



uOttawa

L'Université canadienne
Canada's university

**FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES**



**FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES**

Ann Evans-Arden

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A. (Education)

GRADE / DEGRÉ

Faculty of Education

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

**Negotiating Tensions and Dilemmas in Classroom Assessment:
Case Studies of Two Secondary School Mathematics Teachers**

TITRE DE LA THÈSE / TITLE OF THESIS

Christine Suurtamm

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Ruth Kane

Marielle Simon

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

**Negotiating Tensions and Dilemmas in Classroom Assessment:
Case Studies of Two Secondary School Mathematics Teachers**

Ann Arden

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the Master of Arts in Education degree

June 11, 2010

Faculty of Education
University of Ottawa

Thesis Supervisor: Dr. Christine Suurtamm

Committee Members:
Dr. Ruth Kane
Dr. Marielle Simon

©Ann Arden, Ottawa, Canada, 2010



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-73751-4
Our file *Notre référence*
ISBN: 978-0-494-73751-4

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■+■
Canada

ACKNOWLEDGEMENTS

I am grateful for the support and encouragement of my thesis supervisor, Dr. Chris Suurtamm. I appreciate her time, advice and enthusiasm for my work. She has generously offered me many opportunities to participate in the mathematics education community by allowing me to work as an assistant on some of her research projects, inviting me to attend conferences and meetings, writing several papers with her in academic and professional journals, and presenting alongside her at academic conferences. These experiences have motivated me to continue on this path of inquiry, both in research and in my own classroom practice. Dr. Suurtamm's passion for mathematics education and her dedication to scholarship serve as an inspiration to me.

I would also like to thank the members of my committee for their time and feedback. Dr. Ruth Kane and Dr. Marielle Simon offered thought provoking and helpful suggestions for my writing. In addition, Dr. Jérôme Proulx read early drafts of my work. In the first semester of my course work, Dr. Janice Ahola-Sidaway encouraged me to pursue this thesis journey and for that I am very grateful.

Thank you to the faculty members of the Mathematics Education Research Unit (MERU) at the University of Ottawa, particularly Dr. Chris Suurtamm, Dr. Barbara Graves and Dr. Richard Barwell, for providing such a rich learning environment for graduate students. I have learned a great deal from the MERU reading groups, workshops, and colloquia. I would also like to thank several of my graduate student colleagues including Arlene Corrigan, Jennifer Hall, Tom Hillman and Adrian Jones, with whom I shared courses and tried out ideas. In particular, Martha Koch has been a wonderful mentor to me as we worked closely together on several projects and share a deep interest in educational assessment.

Special thanks to my family, including my parents, Helen & John Evans, and my brother, John, for instilling in me a love of learning and the confidence to pursue my goals. As well, I simply could not have completed this project without the immeasurable hours of loving childcare they provided. Thanks as well to Genevieve Arden and the rest of the Arden family for their support and encouragement. Thanks especially to my husband, Bill Arden, for his unwavering support and patience, and to my children, Jamie and Caitlin, who bring me great joy and perspective.

I would like to thank both Principals I have worked with (as a secondary mathematics teacher myself) during the time of this project, Patsy Agard and Brett Reynolds. They were both been incredibly supportive of my participation in academic and professional activities which occasionally took me away from my own classroom.

Finally, I would like to thank the two teachers who participated in this project. Without them this research would not have been possible.

ABSTRACT

This qualitative study describes tensions and dilemmas that two secondary mathematics teachers experience in their classroom assessment and how they negotiate these tensions and dilemmas in their practice as they try to implement current reforms. The focus on tensions and dilemmas gives particular insight into the complexities of “doing” classroom assessment in a manner that is aligned with the current reforms and is helpful for identifying areas that need further work in both the research and teaching communities in order for these reforms to be more widely implemented. The tensions and dilemmas that the two participants described included issues in grading and reporting, colleagues resisting current reforms in classroom assessment, difficulties interpreting and implementing policies from different levels, students’ poor work habits and study skills and issues in the design and use of performance tasks. These issues were negotiated through discussions and collaboration with other colleagues that support assessment reforms, a general feeling that assessment reforms support and are important to improving students’ learning, and a great deal of personal reflection and constant development of assessment practice.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
ABSTRACT.....	iv
LIST OF TABLES AND FIGURES	viii
CHAPTER 1: INTRODUCTION.....	1
PURPOSE OF THE STUDY	3
RESEARCH QUESTIONS	4
SIGNIFICANCE OF THE STUDY	4
THESIS OVERVIEW	5
CHAPTER 2: REVIEW OF THE LITERATURE	7
EVOLVING PERSPECTIVES ON CURRICULUM, LEARNING AND ASSESSMENT	7
Past Perspectives on Curriculum, Learning and Assessment	8
Current Perspectives on Curriculum, Learning and Assessment.....	9
REFORM-ORIENTED CLASSROOM ASSESSMENT.....	12
Purposes and Uses of Classroom Assessment	12
Assessment for Learning (Formative Assessment)	13
Integration of Assessment and Instruction.....	13
Using a Broader Range of Tools and Strategies	14
Using Assessment to Improve Students' Learning.....	15
Reform-Oriented Views on Grading and Evaluation	15
CHALLENGES IN IMPLEMENTING REFORM-ORIENTED CLASSROOM ASSESSMENT	16
Making Sense of Assessment Reforms.....	17
Assessment Practices are Tied to Teachers' Beliefs.....	19
Assessment Serves Multiple Purposes.....	20
Using New Forms of Assessment.....	20
Complexity of Giving Feedback.....	22
Issues in Grading and Reporting.....	23
SUMMARY.....	25
CHAPTER 3: CONCEPTUAL FRAMEWORK.....	26
TENSIONS AND DILEMMAS IN TEACHERS' PRACTICE	26
FRAMEWORK FOR TENSIONS AND DILEMMAS IN CLASSROOM ASSESSMENT PRACTICE.....	29
Teachers' Assessment Decision Making	30
Tensions between Internal and External Influences on Assessment Practices	31
Dilemmas in Classroom Assessment.....	33
Negotiating Tensions and Dilemmas in Practice.....	33
Paradigm Shift	34
SUMMARY.....	34
CHAPTER 4: RESEARCH DESIGN.....	36
OVERVIEW OF THE RESEARCH	36
RATIONALE FOR A QUALITATIVE MULTIPLE CASE STUDY.....	37
MY ROLE AS A RESEARCHER	38
INVITING PARTICIPATION.....	41
PARTICIPANTS.....	43

DATA COLLECTION AND INSTRUMENTS	44
First Interview	47
Classroom Observations and Debriefing Interviews	49
Final Interview	51
DATA CONSTRUCTION, ANALYSIS AND INTERPRETATION.....	52
CREDIBILITY AND VERIFICATION	59
CHAPTER 5: THE CASE STUDIES	61
CURRICULUM AND ASSESSMENT IN THE CONTEXT OF ONTARIO.....	61
Revisiting Ontario's Assessment Policy.....	65
CASE STUDY #1: AMY	67
Description of the School and Teacher's Background	67
View of Classroom Assessment	69
Description of Classroom Assessment Practices	71
Assessment for Learning (Formative Assessment)	72
Assessment of Learning (Summative Assessment)	76
Tensions and Dilemmas in Classroom Assessment Practice.....	81
Policies Related to Late and Missed Assignments	81
Issues in Grading and Reporting.....	84
Colleagues Resisting Current Reforms in Classroom Assessment.....	89
Issues in Using Assessment Effectively to Improve Learning	90
Negotiating these Tensions and Dilemmas in Practice.....	93
Policies Related to Late and Missed Assignments	93
Issues in Grading and Reporting.....	94
Colleagues Resisting Current Reforms in Classroom Assessment.....	96
Issues in Using Assessment Effectively to Improve Learning	97
Summary	98
CASE STUDY #2: CARINA.....	101
Description of the School and Teacher's Background	102
View of Classroom Assessment	104
Description of Classroom Assessment Practices	106
Assessment for Learning (Formative Assessment)	106
Assessment of Learning (Summative Assessment)	113
Tensions and Dilemmas in Classroom Assessment Practice.....	120
Students' Poor Work Habits and Study Skills.....	120
Issues in the Design and Use of Performance Tasks	124
Issues in Collecting, Recording and Analyzing Assessment Data	127
Colleagues Resisting Current Reforms in Classroom Assessment.....	130
Negotiating Tensions and Dilemmas in Practice.....	133
Students' Poor Work Habits and Study Skills.....	133
Issues in the Design and Use of Performance Tasks	137
Issues in Collecting, Recording and Analyzing Assessment Data	139
Colleagues Resisting Current Reforms in Classroom Assessment.....	142
Summary	143

CHAPTER 6: DISCUSSION OF FINDINGS.....	146
ADDRESSING RESEARCH QUESTION #1: WHAT TENSIONS AND DILEMMAS DO TWO SECONDARY MATHEMATICS TEACHERS EXPERIENCE AS THEY TRY TO IMPLEMENT CURRENT ASSESSMENT REFORMS?.....	146
Assessment Decision Making	147
Tensions Between Internal and External Influences on Classroom Assessment Practices	150
External Influence #1: Institutional Presses on Classroom Assessment Practice.....	150
External Influence #2: Stakeholder Presses on Classroom Assessment Practice.....	154
Internal Presses on Classroom Assessment Practice	157
Dilemmas in Classroom Assessment	159
Summary of Tensions and Dilemmas in Classroom Assessment.....	162
ADDRESSING RESEARCH QUESTION #2: HOW ARE THESE TENSIONS AND DILEMMAS NEGOTIATED IN CLASSROOM PRACTICE?	163
Summary of Negotiating Tensions and Dilemmas in Practice	168
KEY FINDINGS	168
Assertion #1: Determining Grades for Reporting Purposes is a Major Source of Tensions and Dilemmas	169
Assertion #2: Teacher Community is a “Double-Edged Sword” in the Implementation of Assessment Reforms.....	172
Assertion #3: These Teachers are Caught in a Paradigm Shift.....	176
Summary	177
CHAPTER 7: CONCLUSIONS	179
CONTRIBUTIONS AND IMPLICATIONS OF THIS STUDY	180
LIMITATIONS OF THIS STUDY	182
DIRECTIONS FOR FUTURE RESEARCH.....	183
CONCLUDING THOUGHTS.....	185
REFERENCES.....	186
APPENDICES	201
APPENDIX A: INTERVIEW GUIDE FOR INTRODUCTORY MEETING.....	202
APPENDIX B: INTERVIEW GUIDE FOR FIRST INTERVIEW	203
APPENDIX C: CLASSROOM OBSERVATION PROTOCOL	204
APPENDIX D: INTERVIEW GUIDE FOR POST-OBSERVATION DEBRIEFINGS	205
APPENDIX E: INTERVIEW GUIDE FOR FINAL INTERVIEW (AMY)	206
APPENDIX F: INTERVIEW GUIDE FOR FINAL INTERVIEW (CARINA).....	207
APPENDIX G: CARINA’S “ASSESSMENT CHECKLIST” – STUDENT VERSION	208
APPENDIX H: CARINA’S “ASSESSMENT CHECKLIST” – TEACHER VERSION	209
APPENDIX I: EQAO ITEMS USED IN AMY’S CLASS	210
APPENDIX J: SAMPLE OF WORKSHEETS FROM AMY’S CLASS	211
APPENDIX K: EDITING QUOTES FROM TRANSCRIPTS	213

LIST OF TABLES AND FIGURES

TABLES

Table 4.1: Summary of background information about each participant	43
Table 4.2: Overview of data	45
Table 5.1: Summary of tensions and dilemmas Amy has experienced in her classroom assessment practice and how she negotiates them	101
Table 5.2: Summary of tensions and dilemmas Carina has experienced in her classroom assessment practice and how she negotiates them	145

FIGURES

Figure 2.1: Historical overview of curriculum, learning and assessment	8
Figure 3.1: Framework for examining tensions & dilemmas in classroom assessment practice	30

CHAPTER 1: INTRODUCTION

“the teacher is in the difficult position of doing a cognitively and conceptually complex job while hiding the complexity and uncertainty – making it look easy to teach” (Clark & Lampert, 1986, p. 28)

This research is a multiple case study (Stake, 2006; Yin, 2003) of two secondary school mathematics teachers’ experiences with classroom assessment in the context of contemporary educational reforms in Ontario, Canada. These reforms in mathematics education are an international effort that reflects the belief that it is important for all students to develop deep mathematical understandings through the investigation of rich and significant mathematics, rather than to develop mere procedural fluency through repeated practice of decontextualized skills (Lampert, 1990; National Council of Teachers of Mathematics [NCTM], 2000). Accompanying assessment reforms reflect a growing call to use evidence of students’ understandings gathered throughout the learning process to improve students’ learning, rather than just to measure students’ achievement at the end of a learning period (Black & Wiliam, 1998a; Gipps, 1994; Shepard, 2001; Stiggins, 2002; Wiliam, 2007).

