedback?
4. Can you name an instance where feedback helped you to perform better in the future? How? Please be specific.
5. Can you name an instance where you didn't pay attention to the feedback (perhaps you would prefer something else or it isn't helpful)? Why? Be specific.

3.5. Analysis: Survey

After collecting the completed surveys, I read them over with a view to sorting students roughly by degree of acceptance of progressive assessment. I focused on the surveys of particular students whom I knew had already conveyed some sort of strong reaction to my assessment strategies, as I believed their survey responses would be useful for trying to explain why they reacted the way they did. In addition to a rough sort by type of reaction, I also read over the surveys and highlighted comments that struck me as particularly interesting or relevant, as perhaps relating to student conceptions of self (Dweck, 1999, 2008), mathematics (Boaler, 1998, 2002a, 2002b; Star & Hoffman, 2005), school (Brousseau, 1997; Herbst & Chazan, 2012), or assessment (Bagley, 2010; Brown, 2011) as described above in chapter 2. In many cases I noted comments that seemed to need more clarification. The results of the survey constitued a data set that I ended up not mining in great detail, as I had already decided by that time to generate my main body of qualitative data via interviews. My goal at this point was really to identify students as potential interview candidates. I selected the six interview candidates based on these two analyses, sorting by reaction type and highlighting interesting comments.

3.6. Interview

Of the six interview participants, three were female and three were male. Given their survey responses and my informal observations of their behaviour throughout the course, I chose these six students because they represented a range of responses to my assessment practices. Roughly speaking, they sort into three groups, with two

students displaying relatively negative reactions, two displaying relatively positive reactions, and two displaying relatively neutral reactions. The population of our school, in the interior of British Columbia, is relatively homogenous, with a vast majority of students coming from white, middle-class backgrounds. Each of the six interviewees is from a middle-class family with parents in professional occupations. One of the boys interviewed is labelled gifted, and one is labelled learning disabled as a result of testing done early in his school career, but neither of these identifications impacted greatly my treatment of them as students in the course, nor my analysis of their interview data. Typically, students who take Principles of Mathematics 12 have done reasonably well in previous mathematics courses, and these interview participants are no exception. Three of them had earned A's in both Principles of Mathematics 10 and 11, and the other three had earned B's.

For the follow-up interviews, I opened each one by highlighting within each participant's survey responses some interesting or unclear comments that I thought would be worth elaborating on. I asked each participant if he or she could clarify or elaborate on these highlights. As such, each interview took its own shape based on the responses of the participant. For example, one student commented negatively several times about the check-plus scale so I asked her to explain in more detail why she felt the check-plus scale was not very useful. Conversationally, I added questions like "What information does the check-plus scale not provide you that you would like?" The first two-thirds or so of each thirty- to forty-minute interview proceeded informally in this manner.

After discussing in more detail each of the highlights from the survey responses, I did address some common questions to each participant. Defining briefly the terms formative and summative assessment, I asked each participant if he or she could identify elements of each assessment mode in the things I did. My intent with this question was two-fold: first, to see if students had a degree of assessment literacy with respect to some of the terms used by teachers, and second, to see if they would identify the same purposes for my assessment strategies as I do. I closed each interview by asking students to imagine they could create a perfect mathematics teacher, and if so, what would that individual do to assess their learning. Both of these common questions were intended to generate data that might tell me more about what assessment is to students.

The first might reveal students' actual knowledge of assessment theory, while the second might reveal how they idealize assessment. The interviews were recorded, and I took notes during the interview.

3.7. Analysis: Interviews

To analyze the interviews, I listened to each one many times and added explanatory notes to the short-hand notes I took during the interview itself. While doing so, I transcribed various excerpts which were relevant to my research question, and began to note the factors which appeared to be influencing the student's reactions. I used the principles of analytic induction to analyze these data (Patton, 2002). Analytic induction is a process that begins with theory-based hypotheses as a lens to examine qualitative data. The analyst recursively codes the data, like in grounded theory (Glaser & Strauss, 1967, cited in Patton, 2002), looking for themes to emerge, but unlike in grounded theory, has some a priori theory to either verify, refute, or otherwise build from. For me, the primary a priori theory is Brown's (2011) four-construct theory of student conceptions. The other theories (Bagley, 2010; Boaler, 1998, 2002a, 2002b; Brousseau, 1997; Dweck, 1999, 2008; Herbst & Chazan, 2012; Star & Hoffman, 2005) about student conceptions of self, school, mathematics, and assessment also informed this analysis. As the analysis progressed through several interviews, I was able to start cross-referencing my notes and to group various elements of the student comments into categories. After all the interviews were individually analyzed, I reviewed the whole set of notes I had generated to distil some common themes I had identified in the student interviews.

4. Analysis: The students

In the descriptions that follow I present the results of each interview, not as a chronological summary of what was discussed, but as a thematic review of each student's comments.

4.1. Eddie (all student names are pseudonyms)

I did not have a definite sense that Eddie had been reacting positively to my assessment strategies throughout the course, given that his quiet demeanour did not lend itself to displaying strong reactions either positively or negatively, but neither did I have a sense that he was on the negative end of the spectrum. His survey did not indicate strongly one way or the other, either. What it did indicate, somewhat surprisingly, was that he had a fair bit to say about what types of information he appreciated getting and how he got it. And what he wrote was articulate and specific, rather than vague and general like I had come to expect from many of my students. One of his concluding remarks was a specific critique of being marked on a rubric scale, because, he wrote, "it ranks you on useless categories like understanding, effort, etc." I found this comment from a student, that understanding is a useless category, quite arresting and calling into question what Eddie's epistemological conceptions of mathematics are (Star & Hoffman, 2005). Trusting that Eddie would have more to say about assessment given his detailed survey responses, and curious about his reaction to rubrics, I asked him to interview. I opened with a request for him to expand on his comments about rubrics.

His elaboration confirmed for me that Eddie takes an overall pragmatic approach to school. He assured me that he felt "the categories themselves are important, but you can't see what you actually did wrong on actual material." This phrase, "actual material," is characteristic of the way Eddie refers to things he has to learn in math: "actual

material" as opposed to "abstract concepts" is the general conception of learning math that Eddie conveys. Star and Hoffman (2005) might say Eddie emphasizes the more traditional side of the "composition of mathematics," "doing mathematics" and "learning mathematics" scales, where the composition of mathematics is more "facts, formulas, and algorithms," the doing of mathematics is more "a process of obtaining results" and learning mathematics is more "a process of memorizing intact knowledge" (p. 28). But while they may reveal a traditional conception of mathematics, Eddie's comments primarily speak to whether assessment information is useful in student terms to help them fix or correct or change their work. Sometimes, rubric information may be too general to help students adjust their concepts or their ways of working. As Eddie said, a simple score in a category of *understanding* "doesn't help you fix a problem because if that's all you get you can't go and see why you didn't understand." Later in the interview, Eddie revealed that he deems the most useful feedback mechanism to be when a student can ask the teacher direct questions. He feels that you "get exactly what you need to know because you can ask the question directly and have an answer." This description of direct verbal feedback, in contrast to Eddie's description of the vagueness of a general rubric, indicate a preference for a Market Pricing relational model (Bagley, 2010). Eddie wants clearly defined expectations that he can meet in order to earn his desired results. In other words he wants information that is in a form he can use, and the direct verbal feedback is easier for him than interpreting written or symbolic feedback. It is not only his conceptions of the subject that make him prefer the specific feedback but also the practical clarity of the information.

The timing of the assessment feedback is also an important factor for Eddie that determines its usefulness. The verbal feedback Eddie so values is information both in a form and at a time that allows him to make immediate changes in his conceptual understanding or his use of procedures. Other comments in Eddie's interview spoke to this factor of the timing of feedback. He offered a suggestion, for example, that I employ my test reflection sheets (a post-test reflection, you may recall) as a pre-test review tool, and that I assign my unit assignments (unit problem and written assignments) earlier because they are often, in his terms, "where it all comes together."

Eddie's ideal of useful information seems to indicate he has an Improvement conception (Brown, 2011) of assessment overall. He spoke about using information in a

way that indicates he feels himself capable of improvement, and feels in control of his performance as well.

It is important to note here that when Eddie talks about using the information he is not just speaking about getting better scores or higher grades. This is significant because it reveals an important factor in his mindset about learning and assessment: he owns his own results as a reflection of his knowledge and recognizes his opportunity and responsibility to adjust what he is doing based on the information he gets back. As we will see in the results of other interviews, this is not a universal or even a majority mindset. Not once in Eddie's interview did the notion of ranking in relation to other students come up, whereas it did come up in every other interview.

I feel there must be a connection between Eddie's sense of ownership over his results, his pragmatic nature, and his lack of concern about comparative ranking. Dweck's (1999, 2008) description of a growth mindset as opposed to a fixed view of ability accords with what I picked up from Eddie in class and during the interview. Eddie seems to want useful information precisely because he feels himself capable of adjusting and growing based on feedback. In Dweck's (1999, 2008) scheme, the growth mindset welcomes descriptive feedback while the fixed mindset seeks merely to gain praise or avoid critique. When Eddie talked about not using assessment information, it only related to whether the information was in a usable form or whether he had the time to make use of it. There was no mention of emotional matters related to feelings of self worth or value in comparison to others. In Brown's (2011) four-construct scheme, Eddie would likely score low on the Affect and External constructs, because it seems that he cares little about whether assessment involves a feel-good aspect, and he seems focused on the assessment information as a feature of his personal learning situation.

Apart from the timing and format of feedback as discussed above, one other factor that influenced Eddie's reaction to assessment information was scheduling. Eddie was quite articulate in discussing the impact of time management on whether assessment feedback can be used or not. He referred to falling behind in new work as an impediment to taking opportunities to correct or rewrite old work, and he repeatedly referenced a busy course load, with too many courses to respond in an ideal way to all of a teacher's feedback. Again, this is evidence of his pragmatic outlook. He might in

some cases respond differently in an ideal world, but time pressures may prevent that ideal. This pragmatic factor does not fit neatly into the theories of student mindsets I outlined in the literature above, but it is a strong feature of Eddie's mindset, nonetheless.

Given that Eddie's commentary included so many criticisms with respect to timing and format, one might conclude that he was a student reacting negatively to my assessment practices. On the contrary, in his interview he expressed a generally positive attitude about the things I did. When asked about his ideal mathematics teacher's assessment approach, he essentially described my techniques, at least with respect to the alternative ways to demonstrate understanding, and with respect to the negotiated grading practices I employ. Interestingly, though, when I defined formative and summative assessments for him and asked him to categorize my assessment strategies, Eddie did so in what I would consider exactly opposite ways. For example, he referred to my frequent, almost daily quizzes, which do not count for marks, as summative assessments, whereas he classified my unit problems and written assignments as formative. Now, it is possible that Eddie simply understood the terms backwards, but I did press him a little in that regard and he did not reverse his categorization. I look on this as evidence of two things: first, students do not in general have high levels of assessment literacy (in fact none of the students, including one whose parents are both public school educators, said he or she knew what the terms formative and summative assessment meant), and second, students do not always perceive teachers' purposes accurately. It is not surprising that students may not see teachers' purposes accurately, because their perspective is as the recipient of the teachers' actions, rather than the initiator. It is exactly these differences in perspective that I am attempting to identify with this research project.

For Eddie, his unique perspective on assessment seems to be as a system for identifying what is right and wrong and most importantly how to fix the wrong. He values assessment feedback that is timely, unique to his own situation, and in a form that he can use relatively easily to correct his understanding or procedures as necessary. Eddie does not take criticism personally, but sees it as information that is potentially useful for improvement. To this extent, Eddie's mindset accords rather well with my teacher's ideal of assessment. I might have preferred Eddie to focus more on conceptual development and a little less strictly on procedural competence, but nevertheless Eddie did, when he

had the time, make use of my feedback to improve his work. Though Star and Hoffman (2005) suggest the traditional conceptions of doing and learning mathematics are indicators that a student would react negatively to reform efforts, it seems that for Eddie the Improvement conception of assessment (Brown, 2011) and the growth mindset (Dweck, 1999, 2008) are more significant factors. I feel that his generally positive reaction to my progressive assessment was a direct result of his feeling that he owns his assessment results and can choose to grow based on feedback received.

4.2. Maia

Maia was the other student that I grouped with Eddie in the positive reaction category. Unlike with Eddie, however, I did have a strong sense of Maia's positive reaction to my assessment strategies throughout the course. In fact, if there was one student in the class that displayed the strongest buy-in to both my assessment strategies and my overall teaching approach, I would say it was Maia. She also, like Eddie, had some articulate comments to make on her survey, including some thoughtful critiques. Maia praised the check-plus scale in her survey, saying that "it tells me what understanding I have of the concept, not simply if I obtained one correct answer." A couple other times in the survey Maia drew this distinction between understanding something conceptually and getting answers correct. Her drawing this distinction indicated a more reform-oriented conception of mathematics (Star & Hoffman, 2005), so these were highlights I returned to in the interview. Maia's thoughtful critique concerned the format of feedback in one particular rubric I used, which indicated only a general area of weakness with no information on what was specifically incorrect. Based on these interesting comments and my sense of her generally positive reactions, it was an easy decision for me to ask Maia to interview.

Late in the interview, when her interview responses confirmed my sense that she bought in quite strongly to my assessment approach, I pointed out to her this perception and asked her directly if she felt I was right, and if so, why. She confirmed that my sense was true, and suggested that it had much to do with her desire to really understand things. The idea of understanding came up repeatedly in Maia's survey and interview. Several times she noted that she really likes to get more than just information about right

or wrong answers. As she acknowledged in her interview, sometimes students can "just plug it into a formula or something like that but you don't really understand what you're doing with the formula." She indicated that she prefers to get feedback that addresses the conceptual understanding beneath procedures rather than just whether procedures are correct. We began her interview with an extended discussion of the check-plus system, in fact, because she said in her survey that she prefers it because of what it conveys about understanding. In the interview she added that she likes how the check-plus system is "not just based on one number." Not only did she express her appreciation of the check-plus system, but she also liked many of the other strategies I used, and for the most part it was because of how they focused on understanding.