In this study I am using *assessment* as defined in the NCTM *Assessment Standards* document (1995) as “the process of gathering evidence about a student’s knowledge of, ability to use, and disposition toward, mathematics and of making inferences from that evidence for a variety of purposes” (p. 3). Characteristics of classroom assessment practice aligned with contemporary assessment reforms include the seamless integration of assessment and instruction, use of a broader range of tools and strategies, and recognition that the main purpose of assessment is to advance students’ learning (Even, 2005). While the call for change is clear in the literature, implementation of these reforms presents a considerable challenge, particularly in the area of classroom assessment (Barnes, Clarke &

Stephens, 2000; Delandshere, 2001; Lock & Munby, 2000; Rousseau, 2004; Shepard, 2001). This study describes tensions and dilemmas that two secondary mathematics teachers experience in their classroom assessment practice as they implement current assessment reforms, and how they negotiate these tensions and dilemmas in practice.

My interest in the area of classroom assessment stems from my experiences as a secondary mathematics teacher during this period of curriculum and assessment reform in Ontario. The provincial mathematics curriculum¹ was revised for the elementary grades (1-8) in 1997 and for the secondary grades (9-12) in 1999, and again between 2005 and 2007, to reflect current thinking in mathematics education (Levin, 2009; Suurtamm & Graves, 2007). While provincial guidelines for assessment have not changed significantly since 1999, there has recently been renewed attention to assessment reform, concurrent with several new publications from the Ontario Ministry of Education [OME]. In 2005, the OME moved provincial directives on assessment from a separate *Program Planning and Assessment* document (OME, 2000) directly into the subject-specific curriculum documents, including in mathematics. This made the directives more visible to teachers. In 2008, the OME released a discussion document called *Growing Success* to address concerns from various stakeholders regarding assessment, and to provide explanation and clarification in several areas. In this document, the OME acknowledges that the implementation of provincial assessment guidelines has been “inconsistent and uneven” (p. 1), that policy statements “need to be unpacked so that they can come alive in the classroom” (p. 2-i) and includes a list of 45 outstanding issues that have “frequently arisen” (p. 10-i) regarding assessment. These experiences are aligned with research in other

¹ In Ontario the word *curriculum* is used to refer to a set of province-wide expectations for each course contained in policy documents that are publicly available. I use the term *curriculum supports* to refer to textbooks and other teaching/learning materials

jurisdictions that suggests making substantive changes in classroom assessment practice is very difficult (Barnes, Clarke & Stephens, 2000; Briscoe, 1993; Even, 2005).

Many of the issues raised in the *Growing Success* discussion document (OME, 2008) reflect the challenges and uncertainty that I have felt as a secondary mathematics teacher during this period of change. In fact, my initial reason for pursuing graduate studies was to gain a deeper understanding of the origins of these reforms in order to develop my own teaching and assessment practice. Even though I attended and led many professional learning opportunities at the school, district, and provincial levels and served as a mathematics department head for two years, I was still unsatisfied with my own understandings and practice of classroom assessment. The tensions and dilemmas I have encountered in my own classroom assessment practice motivated me to explore the experiences and perspectives of other teachers.

Purpose of the Study

The purpose of this study is to describe the nature and content of tensions and dilemmas that teachers experience in their daily classroom assessment practice as they implement current reforms in teaching and learning, and how they negotiate these tensions and dilemmas. The notion of *teaching dilemmas* (Adler, 1995; Ball, 1993; Berlak & Berlak, 1981; Enydey, Goldberg & Welsh, 2006; Lampert, 1985; Talanquer, Tomanek & Novodvorsky, 2007; Windschitl, 2002) has been used in the literature to describe situations in which teachers face multiple equally viable yet often unattractive alternatives that leave teachers feeling conflicted. Given that teachers' learning, thinking and change are difficult phenomena to study, Tillema and Kremer-Hayon (2005) argue that dilemmas constitute a "powerful conceptualisation" (p. 203) to connect teachers' professional thinking and practice. Since tensions and dilemmas represent core teaching problems and can reveal

important aspects of teachers' thinking about their work, they can be a productive way to investigate teaching practice.

Research Questions

The design of this study was guided by the following two research questions:

1. What tensions and dilemmas do two secondary mathematics teachers experience as they try to implement current assessment reforms?
2. How are these tensions and dilemmas negotiated in classroom practice?

These questions are addressed through case studies of two secondary mathematics teachers in Ontario. Data included individual interviews, classroom observations followed by debriefing interviews and classroom artefacts that the participants shared with me.

Significance of the Study

Despite the central place of assessment in current educational reforms, there have been few studies that describe the experiences of secondary school teachers who are trying to implement classroom assessment practices aligned with current thinking about mathematics teaching and learning. The literature on classroom assessment contains many calls for reform in teachers' practice (e.g., Black & Wiliam, 1998a, 1998b; O'Connor, 2007a; Stiggins, 2004; Wiggins & McTighe, 2005), but there are relatively few studies that describe teachers' experiences with the challenges of implementation. As Gordon Brown (2008) points out, "There is much written about how assessment ought to be understood and used, but much less about how people actually use it" (p. viii). Most of the empirical literature on classroom assessment focuses on elementary and middle schools, which often have different issues than high schools (Senk, Beckmann, & Thompson, 1997).

This research addresses calls in the literature for qualitative studies of assessment at the classroom level (Lock & Munby, 2000; Pellegrino & Goldman, 2008) and for research

efforts that are “embedded in the dilemmas of practice” (Shepard, 2000, p. 13). There is particular interest in case studies of teachers who are trying to transform their practice (Windschitl, 2002). Within the professional teaching community there is also great interest in this topic: for example, the website for *Mathematics Teacher*, the NCTM journal for secondary school teachers, currently lists assessment as one of a few topics that “are important to readers but are rarely discussed in submitted manuscripts” (NCTM, n. d.). Case studies of teachers’ assessment tensions and dilemmas, with attention to important contextual factors, can stimulate important conversations about classroom assessment within both the teaching and research communities.

The participants in this study are different from many of those represented in other studies in two important ways. First, both are experienced and highly qualified secondary mathematics teachers who support and are trying to implement a vision for curriculum and assessment aligned with current reforms. Also, the jurisdiction in which they teach, Ontario, has had reform-oriented province-wide mathematics curriculum and assessment policies for the last decade.

Thesis Overview

This thesis is organized into seven chapters as follows. Chapter 1 provides an overview of the study along with the research questions and the significance of the study. Chapter 2 contains a review of both theoretical and empirical literature on classroom assessment, including an overview of evolving theoretical perspectives on assessment, a description of reform-oriented assessment, and a discussion of some of the challenges in implementing these reforms. A conceptual framework for tensions and dilemmas in classroom assessment practice is presented in Chapter 3. The research design is outlined in Chapter 4 including an overview of the process of inviting participation, data collection and

instruments, data construction, analysis and interpretation, and consideration of credibility and verification. I also present a rationale for pursuing the research questions through a qualitative multiple case study and discuss how I view my role as a researcher. Chapter 5 contains descriptions of the two case studies, beginning with an overview of classroom assessment in the context of Ontario. Each case study provides contextual information about each school and teacher, their views on classroom assessment and a description of classroom assessment practices. There is then a description of the tensions and dilemmas that each teacher experiences in her classroom assessment practice and how these are negotiated. Chapter 6 includes a discussion of the findings from this study across both cases. In addition to directly addressing the research questions, I put forward three assertions based on the case studies. Chapter 7 provides concluding comments, describes the implications and limitations of the study, and makes suggestions for further research.

CHAPTER 2: REVIEW OF THE LITERATURE

This chapter provides an overview of the literature on classroom assessment. In this chapter, I outline past and current perspectives on curriculum, learning and assessment in order to provide a rationale for current calls for reform. I then describe features of reform-oriented classroom assessment and some of the challenges that are emerging as teachers try to implement these reforms.

Evolving Perspectives on Curriculum, Learning and Assessment

American researcher Lorrie Shepard (2000, 2001) has put forth an influential framework of learning, curriculum and assessment to illustrate how current educational reforms are different from traditional practice. This framework describes traditional classroom practice as situated in what Shepard calls the *dominant twentieth century paradigm*, based in behaviourism, social efficiency and scientific measurement. She refers to current reforms in the twenty-first century as being situated in an *emergent paradigm* based in social constructivism, a reformed vision of curriculum and classroom assessment. Shepard argues that some of the difficulties that are currently being experienced by teachers arise due to the concurrent use of instructional practices that are based in the emergent paradigm and assessment practices based in the old paradigm. A diagram of this framework is included as Figure 2.1.

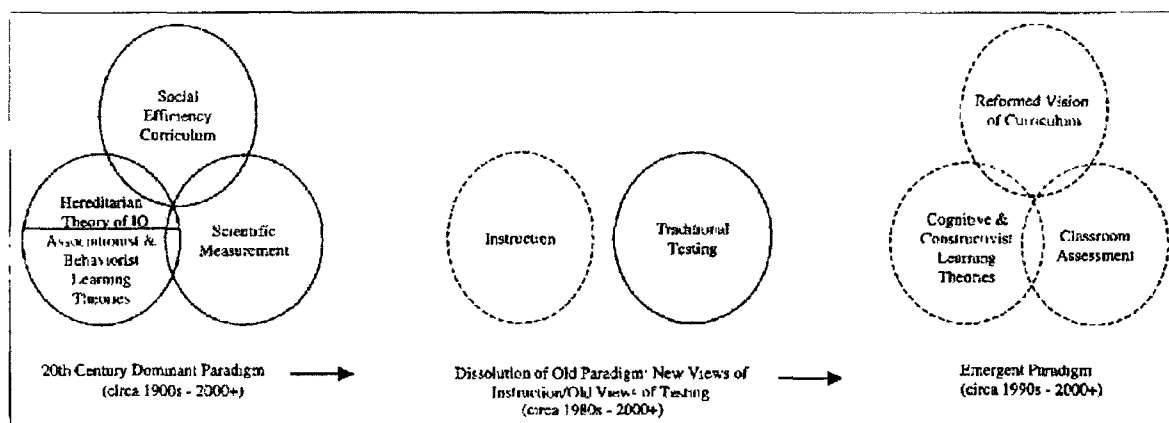


Figure 2.1: Historical overview of curriculum, learning and assessment (from Shepard, 2000, p. 5)

The next two sections provide an outline of these paradigms.

Past Perspectives on Curriculum, Learning and Assessment

The *dominant twentieth century paradigm* was based in an epistemology of objectivism, which views knowledge as unchanging. In this view, there is “one truth” or “real world” that exists and can be uncovered through the methods of scientific inquiry (Windschitl, 2002). Intelligence is viewed as hereditary and as an innate and fixed trait that can be precisely measured. In schools, this led to IQ tests being used to sort students by ability and academic content was reserved only for the elite. The curriculum is viewed as a distinct body of information that can be transmitted to learners and was differentiated for students based on their predicted roles in society. This paradigm is based in associationist and behaviourist learning theories which call for the explicit teaching of small pieces of knowledge in a linear and sequential manner, with frequent testing of small chunks of information (Shepard, 2001). An underlying assumption of this view is that skills are learned in components that are best practiced in isolation for mastery, and that complex skills can only follow the acquisition of basic skills (Gipps, 1994). Teachers and textbooks

are viewed as the sole sources of authority in the class and lectures and demonstrations are frequently used to deliver material to students. There is an emphasis on the “right answers” and “right ways” to solve problems (Windschitl).

The role of assessment in this paradigm is to measure if the learners received the information that was presented. Dominant forms of assessment include “objective” tests, exams and quizzes where students are primarily required to demonstrate low-level recall of facts and application of procedures and algorithms, which are easy to quantify (Shepard, 2001). These tests align with behaviourist learning theories in their focus on small pieces of information that closely match the presentation style during instruction. Assessments tend to be uniformly administered and there is an emphasis on the individual rather than group or whole class cognition (Gipps, 1994). Grading is viewed as a process of scientific measurement, although there is evidence that teachers’ assessments are not aligned with best practices in measurement theory (McMillan, 2003). Gipps argues that the psychometric principles informing this paradigm are still evident in many classrooms through such practices as the emphasis on ranking students, rather than on achievement of curriculum standards. In mathematics this has meant that assessment has frequently been used as a filter to “select students out of the opportunities to learn” (NCTM, 1995, p. 83), rather than to improve students’ learning.

Current Perspectives on Curriculum, Learning and Assessment

Over the past two decades, this view of curriculum, learning and assessment has been increasingly challenged. Technological advances have made vast amounts of information widely available through new methods and at increasingly higher speeds of communication, and these have led to fundamental social and economic changes. Gipps (1994) argues that these changes have created the need to reconsider the basics of schooling

to focus on thinking, reasoning and learning and ensure that students become flexible, adaptive learners, rather than merely acquire information and procedures. In addition, new theories of learning indicate that the complex skills of thinking and reasoning should not wait until basic skills have been mastered (Gipps). Rather, the use of rich, authentic problems can motivate students to understand the need for skills. Epistemological assumptions based in objectivism have been challenged by postmodernism, where truth and knowledge are viewed as value-laden and socially constructed rather than static and unambiguous. This has resulted in reconsideration of what it means to know in each of the disciplines (Shepard, 2001; Windschitl, 2002).

These conditions have resulted in an *emergent paradigm* for curriculum, learning and assessment based in social constructivism² where learning is viewed as occurring through an active process of sense making through socially mediated experiences (Shepard, 2001). Shepard characterizes this paradigm as emergent because “it is not fully developed theoretically and, surely, not adopted in practice” (p. 1073). Within this paradigm, intellectual ability is viewed as socially and culturally developed rather than a fixed, hereditary trait, so there is strong emphasis on giving all students high quality opportunities to learn. The focus is on developing principled or “robust” understandings (Shepard) rather than acquiring isolated facts and skills and there is a shift away from the notion of teaching as something that is done *to* students towards a more collaborative relationship between teachers and students (Geijsel & Meijers, 2005).

² There has been widespread debate over whether learning is essentially an individual or a cultural process (Cobb, 1994, Lerman, 2000, Sfard, 1998, Steffe & Thompson, 2000). In Shepard’s use of the term *social constructivism* she borrows from cognitive, constructive, and sociocultural theories, although she acknowledges that “these camps are sometimes warring with each other” (2000, p 6). Stobart (2008) supports this use and argues that a social constructivist position “seeks to hold in balance learning as a cultural activity and as individual meaning-making. In this, learning is less about idiosyncratic personal interpretations than about personal adaptation of socially created knowledge and meanings” (p 152). He does acknowledge that recent shifts in conceptions of learning and assessment tend toward a more sociocultural emphasis.

Over the past two decades, there have been efforts to articulate the important learning goals, thinking processes and methods of inquiry in many of the school subject disciplines. In mathematics, the NCTM, a large North American mathematics education organization, has made significant contributions to the development of a “reform” view of mathematics (1989, 1991, 1995, 2000) which have significantly influenced the recent curriculum and assessment reforms in Ontario. This vision aims for all students to engage actively and confidently in complex mathematical tasks, and to communicate their ideas and solutions effectively, both orally and in writing. There is a shift away from a focus on procedural proficiency towards the investigation of important topics in mathematics through a variety of rich and challenging problems. Problem solving occupies a central place in the teaching and learning of mathematics with students and teachers engaged in extended investigations of complex problems (Boston & Smith, 2009; Ernest, 1991). In this view, the “dynamic and exploratory” (Henningsen & Stein, 1997, p. 524) nature of mathematics is emphasized through the use of technology, concrete materials and other mathematical tools to investigate problems. There is a new focus on the processes of mathematics that are fundamental to problem solving such as modelling, reasoning and sense making, reflecting and communicating. These goals call for significant changes in the role of the teacher and the culture of the classroom: rather than the teacher and the textbook being the primary mathematical authorities in the classroom, students are encouraged to look to each other as they discuss strategies and solutions, and justify their thinking mathematically. Reaching these goals requires a fundamental reconsideration of the types of tasks that students engage in, the sorts of classroom interactions that promote learning, and essentially what it means to *do* mathematics (Drake, Spillane & Hufferd-Ackles, 2001; Wood, Williams & McNeal, 2006).

Rethinking teaching and learning to align with these goals has important implications for classroom assessment. As Delandshere (2002) points out, “when we attempt to rethink one aspect of our educational practice, the whole enterprise comes into question. It is not possible to rethink educational assessment while leaving other educational practice intact and vice versa” (p. 1481).

Reform-Oriented Classroom Assessment

This section provides an overview of classroom assessment that is aligned with the emergent paradigm. First, I define several purposes for assessment along with some distinctions that have been made in terminology. Next, I discuss three main components of classroom assessment in this paradigm. Finally, I describe a reform-oriented view of grading and evaluation.

Purposes and Uses of Classroom Assessment

In the recent assessment literature there have been several terms used to emphasize different purposes and uses of assessment (Newton, 2007). *Formative assessment* or *assessment for learning* have been used in the literature to contrast the learning purpose with the judgmental or evaluative purpose referred to as *summative assessment* or *assessment of learning* (Gipps, 1994; Sadler, 1989). Stiggins (2002) makes the distinction that *assessment for learning* promotes greater learning while *assessment of learning* determines the status of learning. These two different assessment purposes are not necessarily distinguished by the forms of assessment, such as quizzes, tests or performance tasks, but rather in the way the assessment is used. Stiggins argues that *assessment for learning* must involve students in the process of assessment, and “advance, not merely check on, student learning” (p. 761). While many of these are not new ideas, supporters of

assessment reform argue that there needs to be more emphasis and attention to formative assessment (Black & Wiliam, 1998a; Wiggins & McTighe, 2005).

More recently, the term *assessment as learning* has been introduced to “reinforce and extend the role of formative assessment for learning by emphasizing the role of the student, not only as a contributor to the assessment and learning process, but also as the critical connector between them” (Earl, 2003, p. 25). The emphasis is on developing students’ metacognition and including students as active, engaged and critical participants in the process of assessment. This term emphasizes the necessity of greater collaboration between teachers and students so that students actively participate in the assessment process, rather than merely being subjected to it.

Assessment for Learning (Formative Assessment)

In this section, I outline three characteristics of *assessment for learning* that are aligned with current assessment reforms: integration of assessment and instruction, using a broader range of tools and strategies, and using assessment to improve students’ learning (Even, 2005).

Integration of Assessment and Instruction

Within the emergent paradigm, classroom assessment is viewed as an essential part of learning and instruction, not merely to rank or certify current or past achievement. Instead, proponents of reform argue that teachers need to consciously integrate classroom assessment and instruction in such a way that assessment is a useful part of the learning process. Assessment should be integrated into all aspects of teaching and learning in a seamless manner, inform teachers’ instructional decisions, and promote quality learning skills in students (Carless, 2007). Wiliam (2007) argues that classroom assessment must

serve as a bridge between teaching and learning and illustrates this by describing how teachers can use “hinge point” questions and plan for “moments of contingency” (p. 1088).

Students need to be involved in actively engaging with the criteria for quality work in their own and other students’ work. This is important to help students develop their metacognitive capacity and become self-regulating learners. Instead of the traditional focus on measuring achievement at the end of a period of study, assessment needs to be used in an ongoing and recursive manner to elicit and interpret evidence of students’ learning so that action can be taken to improve learning (NCTM, 1995).

Using a Broader Range of Tools and Strategies

In a reform-oriented view of classroom assessment there is recognition that the complex nature of learning requires a broader range of assessment tools and methods (Delandshere & Petrosky, 1998; Gipps, 1999). The NCTM (1995) calls for “a shift toward using multiple and complex assessment tasks, projects, writing assignments, oral demonstrations, and portfolios, and away from sole reliance on answers to brief questions on quizzes and tests” (p. 29). Assessments must elicit active thinking rather than passive recall, and generate more and deeper information about students’ understandings (Gipps, 1994). They need to focus on learning strategies and processes, not just outcomes or products (Shepard, 2001). Newer assessment strategies include more authentic and performance-based methods, including rich tasks, portfolios, journals and interviews (McMillan, 2004; NCTM). A greater diversity of assessments can help ensure the development of robust understandings, since some different forms are appropriate for different learning objectives, and repeated practice with the same item-types can artificially inflate scores due to students’ familiarity with routines rather than increased achievement (Shepard). Newer assessment tools such as rubrics, observation tools, checklists, and the

use of exemplars can be used to engage students in self- and peer assessment so that they develop an understanding of the criteria of good work.

Using Assessment to Improve Students' Learning

While assessment can serve many purposes, in a reform-oriented view, the main function of classroom assessment is to improve students' learning. Wiliam (2007) argues that "focusing on the use of day-to-day formative assessment is one of the most powerful ways of improving learning in the mathematics classroom" (p. 1091). Recently, Leahy, Lyon, Thompson and Wiliam (2005) and Wiliam (2007) put forth five essential elements of assessment for learning: engineering effective classroom discussions, questions and learning tasks that elicit evidence of learning; providing feedback that moves learners forward; clarifying and sharing learning intentions and criteria for success; activating students as the owners of their own learning; and activating students as instructional resources for one another. Wiliam connects these strategies to an increased focus on students' self-regulation of their learning processes. He argues that "the task of the teacher is not necessarily to teach, but to create situations in which students learn" (p. 1087).

Reform-Oriented Views on Grading and Evaluation

While the emphasis of current assessment reforms is largely on *assessment for learning*, there have also been calls to reconsider the methods used in *assessment of learning* – which often take the form of end-of-unit assessments and final grades (NCTM, 1995; O'Connor, 2002; Senk, Beckmann & Thompson, 1997). Proponents of assessment reform argue that grades should exclusively reflect evidence of academic achievement as opposed to other behavioural factors such as attendance and effort (McMillan, 2003; O'Connor, 2007a). In addition, traditional calculation of grades by arithmetic averaging is being increasingly challenged by the argument that grades should reflect students' most

recent and most consistent performance and the use of zeros as a penalty for late or missed work is strongly discouraged (McMillan, 2004; O'Connor, 2007a, 2007b). Conceptions of the objectivity of grading software have been called into question (Guskey, 2002) with a new emphasis on teachers' professional judgment (Joughin, 2009; McMillan, 2004).

Gipps (1994, 1999) suggests that assessment in the emergent paradigm is aligned with the interpretivist perspective often adopted in qualitative research. Delandshere (2002) argues

As different theoretical conceptions of learning and knowing call on different inquiry traditions, the ways of knowing also change, along with the methods and procedures used. What constitutes data and evidence also has to be redefined. If other, more dialogic and deliberative means of developing understandings are considered, the process becomes less technical, more interpretive, and more consistent with these theoretical perspectives (p. 1480).

Traditional notions of reliability, validity and generalizability are being discarded in favour of qualities such as trustworthiness and authenticity. This view seeks to bring students into ownership of the process and recognizes the complexity of the interactions between students, teachers and assessment. Rather than using objective measures such as counting miscues, teachers need to make sense of the whole range of evidence that they have of students' learning which includes looking for patterns, checking for contradictions, and gathering more information in situations where there is insufficient evidence to confidently make decisions about teaching and learning (Shepard, 2001). Delandshere and Jones (1999) characterize this as a process of "continual judgements, rather than simple measurement" (p. 219) of the quality and validity of understandings demonstrated. These changes in conceptions of assessment further demonstrate the need to use different methods and tools for classroom assessment.

Challenges in Implementing Reform-Oriented Classroom Assessment

In many jurisdictions there is evidence of shifting instructional practices but changes in classroom assessment appear to be emerging more slowly (Black, Harrison, Lee, Marshall & Wiliam, 2003; Delandshere & Jones, 1999; Tierney, 2006). This includes Ontario (Lock & Munby, 2000; Suurtamm, 2004; Suurtamm & Graves, 2007). Earl (2003) goes as far as to argue that the reform-oriented view of assessment *for* and *as* learning “is almost non-existent” (p. 26). As Delandshere (2002) argues, despite the widespread calls for reform in assessment practices, traditional conceptions still inform assessment practices in many classrooms. Assessment still tends to be predominantly evaluative rather than formative, and there is not yet widespread evidence of assessment as a collaborative process, with the design and marking of assessments still largely done by teachers. Even where new assessment methods (such as portfolios and rubrics) are being used, they often serve traditional purposes so that changes in practice are more superficial than the profound reconsideration called for by advocates of reform (Ball, 1990; Cohen, 1990; Hargreaves, Earl & Schmidt, 2002; Senk et al., 1997). This section draws on the literature, including both theoretical arguments and empirical studies, to describe some of the reasons for these challenges.