At one point, Maia expressed in particular how much she liked the written problems, because they "conveyed that understanding is important." Overall, Maia seemed specially aware of a focus on understanding in my strategies. She made a subtle distinction between the check-plus codes and letter grades, for although she acknowledged the parallels between the symbols, she stated that "sometimes if a teacher puts … an A, a B, or a C on something, you think that's your grade right away based on your answers to the questions, but it's not really saying whether you understand it or not." When I asked her at one point what she thought I was trying to accomplish overall, she said that I seemed to be "trying to see if we really understand it." She mentioned, for example, the daily quizzes, which to her showed I was trying to make sure they were "absorbing stuff from the day before – not just moving on." This emphasis is a good thing, in her view. Strikingly, when I asked Maia if she felt like understanding was in general a strong focus of mathematics courses, she said, "Not really, not so much." This discussion underscores for me that assessment practices send strong messages to students about what matters, and it strikes me that, at least for Maia, traditional practices may not be sending messages that suggest understanding is important. The fact that Maia is so aware of and so appreciative of this focus on understanding speaks volumes about her overall mindset. I feel that her highly positive reaction to my assessment approaches is most directly a result of her valuing understanding so highly (Star & Hoffman, 2005; Boaler, 2002a), a value that I share as a teacher in my idealization of assessment.

Here again I feel that, like Eddie, Maia is displaying a growth mindset, with a strong Improvement orientation (Brown, 2011), except that her mindset has an added dimension of more heavily favouring understanding over procedural competence. She sees assessment information as feeding back to her a measure of her understanding, and that understanding is something that can grow and get stronger. At several points in the course, Maia took the opportunity to revise her work and resubmit it based on feedback, and again, like Eddie, this was less about just squeezing more marks out of the system and more about a genuine desire to improve on her level of understanding. Not coincidentally, Maia was also by far the most dedicated student in the course with respect to taking thorough notes. She frequently asked me to review her notes to see if they made sense and were complete, something I gladly did to help solidify her already considerably strong level of understanding. It should be noted that these interchanges were not explicitly connected in any way with the "for marks" assignments, so Maia's motives were obviously not about chasing down marks. Ironically, though, because of the open nature of my interview process for assigning letter grades, Maia's thoughtful note-taking could influence my judgement of her performance.

Although Maia did often take opportunities to correct and resubmit work she did note, similarly to Eddie, that certain forms of feedback were more useful than others. Maia also was able to articulately identify a time when the format of the feedback I gave her was not useful and prevented her from making an adjustment. On one assignment I had indicated some graphs were incorrect, but did not show where in her work she had errors which resulted in the incorrect graphs. She said that "if you actually circle the part of the work that started to go off to the wrong track then I think I'd be more inclined to go back and see if it was just a fluke or if I actually don't understand something." Here again the notion of understanding comes up, but this excerpt also reveals that for Maia, like Eddie, certain forms of feedback information make it easier to adjust her concepts or fix her work. Maia also spoke about the timing of assessment strategies and gave a similar suggestion to place the end of unit assignments a little earlier in the sequence of learning.

While much of Maia's reaction to my assessment strategies was positive, and demonstrated her willingness and desire to make use of assessment information to improve, she did acknowledge in her interview that she does not always choose to

optimize her learning or understanding based on feedback. Basically, she admitted, there are times when she is happy enough with the score or letter grade on work so she chooses not to adjust anything despite the fact that she perceives some deficiencies. I found this admission, from a student who otherwise has a particularly admirable focus on understanding, to be very revealing of the strong influence of grades on student decision-making. Maia acknowledged that she does not necessarily want to fix things if she is getting a grade she is happy with.

> *Maia:* Sometimes even if I don't understand something I can still get a good mark by using formulas and so on, and probably just because I got the good mark I wouldn't go back and re-do it.

With this admission she noted that sometimes grading, by satisfying a student with a good-enough result, short-circuits learning. What these comments reinforce for me is the notion that there is sometimes an economy in the classroom based not on the true goal of conceptual development, but rather on grades for their own sake, and this economy touches even the most dedicated of students. Even a strong Improvement orientation (Brown, 2011) and growth mindset (Dweck, 1999, 2008) is not always enough to ensure a student responds to assessment information in an ideal way. Instead, there are also norms (Herbst & Chazan, 2012) with respect to achievement that each student has come to expect, including a norm of "good enough."

Maia, in my view, is certainly one of the most dedicated of students, perhaps one of the most dedicated I have ever taught. Her overall mindset regarding assessment is influenced heavily by her desire to understand, and to know that she understands. Given that this emphasis aligns so closely with my own ideal for assessment as a teacher, it did not surprise me that Maia perceived my efforts in a positive light, nor that she was so aware of my underlying purposes. When I asked her to categorize my assessment strategies as formative or summative, she sorted things out exactly as I myself would, despite the fact that she said she had never heard those terms specifically before. She was hesitant and unsure of her sorting, which reveals that she was not confident that she perceived my efforts accurately, but she nevertheless identified my purposes. Maia even astutely pinpointed my underlying purpose for not using letter grades and regular scores as frequently as students are used to.

Maia: But I think it's a good thing that you're doing that because often with a percentage, if someone's like a B-level student or something, they see that they're getting a B and they don't want to fix anything because they know they're doing reasonably well, but with the system that you use, then they'll actually look at the feedback that you gave them.

Interestingly, as we will see in later interviews, while this is my purpose as Maia points out, it does not always play out in practice; some students do not in fact look at the feedback I give them, but rather just complain that they do not know whether it is an A or B. But for Maia, the system I use on the whole makes sense and is good, because, again, it aligns nicely with her ideals as a student. She displays a growth mindset (Brown, 2011; Dweck, 1999, 2008), with an emphasis on understanding (Star & Hoffman, 2005), so she is looking for and values assessment information that provides her with a measure of how well she understands.

4.3. Jim

Jim was a student whose attitudes about my assessment practices throughout the course were hard to pin down. His past performance had included A's in mathematics, but I was assigning him B grades, so I felt he was disappointed with that to some extent. But he did not take advantage of opportunities to re-write and resubmit tests or assignments. I was not sure if Jim was indifferent to his standing in the course, or satisfied with his achievement, or whether some other attitude was at work in his mind. In his survey, Jim revealed a strong preference for more traditional forms of feedback, including regularly posted marks. This preference would indicate dissatisfaction with my approach, as I generally eschewed such practices. But he also stated that "verbal and written feedback are more helpful." Thus, Jim's survey suggested a conflicting set of ideas at work in his mind. I felt it worth exploring this tension. Moreover, Jim also made a very interesting comment in his survey about his ignoring feedback in English class because he can "write well enough and will never improve." I wanted to discuss in more depth this notion of improvement and ability.

In his interview, Jim's responses contributed further to my sense of a set of conflicting ideas in his mind. In part, the conflict was a result of his ideas coming out only

in short, disconnected bits because it was painful trying to get him to say more than a few words in response to questions. It was an opposite situation to Eddie. Whereas Eddie was usually quiet in class but verbose in the interview, Jim was usually verbose in class, but getting him to talk in the interview was like pulling teeth. Nevertheless, he said enough for me to get a sense of his unique mindset with respect to assessment, and to identify some of the factors that influenced his reactions to my assessment strategies. His mindset consisted of a preference for familiar forms of feedback, but nonetheless he was capable of acknowledging benefits of non-traditional feedback. In fact, what I found most striking overall about Jim's interview was how perceptive he was about the benefits of changes to traditional assessment systems, but how he still remained staunchly attached to traditional systems. He seemed to acknowledge grudgingly that he could get something out of a more progressive assessment program, but he did not engage in order to receive those benefits.

Jim's preference for traditional feedback (i.e., letter grades and percentages) came out in his survey and right at the beginning of his interview. While, as I said, he was not verbose in his justification, he did claim that traditional, posted marks, for example, help "you know where you're at." He referred to another teacher's practice of having students keep a record of their own marks, where every point on every assignment is weighted the same so that "you just write the total and divide it by what you have so every day you know what mark you're at; so that's just nice." This notion that it is "just nice" suggests to me an emotional comfort level with knowing a score, and indicates a certain degree of Affect orientation in Jim's conceptions of assessment (Brown, 2011). I see this comfort level as quite disconnected from Maia's desire to know if she really understands. In fact, Jim acknowledged that knowing "where you're at" in this context is not about knowing your degree of understanding. When I questioned him as to whether knowing where he is at means knowing what degree of understanding he has obtained, he stated that "your mark doesn't necessarily reflect what you actually know." In this discussion, Jim noted a difference between what a mark conveys and what verbal or written feedback conveys. The verbal or written feedback helps you improve more, by identifying areas that need work, while the score simply conveys a snapshot of right and wrong. When asked about his preference directly, Jim suggested "a combination is the best way to do it," which seems entirely reasonably, but he went on

to say “maybe not for learning necessarily, but it’s good for students because then you know what you’re at.” He did not elaborate much on why the combination may not be best for learning, other than to reiterate that verbal and written feedback is best for learning. Overall, this interchange reveals, again, Jim’s preference for traditional feedback, a preference he cannot quite explain, other than that it is “good” and “just nice.”

While Jim may find it hard to put a finger on why he feels the way he does, I nevertheless get a strong sense that there are two purposes at work in his mind with respect to schooling. There is a connection here to the issue that showed up with Maia about the economy of the classroom. Jim’s comments suggest that at least subconsciously he sees there being a divide of sorts between the purpose of learning and the purpose of getting the highest score possible. The way he described the running total of marks in the other teacher’s class sounds not unlike the score you earn while playing a video game, updated in real time after every event that takes place. Though it does update in response to things you do, it is nevertheless an external thing, rather than an intrinsic feature of one’s understanding (Brown, 2011).

It would be one thing if Jim’s mindset on scores was just a small part of an overall balanced view of assessment as measuring real learning, but it strikes me as much more than that in his case. At another point in the interview, Jim discussed how learning things in school is different than learning things elsewhere, “’cause in school you have to give a mark – that’s just how our whole school system’s based.” So Jim sees the grade or score as a distinctive feature of learning in school that is absent everywhere else: “anything that you learn in your childhood – like learning to ride a bike – there’s no grade or mark involved.” And when I asked what the point is of basing the whole system on the mark, Jim astutely observed that “it puts people against other people because they have to have a way to see how you’re doing in comparison to perfect or other students.” Note how closely this comment aligns with the founding purpose of the tradition within mathematics assessment that I described in the literature review: the point in Jim’s eyes of marks and grades is to rank and sort. It is significant that he distinguishes between “comparison to perfect” and comparison to other students, because there seems to be tacit awareness of some ideal level of performance that a student is being judged against. But recall how Angoff (quoted in Wiliam, 2010, p. 254)

highlighted the norm-referenced evaluation lurking behind criterion-referenced evaluation, and how Gipps (1994) suggested any evaluation can become high-stakes when ranking and sorting plays a part in it. The ideal "perfect" that Jim refers to is conflated in his mind with comparison to other students, so it is just another element of a competitive system.

Jim candidly pointed out this competitive purpose in school, but I sensed throughout the interview that he was not always appreciative of how it played out in practice. When I asked him how he felt about it, Jim said "some days good, some days not so good." It is good for him when he does well, obviously, but not so good when he does not do as well. This comment reveals the emotional impact grades have on Jim (Brown, 2011), an impact I witnessed directly when I assigned him grades that he was not happy with. But Jim, despite being unhappy at times with his standing in the course, almost never took opportunities to rewrite any tests or assignments. These choices are indicative of the maladaptive behaviours associated with the Affect and External constructs in Jim's conception of assessment (Brown, 2011). But his choices surprised me somewhat, given Jim's desire to do well in the marks-based system, so I probed for some explanation in his interview.

Initially, Jim indicated that he was "just lazy." In fact, he described himself as lazy several times in the interview. While I do not want to dismiss the notion of laziness outright, I feel it is too vague a term to be helpful in describing student mindsets, because I feel that my students' behaviours are the results of their choices, whether they consciously perceive their decision-making process or not. For example, recall Maia's choice not to revise work when she is happy with the first mark obtained. With Jim, he was acknowledging he was not happy with a result, but still did not choose to rewrite, and I sensed laziness did not truly account for this choice. Probing a little more, I discovered that for Jim a rewrite represents an unfair advantage in the competitive scheme. What you achieve on the first attempt is "not the same as when you figure it out after," Jim said, so he felt it was somehow unfair for him to raise his mark that way. Here, again, there is a disconnect in Jim's mind between a mark or grade and the understanding it is supposed to be a measure of. He admitted that one's understanding can improve over time, and that a rewrite can demonstrate that increased understanding, but he still maintained that the mark that reflects that improvement is not

the same as what the mark should be if that level of understanding is demonstrated on the first try.

It is not bizarre for Jim to feel this way, as it seems to be a part of educational culture, wherein grades are part of a reward and punishment system as opposed to a report of degree of understanding. These elements of the system would be well-established norms (Herbst & Chazan, 2012) for many people regarding assessment. The reward and punishment approach to grades often coincides with the ranking and sorting tradition that sometimes influences teachers' assessment decisions. Certainly in conversation with many of my colleagues I have heard comments similar to Jim's, and those colleagues often have rather interesting systems of adjusting grades based on whether the performance is a first try or not. Quite common, for example, is allowing only students who failed a test to re-write, and then only recording a score of fifty percent, regardless of the actual demonstration of understanding.

But there is more to Jim's dissatisfaction with rewrites than simply the breach of a didactic contract (Herbst & Chazan, 2012). Despite the fact that Jim admitted one's understanding can improve and be demonstrably better on a second try, I sensed that at least at some level he felt this was not one's real level of understanding anymore. At another point in the interview, we discussed the notion of improvement, because Jim had written the comment in his survey that he "will never improve" in writing English papers. In the interview, Jim elaborated on this comment by saying that his ability to write is something that comes with growing up, and "one class won't change it." This comment suggests strongly a fixed-ability mindset, as described by Dweck (1999, 2008). My feeling is that Jim's reluctance to rewrite is because he feels the first mark is a true picture of some fixed ability, while a revised mark is a distorted picture in some way. Jim did note that "effort alone could make [his writing] better," but this admission only reinforces my view that Jim has a fixed-ability mindset. As mentioned, Dweck describes students with this mindset as believing that you either can do it with little effort because of innate ability, or it is probably not worth struggling over.

With a fixed-ability mindset, getting anything but high marks is demoralizing, because it suggests a deficiency in ability (Dweck, 1999, 2008). My sense was that, for Jim, the B grades he was earning in my class made him feel like he did not have high

ability in math, but he chose to blame his performance on mere laziness. I addressed exactly this notion in the interview with Jim, asking him why, if he wanted an A but was not earning one, he did not take opportunities to rewrite assignments and tests. He mentioned here the fairness issue I described above, but he also spoke dismissively about his "wanting an A," saying that "the only reason I cared about my mark in this course is I needed an 89 average over four courses to get into engineering at UBC." This comment gives some justification for Jim's emotional reaction to his level of performance, but does not explain why he did not take opportunities to improve his grade. In fact, I would have thought that with a specific university admissions goal in mind, he would be particularly motivated to do all he could to achieve a higher mark. This again suggests a conflicting set of ideas at work in Jim's mind.

When I got to the point in the interview where I asked Jim about his ideal mathematics teacher, he responded with a startling question of his own. The first thing he asked was, "Would the teacher have to give marks?" I suggested it was up to him to define the ideal, and asked if it would be more ideal if he or she did not have to give marks. He stated emphatically that it would be better if the teacher did not have to give marks. Again, despite his earlier assertion that he prefers regularly posted marks, here Jim is acknowledging the benefit of a really progressive ideal, doing away with grades and ranking altogether. However, when Jim described this ideal a little further, he stated that there would have to be some sort of "scale in place" for the students' benefit. I find it very interesting that Jim idealizes an assessment scheme without grades, but that he cannot actually visualize a system without some kind of grading. This suggests to me that he at least tacitly acknowledges a negative influence of grades on learning overall, but that the idea of grades is nonetheless so ingrained that he cannot imagine schooling without them.

Jim's conflicting ideas about assessment seem to stem from his apparent fixed-ability mindset (Dweck, 1999, 2008) and his accompanying sense of the necessity of a grading regime. He feels that grades are a necessary part of the competitive system to sort people out based on their innate abilities, and so he prefers clarity within a grading system, despite the fact that such information is not always a positive reflection of his own ranking. The conflict arises from Jim's subtle, almost unacknowledged, sense of dissatisfaction with this system, and his ability to see that this system is not always

connected to true learning. These ideas help to explain why I could not fully pin down Jim's attitude towards my assessment approach. He acknowledged that my system seems to be based "more on understanding," which would accord with his dissatisfaction with competitive grading, but his overall preference is for the familiar competitive scheme that he has, perhaps reluctantly, bought into, which is a scheme I am deliberately trying to break down. It is not exactly clear why, but Jim seems to distrust the Communal Sharing (Bagley, 2012), or consensus building, element of my grading scheme, which suggests perhaps that he wants the concreteness of a familiar Authority Ranking structure. Jim's conceptions of assessment seem to include a complicated mix of External and Affect factors, and little Improvement orientation (Brown, 2011), which accounts for a generally maladaptive behaviour pattern. But the interesting thing is that Jim, when pressed, does acknowledge some benefits of progressive ideals.

4.4. Adam

Adam's survey stated a strong preference for traditional letter grades and scores, much like Jim also showed. In fact, he wrote that it was "the only method that makes logical sense." He also wrote that he dislikes any feedback that has "room for interpretation" and any level of "abstraction." In particular, he singled out my check-plus system of codes and called them "too vague." These survey comments suggested Adam had some fairly strongly-held opinions about the nature of assessment results, so I thought it would be worth exploring his perspective. Interestingly, though, I was a bit surprised by the negative tone of his survey comments with respect to my strategies specifically, because I had not earlier perceived a negative attitude in Adam. He also, like Jim, did not avail himself of opportunities to re-write or improve his results and did not seek clarification from me of his results, so I assumed he was at least moderately satisfied with his achievement and that he generally understood the nature of the feedback I had been giving. His survey results, though, suggested otherwise, which I felt was another good reason for me to interview him.

Early in Adam's interview, I asked him to clarify what he meant in his survey about preferring feedback where there is not much room for interpretation. He explained that he likes to know "exactly what I got, like such and such a score out of total possible."

When I asked him to elaborate on why he prefers such feedback, he initially stated that he liked "something definite" and that it was "just something that helps with my learning." But Adam could not elaborate on how exactly the definite score helped with his learning. Seeking further insight, I asked about an example of something that was not a definite score, and what it lacked compared with the definite score. Adam spoke about my non-numerical feedback codes, which he said are confusing and prevent him from putting "a definite tally" on how he did. A numerical score, he claimed, gives him a "more fulfilled understanding of – how you see it – my comprehension of the subject." The parenthetical reference to "how you see it" is illuminating here, suggesting that Adam values the External aspect of assessment highly (Brown, 2011) placing more weight on the teacher's evaluation and less on the intrinsic development of confidence in one's own growth.

Like Jim, Adam was very firm in his preference for the familiar score, but was also unable to clearly justify that preference. I perceive several different facets to this preference that are worth mentioning. First, Adam did indicate that part of his preference was based on his perception of the clarity of the information: "being able to see that I got a 15 out of 20…makes it very clear to me how I did on the test." A score or grade, because of familiarity, is at least clear to students. It fits within their established norms for assessment results (Herbst & Chazan, 2012). I would concur that a score or grade communicates clearly, but exactly what it clearly conveys is not always certain. As Jim acknowledged, the score or letter grade is not always clearly communicating degree of understanding. Adam also acknowledged this fact, when I put the question to him. He admitted that, for example, a 15 out of 20 does not necessarily say much about understanding, especially when you compare two different scores, both of which are 15 out of 20, but which are achieved by getting different questions correct. Revising his opinion about the meaningfulness of the raw score, Adam then suggested that perhaps useful feedback would be a score broken down by learning outcomes or topics, "a certain tally out of those, that would probably be best." Here again, like Jim, Adam will acknowledge and even suggest a benefit of more progressive assessment – in this case, disaggregated scores – but that only comes with forced reflection after I pushed a little with my questioning. The fact is that Adam's initial feeling is that a raw score is a logical and clear indicator. I would contend that it clearly indicates relative rank, either relative to

others, or relative to prior personal performances, as the case may be. In other words, students like scores or letter grades because they help them determine whether they are doing better than others, if they have a competitive mindset, or good enough for their own standards, if they are less competitive.

What is interesting is that any scale or ranking system could satisfy Adam's desire for this clearly communicated relative rank, including, in fact, my check-plus scale. In the interview, Adam commented that he "just didn't understand the value of each one," meaning each code in the scale. When I asked Adam why he did not simply seek clarification from me on what my feedback system meant, he mentioned two different reasons. First, he admitted that "maybe it's because my mark was already fairly low and I didn't want to see exactly what I had." Notice that, despite his assertion to the contrary, my feedback system must have conveyed something to him, or else he would not know his mark was low. He went on to add that he feels a certain amount of pressure from his parents to be a straight A student because they drive him in to school from out of town every day.

> *Adam:* I'm expected to do well in school because of all the effort my parents put in, and if I know exactly what I got and I come back to them, I could be like…well, it's just a world of hurt.

What seems to have motivated Adam in his decision to remain in the dark about what my assessment feedback meant was a classic avoidance of pain. His conception of assessment seems also to include a strong Affect component (Brown, 2011). Thus, the score or grade is serving as, again, a reward or punishment and students are using "a calculus of pleasure or pain" (Jevons, 1879, quoted in McCormick, 1997, p. 21) that has little to do with learning or understanding. While both Jim and Adam expressed a distinct preference for the familiar score, there seem to be some nuanced differences in why each of them holds that preference. Another area, though, where their comments coincided was in the perception that only an A grade is a worthy achievement. While I would put Jim's attitude down to having a fixed ability mindset (Dweck, 1999, 2008), it seems from Adam's comments that at least in part his perspective is influenced by parental pressure.

But avoiding the pain of full awareness was not Adam's only reason for not clarifying with me my check-plus scale. He also mentioned he was "too occupied with everything else to take the effort, even though there probably wasn't much effort needed to be exerted." By "everything else" here he presumably means other school work, but also extracurricular activities, which for him included a particularly busy soccer season. We did not clarify just then what the "everything else" means, but we did talk later on about his high level of involvement in soccer, which kept him very busy during the fall semester in which he took my course. This admission echoes some of Eddie's comments on needing to make efficient use of time. For Adam, it appeared to be too great an investment of his time and energy to clarify my feedback system, especially when coupled with his desire to remain blissfully ignorant of exactly how low his mark was. This factor that influences student decisions seems parallel to the efficiency impulse that affects teacher decisions regarding assessment. For students too there are pragmatic considerations about time management and so on, which makes the familiar system of scores and letter grades appealing because it costs little or no time to learn or interpret. Again, this pragmatic factor, like with Eddie, does not correspond specifically to a feature of the conceptions of self, mathematics, school, or assessment discussed earlier (Bagley, 2010; Boaler, 1998, 2002a, 2002b; Brousseau, 1997; Brown, 2011; Dweck, 1999, 2008; Herbst & Chazan, 2012; Star & Hoffman, 2005), but it is an element of student mindsets that influences their reception of assessment in the classroom.

Yet another facet of Adam's preference for familiar scores is that he places more value on tasks in school that provide familiar scores than on tasks that are accompanied by non-traditional feedback. He acknowledged that my unit problems and written assignments, which were returned with written comments and a check-plus code, somehow mattered less to him than tests, which carried scores. But he found it hard to explain why, apart from saying "it's probably just going back to … things we know, so looking at a test and seeing what I got there just outweighs getting, like, a check." Again, this difficulty to explain suggests to me that, despite Adam's claim that traditional feedback makes *logical* sense, it is not a matter of logic at all. In truth, for most students, these issues tend to be unexamined, and grades and scores, rather than being rational, are instead merely an expected and familiar part of their established norms (Herbst & Chazan, 2012).

The expectation of the familiar score or grade is so ingrained that an assignment without it seems less important. Moreover, if conventional assignments or tests also did not include scores, it would be quite a shock. I asked Adam how he would react to getting only a check-plus code on a test, and he confessed, “I don’t think I’d be too pleased – I can see that happening, but I don’t know how I would respond to it.” To a great extent, Adam expects scores in school, and seemed unable to contemplate school without them. Removing them would be a breach of the didactic contract that he perceives to be in place, which would be a shock to his system, as it were (Herbst & Chazan, 2012).

At one point in the interview, I pushed a little on this issue to explore learning things outside of the context of letter grades. I asked Adam to think of a time when he learned something without getting scores or letter grades. He struggled to do so initially, apparently thinking I meant only in school contexts, so I told him that he could think about life in general. I prompted him to think about how he learned things in soccer. As a very dedicated and successful soccer player, Adam was able to speak quite clearly to the manner in which he learns something new in soccer, such as perhaps a new defensive formation that his coach implements. He spoke about the types of on-the-spot feedback received on whether he is implementing the pattern correctly. Sometimes there is just praise, but more usually specific commentary related to what is being done right or wrong, and in practice situations, the coach will even stop everything and point out right or wrong elements, move players around, and then restart the drill or play.

After this detailed discussion of receiving corrective feedback, I pointed out to him that he never receives a score when he learns something new in soccer. Though he acknowledged that this is true, and he admitted that meaningful learning can happen, even in school, without receiving scores, he confessed to being “rattled” when we tried to return to the line of interview questions. I asked him to explain and he said, “I have thought about these things before but never talked about them, and now I have just a whole other way of thinking going on…now I’d like to know how this could work.” Pursuing this line of thinking, I confessed to Adam that I would like to teach and assess mathematics as a coach teaches soccer, and I asked him if he thought that could be possible in school.

Adam: It would be possible but I don't think there would be a whole lot of acceptance to it from students, because it's not what they're used to.

Thus, the conversation circled back around to the notion of familiarity and expectation, the norms, of scores and letter grades. At another point in the interview, Adam, while attempting to describe what he perceives to be my purpose in using alternative assessment techniques, suggested I am trying to "get away from grades because you don't use grades for anything apart from school."

Inter.: How do you feel about those goals? What do you think, are they worthwhile?

Adam: I think they are, but implementing them into teenagers' minds after so many years of grades and percentages would be difficult or tricky for most to accept.

Throughout Adam's interview this thread kept appearing, woven into many of his answers. Behind students' preference for traditional scores and grades is a desire for and expectation of the familiar. Again, though, like Jim, Adam was able to see the benefits of the non-traditional, when he was presented with some of the rationale. Nevertheless, his mindset during the course and coming into the interview was dominated by his preference for familiar feedback. The other factors which influenced Adam's reactions to my progressive assessment techniques were the parental pressure he felt, and time pressures that he felt due to extracurricular commitments.

4.5. Julie

Julie's survey was of particular interest to me. Before I read it, I was curious to see what she would end up writing, because I had a strong sense throughout the course that she, of all students, felt most negatively about my progressive assessment strategies. Her survey comments, indeed, conveyed a strong negative reaction to my check-plus scale and to receiving infrequent updates on letter grades and percentages. The negative tone of her reactions, though, was less interesting to me (since I expected it) than some of the reasoning she displayed. With respect to the check-plus system, she suggested it "would be better if we knew what each symbol represented and its value." I thought it interesting that she drew a distinction between what a code represents and its

value, which I presumed meant its letter grade equivalent. Julie also wrote that "a letter grade and percentage would contribute nicely." I wondered what she meant by "contribute." Her survey comments in general emphasized letter grades and percentages. She claimed that "all that really matters is how I do on tests and my report cards." These comments, focused on grading, helped to explain why I perceived a negative attitude earlier in the course, and I thought it worth exploring Julie's perspective. In addition to the overall focus on grades, Julie also included a closing comment on her survey that referenced her busy schedule, which kept her from making use of my quizzes as useful formative feedback. Julie, like Adam and Eddie before her, were influenced in their decisions by busy student lives. This too was a factor I wanted to discuss with her.

Though it was not a dominant theme in Julie's interview, she did touch on the issue of a busy schedule that kept her from taking full advantage of assessment feedback. She noted that it is "not like a good excuse, but I have something almost every day outside of school – hockey almost every day last semester, and other times it was piano, or chores." Like Adam, Julie was a committed athlete and though she did take opportunities to rewrite and improve her grades, she rarely came to see me at lunch for help with the work. She would rather just re-do and re-submit, which did not always result in much improvement. This is an example of maladaptive behaviours (Brown, 2011), but related to practical constraints in Julie's mindset more than to psychological conceptions of assessment.