Making Sense of Assessment Reforms

In a review of research on educational policy implementation, Spillane, Reiser and Reimer (2002) argue that the apparent resistance to adopting current reforms is not due to a lack of effort on the behalf of teachers, but rather due to the complexities of making sense of what is being asked of them. While the slow nature of change in classroom practice makes it appear that teachers are reluctant to change their practices, Spillane et al. argue that, on the contrary, teachers and school administrators frequently work hard to implement higher-level policies. They contend that messages about reform contained in policy

documents are difficult for educators to make sense of because of factors related to teachers' previous knowledge, beliefs and attitudes which can result in "multiple interpretations of a single policy by implementing agents" (p. 420). This point is echoed by Sherin, Mendez and Louis (2004) in their discussion of the differences between the "written curriculum" and the "enacted curriculum" (p. 210). During the process of interpretation, teachers often focus on superficial similarities between new messages and old practices rather than critically important differences. This is exacerbated by policy documents that often falsely present changes in teaching and learning as "unproblematic and well understood" (Gipps, 1994, p. 126). It is often not clear to teachers and others exactly what should be changed in practice, why these changes should be made, and how this should be accomplished (Delandshere, 2002; Fullan, 2001).

Delandshere (2002) points out that the reconceptualization of learning in the emergent paradigm has consequences for assessment: "In the move from to know, to knowing – that is, from a state of having (knowledge) to an action involving participation, transaction, and transformation – the traditional notion of assessment becomes elusive and the phenomenon under consideration much less tangible" (p. 1475). As Black and Wiliam (1998a, 1998b) point out in their influential review of literature on assessment and learning, it is clear that classroom assessment practices have an impact on student achievement but it is less clear what constitutes effective classroom assessment. Each teacher needs to reconstruct his or her own practice, yet it is impossible to prescribe classroom assessment practices aligned with current reforms due to the more interpretive conceptualization of assessment. The integration of assessment and instruction results in teachers simultaneously consider the needs, answers and other cues from their students, as well as a myriad of possible responses and actions on a moment-to-moment basis to improve students' learning

(Leahy et al., 2005). Clearly, making sense of the calls for reform is more than simply decoding messages from policy documents.

Assessment Practices are Tied to Teachers' Beliefs

Studies of teachers' classroom assessment practices indicate that they are closely tied to personal beliefs about education and learning. In a study of classroom assessment and grading decision making with 27 secondary mathematics and English teachers in Virginia, U.S.A. (selected as a maximum variation sample from a larger sample), McMillan (2003) found that teachers' assessment decision making is strongly influenced by tensions between internal and external influences on classroom assessment practice. In particular, teachers' beliefs and values were "clearly cited as the most important influence on assessment decisions" (p. 36). These beliefs are often highly resistant to change (Gill, Ashton & Algina, 2004). One example of this was a case study of a Grade 8 science teacher in Ontario who was trying to implement assessment reforms (Lock & Munby, 2000). Even with a great deal of support and resources, the teacher became frustrated and abandoned these practices during the year. The researchers concluded that a primary reason for this was conflict between the teacher's beliefs and the assessment practices he was trying to implement. Even though this teacher was initiating changes of his own volition, he held largely traditional beliefs about teaching and learning. He did not see strong connections between assessment and teaching, learning or planning and tried to implement new assessment practices without fundamentally reconsidering these other aspects of his teaching practice. This resulted in feelings of frustration and a greater workload. The researchers concluded that changing assessment practices was a very complex process because of these deep connections with other aspects of teaching practice and that efforts to

examine teachers' beliefs about assessment must be connected to their views on teaching and learning in general.

Assessment Serves Multiple Purposes

One of the most frequently mentioned tensions in the literature is that of the multiple and sometimes conflicting purposes of assessment. In articles theorizing classroom assessment, Harlen (2005) and Black and Wiliam (1998a), all from the U. K., point to tensions inherent in the use of assessment information for both formative and summative purposes. Teachers must manage several conflicting roles including that of facilitator and examiner, mentor and reporter (Katz, Earl & Olson, 2001; Stobart, 2006). Some authors and researchers argue that formative and summative assessment should be kept firmly separated (e.g., O'Connor, 2007a) since students should not be penalized for efforts early in the learning process, but others argue that they can be mixed (e.g., Harlen), which can create confusion for teachers. Wiliam (2007) points out that since assessment serves many purposes, teachers can find that different purposes require different levels of aggregation of assessment data. For example, reporting on students' achievement of particular standards in a mathematics course requires compilation and analysis of different data compared to reporting on students' learning skills.

Using New Forms of Assessment

Another challenging area is the design and use of non-traditional forms of assessment which are "rich (complex) and interpretive (potentially subjective)" (Windschitl, 2002, p. 148). Windschitl points out that the evaluation tools (such as rubrics) for these types of performances must be flexible, yet contain clear enough criteria to be useful to students which is a difficult task to balance. If these tools are too general then they are not helpful to students, but if they are too specific they risk diminishing the rich nature

of the task and encouraging superficial learning to merely complete the task (Stobart, 2006). This latter phenomenon has been termed “criteria compliance” rather than learning (Torrance, 2007). In a study of assessment and grading practices in 19 American mathematics classes in five high schools in three states, Senk et al. (1997) suggest there are two key factors in teachers’ use of new assessment forms: knowledge and beliefs, and available instructional materials. They found that many teachers have limited knowledge of new assessment forms and lack the confidence to use them in their classrooms, and that newer forms of assessment have been shown to take longer to create and to assess. They also argue that there needs to be more guidance for teachers in balancing old and new forms of assessment.

A recent American study on the classroom assessment practices of 262 secondary school mathematics teachers (who were NCTM members) in nine states revealed that while there was some evidence of multiple assessment methods, there was a strong reliance on traditional paper-and-pencil tests (Ohlsen, 2007). Teachers did report using other assessment strategies such as projects, presentations and performance tasks at least some of the time, but half of the teachers reported that they used performance assessments “little” or “not at all”. It is suggested that teachers may be hesitant to use performance assessments because they lack confidence, have not been successful with these types of assessments in previous attempts, find these take more class time than traditional assessments and take longer to mark, and involve an element of unpredictability due to the open-ended nature of the questions.

Suurtamm (2004) conducted case studies of five Ontario secondary mathematics teachers who were using authentic assessment practices³ in their classrooms. These teachers believed that assessment was closely connected to other aspects of mathematics teaching and learning. As they incorporated new instructional strategies they found that traditional assessment strategies did not sufficiently allow students to demonstrate the full range of their learning so they began to expand their assessment repertoire. While these teachers were successful in implementing authentic assessment practices, they did experience tensions and dilemmas in their practice. For instance, their teaching style was “dichotomous” (p. 501) since they were using innovative instructional and assessment strategies within a largely traditional mathematics curriculum setting. The teachers often felt “caught in the middle” (p. 505) since they needed to prepare students for a traditional examination, but also felt compelled to instruct, and thus assess, in a variety of ways. The three primary dilemmas that these teachers faced included balancing traditional expectations in the curriculum with authentic assessment activities, matching new assessment techniques with traditional reporting methods, and dealing with feelings of isolation and alienation from colleagues who were using more traditional methods.

Even (2005) argues that while the literature on assessment reform often focuses on new forms of assessment, there is little guidance for teachers on how to use these richer assessment data in their decisions about assessment and instruction. She suggests that “the new tools continue to serve a traditional assessment purpose that focuses on certifying students’ attainment at the end of a period of instruction” (p. 58).

Complexity of Giving Feedback

³ Suurtamm (2004) defines authentic assessment to be “assessment that involves students in tasks that are worthwhile, significant, and meaningful and that resemble learning activities” (p. 499)

Another issue is the complex relationship between formative assessment and feedback. Providing effective feedback is difficult because it is highly situational and depends on “many interacting factors: motivation; the complexity of the task; the expertise of the learning; and the level and quality of the feedback” (Stobart, 2008, p. 160). Feedback is often ineffective, either because it is unclear, given too late for students to use productively, or students are not required to do anything with it (Carless, 2007; Stobart, 2006). In a widely cited review and meta-study on feedback interventions, Kluger and DeNisi (1996) found that while on average feedback improved performance, over one-third of feedback interventions resulted in a decrease in performance. Nicol and Macfarlane-Dick (2006) point out that “feedback messages are invariably complex and difficult to decipher, and that students require opportunities to construct actively an understanding of them” (p. 201).

Issues in Grading and Reporting

Significant challenges have emerged around grading and reporting. Earl (2003) points out that classroom assessments often “don’t give much indication of mastery of particular ideas or concepts because the test content is generally too limited and the scoring is too simplistic to represent the broad range of skills and knowledge that has been covered” (pp. 22-23). This is often because of measurement conceptions from the old paradigm which result in the most easily judged and highly reliable aspects of a particular construct being assessed (Harlen, 2007; Stobart, 2006). Brookhart (2004) points out that in the design of assessments there is an underlying tension between the authenticity of any given task and ensuring that this represents a small enough chunk of instruction that students can manage and learn. The use of more open-ended performance tasks requires adjustments to scoring techniques, and that the wide variety of assessment information that

teachers collect and use makes constructing a useful summary of a student's performance difficult (Zeibarth, 2003). Even though there is increasing skepticism over the fairness and accuracy of grades and marks, many of the traditional features of assessment of learning are widely accepted by parents and members of the public (Earl, 2003). Proponents of reform suggest that rather than averaging, there should be more consideration of consistent performance, with more attention to recent evidence (Guskey, 2002; Harlen, 2007; O'Connor, 2007a). Guskey argues that there are serious shortcomings in the use of computerized grading software which "lead educators who use them to believe that mathematical precision necessarily brings greater objectivity and enhanced fairness to grading" (p. 776).

One of the biggest challenges in the emerging paradigm of assessment is that of reconceptualizing reliability (Gipps, 1994). Harlen (2007) points out that traditional notions of reliability continue to restrict what is assessed. For example, on external and large-scale assessments, items that require students to be creative in their responses or to present arguments "are rarely considered and would be unlikely to survive the pilot trial used in developing and selecting items for external tests and examinations" (p. 32).

In a study of secondary school teachers' (Grades 6-12) classroom assessment practices, McMillan (2003) found that teachers' decision making about assessment and grading was a "highly individualized, idiosyncratic process" (p. 38) and noted that teachers had difficulty providing a rationale for their own specific assessment and grading practices or justifying what they did. He noted that teachers' own beliefs about learning and assessment were most often cited as the primary influence on decisions, despite a lack of grounding in measurement theory. Teachers' preferred assessment practices were aligned with their own general philosophy of education, but external pressures sometimes resulted

in the use of assessment practices that were not consistent with their own beliefs. McMillan suggests that assessment should be conceptualized as “more subjective and even intuitive than is implied by the use of technical measurement concepts” (p. 39).

Summary

While the aims of reform-oriented classroom assessment are supported and encouraged by the educational research community, there is growing recognition that this is not easy for teachers to do (Even, 2005). Shepard (2001) argues that “Although contemporary rhetoric implies that a shared understanding exists about what it means to use assessment data to improve instruction, examples offered suggest considerable ambiguity” (p. 1093). A recent large-scale study of the implementation of the mathematics curriculum in intermediate grades (7-10) in Ontario (Suurtamm & Graves, 2007; Suurtamm, Koch & Arden, in press) shows some evidence of changing assessment practices, but of the 1096 teachers who participated in a questionnaire, only 9% indicated that they *did not* feel that they needed any further professional development on the topic of assessment.

The NCTM (1991) points out that “Because teaching mathematics well is a complex endeavour, it cannot be reduced to a recipe for helping students learn” (p. 22). The creation of assessment tasks is time-consuming and difficult, and the interpretation of the ensuing data is complex (Even, 2005; Senk, et. al, 1997). Black et al. (2003) argue that implementing current assessment reforms

requires personal change. It means changing the way a teacher thinks about their teaching and their view of their role as a teacher. Since the way a teacher teaches is inextricably linked with their own personality and identity, ultimately it means changing yourself (p. 80).

CHAPTER 3: CONCEPTUAL FRAMEWORK

This chapter introduces a conceptual framework that describes my approach to examining tensions & dilemmas in classroom assessment practice. First, I define tensions and dilemmas in teaching and discuss why these are viewed by many researchers as fundamental and interesting aspects of teachers' work. I also explain why I view tensions and dilemmas as important sites for inquiry to further develop the emergent paradigm for curriculum, learning and assessment. Next, I present and discuss the conceptual framework, which is adapted from McMillan's (2003) framework for *assessment decision making*. Finally, I discuss how this framework will be used in the analysis and discussion of this study.

Tensions and Dilemmas in Teachers' Practice

As noted in the introductory quote to this study, teaching is widely regarded as a "cognitively and conceptually complex job" (Clark & Lampert, 1986, p. 28). Classrooms are *multidimensional and unpredictable environments in which many events occur simultaneously that require immediacy in decision making and responses* (McMillan, 2004). The current calls for reform in teaching and learning make this work considerably more complex since they contain divergent goals, focus on different aspects of teachers' work and can sometimes be contradictory (Davis & Sumara, 2007; van Veen, Slegers, Bergen, & Klaassen, 2001). Sfard (2003) even argues that the role of the teacher "was never as difficult as it is today" (p. 386). The implementation of current reforms requires fundamental changes in the nature of teachers' work, but it is not always clear to teachers exactly what or how their practices should change (Geijssels & Meijers, 2005).

Recently, there has been a great deal of interest in teaching dilemmas as an important phenomenon and a productive lens for studying problems of teaching (Adler,

1997, 1999; Ball, 1993; Crockett, 2002; Enyedy et al., 2006; Lampert, 1985, 2001; Pope, Green, Johnson, & Mitchell, 2009; Talanquer et al., 2007; Tillema & Kremer-Haydon, 2005; Windschitl, 2002; Woods et al., 1997). Writing about her own dilemmas of teaching elementary school mathematics, Deborah Ball points out that dilemmas are a fundamental aspect of teaching, not just of teaching in ways aligned with current reforms, although implementing reform-oriented practices can bring out new dilemmas. She suggests that teaching dilemmas often arise from competing but worthwhile aims. In a study with 22 pre-service teachers, Talanquer et al. found dilemma analysis to be a valuable way to examine teachers' thinking and concerns about their practice. Similarly, in work with teacher educators, Tillema and Kremer-Haydon argue that dilemmas are a "powerful conceptualization" (p. 203) for examining teachers' beliefs about teaching. In their study with 35 teacher educators, Tillema and Kremer-Haydon found dilemmas to be helpful in getting teacher educators to articulate their practice, scrutinize their own beliefs, and make their perspectives explicit.

In this study, I follow Talanquer et al. (2007) in defining *teaching dilemmas* broadly as "problem spaces created in the minds of teachers as they engage in the practice of teaching" (p. 401). Tillema and Kremer-Haydon (2005) make a more fine-grained distinction between *tensions* as "the emotional awareness of a possible conflict" (p. 204) and *dilemmas* as being "more narrowly focused on actual situations of teaching" (p. 205). Here, I am interested in *both* tensions and dilemmas, and will clarify my approach to this distinction later in the chapter. Woods et al. (1997) view teaching dilemmas as "lived experiences" (p. 19) that are often connected to dissatisfaction with previous decisions or anxiety about future choices that may arise in the classroom. They stress that for an episode or situation to be considered a teaching dilemma, teachers must be aware of the fact that

they are “confronted by choice, which they are free to make, although at times it may be difficult to do” (p. 19).

Examining tensions and dilemmas in teaching practice can help articulate teachers’ concerns and problems (Talanquer et al., 2007). Tillema & Kremer-Haydon (2005) point out that dilemmas can be a “useful trigger” (p. 204) for thinking and acting. They argue that dilemmas are a bridge between teachers’ cognitions and actions, so the study of dilemmas can allow close examination of teachers’ beliefs and orientation towards practice. Enyedy et al. (2006) define teaching dilemmas as “conflicts in which there are multiple, equally viable and often unattractive alternatives” (p. 72) and suggest that dilemmas are more often managed than solved. Windschitl (2002) argues that this *negotiation* of teaching dilemmas is of particular interest since it involves a high level of professional thinking in considering possible alternatives for action and deciding on the best course of action.

I view the tensions and dilemmas that teachers experience in their classroom assessment practice as important sites for developing what Shepard (2000, 2001) termed the *emergent paradigm* for curriculum, learning and assessment (discussed in Chapter 2). As Shepard argues, in this current period of reform, some teachers are experiencing difficulties because their instructional practices are based in the emergent paradigm while their classroom assessment practices are still largely drawing on the old paradigm. Several studies have illustrated this point by describing the experiences and perspectives of teachers who are trying to incorporate new elements into their classroom instruction, but still using traditional assessment practices (e.g., Delandshere & Jones, 1999; Lock & Munby, 2000).

For this study, the two participants were selected because their beliefs and practices about *both* instruction and assessment appear to be based in the emergent paradigm (participant recruitment and selection are discussed in Chapter 4). Given this alignment, it

will be interesting to see what tensions and dilemmas the participants describe in their assessment practice. Examination of these tensions and dilemmas can point to areas for future research and professional development to support teachers using classroom assessment practices aligned with current reforms.

Framework for Tensions and Dilemmas in Classroom Assessment Practice

In this section, I present a framework for examining tensions and dilemmas in classroom assessment practice that integrates previous research on teachers' assessment decision making (McMillan, 2003, 2004), influences on assessment practice (Saxe et al., 1999) and negotiation of dilemmas (Tillema & Kremer-Haydon, 2005; Windschitl, 2002). The framework is illustrated in Figure 3.1 and each part is discussed below. While this discussion is presented in a linear manner, I must emphasize that I view all aspects of the framework as mutually interactive.

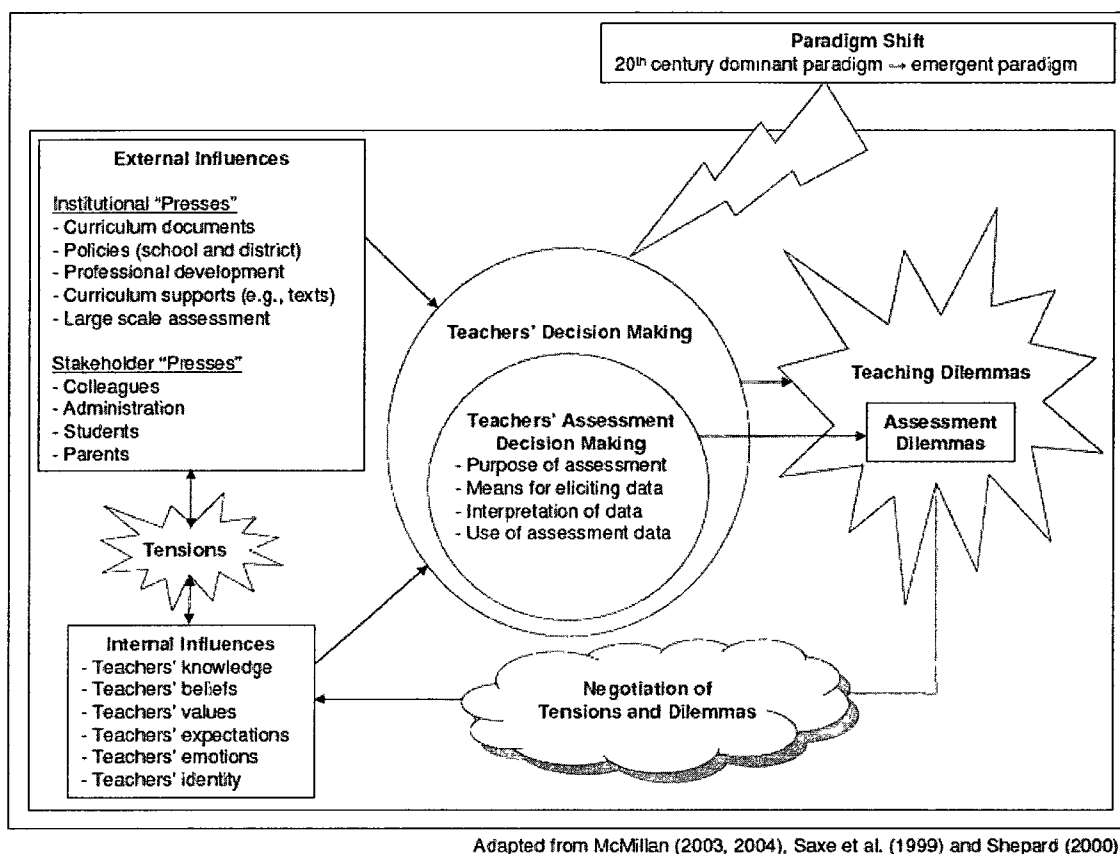


Figure 3.1: Framework for examining tensions & dilemmas in classroom assessment practice

Teachers' Assessment Decision Making

At the centre of this framework, illustrated in Figure 3.1, is James McMillan's (2003) notion of teachers' *assessment decision making*. He suggests that effective teaching is guided by a continuous process of decision making, including many decisions about assessment. In the diagram, I have represented assessment decision making as a subset of teachers' decision making. I see a strong connection between this focus on decision making and Woods et al.'s (1997) conception of teaching dilemmas being closely tied to choices that teachers have to make. McMillan defines assessment decision making as "a process in which teachers balance the demands of external factors and constraints with their own

beliefs and values to determine classroom assessment practices” (p. 42). As conceptualized in the emergent paradigm, assessment and therefore assessment decision making are important aspects of teaching practice, fundamentally integrated with other aspects of teaching and learning (Brookhart, 2003; Shepard, 2000; Suurtamm et al., in press; Wiliam, 2007). As McMillan himself points out, “it may well be that classroom assessment decision making is both an assessment decision *and* an instructional decision” (p. 38, emphasis in original).

McMillan (2004) identifies four components of classroom assessment, which point to four types of assessment decisions: purpose (“Why am I doing this assessment?”), measurement (“What techniques should I use to gather information?”), evaluation (“How will I interpret the results?”) and use (“How will I use the results?”) (p. 9). In the framework, I re-named two of these components for clarity: I replaced *measurement* with *means for eliciting data* and *evaluation* with *interpretation*. I believe these terms are consistent with McMillan’s intended meaning and avoid confusion with other uses of *measurement* and *evaluation*. While McMillan does not explicitly connect these four components of assessment (2004) to his framework for assessment decision making (2003), I have incorporated them into the framework for this study since they may be helpful in identifying the types of assessment decisions in which teachers experience tensions and dilemmas. This will be a point of discussion in the cross case analysis of this study.

Tensions between Internal and External Influences on Assessment Practices

McMillan (2003) argues that the assessment decision making process is “characterized by tensions” (p. 35) between internal and external influences. His framework lists several influences of each type, determined from a qualitative study with a subset of 28 teachers from a much larger sample in Virginia, U.S.A. This study found that

the greatest influence on teachers' assessment practice was their own beliefs and values, as well as their knowledge and expectations. External influences on teachers' assessment decision making included accountability testing, district policy and parents.

This characterization of tensions between internal and external influences on assessment decision making is very similar to a framework for conceptualizing developments in teachers' assessment practices put forth by Saxe et al. (1999). They describe external and internal "presses" on teachers' assessment practices as being in constant tension and identify two main categories of external presses: institutional and key stakeholders. Institutional presses include standards, curriculum materials, district testing and professional development; while stakeholder presses include parents, administrators, colleagues and students. The internal presses include teachers' values, interest in trying new assessment strategies and tools, and satisfaction with their own assessment practice. These "presses" act to support or inhibit change in teachers' assessment practices.

The framework for this study draws on both McMillan (2003) and Saxe et al. (1999), as illustrated on the left side of Figure 3.1. Within the category of external influences I have used Saxe et al.'s organizing titles of institutional and stakeholder presses. To the list of internal influences on assessment practices, I add teachers' emotions and identity since these have also been shown to influence teachers' practice and adoption of current reforms (Beijaard, Meijer & Verloop, 2004; Day & Leitch, 2001; Enyedy et al., 2006; Geijssels & Meijers, 2005; Hargreaves, 1998; Hodgen & Askew, 2007). Within this study I will explore whether the participants experience these (or other) influences as being in tension and how they negotiate these influences and tensions in practice.