The majority of Julie's interview, though, focused on issues related to letter grades. Early in the interview, Julie stated that her preference for letter grades stems largely from the fact that she is "really used to having that" as usual feedback on school work. As I have noted before, there is a sense of familiarity with traditional scores and letter grades, which is why students expect them and notice their absence. This factor parallels the teacher's sense of tradition in making assessment decisions. It is not so much that the score or letter grade conveys specific information that students find useful, but that it conveys the kind of information they are used to getting. As noted earlier in Adam's interview, students want scores or letter grades because it costs them little time or energy to interpret such feedback. Thus students are influenced also by a sense of

tradition and a need for efficiency in how they respond to assessment, similarly to the influences felt by teachers.

Julie's preference for traditional scores and letter grades, though, was also heavily dependent on her perception of competition in school. When I pressed her to elaborate on her survey comment that a letter grade and percentage "would contribute nicely," one of the factors she alluded to was that the letter grade conveys relative ranking in the class. She said an A or B indicated you were doing well, to which I asked in what ways she meant "doing well." Julie's first reaction, after a pause in which her facial expression indicated she felt the question was strange, was that A or B meant "above average." She said this as though it was obvious and hardly needed stating. There was no mention of the grade or score conveying a level of understanding. Rather, Julie suggested later that in fact grades were the prizes to be won in the game of school: "I know it shouldn't be, but school is a competition for grades." Though she prefaced this comment with an admission that it somehow is wrong, Julie's tone at this point was matter-of-fact, as though the issue was beyond dispute. But when I pushed for some explanation as to how exactly students are in competition, the only real outcome she could point to was university acceptance: "no university is going to take me if I have C's on my report card." In fact, Julie claimed that university acceptance was "all that really matters." She explained that above average school grades meant access to university, which was a ticket to better jobs and therefore success. Despite the doubtful veracity of her comment regarding C's, it is interesting to note that the only "scarce resource" for which students can be said to compete is seats in universities, and perhaps the scholarship money that helps them get there. I found myself wondering, though, if I was really hearing Julie speaking here or was instead hearing the mantra she may have heard from her parents over the years. Regardless, it is a factor that influences student reactions to assessment, this notion that assessment results are the means to success in worldly terms. There is a mix of factors at play in these comments by Julie, including established norms about grades (Herbst & Chazan, 2012) and an External orientation (Brown, 2011) in how she conceives of assessment.

Notice, though, as I did, that this whole discussion says nothing about being successful because you have learned something. I probed to see if Julie was aware she was leaving out what I believed to be a crucial part of the equation (if not the only part).

Quite strikingly, Julie only very reluctantly admitted that grades should reflect understanding and it is the underlying knowledge and ability that actually allows for success. I have to admit that at this point in the interview my role as researcher was probably overcome by my sense of indignation about Julie's singular focus on grades for their own sake. My line of questioning was a rhetorical attempt to force the admission out of her that the understanding is what matters. I asked what a mark is supposed to convey, to which she replied, "just to show us how well we're doing in this class." I asked again what she meant by "well," and again there was an interesting pause, but she replied, "what you get on tests." Still pushing, I asked what the scores on tests were supposed to show, or what the tests were supposed to do in general, and then, finally, after I rephrased my question a couple times, she said, "yeah…tests how well you know stuff." It was very interesting to see in her body language and hear in her tone of voice a real reluctance to make this admission. Julie clearly does not perceive assessment results to be something central to her own improvement, but rather to be something external to her, serving purposes of comparative ranking and sorting for entrance to further levels of schooling (Brown, 2011).

This disconnect between grades and understanding was perhaps the overarching theme of Julie's entire interview. We returned to the idea from various angles throughout our conversation. In the early part of the interview when I was trying to explore what connection she saw, Julie stated that "we should comprehend, but cramming does play a large role in school." Again, note that though Julie acknowledges what should be, she nevertheless seems to be justifying or at the least apologizing for the fact that school is not really about understanding. She claimed that she is not the only student who feels this way either, and that the common student outlook differs from my emphasis.

> *Julie:* Students as a whole feel that marks and letter grades are of importance whereas you strongly feel that whether we understand is important.

During the subsequent discussion about university acceptance and the need for good grades, Julie made a startling claim that "if given the choice between having a semi-lower mark with understanding and having a higher mark, you'd choose the higher mark." It is not so startling to say students would take the higher grades, but that Julie

would choose to set up this false dichotomy; it strongly suggests a perception in her mind that grades and understanding do not correlate. In keeping with the tone of apologizing for the facts of school life, Julie went on to discuss achieving good grades by cheating, which, though she claimed, “I’m not saying I do this,” she nevertheless all but justified because, again, it is the grades that are “all that really matter.” Astonished, I pushed again here to see if she might admit that gaining entrance into university by cheating might mean that the cheater lacks the requisite knowledge and skills to succeed. Once more, she did not initially acknowledge that cheaters would not prosper. It seemed as though Julie felt that the grades that gain you admittance to the next level do not need any accompanying knowledge or skills. Julie’s mindset includes a conception of mathematics that emphasizes procedures and memorization (Star & Hoffman, 2005; Boaler, 2002a, 2002b), mixed with an External orientation (Brown, 2011) that downplays even the significance of learning those things. It is a strongly maladaptive mix of elements.

Towards the end of the interview, when I had moved beyond the lawyer-like tone I had fallen into, I was surprised by the fact that Julie was quite aware of my underlying purposes in choosing progressive assessment techniques. In response to the question about what she felt I was trying to accomplish, Julie noted that I was trying to make the students “learn and comprehend the matter … and not overly obsessed about letter grades.” It struck me that she was well aware of why I did not give frequent letter grades, but she obviously was not persuaded that it was a worthy goal. However, not long after, she did admit that letter grades and understanding “should be connected.” I am not sure if by this time she was merely saying what she thought I wanted to hear. Certainly earlier in the interview her comments suggested that she did not in fact feel there needed to be any such connection, and that grades were a valid student goal for their own sake.

In the end, it is not surprising that Julie objected to my techniques. Whereas my ideal of assessment, based on “a rich and practical theory of the individual learner” (Ginsburg, Jacobs, & Lopez, 1993, p. 158) as noted in Chapter 2, emphasizes individual, non-comparative, descriptive feedback for the purpose of promoting deep conceptual understanding, Julie apparently does not value the goal of deep understanding nor the form of such feedback. For her, rather, school seems to be about getting grades in whatever way possible, and actually understanding the material seems to her like the

more difficult and hence least palatable means of doing so. I would conclude that Julie, being aware that my main goal was promoting understanding, and that I rewarded deep understanding with high marks in the end, was displeased that this was the only means at her disposal of achieving what she had said was “all that really mattered.”

4.6. Kara

Much like with Julie, I felt that Kara was displeased with my progressive assessment techniques through much of the course. To my surprise, however, her survey revealed that she in fact was starting to appreciate what I was doing. She wrote that “the check-plus system has grown on me.” Given my interest in factors influencing students’ reception of progressive assessment, I found it compelling to investigate the case of a student whose attitude had changed from dislike to acceptance over the semester. Kara also praised the test reflection system in her survey, “because it maps out exactly what needs work.” But her survey comments were not entirely positive, as she noted that the quiz system I used was not particularly helpful. Kara also made a similar comment to Julie’s about letter grades, saying “what my mark will translate to on a report card…all-in-all that’s all that really matters.” So Kara’s survey suggested she was warming up to my progressive techniques, but that she still held some attachment to traditional systems. I was keen to see what I could discover about what lay behind Kara’s attitudes, and I found that her thoughtfulness and depth of response made the interview process very worthwhile.

On the whole, there was actually a stark contrast between Kara and Julie’s ideas despite the fact that prior to the interviews I had presumed they would exhibit similar thinking. They had seemed similar in their antipathy to my methods, but it turned out that Kara’s objections were first of all more reasoned, even where they were similar to Julie, and second were less hardened – Kara seemed more open-minded about the benefits of progressive assessment. An example of the distinction between Kara and Julie is where Kara explained to me her “all that really matters” comment. Whereas Julie’s comments struck me as relatively unexamined, a parroting of parental or social banalities, Kara was able to clarify what she meant when she said “in the long run, no matter how deep your learning is, that’s kind of what matters.” She added that “when other people are looking

at it they don't see how much I've learned," referring specifically to university admissions and scholarship departments, but rather that what shows is "a number on a piece of paper to the people that are going to be judging whether they want you at their school or whether they are going to pay you to go to their school." And this same requirement of reporting a number applies whether the teacher is more progressive or more traditional, Kara noted. The idea here is similar to Julie's External orientation (Brown, 2011), but note how more nuanced is the reasoning. And note also that Kara hints at the connection that is supposed to exist between learning/understanding and the grade.

Despite the fact that she saw there should be a connection, though, some of Kara's other comments indicated that perhaps she did not see grades strongly connected to understanding. At one point when she was describing how she thought assessment should be done, she noted that in early grades they should eschew grades in favour of "a scale of how much you learned." Clearly she does not see the letter grade scale as such a scale, which is interesting. At another point, Kara revealed what she feels is the real purpose of grades: "we have them because you need to know how you rank beside other people." So, for Kara, grades are at root a comparative tool, and not a scale that lends itself to a less competition-laden description of "amount learned." In fact, it is the de-emphasis of competition that was one of the reasons Kara gave for why my check-plus scale had grown on her. She said it "highlights what you did well better and is more positive looking than with numbers or letters." This comment suggests numbers or letters, perhaps because of their familiarity as a ranking tool, are more naturally devices of comparison with others. Kara's insights suggest some possible benefit to breaking with the norms (Herbst & Chazan, 2012) associated with grading.

Unlike Julie, Kara was also able to give more thoughtful analysis of what makes feedback useful. She seemed to prefer feedback that describes learning and specifically how to improve. For her, the preference for letter grades was based on the fact that letter grades are part of the system by which she judges whether she is mastering material. In the interview, I told her that I had felt she wanted more traditional feedback earlier in the course and was not happy about not receiving frequent updates about letter grades. She did admit that "it was confusing at first" and talked about my check-plus scale as "another scale to learn," but she revealed that the negative attitude I had picked up on was more directed at the difficulty of the course and her awareness that she was

struggling to learn the mathematics. Her initial negative reactions were also coloured by her impression that I was going to be like an English teacher she had the previous year, who gave descriptive feedback only and, according to Kara, only assigned an actual grade once during the year, at the end. She strongly disliked not knowing where she stood in terms of her report card performance in that course, but confessed that I ended up providing enough clarity with my unit test scores and end-of-term interviews. Kara also noted especially that the test reflection sheets were a valuable means of describing what she knew and did not know, which was what she said is the kind of feedback she really appreciates so that she can address deficiencies in understanding. Though she has some elements of an External orientation in her conception of assessment, Kara also has some Improvement orientation (Brown, 2011). Interestingly, though, Kara had to warm up to the test reflections, finding them "at first annoying, but then easier to pinpoint what I had to focus on for that unit." She stated that the tests with the reflection sheets were a good balance of useful feedback, because "the score gives you a heads up relative to the other units which ones you did better in, but then the test reflection examines within the test what things you need to focus on."

Notice that in Kara's comments, she indicates a desire to really know the material. Feedback she found useless was feedback that indicated a lack of knowledge with no clarity about how to improve. Kara mentioned the quizzes in particular, because she said they most often simply pointed out what she already knew: that she had not yet grasped yesterday's concepts. She seemed to think the quizzes were primarily a tool to promote homework completion, and she felt this emphasis was subtly unfair to the busier students.

> *Kara:* I will get my homework done within a couple days but not always the night after the lesson so sometimes it's frustrating when people who do have time get all the quiz questions right.

This comment highlights the factor of time management and the student's need to compartmentalize her student life for efficiency's sake. But something else seems also to be at work in this comment, because, despite the fact that I do not really count the quizzes towards the term grade, Kara felt they were in some way punishing her. One possibility is that competition is the central factor, but it would be competition for its own

sake, rather than for some scarce resource in this case, because, again, the quizzes are meant to be formative assessments. I would suggest though, because of Kara's other comments that indicate her desire to know what she is missing and how to improve, that the main factor is the emotional sense of failure or inadequacy that the quizzes expose for her. This is some evidence of an Affect orientation in Kara's mindset (Brown, 2011).

I was interested in this idea of the emotional impact of feedback so I followed it up with Kara. She acknowledged that there is "a delicate balance with the emotions" when it comes to assessment feedback. She described a fine line between tipping someone towards more effort and tipping someone towards giving up. As an ideal, she proposed "bit-by-bit feedback" which helps to give a sense of "an achievable goal." I think this commentary again highlights Kara's preference for feedback in chunks that one can react to and do something with, to address deficiencies in understanding. The emotional impact then, for her, must be when it seems impossible to address the weakness, either because the gap in understanding is too wide or the volume of material is too large.

Towards the end of Kara's interview, when I addressed to her the common set of general questions, it became even clearer to me that Kara does have a focus on understanding, and that she idealizes one-on-one assessment and feedback. Echoing Jim's comment, Kara also stated, in response to the ideal math teacher question, that it "would be more ideal if they didn't have to give grades." Kara also seemed well aware of my goals, noting that I seemed to be trying to "help us help ourselves," and that I wanted students to "actually get a very deep understanding." In view of our shared emphasis on understanding, Kara spoke quite appreciatively of the end-of-term interview process, saying that it contributed strongly to her sense of buy in to my assessment program by the end of the course. She said she appreciated how I communicated where they were at, and that, in fact, the interview process made it seem like the students "were involved somehow more in the mark." Strikingly, Kara claimed that with traditional marking schemes, sometimes "it almost feels like you have nothing to do with your mark." While we did not discuss in any more depth what Kara meant by this, perhaps it indicates, along with the bulk of her interview comments, that Kara appreciates a sense of control over how she is doing in terms of both what she is learning and what her grades are going to be. Kara's suggestion that students often feel disconnected from their marks

reinforces Bagley's (2010) claim that Authority Ranking relational models dominate in assessment interactions. Kara seems to prefer the Communal Sharing elements of my assessment scheme.

Overall, Kara's interview ended up being one of the most interesting for the depth of reasoning she displayed and the range of factors she revealed that influenced her reactions to my progressive assessment efforts. She displayed elements of both External and Affect orientations, but an Improvement orientation in the end won out, which allowed her to display adaptive, growth-oriented behaviours overall (Brown, 2011). Whereas I had assumed she was flatly displeased with my techniques because they were non-traditional, as it turns out her negative attitude was more about an awareness that she was struggling to master Mathematics 12. Over time, as my approach allowed her to identify and address deficiencies, and thereby improve her understanding and grade, Kara displayed an appreciation for opportunities to rewrite and resubmit tests and other work. She ended up justifiably proud of achieving a solid B in the course, and demonstrated comprehensive enough knowledge overall to earn one of the highest scores on the final exam.

4.7. Summary

As is evident from the results of the interviews as described above, the six students involved in this study had varying reactions to my progressive assessment strategies. Some students were generally positive about the assessment techniques I employed, others were for the most part negative, and at least one showed some change in attitude from negative to positive during my course. Each student's unique attitude and response to my assessment techniques was based on a mix of factors in his or her student life and mindset. But, as I have already mentioned at places in the descriptions above, links between the student mindsets are apparent. In the next chapter, I will draw on those links to categorize the factors influencing student reactions to my progressive assessment techniques.

5. Results and Analysis: Across the spectrum

Having discussed each student's comments in light of the conceptions of self, school, mathematics, and assessment, it is now possible to assess collectively which of the various conceptions are stronger influences on my students' reception of assessment. Moreover, some common factors have become apparent that are not accounted for in the theories I have used as a basis for analytic induction.

While the previous chapter dealt with the interview responses on an individual basis in order to help reveal how the students receive assessment uniquely, this chapter will make explicit the common threads that run through the interviews, in order to help describe what assessment is to students more generally. As a basic structure to lay over the data, I will use a scale of receptivity. The students individually displayed varying levels of receptivity, or openness, to my progressive ideals, and I will attempt in this chapter to group the theoretical factors in student mindsets that help to situate them along the receptivity scale, from receptive or positive responses at one end to unreceptive or negative responses at the other. In essence, this spectrum is similar to Brown's (2011) contrast between adaptive and maladaptive responses or learning behaviours. But I will also describe the factors of student mindsets that appear across the receptivity scale, appearing in students at both ends of the spectrum. A cluster of these features of student mindsets are practical considerations regarding efficiency that do not appear in the theories discussed in chapter 2. How students receive assessment cannot solely be explained in terms of psychological conceptions.

5.1. Positive Reception

The students who received my progressive assessment efforts positively were characterized most strongly by an Improvement orientation (Brown, 2011). Their conception of assessment was as a means to obtain useful information for improving

their understanding or ability in the course. Eddie and Maia both emphasized this conception in their interview comments, but so too did Kara. Kara's case is instructive, given that my impression before the survey and interview was that she received assessment negatively. But her interview comments revealed that by the end of the course, once she had adapted her learning behaviours based on the consensus-building approach we were using, her reception of assessment in my course was positive. My research confirms, then, what Brown (2011) notes, that an Improvement orientation promotes adaptive learning behaviours.

Kara's positive response to assessment was based heavily on her appreciation of the consensus-building approach. Maia too, and Eddie to a lesser degree, seemed to prefer the Communal Sharing (Bagley, 2010) relational model as it played out in my classroom. Kara noted how my assessment scheme and the end-of-term interview process helped her feel connected to her mark in a way she did not always see in other courses. Though Maia did not discuss the end-of-term interviews explicitly, her interview comments and general classroom behaviour did reveal a commitment to the shared goal-setting, specifically in the area of conceptual understanding. Her approach to note-taking, which involved seeking feedback from me to ensure she was accounting for the concepts correctly, shows remarkable initiative on behalf of a student in taking ownership of her learning, which is a key feature of the Communal Sharing model (Bagley, 2010). Eddie, too, showed a tendency to own his results, and to build consensus on his degree of understanding. His preference for direct feedback in conversation with a teacher is evidence of this tendency.

Owning results and seeking consensus on understanding are also a reflection of an incremental conception of intelligence (Dweck, 1999). In different ways, each of these three receptive students demonstrated a growth mindset (Dweck, 2008). Eddie sought the kind of direct feedback he desired because it helped him improve. Maia believed her understanding was capable of growing over the course of a unit of study, and indeed focused her energies on that growth as the central purpose of learning. Kara, too, in her way, appreciated the test reflection sheets because they helped her focus attention on the areas in which she could improve the most. She noted that for her there was a strong sense of accomplishment in how her understanding grew over the course of the semester.

It should come as no surprise that these three elements (Improvement conception (Brown, 2011), preference for Communal Sharing (Bagley, 2010), and growth mindset (Dweck, 2008)) characterize the receptive mindset, given how these elements mutually reinforce each other. It is not that they are just three ways of saying the same thing, for it is possible, for example, to see one's intelligence as something that can grow without perceiving the purpose of assessment as contributing to that growth. But consider again the magnet metaphor, picturing how the three elements exert aligned forces. Having a growth mindset regarding intelligence makes it more likely one would see the purpose of assessment in terms of improvement. Similarly, seeing the primary purpose of assessment as related to improvement requires something to be improved. And the improvement of understanding as a goal of assessment is aided by the development of shared responsibility between student and teacher.

Students at the receptive end of the spectrum, then, see assessment as a process for acquiring potentially useful information that helps them improve in their understanding. This perception made Maia, Kara, and Eddie open to my efforts to the extent that I did indeed provide them with such useful information. And they used the information thus provided, adapting their learning behaviours in order to deepen their understanding.

5.2. Negative Reception

On the negative end of the spectrum, students displayed more maladaptive behaviours, including avoidance and complaints, reflecting their primarily Affect- and External-oriented conceptions of assessment (Brown, 2011). Adam avoided clarifying his results and adjusting his approach because he feared "a world of hurt" from his parents. Jim, both in his body language and in comments related to how he feels about competition, revealed a strong emotional component to his lower than usual results in my class. The affective response, though, seems intimately connected to an overarching External orientation, in that students perceive assessment results as coming from outside of themselves, handed down as judgements, and serving the needs of social structures rather than the individual learner. None of Adam, Jim, and Julie articulated explicitly this perspective in their comments, but each did focus on results for some

external purpose, whether it be earning parental approval, scholarships, or admission to university. These external purposes, it should be noted, were mentioned also by Kara and to some extent Maia, so it seems that the External and Improvement orientations are not mutually exclusive. The unreceptive mindset, then, is not so much characterized by the External and Affect conceptions as it is by the lack of an Improvement conception.

The lack of an improvement conception would seem to indicate a fixed-ability mindset (Dweck, 1999, 2008). I only found significant evidence of this in Jim's comments though, especially in his reference to not improving his writing and in his confused response to earning higher scores on rewrites and second attempts. There was some indication in Julie's comments that she saw assessment results as separate altogether from the understanding they are supposed to measure, which might indicate an entity view of intelligence, but it is not entirely clear.

What is clear is that the unreceptive mindset holds a preference for traditional evaluation schemes as found in an Authority Ranking (Bagley, 2010) relational model. All three of Julie, Adam, and Jim professed a strong desire for the supposed clarity and certainty of traditional scores and letter grades delivered by the teacher, which, as Adam said, conveys to them "how the teacher sees" their level of success in the course.

Again these features make sense as a cluster of traits in a generally unreceptive student, who, lacking an Improvement orientation (Brown, 2011), receives assessment as information about one's ability, delivered by the one person (the teacher) instilled with the authority to pronounce such judgements. For such a student, tradition holds a lot of weight because they have developed strongly held expectations about how school works based on their immersive exposure to the system of ranking and sorting, which, as Jim said, is "just how our whole school system is based." The choices made by students at the unreceptive end of the spectrum confirm Brown's (2011) description of maladaptive learning behaviours: avoidance, disengagement, complaints, and so on. In light of the psychological conceptions, these choices make sense given that the students will either be receiving assessment as a negative judgement of low ability, which provokes avoidance behaviours, or will perceive a general lack of the kind of information they have come to expect, which provokes the complaints about the progressive system. The non-evaluative information that is provided to them does not fit with their psychological

expectation of how assessment is supposed to be, which makes the information almost irrelevant.

5.3. Themes Across the Spectrum

While it is possible to identify differences between students that place them somewhere along the receptivity scale as described above, there are also some threads that kept recurring in the data that seem to be features of student mindsets whether they are receptive or unreceptive.

5.3.1. Norms

In all six interviews, the participants revealed that they held common expectations regarding assessment in mathematics classrooms (Brousseau, 1997; Herbst & Chazan, 2012). The expectations were the same whether the student was receptive or unreceptive. In all cases, the more progressive elements of my assessment scheme, including avoiding traditional scores and letter grades and using an interview process to arrive at a consensus for which grade to report, were perceived to be abnormal in some way. This comes as no surprise, given the very definition of progressive. What is worth noting here, then, is how the aforementioned psychological conceptions set up students to respond to the breach of didactic contracts (Herbst & Chazan, 2012) that they presume hold in mathematics classrooms.

The unreceptive student is generally opposed to the breach of contract. Witness Julie's unspoken disapproval in her pregnant pauses before she answers questions. She seems taken aback by questions that call into question the typical or traditional. Adam and Jim, too, stated a strongly held preference for traditional forms of feedback, but were not able to clearly articulate reasons why. The inability to explain suggests that the traditional forms are just "normal" to them. Interestingly, both Adam and Jim (but less so Julie) did acknowledge, when pushed, some benefits of breaking with the norms, although it was somewhat reluctant. As mentioned above, their External and Affect orientation (Brown, 2011) seems to make the kind of feedback received in progressive assessment largely irrelevant to them.

The receptive student, on the other hand, seems open to or even appreciative of breaking with the norms related to assessment. Though they hold the same expectations as the unreceptive student, their mindset makes them distrust the tradition, and welcome a change. Consider Maia's revelations about how her normal experience in mathematics conveyed little emphasis on understanding, which made her appreciate my assessment, given her Improvement conception (Brown, 2011) and focus on understanding. Kara also displayed appreciation for my "abnormal" approach when she commented on how the interview process helped her feel more connected to the results she was earning.

5.3.2. Practical Constraints

While the themes collected above all confirm or otherwise illuminate elements of the theoretical framework I laid out in chapter 2, there was one collection of threads that appeared in the interview data that was not accounted for by the theories of psychological conceptions. In every interview, the student referenced some element of his or her student life that in practical terms prevented a fully receptive response to progressive assessment. This was true for the most receptive students like Maia and Eddie as well as for the unreceptive students like Julie. These practical constraints can be grouped into three categories: time constraints, efficiency, and practical considerations about grades. I call these features "constraints" because they seem to be factors of student life that influence decision-making about assessment regardless of the student's unique mindset as discussed above.

Time Constraints

It may be prosaic and overly obvious to say it, but sometimes students simply have too much on their plate for them to demonstrate buy-in by accepting and responding to a teacher's assessment feedback. Within my classroom situation, there are certain required assessments students complete, but I try to allow for optimal performance by permitting re-writes and re-submissions, and providing feedback on how to improve the work to help students address conceptual gaps and errors. Again, this is all part of my idealization of assessment as an individualized process, in which I provide personally-tailored information to help and allow students to improve. But students, more often than not, do not utilize the information nor even the opportunity, which breaks the

integrity of the ideal I have formed of the assessment process. A very straightforward reason for this decision is often a lack of time.

All of the students I interviewed in some way or other admitted that they sometimes do not use assessment feedback because they are making a time-management choice. The busiest students, Adam, Julie, and Kara, had the most to say about this issue, but each of the other three also commented at least once about not having enough time to do one's absolute best on everything. Of course various reasons and motivations lie behind a student's choices, but as a practical issue, time management plays out similarly in each student's behaviour. If the ideal of assessment as I envision it involves each student responding to my tailored feedback, when given, to improve work and enrich understanding, in some way each student opted not to buy in at all times to this ideal. At least some of those times, at least in part, these decisions were based on the students having too many other demands on their time.

What does this say then about how students receive assessment? This factor reveals that the assessment process is at one level a mere practical concern to students, a part of their student life along with a load of other courses as well as extra-curricular activity, employment, and socializing. The fact that this reality came up in each student's comments suggests it is a common issue affecting students generally. It is not always a lack of acceptance or a negative attitude that keeps students from responding to assessment feedback as the teacher intends, but rather, as Eddie pointed out, given that he has "a lot to balance at one point…[he] can't spend enough time to completely fix everything that the feedback is telling [him] about."

I was not and am not surprised by finding this result in my interviews. In fact, as I said earlier, it hardly needs stating. But it is notable how universal this issue appears, and it is actually worth stating because of how it connects to the practical concerns of a teacher too. One of the factors noted in the literature review that affects teacher decision-making with respect to assessment is the notion of what will be efficient in the teacher's busy life. As a teacher, then, it is worth keeping in mind not just one's own practical concerns, but also those of the students. Sometimes efficient choices in the teacher's life – say, to put off marking until a convenient evening or the weekend, which often means significant delay in feedback – impact the time management issues for

students. Perhaps more immediate feedback or direct feedback in situ would be feedback that students on the whole would make more use of. Or, as Kara pointed out, perhaps the timing and volume of feedback needs to be adjusted so that it comes "bit by bit" and is not too overwhelming for the students' busy lives.

Efficiency

Time constraints are one factor within a broader umbrella issue of efficiency. It is worth noting other factors of efficiency that influence student decisions. My interview results also revealed efficiency issues within a course itself, as opposed to issues that impact the students' choices from their lives in general. For example, consider Adam's choice to focus mostly on the feedback I gave him on tests and quizzes as opposed to the other important assignments within each unit. Or consider Kara's choice to address misconceptions in the one unit she found most difficult rather than attempt rewrites on everything. Again, though different motivations may lie behind these choices, they each represent different facets of efficiency. For Adam, the test results were familiar, and hence, he believed, understandable, so it made more sense for him to address his performance on those tasks rather than try to interpret and respond to the non-traditional feedback I provided on the other assignments. And for Kara, though she might benefit from rewriting tests and assignments in all units, it made most sense to focus on the one that was most difficult for her, so as to improve it as much as possible.