Dilemmas in Classroom Assessment

In addition to tensions that result from competing influences on teachers' assessment decision making, I am also interested in the dilemmas of practice that arise as teachers carry out their work. As discussed earlier in this chapter, dilemmas are important phenomena of interest in studying teaching, not just in assessment. While some theoretical work has pointed to potential dilemmas in classroom assessment (e.g., Gipps, 1999; Katz, Earl & Olson, 2001), there have been few empirical studies that address dilemmas in classroom assessment (e.g., Even, 2005; Higgins, Hartley & Skelton, 2001; Suurtamm, 2004; Webb & Jones, 2009). In the framework I am proposing here, I include assessment dilemmas as a subset of teaching dilemmas as illustrated on the right side of Figure 3.1, since not all teaching dilemmas relate to classroom assessment.

Negotiating Tensions and Dilemmas in Practice

In addition to exploring the tensions and dilemmas that the participants experience, I am also interested in the negotiation of these tensions and dilemmas. It has been suggested that examination of these negotiations is of particular interest since these can provide significant insight into teachers' beliefs and thinking (Enyedy et al., 2006; Tilemma & Kremer-Haydon, 2005; Windschitl, 2002). Following Wenger (1998), I use *negotiation* to include interpreting and acting, doing and thinking, and understanding and responding. As illustrated by the double arrows at the bottom of Figure 3.1, I suggest that how teachers negotiate their dilemmas both *influences* and *is strongly influenced by* internal factors such as their beliefs, values, emotion and identity. Again, I need to emphasize that I view the negotiation of tensions and dilemmas as simultaneous to the experience of these tensions and dilemmas.

Paradigm Shift

At the top of the framework, I have included a reference to the paradigm shift in curriculum, learning and assessment that Shepard (2000, 2001) has identified. It is of great interest to me is how these two teachers' experiences of tensions and dilemmas relates to the current shift from the dominant twentieth century paradigm to the emergent paradigm, particularly in assessment.

Summary

In this chapter, I have argued why tensions and dilemmas are important phenomena to study in teachers' practice, that examination of their negotiation can provide insight into teachers' beliefs and thinking. In addition, I outlined the conceptual framework for this study. At the centre of this framework is McMillan's (2003) notion of teachers' assessment decision making, which he suggests is "characterized by tensions" (p. 33) between external and internal influences on teachers' assessment practice. I distinguish between tensions as the constant presence of these different and sometimes conflicting influences, and dilemmas as more specific situations that arise in teachers' practice and need to be resolved or managed.

This framework, presented in Figure 3.1, will guide the analysis presented in both the individual case reports and cross-case analysis. In the individual case studies (Chapter 5), I will describe the teachers' tensions and dilemmas together, along with the ways they negotiate these. In the cross-case analysis (Chapter 6), I will consider how the tensions and dilemmas experienced by the participants in this study compare to the four types of assessment decisions identified by McMillan (2004). Next, I will examine the tensions between internal and external influences on teachers' assessment decision making (McMillan, 2003; Saxe et al., 1999), as well as the sorts of dilemmas the participants

experienced during the study. Finally, I will describe how the participants negotiate these tensions and dilemmas.

CHAPTER 4: RESEARCH DESIGN

This chapter provides an overview of the research design for this study. This includes a rationale for choosing a qualitative approach to this research, for selecting a multiple case study design involving two cases, and a discussion of my role as a researcher. I also describe how the two teachers, Amy and Carina, were invited to participate; some aspects of the two participants themselves; my approach to data collection and instrument design; and the process of data construction, analysis and interpretation. I end with an argument for the credibility of this inquiry. This section is fairly lengthy, but I believe that it is very important to explicitly address my own perspectives and explain some of the decisions I made during the research in order to allow the reader to judge the integrity of the study (Milne & Oberle, 2005; Schram, 2006; Stake, 2006; Whittemore, Chase & Mandle, 2001).

Overview of the Research

This research is a multiple case study (Stake, 2006; Yin, 2003) of two secondary school mathematics teachers' experiences with classroom assessment in the context of the current educational reforms in Ontario, Canada. The focus is on participants' perspectives on and experiences of tensions and dilemmas in their classroom assessment and how they negotiate these tensions and dilemmas in their day-to-day practice. The design of this study is guided by the following two research questions: What tensions and dilemmas do two secondary mathematics teachers experience as they try to implement current assessment reforms? And, how are these tensions and dilemmas negotiated in classroom practice?

The methods used in this study included interviews, classroom observations, and analysis of artefacts shared by the participants. An initial interview was conducted to get a sense of the teachers' views on and experiences with classroom assessment. This was

followed by two or three classroom observations to get a sense of each teacher's context. Two debriefing interviews were conducted with each teacher, each following a classroom observation, to explore teachers' intentions and thinking about classroom assessment during each lesson. A final interview was conducted with each teacher to further explore her experiences and allow me to seek clarification of my own interpretations.

Rationale for a Qualitative Multiple Case Study

Since I am interested in the "specifics of particular cases" (Denzin & Lincoln, 1998, p. 10), a qualitative approach is most appropriate to this inquiry. The purpose of qualitative research is "not necessarily to map and conquer the world but to sophisticate the beholding of it" (Stake, 1995, p. 43) and the goal is "not to reveal 'truth' but to generate insights" (Milne & Oberle, 2005, p. 413). These views align with my goals for this research. I believe that classroom assessment is a complex and multifaceted yet very personal endeavour, especially in light of current curriculum and assessment reforms, which makes it well suited to qualitative inquiry. My intention is not to evaluate or compare classroom practices to descriptions in policy documents, but rather to gain an understanding of each participant's unique perspective and challenges. A multiple case study will allow a close examination of two teachers' lived experiences in their own particular contexts and the meaning(s) they make of those experiences.

Yin (2003) defines a case study as an empirical investigation of a contemporary phenomenon within its natural context using multiple sources of evidence. Qualitative case study research is generally regarded as more exploratory than confirmatory (Gerring, 2007; Hancock & Algozzine, 2006) and consideration of context is highly important (Creswell, 1998; Stake, 2006; Yin). In multiple case studies, individual cases are bound together by interest in a phenomenon, condition, function or program, which Stake calls a "quintain"

(p. 6). In this research I follow Stake's (1995, 2006) approach to case study research with a stronger interest in "seeking understanding of human experience" (1995, p. 38) than in identifying cause and effect relationships. Conducting two case studies allows for broader insight into teachers' experiences than one case study allows, while recognizing that as the number of cases increases, the depth for single cases decreases (Creswell, 1998; Yin, 2003). Several other related qualitative studies which influenced my thinking about both theoretical and methodological considerations had between two and four cases (Delandshere & Jones, 1999; Enyedy, Goldberg & Welch, 2006; Romano, 2006).

My Role as a Researcher

In qualitative inquiry the role of the researcher is fundamental to all aspects of the study. As Stake (1995) contends, "Phenomena need accurate description, but even observational interpretation of those phenomena will be shaped by the mood, the experience, the intention of the researcher" (p. 95). The essentially interpretive role of a qualitative researcher not only affects the analysis and interpretation of data, but the actual data themselves. Schram (2006) argues that "Qualitative data can be thought of as the objects and events that a researcher perceives and describes" (p. 11) and reminds researchers that there is an "inevitable way in which all data are filtered through your own reasoning and that of your research participants" (p. 11).

Researchers need to make choices about many aspects of fieldwork including the degree of involvement and participation, familiarity with participants and the setting, positioning, and role awareness (Schram, 2006), and it is important to make these decisions and their rationale apparent to the reader of qualitative research (Stake, 1995). There is, however, no prescription for managing these choices (Stake). For example, writing about interviewing in qualitative research, Seidman (2006) points out that over-sharing on the

part of a researcher can distort an interview and distract participants, but also notes that some sharing of experiences may encourage a participant to continue sharing.

With regard to my role as a researcher in this study, it is important to state that along with the participants, I am currently a secondary mathematics teacher in Ontario. This was helpful during the transcription process since I was familiar with resources, policies, acronyms and some of the people mentioned in the interviews. Beyond this, it is important to state that a great deal of my professional energy as a classroom teacher, particularly since I began graduate studies in 2006, has been focused on developing my own assessment practice. I have led professional development sessions at the school, district and provincial levels. My own experiences inevitably inform my interpretations of what the two participants in this study shared and the meanings they make of their experiences.

The fact that I am a teacher had implications for the recruitment of participants for this project. Five teachers expressed an interest in participating in this study and I have professional relationships with all of these teachers through such activities as our joint participation in several professional development initiatives, membership in teacher organizations and conferences that we have attended.

Since I already had professional relationships with the two participants, Amy and Carina, we quickly established a good rapport during the interviews. This was very important because I was interested in their experiences negotiating tensions and dilemmas in classroom assessment, which is a highly personal aspect of their teaching practice. The emotional nature of assessment for teachers was highlighted to me during the data collection phase of this study when some aspects of provincial assessment policies were raised in the media, resulting in several editorials and letters from readers for weeks

afterwards. Teachers at my own school were “buzzing” with reactions and I encountered several heated debates in the staffroom. Both participants in this study, at different schools, also acknowledged that some aspects of classroom assessment were “hot topics” for teachers. I believe that Amy and Carina trusted me due to our previous work together and because of my “insider” standing as a current classroom teacher and I believe they spoke quite freely about their own struggles. During the study, both teachers brought up some experiences we had shared and referred to several specific resources and initiatives that they knew I was aware of and had used or participated in myself. My participation along with Amy and Carina in some professional activities meant that I had access to additional context beyond what they told me during the interviews and what I observed during the classroom observations. I believe that this led to greater openness during the interviews since they could refer to our shared experiences without having to offer long introductory (and possibly prohibitive) explanations, but it also meant that I had to be constantly aware of our taken-as-shared understandings. During the interviews, I tried to be aware of this and frequently asked for descriptions and clarifications of some contextual information that I already had a sense of. In preparation for the fourth and final interview with each participant, I reviewed all the data and actively sought to clarify any taken-as-shared understandings.

Since I did not want my own specific experiences of tensions and dilemmas to be overly suggestive to the participants, I explained in the first interview that I had experienced many challenges in my own assessment practice that stimulated my interest in this research study. I made a concerted effort to avoid bringing up my own challenges but occasionally, and only after they had been mentioned by a participant, I did state that I had also experienced a similar challenge in my own practice. This felt to me a natural part of

the conversation since both Amy and Carina already knew me as a secondary mathematics teacher.

During the classroom observations, I tried to remain as unobtrusive as possible and I explained this to both participants during the initial interview. We decided that I would sit in an empty desk in the back of the classroom, and not participate in the lesson. Upon arriving for the classroom observations, the teachers both asked me to introduce myself briefly, although both had told their classes I was coming. After this I took my seat and tried to keep a low profile by minimizing my body movement, avoiding direct eye contact with the teacher and students, and writing quietly.

Inviting Participation

For this study, I wanted to examine the experiences of current secondary mathematics teachers who had been teaching for at least three years. I was interested in the perspectives of experienced teachers because their challenges are often different to those that new teachers face. In addition to approval from the University of Ottawa's own Research Ethics Board, I needed to obtain permission from the school districts I approached since I wanted to interview teachers in their place of work and conduct classroom observations. After this permission was received, the districts required that I contact individual school principals to gain their permission before speaking to any teachers. I contacted several high school principals and requested a short face-to-face meeting to describe the research study and deliver recruitment letters. These letters had my contact information for interested participants. The letters were distributed by the principal into the mailboxes of all mathematics teachers at his or her school.

Through my experiences as a secondary mathematics teacher in many initiatives at the school district, regional and provincial levels, in a few different areas of Ontario, I had

several teachers in mind who I thought who would be good candidates for participation. By this, I mean that I had been present at meetings or workshops where I had the impression that they were trying to implement current curriculum and assessment reforms, were well respected within the math community and appeared to be comfortable sharing their perspectives with colleagues. This is in line with Stake's (1995) advice of developing "a connoisseur's appetite for the best persons, places, and occasions. 'Best' usually means those that best help us understand the case, whether typical or not" (p. 56). While abiding by the school districts' recruitment policies, I employed "purposeful sampling" (Creswell, 1998; Seidman, 2006) in contacting principals of schools where some of these potential participants were teaching. I hoped that some of these teachers would contact me to express interest in participating in the project, although I was prepared to invite participation from teachers I did not previously know. As it turned out, I knew all teachers who contacted me to express interest in the project.

As a result of this purposeful distribution of recruitment letters in five schools, five teachers expressed interest and all five met the criteria of being current secondary mathematics teachers with more than three years experience. Two of these teachers indicated they were interested in the project and would take part if I needed more participants but said that due to time constraints they would prefer not to. I thanked both teachers for indicating interest but stated that I did not wish to place any further demands on their time. A third teacher agreed to participate, but had to withdraw several weeks later due to medical issues. She did not sign a consent form and no data were collected. This left two teachers who were interested and able to participate in the study.