Again, this factor highlights that, for students, assessment is a practical concern, perhaps less so than an ideal one. The ideal might be to address every deficiency and misconception, to deal with each weak area, to strive for all-encompassing perfection, but, frankly, students know that this is impossible so they sometimes choose to cut their losses, so to speak, and maximize the rewards within a more limited scope.

Practical Significance of Grades

No discussion of "maximizing rewards" within a classroom could ignore the significance of grades to students. All of the students I interviewed, in some way or another, displayed a motivation to acquire higher grades, or to minimize risks of lower grades. To be sure, some students with a stronger External orientation (Brown, 2011) were more focused on grades than others, with Julie, Adam, and Jim the most grade-

dependant. But even Maia, who remember I felt was one of the most dedicated learners I have ever taught, confessed to short-circuiting learning sometimes if she was happy with the mark she had earned.

The significance of grades in students' reception of assessment played out, as I mentioned earlier, at both a practical as well as theoretical level. On a practical level, grades have significance because of their familiarity as a feedback system. In the case of Julie, Adam, and Jim, especially, grades were preferred because, thanks to familiarity, it was easy for them to interpret such results and thereby make decisions about how to respond. Again, recall Adam's point about how test scores weighed more heavily in his mind, despite the fact he knew at some level that the written assignments and unit problems were equally important. Adam claimed that grades made logical sense to him, implying that somehow the grades had more theoretical meaning, but it seemed evident to me that it was mere practical familiarity that was the basis for his preference. Thus, on a practical level, assessment in students' eyes involves grades and scores received for work completed – grades are the norm they have been trained to expect in an intuitive, almost unexamined way. Quite apart from any meanings associated, students find it strange not to get grades and scores on work they submit because, practically speaking, they have always got them before.

Grades also have a practical impact for students as they look towards future schooling. As most students mentioned during the interviews, university entrance and scholarships hinge on high school achievement. For my senior level students, these practical concerns are immediate and significant. As such, students will base decisions they make on a target grade they are aiming for, and will either do extra to reach it, or not do more if they are already achieving it. Thus, what looks like buy in to the unspoken ideal of assessment, or what looks like opting out, may be in fact motivated by the grade incentive as opposed to my teacher's ideal of deepening understanding or maximum possible achievement.

What this says about how students receive assessment is that it may be nothing more than a practical hurdle between them and future goals. Again, distinct from any meaning associated with letter grades — that is, any level of knowledge or understanding that the grade is supposed to reflect — letter grades can and often do

become an end in themselves, based on students' perceived requirements for some future plans. Even Maia revealed some effect of this External conception of assessment (Brown, 2011). To put it in the extreme, this factor is what plays out in the stories of students buying grades, such that real economics invades the classroom economy, and individuals on both sides of the exchange are willing to completely forego the proper relationship between learning and grades. But in more subtle ways the same factor is at work in my students when they accept, as even Maia confessed to doing, grades they do not feel they deserve.

5.3.3. Theoretical Constraints

Much like the practical constraints above, there are also common theoretical constraints that impact students' choices whether they are generally receptive of progressive assessment or not. These factors include a curious theoretical disconnect between grades and understanding, and a strong conception of school as a competitive context.

Meaning of Grades – Disconnect Between Grades and Understanding

The notion that Maia would feel she deserves or does not deserve a particular grade shows that, obviously, grades have a meaning beyond their use as a ticket to future schooling. In my review of the results of my interviews I could not help but notice how nuanced were the students' attitudes about what grades mean. Even within one student's mindset, I could perceive sometimes several different notions at play about grades and what they mean or reflect. Recall Jim's mix of conflicting ideas, which included notions of rewards, effort, reflections of fixed ability, and summaries of understanding, among other things. Thus, beneath the merely practical concerns of grades in students' minds, there are important ideas or theories about what grades are that influence their attitudes and thereby decisions.

The main issue I noticed in all interviews was an apparent disconnect in students' minds between grades and the actual concepts being learned. In my idealization of assessment, grades are almost an afterthought, a required code I have to attach to each student's report to indicate (rather inaccurately and coarsely, I feel) what level of understanding and achievement he or she reached in a term. But the significance of the

grade, such that it is, is a reflection of understanding. Does this student comprehend excellently? Very well? Just above an acceptable standard? And so on. In truth, my ideal would be to do without grades, and to provide more disaggregated and meaningful information to students, parents, and others, which is why I try to downplay the significance of grades as much as possible. But the truly startling thing I discovered in my interviews is how tenuously, at best, students will connect grades and understanding.

At the extreme end of the spectrum, of course, is Julie's apparent but mysterious reluctance to even acknowledge that grades are supposed to reflect what you know and understand. But recall that even Maia, who of all the students most desired deep and lasting understanding, stated matter-of-factly that grades do not always accurately reflect understanding, and in fact, that understanding is not even always emphasized as the goal of schooling. Now, for Maia, this appears to be a weakness of some school experiences, whereas for Julie, she seems to want the disconnect to be there, but the point is that in almost every interview some acknowledgement of the grade-understanding disconnect was made.

One of the more striking insights in the interviews occurred when I discussed with Jim and then Adam the fact that grades have nothing to do with learning in most contexts outside of school. The insight was strong for both the students and me. Adam, remember, confessed to being "rattled" when we discussed how he learns complex things in soccer without ever being graded. I must confess to being somewhat rattled when Jim pointed out that "pretty much everything you learn outside school" does not involve grades. If learning can occur outside of school without grades at all, and school is about the only place where we assign grades for learning, is it really strange that students might perceive a disconnect between the purpose grades serve and the actual learning that does or does not happen?

Recall how Kara, in a typically insightful comment, pointed out that the disconnect is so strong sometimes that she does not even feel like her grades have anything to do with her. Consider the significance of this comment. If Kara feels no connection to the grade she has been given, and the teacher is the only other person involved in the grading interaction, then I can only conclude that Kara must, if even subconsciously, feel the grade is connected solely with the teacher giving it. This flips

my ideal as stated above somewhat on its head. Whereas I as a teacher feel grading is about the student, perhaps Kara's comment suggests students sometimes feel grading is about the teacher. And really, I would not necessarily argue otherwise. I acknowledge that grading is subjective, in its very nature, and also, as I said earlier, wish to downplay the role it has in the learning situation. Ultimately, my ideal is for students to own their results, whether they be summary grades or formative comments and feedback, so that they can use them to improve and deepen understanding. The fact that students all note varying degrees of disconnect between understanding and grades is of some concern to me. This attitude is likely promoted by traditional, Authority Ranking (Bagley, 2010) models, and is connected most strongly, again, with the students who have an External orientation (Brown, 2011). But even Maia displayed some aspect of this attitude, so the attitude seems to cross the boundaries of the conceptual models outlined in the literature review, and reflects something more universal about the meaning of grades in student experience.

Perceived Competition or the Comparative Impulse

If students give some other sort of meaning to grades than a measure of understanding, what could it be? I do not think that students see grades as having no meaning, though I do feel that in some ways the meaning they attribute to grades is meaningless, so to speak. Apart from something arbitrary, which Kara's earlier comment hints at, the most common notion students mentioned for the meaning of grades is as a mere ranking tool. Jim stated that he feels schools "have to have a way to see how you're doing in comparison to…other students." Julie also highlighted the ranking and sorting purpose, and Kara stated openly that "you need to know how you rank beside other people." I note that both Jim and Kara use words like "have to" and "need," as though there is some imperative about ranking and sorting that stands beyond question. I suppose it is not a stretch to assume the imperative to which they refer is the need, mentioned in the literature review earlier, to sort for jobs and future placement in schooling, but it should be noted that they do not explicitly say so here. Moreover, it is not actually that difficult to imagine a situation where, for example, university access is universally available and thus sorting is not required per se. What strikes me is that the students' comments suggest a conception of ranking and sorting as an inherent

imperative, required by the nature of assessment and learning itself, rather than a requirement from outside.

Certainly it seems that at least some students have an inherent comparative impulse within the learning environment that governs their reactions to assessment information. Again, the most striking evidence of this impulse was given by Julie, when she said that earning 99% would be less exciting if she found out everybody else had earned 100%. Jim and Kara, too, though, displayed a strong comparative impulse in some of their answers. Kara, remember, pointed out that the daily quizzes were unfair to the busier students because the ones who can spend all their time on school get all the questions right. Jim felt that, despite his acknowledgement that a rewrite could reflect improved understanding, a mark for a rewrite should not be the same as a similar performance on the first try. These attitudes about fairness strongly suggest a comparative response to assessment information. This impulse starkly contrasts with the ideal of assessment I promote as a teacher, which aims to make each student's feedback unique and private to him or her, for personal benefit, not public comparison.

The comparative impulse does not seem to be universal, however. Neither Maia nor Eddie made any mention of grades in comparative terms. Perhaps they have the comparative impulse but it did not come out in the interview, though I feel that even if they do have some comparative tendencies, they are weaker. The overall tone of their interview comments conveyed a strong focus on personal achievement and self-efficacy. Thus, though Maia and Eddie noted that grades may be disconnected from understanding, they see the disconnect as a failure of a system that should in fact connect grades and understanding. Maia's and Eddie's perspective, remember, is coloured by their strong Improvement orientation (Brown, 2011), which makes them see achievement in school in personal rather than comparative terms.

The result of a strong comparative impulse, though, is that students perceive school to be competitive. Now, competition implies a scarce resource being sought by multiple parties, who cannot all claim it in its entirety. This notion does not actually describe what real learning is about. Concepts to be understood do not get consumed as they are learned, so any number of students can claim them in their entirety if they so choose. And, if grades are assigned according to criteria, and not distributed in a fixed

manner (e.g., bell curve), all students can theoretically earn the maximum possible mark. Thus, in contemporary learning environments with criterion-referenced assessment, competition, per se, is a myth. But half of my interview participants specifically referred to school as competitive.

On the one hand, these students could be referring to situations where outmoded grading schemes are being applied, and teachers are actually applying loosely bell-curvish assessment to their classes. And, likewise, they are certainly referring at least in part to university decision-makers, which, in practice (or, at least, in students' perceptions of practice), offer limited seats to only the highest achievers, thereby instilling from outside a competitive element to achievement in schools. But I think, on the other hand, that at a simple level the comparative impulse is working in students to make the competition an end in itself. That is, they wish to be better than others for the simple psychological expedient of feeling better than others. If grades are perceived to be about ranking and sorting, do students want A's or B's because they reflect excellent or very good understanding? Or do they want A's or B's because they simply show others that you are better than most? Julie's comments seem to imply the latter.

I think it is also possible that some motivation for the comparative impulse comes from outside the student. Julie's comments about university entrance and "success" in life struck me as parroted notions of societal pressure. Adam also shared his personal feelings of parental pressure. Both of these sources may promote in students a focus on grades and achievement for their own sake, and a sense of comparison to others.

The comparative impulse is one of the attitudes that underlies valuing grades for their own sake. It is an attitude about the purpose of school that contrasts quite strongly with the attitudes displayed by students like Maia, and to a lesser extent, Eddie and Kara. Remember that Julie noted a disconnect between grades and understanding and seemed reluctant to even admit that they should be connected, which suggests to me that her sense of purpose in school is quite divorced from understanding for its own sake. Maia, on the other hand, noted the discrepancy between grades and understanding as a flaw in the system. Her ideal, shared to a lesser degree by Eddie and Kara, is that understanding thoroughly is the purpose of learning. This ideal, though it does not preclude the comparative impulse, is nevertheless at odds with it, because it

requires an honest look at oneself – where is my understanding at, where can I take it, and how do I do so?

5.4. Summary

The factors discussed in this chapter help to reveal why some students appear to buy in to progressive assessment more strongly than others. The buy in I was and am looking for as a teacher is really students making decisions to use the assessment information I give them as I have intended it to be used. But these interviews have revealed that both practical and theoretical issues sometimes get in the way of this ideal, that students make what seem to me to be sub-optimal decisions. Whereas I feel that my detailed feedback should be useful and, in fact, always used, to improve understanding and performance, students sometimes do not have the time or do not find my presentation of the information efficient enough to make use of. Moreover, my ideal is based on a theoretical conception of the purpose of learning and a capability for growth that students, frankly, do not always share. And my ideal downplays the significance of grades, which flies in the face of both practical and theoretical realities students seem to hold to.

6. Further Theoretical Considerations about Students' Perceptions of Assessment

The fore-going analysis of where students lie on a receptivity scale, as well as the practical and theoretical constraints that affect students no matter where they are on the scale, points to one fact that became increasingly clear to me as I analyzed my data: students' reactions to assessment involve decisions about what to do with the information fed back to them. The psychological conceptions outlined in chapter two help to explain why students react the way they do, and the constraints outlined in the previous chapter show some external factors that play a part, but in any case the evidence of internal or external factors is the decisions made and the behaviours that result from those decisions. As such, I have found myself intrigued by ideas related to decision theory (Hansson, 2005), because it has become apparent to me that what I mean by "varying student reactions" is really "different decisions students make in light of assessment." Thus, the factors described in the previous chapter serve as elements of a decision making model enacted by students.

6.1. Decision Theory – Bounded Rationality

Throughout my interviews and my review of the results, I became increasingly aware that what I was encountering was a complex set of choices students make moment by moment. It occurred to me that my analysis, in a small, amateurish sort of way, was imitating the kind of work Levitt and Dubner (2005, 2009) do to describe interesting human behaviours in their *Freakonomics* series of books. That is, just as Levitt and Dubner attempt to ascertain the motivations behind strange human behaviour, and do so by analyzing the unusual incentives people respond to, I was touching on a process by which students choose among various possible responses based on their unique views of the incentives involved. Now, of course, my research is far less ambitious than Levitt and Dubner's (2005) stated claim "to explore the hidden side of …

everything" (p. 12), and I have not applied the sophisticated economic tools that Levitt, a professional economist, has at his disposal. But the concept of analyzing student behaviour in a similar theoretical vein appeals to me.