Participants

The two teachers who agreed to participate are Amy, who has six years experience, and Carina, who has twenty years experience. Both teachers work in the same school district but at different schools, and neither teacher works at the same school as I do. At the time of the study Amy was teaching Grades 9 and 10 mathematics courses and Carina was teaching Grades 10 and 12 mathematics courses. Both teachers hold qualifications in intermediate and senior mathematics, which in Ontario means that they are qualified to teach mathematics in Grades 7 through 12. Both teachers also hold an Honour Specialist qualification in mathematics which is a designation granted by the Ontario College of Teachers [OCT] after teachers complete an additional qualification course. This course is focused on “developing the capacity for curriculum leadership” (OCT, n.d.) and to gain entry teachers must have taken at least six full-year university courses in mathematics (although some providers of this course require 9 full-year courses) and have at least two years of teaching experience. A summary of each participant’s qualifications, experience and teaching assignment is included in Table 4.1.

Table 4.1: Summary of background information about each participant

Name (Pseudonym)	Gender	Teaching Qualifications	Experience (years)	Teaching Assignment
Amy	F	Mathematics and science in intermediate/senior divisions (Grades 7-12) Honours Specialist (Mathematics)	6	Grades 9 & 10 math
Carina	F	Mathematics and science in intermediate/senior divisions (Grades 7-12) Honours Specialist (Mathematics)	20	Grades 10 & 12 math

As part of the informed consent, confidentiality and anonymity were assured in reporting. Any information which might identify an individual, school or district has been removed and participants are referred to by pseudonym. In choosing pseudonyms and describing the participants and their contexts, I had to find a balance between doing justice to the participant while protecting their identity (Seidman, 2006). Due to the school district's requirements for recruiting, the principals at both schools were aware of each teacher's participation, but I did not have any contact with the principals other than the meeting to ask for their permission to distribute letters of invitation and then informing them via email of the teacher who had agreed to participate. Both teachers understood that their principals knew of their participation in the project.

In this study I do not claim that the two participants are representative of all secondary mathematics teachers. Following case studies by Delandshere & Jones (1999) and Watson (2006), I believe that inquiry into Amy's and Carina's experiences reveals some of the complexities of assessment in the secondary mathematics classroom in the context of current reforms. My intention with this research is not to generalize about all teachers' assessment practice, but rather to provide a description of the perspectives and experiences of two teachers.

Data Collection and Instruments

In this study, multiple sources of evidence were sought, analyzed and interpreted. These included individual interviews, classroom observations followed by debriefing interviews, classroom artefacts the participants shared with me and other publicly available materials. An overview of the data is provided in Table 4.2. The data collection for Carina's case took place over 5 months while Amy's case took place over 1 month. The interview guides and observation protocol are included as Appendices A-F and will be discussed later

in this section. I have included samples of artefacts from the lessons I observed in Appendices G-J. Since some of the classroom artefacts could potentially identify the participants I did not include all the artefacts shared by the participants.

Table 4.2: Overview of data

	Amy	Carina
First Interview	• Data: interview transcript; publicly available classroom artefacts	• Data: interview transcript; classroom artefacts
Observation #1 & Debriefing Interview	• Observation of Grade 9 class (75 minutes) • Data: field notes; transcript of debriefing interview; classroom artefacts	• Observation of Grade 10 class (75 minutes) & Grade 12 class (75 minutes) • Data: field notes; transcript of debriefing interview; classroom artefacts
Observation #2 & Debriefing Interview	• Observation of Grade 10 class (75 minutes) • Data: field notes; transcript of debriefing interview; classroom artefacts	• Observation of Grade 12 class (75 minutes) • Data: field notes; transcript of debriefing interview; classroom artefacts
Final Interview	• Data: interview transcript	• Data: interview transcript
Overview of Interview and Observational Data	• Total duration of interviews: 1:53:55 • Interview transcripts: 49 single-spaced pages • Total observation: 2.5 hours (2 classes)	• Total duration of interviews: 4:34:54 • Interview transcripts: 105 single-spaced pages • Total observation: 3.75 hours (3 classes)
Other data (publicly available)	• school profile • Ontario College of Teachers qualification record • provincial mathematics curriculum document (containing curriculum and assessment policy) • district assessment policy	• school profile • Ontario College of Teachers qualification record • provincial mathematics curriculum document (containing curriculum and assessment policy) • district assessment policy

As Stake (1995) points out, interviews and qualitative field observations are very different. During observations, researchers generally try not to interfere or control what happens and hope that events and interactions occur as though the researcher was not there,

while what is covered in interviews is “targeted and influenced by the interviewers” (p. 66). In this study, the classroom observations allowed for a shared experience between the researcher and the teacher which could be followed up on during the ensuing debriefing interviews. The observations allowed me to get a sense of each participant’s context and a description of the classroom activities. The debriefing interviews that followed each classroom observation were an important part of the data since they allowed the teachers to share their intentions and thinking around classroom assessment in each lesson since many aspects of teachers’ classroom assessment practice are not visible to observers. Teachers make many instructional and assessment decisions while planning lessons outside of the classroom, as well as “on the fly” during class time in response to individuals or to the classroom collective. Teachers’ experiences of tensions and dilemmas, as well as their negotiation may not be visible to an observer in the classroom. As well, much of teachers’ thinking about classroom assessment, as well as some of their practices including marking, grading, designing lessons and assessment strategies and tools are done outside of the classroom. The debriefing interview allowed me to ask each participant about the types of things they were thinking about during planning, decisions they made during the class, and their rationale for different aspects of the specific lesson.

The initial interview allowed me to get a general sense of the participants’ experiences and to build a relationship so that the classroom observations and ensuing debriefing interviews were more comfortable for the participants. The final interview allowed me one more opportunity to talk to each teacher about her assessment practice and clarify issues from previous meetings. This last interview occurred in June for both participants so they were able to share their thoughts about the whole semester, including

the final phase of summative tasks, exams and a provincial assessment. This look back at the semester allowed for some triangulation of the data with previous interviews.

In this study, it was important for participants to be comfortable during the interviews in order to discuss their experiences and challenges with classroom assessment (Dilley, 2000; Schram, 2006). I asked participants where they would prefer to have the interviews and both Amy and Carina indicated they would like to conduct the interviews in their classrooms. All four interviews with Carina occurred in her classroom, as did two of Amy's interviews. Two of Amy's interviews occurred in a different classroom since there were students completing work in her classroom. All of the interviews were audio recorded and transcribed so I could work with the words of the participants (Seidman, 2006).

In the design of the interview protocols I began with close-ended questions in order to establish conversation and put the respondent at ease (Dilley, 2000). I then moved to open-ended questions to elicit descriptions of each participant's experiences and intentionally wrote these in a conversational style to try to avoid conveying a sense of a clinical interview (e.g., I wrote "In relation to assessment, what sorts of things have you tried in your classroom over the past few years?" rather than more formally worded questions). For several of these questions I listed several possible follow up prompts. While the interview guides were very similar for both participants, I approached the interviews with the view that each participant had unique experiences and stories and I did not expect the interviews to unfold in identical manners (Stake, 1995). Before conducting these interviews, I rehearsed the questions in pilot form with a colleague, but did not collect any data from these pilot interviews.

First Interview

Both participants contacted me via email to express interest in the project. I had intended to hold an introductory meeting with each participant before the first interview but since I already knew both Amy and Carina, I offered to combine the introductory meeting and first interview. I did not feel that the initial meeting was necessary for the project in the same way that it would have been to begin building a relationship with participants that I did not previously know, and I wanted to be respectful of Amy and Carina's time.

Before turning on the audio recorder during the first interview I explained the project, went over the informed consent forms and the participants signed these. I asked questions about the length of participants' teaching experience and qualifications in order to confirm that they were eligible for the study, as well as what courses each participant was currently teaching and had previously taught. The interview guide for this part of the interview is included as Appendix A.

The rest of this first interview was semi-structured in nature and focused on exploring each participant's perspectives on and experiences with classroom assessment. The interview guide for this part of the interview is included as Appendix B. The first section contains two questions to provide information about the participant's background and her experiences with assessment as a student, which are very powerful models for her own practice (Garet, Porter, Desimone, Birman & Yoon, 2001; Loucks-Horsley, Hewson, Love & Stiles, 2003). The questions in the next section ask about the participants' assessment practices and any changes they have experienced. I believe that these questions allowed teachers' experiences with tensions and dilemmas to emerge more naturally during the discussion of these questions than if I directly asked the participants to specifically describe their tensions and dilemmas. Questions four and five (asking how participants get a sense of students' understandings and determine grades) drew on questions posed by Suurtamm &

Graves (2007) in their large-scale study of curriculum implementation in Ontario. These questions were also suggested by Wiliam (2007) since they can highlight tensions between teachers' assessment and grading practices. The final section asks about challenges and supports that the participant has experienced. Asking directly about challenges addresses my first research question about what tensions and dilemmas teachers experience. It also allowed me to address the second research question about how these tensions and dilemmas are negotiated since I could follow up with questions about coping strategies. Asking participants about supports they received allowed exploration of specific strategies that helped with the implementation of current assessment reforms.

This first interview with Carina was conducted in her classroom which was her preference since it allowed her access to her materials (e.g., her assessment binder) and equipment (e.g., her laptop). Amy's first interview was conducted in a classroom down the hall since some students were finishing writing a test in her classroom and there were three interruptions to the interview as students came to ask her questions. I turned the recorder off during these conversations.

Classroom Observations and Debriefing Interviews

In the weeks following this first interview I visited each teacher's classroom twice to observe lessons in order to gain a better understanding of her specific context as well as to have shared lesson experiences that could serve as the basis for conversations about classroom assessment. These visits were arranged at the invitation of each teacher, when she felt she was doing something during the period that she wanted me to see and then discuss. For Amy, this included a class where she was using whole-class interactive assessment strategies and another where her class was involved in problem solving. For Carina, this included classes where she was gathering data for her Assessment Checklist

(explained in Chapter 5), and formative quizzes. During these observations I took handwritten field notes and the observation protocol is included as Appendix C. I did not make audio or video recordings since I thought this might deter potential participants. My intention for these field notes was to help provide what Stake (1995) calls “a relatively *incontestable description* for further analysis and ultimate reporting” (p. 62, emphasis in original). My focus was on the teacher, rather than the students, and I noted a description of the classroom as well as aspects of the lesson written on the interactive whiteboard, comments from the teachers, activities that the teacher asked the students to engage in, and comments made by the teachers to whole group, small groups and individual students.

Each classroom observation was followed by a short conversational debriefing interview with each participant and this interview guide is included as Appendix D. Unlike the classroom observations, the debriefing interviews were audio recorded. Carina invited me to observe two consecutive class periods during one afternoon and this was followed by a single debriefing interview. The second occasion that I observed Carina’s class, I saw one lesson and the debriefing interview was held at the end of the period. Both debriefing interviews were held in Carina’s classroom. Both times I visited Amy’s classroom I observed one class period and the debriefing interviews were held right after the class. One of these interviews was conducted in Amy’s classroom while the other was carried out in a classroom down the hall. In these debriefing interviews I asked the participants what their impressions were of the class and what they were thinking about during the lesson in relation to classroom assessment. I was particularly interested in how each teacher got a sense of her students’ understandings during the lesson, if any changes were made during the lesson and how the lesson related to upcoming assessments. I also sought clarification on contextual issues such as the use of interactive whiteboards in both classrooms.

Final Interview

Final interviews for each teacher were held during the exam period in June. This was partly due to mutual time constraints as well as to allow me time for preliminary analysis of previous data to inform the final interview. This ended up being a very good time of year to conduct the final interviews since both teachers were negotiating the tensions and dilemmas associated with exams, final tasks, grading and reporting, and had current and specific examples to share with me. These interviews were semi-structured and audio recorded and the questions were planned after reviewing all the data for each case. I asked each participant to describe some of her specific experiences with assessment over the past few weeks and clarified issues from previous observations and interviews.

Other Data

In addition to the interviews and classroom observations, I collected classroom artefacts that the teachers shared with me. These included planning and assessment tools, quizzes and class problems. In addition, I accessed some of the materials used in one of Amy's classes through the internet including classroom resources released by the provincial Ministry of Education, and sample items from the provincial assessment in Grade 9 mathematics run by the Education Quality and Accountability Office [EQAO].