The economic theory behind Levitt's work is a modified version of the fundamental neo-classical economics theory known as the rational utility maximization hypothesis, or RUMH (McCormick, 1997). The RUMH in its basic form is an economic construct that says that consumers are rational beings, with tastes that do not change and of which they are aware, who calculate costs and benefits in order to make decisions that maximize utility. The use of the RUMH involves mathematics, specifically analysis, to identify demand functions that can be optimized using calculus. One of the leading theorists of the RUMH referred to it as "a calculus of pleasure and pain" (Jevons, 1879, quoted in McCormick, 1997, p. 21). This notion became a core concept in economics from the late 1800s through much of the twentieth century, but it was not without its critics (McCormick, 1997; Jones, 1999). Many neo-classical economists have attempted modifications of the theory, to account for the obvious fact that people do not always maximize benefits in the decisions they make. Levitt's particular flavour of the theory is known as "bounded rationality" (Jones, 1999). In bounded rationality, the consumer is acknowledged to make less than purely optimal decisions, but is said to make decisions that are nonetheless "optimized" within a given set of constraints. Thus, Levitt and Dubner's (2005) claim that economics is really all about incentives, and their notion that the work of the economist (and of the social critic, which they have become) is to identify the incentives that are the bounds on the consumer's (or citizen's) decisions.

Now, Levitt is not without his detractors, among both economists and social commentators (Foote & Goetz, 2005; Glenn, 2006), and the RUMH itself is called into question by many (Hodgson, 2012; Jones, 1999), so I do not want to claim that this is a theory that clarifies everything about my students' decisions. But I do feel that, in principle, the idea of rational decision-making, optimizing as it were, helps me make sense of student reactions to assessment. As one introductory text on decision theory puts it, "decision theory is concerned with *goal-directed behaviour in the presence of options*" (Hansson, 2005, p. 6, emphasis in original). My inquiry is into the behaviour students exhibit in response to assessment, attempting to illuminate the explicit or

implicit goals that direct their choices. In this light, grades of course serve as a significant incentive, but even the factors of time constraints and other efficiency concerns would appear, in the parlance of decision theory, as decision variables within the optimization model. Parental and societal expectations, university and scholarship goals, individual factors such as past success – all of these are variables that would affect the model.

Remember that much of the impetus for my research comes from my wish that students would do more with the efforts I was putting in for their benefit. Rational utility theory, even just in a metaphorical sense, helps me see that my notion of their benefit might simply not accord with theirs. That is, what I might describe as the function to be optimized in the learning situation is not likely to be identical to the function students would be using in their decision-making. On the one hand, different models can include the same variables but assign different weights to them, as in Julie's comments about how students put more weight on grades while I "feel very strongly that understanding is important." It is not that I place no weight on grades, or that students disregard understanding entirely, but rather that our version of the optimization function weights these factors unequally. On the other hand, it is also possible for one model to include a variable that a different model excludes. In fact, if you look back at my interview results with both of these possibilities in mind, you will see that some students more than others have a mindset that coincides with my idealization of assessment. I would say in light of bounded rationality that these students would be acting on incentives that more closely align with my view of the appropriate costs and benefits within the classroom.

Even the criticisms aimed at rational utility include ideas that hold some significance for my interview results. One of the criticisms of rational utility maximization in recent years points out that the RUMH does not make much sense of the people who pursue higher rank or status for its own sake, rather than for the utility benefits that come with it, and that in fact sometimes people choose elevated status despite small but measurable material costs associated (Huberman et al., 2004). I think that in schools grades contribute to exactly this phenomenon. As discussed above, grades become for many students an end in themselves, or at least an indicator merely of relative superiority, which is an end in itself.

6.2. Decision Theory – Prospect Theory

The idea of pursuing ends other than mere economic utility is a foundation of the Nobel-prize-winning work of Daniel Kahneman, who pioneered, along with Amos Tversky (2011, 1984, 1974), the ideas of prospect theory that challenged the rational model of judgement and decision making. In this work, Kahneman and Tversky (1974) showed that in fact people do not always act or decide rationally but are instead influenced by intuitive thought processes that result in either heuristic decisions or judgement biases. In heuristic decisions, the agent short cuts a truly rational decision-making method, because the required cognitive demand is too high, and replaces it with an intuitive judgement instead. Our intuitive mind is a "machine for jumping to conclusions" which evolved over time as a necessity for responding to risky situations in a dangerous world (Kahneman, 2011, p. 79). Kahneman (2011) refers to the heuristic process as replacing a hard question with an easy one, and taking the resulting judgement or answer as an answer to the hard question. As an intuitive process, this heuristic decision often goes unnoticed, and we assume in fact that we are rationally answering the hard question to begin with. In terms of my research, I feel that a simple example of such a heuristic judgement at work in students is when they ascribe worth to a school task only if it is "for marks." That is, they are answering the hard question, "Is doing this task of value to me?" by answering instead, "Is this task for marks?" But a more nuanced example is when students decide to try harder in courses they enjoy rather than in courses in which they might benefit more from increased effort. Here the hard question is "Which course should I work harder in?" whereas the easy question is "Which course do I enjoy more?"

Heuristic decision making happens because of systematic biases in our intuitive thought processes. Two of the most prevalent intuitive biases are biases of availability and affect. The availability bias is that we expect an event or outcome to be more likely than it is if it easily comes to mind when we are asked to consider it (Kahneman, 2011, p. 129). This is the bias that makes people believe, for example, that shark attacks are quite likely if there have recently been several prominent news stories about them. The affect bias describes the tendency to judge based on the strength of our emotions (Kahneman, 2011, p. 138). Both of these biases underlie heuristic decisions students

make with regard to assessment. Adam's choice to focus on test feedback as opposed to feedback on writing assignments could be an availability heuristic because of the familiarity of scores versus descriptive feedback. Despite the fact that the descriptive feedback, being descriptive, is meant to be more helpful for adjusting his understanding or quality of work, the score, because its significance in school contexts comes easily to mind, is perceived as more meaningful and even helpful. Jim and Kara's comments displayed some affect bias when they revealed emotional reactions to assessment feedback that did not coincide with the intent or the practical implications of the assessment circumstances.

Another fundamental judgement bias that Kahneman describes is an imbalance in people's reaction to losses versus gains. In another seminal work defining prospect theory, Kahneman and Tversky (1984) showed that, systematically, losses loom larger than gains in how people made decisions in simple gambles. That is, in studies where the very same gamble, in economic terms, is described in one scenario as a gain while in the other as a loss, the subjects routinely showed more aversion to the loss scenario than they showed attraction to the gain scenario. Under rational utility, these two scenarios should be identical, and rational agents would show no distinction in choice between them. Subsequent research shows that losses loom larger than gains in a wide array of real world decision making scenarios also (Kahneman, 2011). When I think about my students in light of this idea, I see that to them the loss of time, or the loss of familiarity, or any other loss they face in progressive assessment may loom larger than the gains in, for example, richer understanding. I think especially of Julie's interview here, which was so puzzling at the time because she seemed to acknowledge only grudgingly that learning should be about understanding. To her, I feel now, a gain in an intangible quality like understanding is simply not as significant as her perceived losses in grades, or university prospects, or even mental effort. Again, it is not so much that students do not value the benefits I have in mind for them, it may rather be that they do not value those benefits as highly as the corresponding losses I am asking them to face.

In fact, this notion that losses loom large applies more broadly in the context of my research, insofar as my entire progressive assessment approach may represent a loss of sorts to students. As I mentioned in the review of literature earlier, the very term progressive implies transition from one set of ideas to another, a paradigm shift. Thus it

represents a breaking of norms that have come to govern the way people perceive and conceive of something. As Kahneman (2011) points out, the establishment and maintenance of norms is the domain of the intuitive side of our minds (p. 71). We are highly sensitive to surprise, which is precisely the experience of something unexpected, something contrary to the norms we have in mind for a given context. As discussed earlier, assessment in mathematics classrooms is a context governed by norms in students' minds, and when I wish to make my assessment progressive, I am disturbing those norms, sometimes deliberately, sometimes unwittingly. Thus my assessment program asks students to face a loss of what is normal, what they have come to expect from mathematics teachers and from mathematics courses. And that loss may loom significantly larger for students than the gains in understanding and appreciation of mathematics that I have in mind for them.

The connection between the concept of didactic contract (Brousseau, 1997) and Kahneman's description of judgement biases in the intuitive mind is clear. The loss I am asking students to face as discussed above is precisely a breach of the didactic contract (Herbst & Chazan, 2012) established by the context of being in a senior academic mathematics classroom. The contract stipulates, for example, that performance in mathematics is gauged by determining the number of correct answers to procedural questions. I ask students to write to explain their thinking, and sometimes do not even tell them whether they are right or wrong when they are performing procedures (instead giving feedback like, "Convince me, or convince yourself. How can you know if you are right?"). The contract stipulates that letter grades and scores are central elements in describing performance in mathematics classrooms, whereas I am deliberately trying to marginalize their use and influence in my course. Recall Adam's poignant comment about whether students are likely to accept some of my progressive ideals:

> *Adam*: Implementing them into teenagers' minds after so many years of grades and percentages would be difficult or tricky for most to accept.

The ideals are "difficult or tricky" to accept because they breach the contract established by "so many years of grades and percentages."

What my ideal of progressive assessment amounts to, then, is redefining the contract in my classroom. As described in chapter 2, the ideal includes an aversion to ranking and sorting, a focus on richer understandings of what it means to do mathematics and to be good at mathematics, and an emphasis on non-evaluative, individualized, descriptive feedback to be used by students for adjusting their conceptions and procedures. The standard contract, it would appear from my students' interviews, allows students to rank and compare themselves, includes an expectation that the teacher does so also, establishes that mathematics is about mastering procedures, and emphasises the use of evaluative and summary feedback frequently and almost exclusively. The varying reactions I see in my students to my progressive ideals represent their varying degrees of comfort with redefining the contract. Certainly the negative reactions can be understood as a reaction to the breach of the didactic contract (Herbst & Chazan, 2012).

Moreover, for each individual, taking on the role of student involves his or her unique interpretation of the contract in place, which is coloured by past experience and the self-concept he or she has formed. For students like Adam, Jim, and Kara, it was strikingly abnormal for them to receive scores and grades lower than an A in mathematics, so much so that Adam did not even want to let his parents know about his standing to avoid "a world of hurt." While it is not strictly against the contract for a teacher to assign low grades in general, for these students it seems to breach their personal contract, meaning both what they expect of the teacher and what they expect of themselves. Their reactions influenced the degree to which they approved of my methods because they could attribute the unexpected results at least in part to my breaching the contract in the first place with progressive methods. Students will also uniquely interpret the didactic contract based on whether they have the growth mindset or fixed ability mindset (Dweck, 2008). As "mindsets" these psychological constructs are grounded in norms and expectations (Brousseau, 1997; Herbst & Chazan, 2012). They operate at the intuitive level, and influence what the student expects of the teacher in an instructional system, but also what the student expects of him- or herself.

6.3. Summary

While chapter 5 outlined some of the practical and theoretical issues of concern to students regarding assessment, this chapter has explored the psychological considerations underlying those issues of concern. My initial consideration was an attempt to explain their behaviour in rational terms, for indeed it does seem at one level that students make conscious choices in response to assessment. But pure rationality does not seem to fully explain their decisions, so I was drawn to the idea of intuitive biases that influence irrational judgements and choices. The notion of intuitive thought, grounded as it is in the mechanism of norms, connects with the concept of the didactic contract, which ties my results back to where I started this investigation, in terms of my personal idealization of assessment.

7. Conclusions

I began this research project in an effort to help me understand why students were not always responding positively to my assessment program, despite the fact that my intention with changing my assessment techniques was to benefit students and their learning directly. Given that I framed this research by the ideal of assessment that I distilled from my reading of assessment literature, any conclusions I draw here are at least in part commentary on that frame. The research might either confirm the frame and provide some insight as to why student behaviour does or does not fit within it, or it might point to some flaw or other in the frame itself based on the perspectives gleaned from the students. I think a little of both kinds of comments can be deduced from the analysis described above.

7.1. My Role in the Interview Process

Before I make any comments about the results of my research, however, I need to acknowledge an interesting component of the research that to some extent colours the conclusions that can be reached. During several of the interviews, and more so after them, it became obvious to me that I was not able to fully retain my objectivity as a researcher in certain lines of questions. Instead, I found myself, as I pointed out earlier while describing the results, assuming the role of teacher, attempting to persuade my student of some point or other. This happened most clearly and strongly with Julie when I tried to get her to acknowledge the connection between grades and understanding, but it also occurred to some extent with Adam and Jim. In the case of Julie, I almost felt I was treating her like a hostile witness in a TV courtroom drama. I clearly had a vested interest in the results of these discussions, and at least subconsciously wanted my students to validate my efforts at progressive assessment.

What this says about my research most clearly, I believe, is that the ideal that was operating in my mind influenced the line of questioning I pursued. I could not remove myself and my idealization of assessment from the context of the interview. The question this raises is what was it I was pushing? Or what was I pushing against? This intrusion of my idealization into the interview does not, in my opinion, preclude drawing any useful conclusions from the research, but it means that I have to factor in the "push" as part of the analysis. In fact, I feel that the presence of the push helped to reveal how strongly held some student conceptions actually were, which is a useful part of the results of the interviews.

In light of the theories outlined in chapter 2, it seems apparent to me now that part of what I was pushing was a justification for redefining the didactic contract. As Herbst and Chazan (2012) say, whenever a breach in the didactic contract is made, there is an expectation of some justification (p. 607). I was pushing just such a justification, and I feel that I was pushing against the students' indignation over the breach in the contract. Furthermore, in light of the theoretical considerations in the previous chapter, I think that I was pushing against a judgement bias – that my students were reacting with their intuitive mind, answering easy questions, while I was trying to push the hard question.

7.2. How Are My Students Receiving Assessment?

If the ideal that I drew from the literature is taken to be the ideal for teachers, what can we say about how students view assessment? As chapter 5 outlined, students' perceptions of assessment are coloured by both practical and theoretical concerns. To students, assessment is just another aspect, in practical terms, of their busy student lives, and perhaps not a central one at that. They have been conditioned to expect certain kinds of assessment feedback, and have developed in many cases a distorted view of the purposes of both learning and assessment, and even of themselves as learners. Based on these perceptions, students often do not react to progressive assessment as the theoretical ideal intends they should.

7.2.1. What Lies Behind Varying Student Reactions?

Basically, my interviews suggest that students will accept the ideal of progressive assessment to the extent they accept a particular view of the nature of learning and of themselves as learners. That is, do they consider understanding at a deep level the goal of learning? And do they consider themselves capable of achieving that goal? Quite simply, an answer of yes to both these questions is really the unspoken contract regarding learning that my ideal presupposes for students. And can I say that these things are a reality for most, even if not all students? Actually, I would have to admit that it seems rather the opposite is true. Most, if not all, students would answer no to one or both of these questions. I had only one in six of my interview participants who seemed close to accepting this unspoken contract.

For students then, a less idealistic mix of goals and self-concepts are the norm. The fact of grading skews most of their attitudes about the purpose of learning and their sense of self-efficacy, while very real practical issues make the ideal difficult or impossible to achieve even if they have the right sense of purpose and belief in themselves.

7.2.2. Why, in Particular, Do Some Students React Negatively to Progressive Assessment Efforts that are For Their Benefit?

The less idealistic mix of goals and conceptions of assessment show that students hold a different set of expectations regarding learning mathematics, so they do not accurately perceive or fully appreciate the benefits proposed by progressive assessment. Instead, they perceive the loss of the familiar, the expectations they have developed over time, as more significant than the gains involved in progressive assessment techniques. The negative reaction some students exhibit is a reaction to the breach of the didactic contract.

The good news is evident in a student like Kara, though, who revealed a marked reversal of opinions regarding my progressive assessment. While she was really worried, by her own admission, at the start of the course, she did come to appreciate some of the most progressive elements of my assessment approach by the end. It is possible to adjust the didactic contract and have students accept it.

7.3. Is the Ideal Truly Ideal?

Though students can adapt to changing expectations, I must also, as mentioned earlier, examine whether the ideal I have adopted might also need adjustment. Clearly the ideal as I have envisioned it is of the hopeless variety, but I remain committed to it as an ideal nonetheless. The idealism in practice needs tempering, for sure, but I do not mind being the one who remains committed to deep understanding for all students. In practice, though, I have numerous ideas for adjusting my approach to try to meet the students more where they are at, including providing more succinct direct feedback in the learning situation, for example. But the fact that I was pushing against something means there is a resistance to the ideal that may over time be worn down.

7.4. Dispelling Myths

Some of the resistance I felt was clearly a result of myths about assessment and learning that persist in the social consciousness. The strongest of these seems to be the notion of competition in the learning environment. As noted, it is a myth that high achievement is a scarce resource that persists from the days of marking on a curve. Another myth, less monstrous but equally prevalent, is that only A's and B's mean success. This myth likely persists as a result of the comparative impulse described above.

My sense is that much work can be done in my own practice personally as well as more broadly within the field of mathematics education to dispel such myths. I for one intend to make the unspoken contract I described above a spoken one with my students in the future. I believe I can also create some activities to do early in a course to develop some assessment literacy and help students better understand the form and function of some of the non-traditional feedback I give them.

I do not believe that my sense of disconnect between my ideals and my students' reception of my efforts will ever entirely go away. But then again, I have been teaching mathematics for over a decade now and I have yet to find the class that universally loves all the topics we address, either. But that does not keep me from addressing the same

topics with them each year, because I know that part of my job is believing in what I am doing for the sake of the students when they do not seem able to.

7.5. Research Considerations

As a teacher, this research project has helped to answer some of the questions I had about why my efforts have been less successful than I had hoped they would be, and has also given me some ideas for adjusting my efforts accordingly. As a researcher, I have come away with new questions and ideas about how I might have investigated these issues differently. The theoretical considerations of chapter 6 open new doors I would like to investigate. I find it interesting to consider the factors influencing student reactions to assessment as variables in an optimization model. Are there other factors to consider? Can they be identified? Can students' weighting of various factors be more directly measured? I am also intrigued by the kind of research Kahneman and Tversky (1974, 1984) did in developing prospect theory, conducting surveys that present subjects with choices between subtly worded alternatives in an effort to identify and quantify biases in thinking. Perhaps their method could be applied to research on student perceptions and attitudes. There are unsolved riddles for me in terms of how students truly conceive of "understanding." My sense is that interesting biases and judgement errors would come to light upon careful examination.

References

Bagley, S. (2010). Students, Teachers and Alternative Assessment in Secondary School: Relational Models Theory (RMT) in the Field of Education. *Australian Educational Researcher, 37(1)*, 83-106.

Bahr, D. L. (2007). Creating mathematics performance assessments that address multiple student levels. *Australian Mathematics Teacher, 63(1),* 33-40.

Black, P., & Wiliam, D. (1998a). Assessment in classroom learning. *Assessment in Education: Principles, Policy and Practice, 5(1),* 7-75.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80,* 139-148.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21,* 5-31.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook 1: The Cognitive Domain*. New York: David McKay Co. Inc.

Boaler, J. (1998).Alternative approaches to teaching, learning and assessing mathematics. *Evaluation and Program Planning, 21,* 129-141.

Boaler, J. (2002a). The Development of Disciplinary Relationships: Knowledge, Practice and Identity in Mathematics Classrooms. *For the Learning of Mathematics*, 22(1), 42-47.

Boaler, J. (2002b). *Experiencing School Mathematics: Traditional and Reform Approaches to Teaching and their Impact on Student Learning*. Lawrence Erlbaum Associates: Mahwah, New Jersey.

Boaler, J., & Staples, M. (2008). Creating Mathematical Futures Through an Equitable Teaching Approach: The Case of Railside School. *Teachers College Record, 110*, 608-645.

Bodin, A. (1993). What does to assess mean? The case of assessing mathematical knowledge. In M. Niss (Ed.), *Investigations into Assessment in Mathematics Education* (pp. 113-141). Boston: Kluwer Academic Publishers.

British Columbia Ministry of Education (2000). *Mathematics 10 to 12 Integrated Resource Package.* Victoria, BC: Ministry of Education.

British Columbia Ministry of Education (2009). *Reporting Student Progress: Policy and Practice.* Victoria, BC: Ministry of Education.

Brousseau, G. (1997). *Theory of didactical situations in mathematics: Didactique des mathématiques 1970-1990.* (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Eds. and Trans.). Dordrecht, The Netherlands: Kluwer.

Brown, G. (2011). Self-regulation of assessment beliefs and attitudes: a review of the Students' Conceptions of Assessment Inventory. *Educational Psychology, 31,* 731-748.

Buhagiar, M. A. (2007). Classroom assessment within the alternative assessment paradigm: revisiting the territory. *The Curriculum Journal, 18(1),* 39-56.

Davis, G., & Gadanidis, G. (1999). *Addison-Wesley Mathematics 12 (Western Canadian Edition) Teacher's Resource Book*. Toronto: Addison-Wesley.

Dweck, C. S. (1999). *Self Theories: Their Role in Motivation, Personality, and Development* . Hove: Psychology Press, Taylor and Francis Group.

Dweck, C. S. (2008). *Mindset: The New Psychology for Success*. New York: Ballantine.

Foote, C., & Goetz, C. (2005). Testing economic hypotheses with state-level data: A comment on Donohue and Levitt (2001). Federal Reserve Bank of Boston Working Paper 05-15, November 22, 2005.

Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics, 18,* 402-423.

Galbraith, P. (1993). Paradigms, problems, and assessment: Some ideological implications. In M. Niss (Ed.), *Investigations into Assessment in Mathematics Education* (pp. 73-86). Boston: Kluwer Academic Publishers.

Ginsburg, H. P. (2009). The challenge of formative assessment in mathematics education: Children's minds, teachers' minds. *Human Development, 52,* 109-128.

Ginsburg, H.P., Jacobs, S. F., & Lopez, L. S. (1993). Assessing mathematical thinking and learning potential in primary grade children. In M. Niss (Ed.), *Investigations into Assessment in Mathematics Education* (pp. 157-167). Boston: Kluwer Academic Publishers.

Gipps, C. V. (1994). *Beyond Testing: Towards a Theory of Educational Assessment.* London: The Falmer Press.

Glenn, D. (2006). A Reputation Under the Gun. *Chronicle Of Higher Education*, *52*(34), A20-A23.

Hansson, S. O. (2005). *Decision Theory: A Brief Introduction*. Stockholm: Royal Institute of Technology (KTH).

Harkness, S. S. (2008). Social constructivism and the believing game: A mathematics teacher's practice and its implications. *Educational Studies in Mathematics, 70,* 243-258.

Herbst, P., & Chazan, D. (2012). On the instructional triangle and sources of justification for actions in mathematics teaching. *ZDM, 44,* 601-612.

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28(3),* 24-31.

Hewitt, D. (2004). Feedback. In E. Simmt & B. Davis (Eds.), *Proceedings of the 2004 Annual Meeting of the Canadian Mathematics Education Study Group* (pp. 105-110). Edmonton, AB: CMESG/GCEDM.

Hewitt, D. (2010). Feedback: Expanding a repertoire and making choices. In Leikin, R. & Zazkis, R. (Eds.), *Learning Through Teaching Mathematics* (pp.263-285). New York: Springer.

Hiebert, J., & Lefevre, P. (1987). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: The Case of Mathematics* (pp. 1-27). Hillsdale, NJ: Erlbaum.

Hodgson, G. M. (2012). On the limits of rational choice theory. *Economic Thought, 1,* 94-108.

Huberman, B. A., Loch, C. H., & Önçüler, A. (2004). Status as a valued resource. *Social Psychology Quarterly, 67(1),* 103-114.

Jones, B. D. (1999). Bounded rationality. *Annual Review of Political Science, 2*, 297-321.

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York: Farrar, Strauss and Giroux.

Kahneman, D., & Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist, 39,* 341-350.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions.* 2nd ed. Chicago: University of Chicago Press.

Levitt, S. D., & Dubner, S. J. (2005). *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything.* New York: William Morrow.

Levitt, S. D., & Dubner, S. J. (2009). *Superfreakonomics: Global Cooling, Patriotic Prostitutes, and Why Suicide Bombers Should Buy Life Insurance.* New York: William Morrow.

McTighe, J., & Wiggins, G. (2001). *Understanding by Design*. Alexandria, VA: Merrill/Prentice Hall.

McCormick, K. (1997). An essay on the origin of the rational utility maximization hypothesis and a suggested modification. *Eastern Economic Journal, 23(1),* 17-30.

Mellin-Olsen, S. (1993). A critical view of assessment in mathematics education: Where is the student as subject? In M. Niss (Ed.), *Investigations into Assessment in Mathematics Education* (pp. 143-156). Boston: Kluwer Academic Publishers.

Mulgrew, A., & Rawe, E. (2001). *Smerging Data: More Than Just Number Crunching*. Edmonton, AB: Alberta Assessment Consortium.

O'Connor, K. (2002). *How to Grade for Learning: Linking Grades to Standards* (2nd Edition). Glenview, IL: Skylight.

O'Connor, K. (2007). *A Repair Kit for Grading: 15 Fixes for Broken Grades*. Portland, OR: ETS/ATI.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks: Sage.

Pegg, J. (2003). Assessment in mathematics: A developmental approach. In J. Royer (Ed.), *Mathematical Cognition* (pp. 227-259). Greenwich, CT: Information Age Publishing.

Piaget, J. (1970). *Science of Education and the Psychology of the Child* (D. Coltman, Trans.). New York: Orion Press.

Popham, W. J. (2008). *Transformative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Raymond, A. M. (1996). The development of preservice elementary teachers' beliefs about and knowledge of alternative mathematics assessment. Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Panama City, Florida. Retrieved from ERIC.

Ridgway, J. (1998). From barrier to lever: Revising roles for assessment in mathematics education. *NISE Brief, 2(1),* 1-9.

Romagnano, L. (2001). The myth of objectivity in mathematics assessment. *Mathematics Teacher, 94,* 31-37.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144.

Schiralli, M., & Sinclair, N. (2003). A constructive response to 'Where Mathematics Comes From.' *Educational Studies in Mathematics, 52,* 79-91.

Shepard, L. A. (2005). Formative assessment: Caveat emptor. Paper presented at ETS Invitational Conference, The Future of Assessment: Shaping Teaching and Learning, New York. Retrieved from ERIC.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*. Reprinted in *Mathematics Teaching in the Middle School, 12(2), 2006,* 88-95.

Smith, J., Herbel-Eisenmann, B., Breaux, G., Burdell, G., Jansen, A., Lewis, G., & Star, J.R. (2000). How do students adjust to fundamental changes in mathematics curricula? In M. Fernandez (Ed.), *Proceedings of the twenty-second annual meeting of the North American chapter of the International Group for the Psychology of Mathematics Education* (pp. 127-132). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Spade, P., & Panaccio, C. (2011). William of Ockham. In Zalta, E.N. (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*. Retrieved from http://plato.stanford.edu/archives/fall2011/entries/ockham/.

Star, J.R., & Hoffmann, A.J. (2002). Assessing students' conceptions of reform mathematics. Paper presented at the twenty-fourth annual meeting of the North American chapter of the International Group for the Psychology of Mathematics Education (PME-NA), Athens, Georgia.

Star, J.R., & Hoffman, A. J. (2005). Assessing the Impact of Standards-Based Curricula: Investigating Students' Epistemological Conceptions of Mathematics. *The Mathematics Educator, 15(2)*, 25-34.

Stiggins, R. (1997). *Student-Centered Classroom Assessment* (2nd Edition). Old Tappan, NJ: Prentice-Hall.

Strickland, K., & Strickland, J. (1998). *Reflections on Assessment: Its Purposes, Methods, and Effects on Learning.* Portsmouth, NH: Boynton/Cook Publishers.

Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education, 9,* 172-192.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Western and Northern Canadian Protocol for Collaboration in Education (WNCP). (2006). *Rethinking Assessment with Purpose in Mind: Assessment for Learning, Assessment as Learning, Assessment of Learning*. Manitoba Education, Citizenship, and Youth. Retrieved from http://www.wncp.ca/

Wiliam, D. (2007). What does research say the benefits of formative assessment are? *NCTM Research Brief, June 2007,* 1-3. Retrieved from http://www.nctm.org/

Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education, 34,* 254-284.