I also collected and analyzed other publicly available documents to further my understanding and description of each teacher's context. These included school profiles, qualification records from the Ontario College of Teachers (provincial regulatory body), provincial curriculum documents which include assessment policy, and school district assessment policy. I did not include any pieces in this manuscript that could identify the teacher, school or district.

Data Construction, Analysis and Interpretation

In the data analysis for this study I have drawn largely on the views of Erickson (1986, 2004) and Stake (1995, 2006) who both emphasize the active and interpretive nature of data *construction* rather than the simplistic notion of *finding* data. Erickson (2004) argues that “Fieldnotes, interview transcripts, and archival records (as well as audiotapes and videotapes) are most appropriately conceived not as “data” in their unreduced form – they are resources for data construction within which data must be discovered” (p. 486). Similarly, he emphasizes the active nature of searching for patterns and themes within this constructed data through repeated review and “progressive problem solving” (p. 486) rather than a straightforward emergence of themes from the data. In this study the data gathering, construction, analysis and interpretation were highly recursive processes. In this section, I describe the details of these processes.

In this research, I collected and transcribed all the data, which allowed me to know the data very well. As Schram (2006) points out, the process of constructing transcripts is often portrayed as deceptively simple but is actually a highly interpretive task due to the lack of clear endings in speech: “even a transcript is the product of ongoing interpretive and ethical decisions about the significance you give to what other people convey as meaningful” (p. 12). I constructed each transcript within a short period of time after each interview to allow me to add in notes about my own impressions (e.g., sarcasm) and non-verbal communication (e.g., shrugging) within parentheses. Milne and Oberle (2005) emphasize the need for “accurate transcription” (p. 416) to ensure authenticity of the data and to this end I listened to all of the interviews in their entirety at least twice, including a re-reading of the completed transcripts while listening to the audio recording. In addition,

Milne and Oberle suggest that the accuracy can be enhanced by the researcher's "familiarity with both the subject area and the terminology used" (p. 416). Since I am a secondary mathematics teacher in Ontario, I was able to make sense of acronyms, titles of projects and initiatives and other terminology specific to the context of classroom assessment in Ontario.

Since I knew there would be four interviews with each participant, I did not intentionally review the data from the first three meetings (beyond the transcription process) until I prepared for the final interview. I did not want to narrow my focus and interpretations for the second and third interviews, although I recognize that it is impossible not to consider the impression one already has of the data (Seidman, 2006). For the final interview, however, I wanted to be sure that I had an opportunity to clarify my understandings of the participants' meanings so I spent time re-reading the transcripts and my field notes several times and writing down points or questions that I wanted to follow up with. This first analysis of the data from the first three interviews for each participant included a summary of background information about each participant, a description of her assessment practices, and a preliminary list of tensions and dilemmas along with some tentative groupings of these tensions and dilemmas. I also made a list of questions to ask in the final interview to clarify my understanding of the participants' experiences and these were incorporated into the final interview guide (so the guides for these final interviews were slightly different for each participant). For example, Amy had indicated she struggled to interpret and use the provincial Achievement Chart (explained in Chapter 5) and I wanted to make sure I more fully understand how she negotiated this in her daily classroom assessment. With Carina, I wanted to ask about her reflections on her Assessment

Checklist, a tool that she had developed for recording students' achievement (explained further in Chapter 5).

After the data collection and transcription were complete, I began a concerted process of data analysis, although as Seidman (2006) argues, in qualitative research it is impossible to completely separate the processes of data collection and analysis. Erickson (2004) points out there is no “imitable model” (p. 492) for analysis of qualitative data. He encourages researchers to try constructing analysis both “from top down and bottom up, until you find an analytic framework that fits your data and you find data that are consonant with your framework.” (pp. 492-493). Sipe and Ghiso (2004) encourage researchers to construct categories in a way that will allow the data to be interpreted and understood “in fresh ways” (p. 482) and to “constantly work to consider alternative perspectives” (p. 482). Seidman recommends physically marking the transcripts with brackets, underlining and highlighting important passages with the view of keeping tentative labels. After this process, he recommends “sifting out the ones that now seem very compelling, setting aside the ones that seem at this stage to be of less interest” (p. 126). Stake (1995) argues that for the more important episodes and passages of text “we must take more time, looking them over again and again, reflecting, triangulating, being sceptical about first impressions and simple meanings” (p. 78). These were all principles that I took seriously as I worked with the data.

During this process, I went through several stages. While this explanation of what I did may appear linear, I want to emphasize that as I moved through the different phrases I constantly considered and re-considered my own choices and interpretations. Data analysis in qualitative research does not only occur while reading transcripts or sitting at a

computer, rather, it is inevitable that the researcher lives with the data, thinking about it in different ways, even during non-research activities.

In the first phase, I created a new electronic file for early analysis with four sections: background and contextual information (such as information about courses taught, beliefs about assessment), assessment practices, tensions and dilemmas, and negotiation. I then re-read all of the transcripts again (including the final interview now), and added information in point form to each section. Working with one case at a time, I reviewed all of the transcripts and added summary points to each of the four sections. As I entered new information, I began to make some loose groupings. For example, I included information about summative assessments together, but where formative and summative assessment seemed to overlap, I made a note of this. At this point, I did not try to fit all the data into groupings. Occasionally, as I considered the data, I felt that a grouping was too specific so I deleted the subheading. Since I was working with both hard copies and electronic copies of the transcripts and field notes, I was able to include quotes using the words of participants to illustrate my point form notes. I found it helpful at this early stage of analysis to keep quotes in the words of participants so I could constantly question whether I felt my own interpretations matched the meanings I believed the participants had intended to share. This meant that the early analysis files were very long (over 30 single spaced pages for each participant). I then compared this analysis with the preliminary analysis done before the final interview (which I had done more than a month previously).

The next step in my process of data analysis, once I had finished this initial data reduction, was to make a copy of this electronic file and begin to search for more tentative groupings and sought to further reduce the data. The four sections appeared to still be an effective way to organize the information. In the sections on background information and

assessment practices, I began to make more stable groupings (such as “tests”, “quizzes”, “observations”). In the section for tensions and dilemmas, I ended up with six main groupings with 16 subheadings for Amy and seven main groupings with 19 subheadings for Carina. At this point, the title for each of these groupings was a single word or short phrase, such as “time” or “grading/evaluation”. For each participant, I organized the final section on negotiation around these same six or seven main groupings.

Since I felt that this was a very “bottom up” approach, I then decided to reconsider all of the data for each case in a “top down” manner (Erickson, 2004). I put aside the point form lists that I had generated, and wrote summaries of what I felt were each participants’ tensions and dilemmas. This time I phrased the titles for each grouping as an issue such as “issues in grading and reporting”. By the end of this phase, I had four main categories of tensions and dilemmas for each participant. For each case on its own, I then took both sets of categories (from the “bottom up” and “top down” analyses) and combined them. This involved some rearranging, but I felt these new categories represented the experiences of the participants. These four categories became the organizing framework for discussion each teacher’s experience of tensions and dilemmas in each case report. Between the participants, two of the categories were the same and two were different. I then used these four categories as an organizing framework for writing about how each teacher negotiates the particular tension or dilemma.

Next, I wrote the individual case reports. The structure for each report includes a description of the school and teacher’s background, the teacher’s view of classroom assessment, a description of classroom assessment practices (grouped into assessment for learning and assessment of learning), tensions and dilemmas in classroom assessment (using the four main categories described above), and how each teacher negotiates these

tensions and dilemmas (again, using the four main categories). These case reports included rich descriptions of the context and physical situation in addition to “good raw material” (Stake, 1995, p. 102) and narrative description. As Stake argues, the contextual information and descriptions of the physical situation are important to

develop *vicarious experiences* for the reader, to give them a sense of ‘being there’ ... There should be some balance between the uniqueness and the ordinariness of the place. The physical space is fundamental to meanings for most researchers and most readers (p. 63, emphasis in original).

It is important to include “thick description” (Geertz, 1973, cited in Stake, 1995) in the case report. This involves using the language of the participants as much as possible⁴. Each of the case reports is written to address the two research questions (What tensions and dilemmas do secondary mathematics teachers experience as they try to implement current assessment reforms? and How are these tensions and dilemmas negotiated in classroom practice?). For readability, each case report includes a table summarizing the tensions and dilemmas that Amy and Carina experienced, as well as how they negotiate these tensions and dilemmas (Table 5.1 and Table 5.2).

I felt it was important to provide summaries to each of the participants as a form of member checking. Since the individual case reports were each 30-40 pages (double spaced), I decided to summarize them in 10 (single spaced) pages out of respect for the participants’ time. I also felt it was more likely that the participants would read and respond to a shorter summary. I eliminated some of the explanations that I thought would seem repetitive to the participants, and reduced the number of quotes that were included. This

⁴ In using the words of participants, Seidman (2006) encourages the deletion of some of the repetitious idiosyncrasies of speech such as “like” and “um” while maintaining the authenticity of what is being said, in order to enhance the readability of the text. See Appendix K for an example of some of the choices I made during transcription to ensure readability.

process allowed me to spend further time refining some of the details of what data to include.

During this phase of writing the case summaries, I began to step back from the cases to perform cross-case analysis (Stake, 2006). I began by writing a summary of similarities and differences between the cases. At the same time, I worked on refining the conceptual framework for the study. While I had an early focus on comparing the data to dilemma frameworks from Windschitl (2002) and Tillema and Kremer-Haydon (2005), after revisiting some previous reading I felt that McMillan's framework for assessment decision making (2003) and his framework of four components of assessment (2004) resonated more strongly with the data. For the cross case analysis, I began by listing the four components of assessment (purpose, means for eliciting data, interpretation and use of assessment data) and considered which of these were the main sources of tensions and dilemmas in Amy and Carina's practice. Following McMillan (2003) and Saxe et al. (1999), I then considered external and internal factors that appeared to influence each teacher's assessment decision making. Next, I considered what sorts of assessment dilemmas the participants experienced during the study. Finally, I compared how Amy and Carina negotiate their tensions and dilemmas in classroom assessment. This analysis is included in the first part of Chapter 6.

In my final consideration of the data, there were three important ideas that I wanted to discuss. Since I felt that these were a synthesis of the research study as a whole, I decided to include them after directly addressing the research questions. Influenced by Stake (1995, 2006) and Yin (2003), along with a case study by Delandshere and Jones (1999), I decided to present these as assertions and these are included in the second part of Chapter 6.

Credibility and Verification

As in most other aspects of qualitative research there are no clear guidelines for establishing the validity of an inquiry (Creswell, 1998; Whittemore et al., 2001), but here I present an argument for the credibility of process and the assertions made in this study. Credibility refers to “the conscious effort to establish confidence in an accurate interpretation of the meaning of the data” (Whittemore et al., p. 530).

The design of this study included four separate visits with each participant, including two or three classroom observations with debriefing interviews immediately following. This allowed for an analysis of the internal consistency of each participant’s perspective (Seidman, 2006). The case study reports were written with the rich, thick description that is essential to qualitative research in order to establish “confidence in its authenticity” (p. 26). I included numerous quotations in the words of the participants to allow the reader to get a sense that the participants grappled seriously with the questions and issues through things such as “the syntax, the pauses, the groping for words, the self-effacing laughter” (p. 25). Inclusion of longer passages allows the reader to see instances where “the interviewer has kept quiet, not interrupted her, not tried to redirect her thinking while she was developing it; so her thoughts seem to be hers and not the interviewer’s” (p. 25). In addition, I have attempted to provide great detail and make apparent my role and point of view (Stake, 1995).

Triangulation is widely regarded as an important aspect of the credibility of qualitative research (Creswell, 1998; Stake, 1995, 2006). Stake (2006) describes the process of triangulation in case study research as “repetitious data gathering and critical review of what is being said” (p. 34) and “being redundant and skeptical in seeing, hearing, coding, analyzing, and writing” (p. 77). It involves the convergence of descriptions and

interpretations over different sources of data over time (Creswell, 1998). In this study, I met each participant four times and collected multiple sources of data so I had opportunities to consider the data and my interpretations within each interview and observation, and across all of the data as part of the analysis. This was a continuous and recursive process to ensure that my interpretations were defensible, that the data matched the assertions, and that the assertions matched the data (Stake, 2006).

Member checking is the other procedure widely associated with establishing credibility in qualitative case studies (Creswell, 1998; Stake, 1995, 2006). This involves inviting the participants to examine the data and to offer comments. Any input from the participants at this stage is regarded as new data and “contribute to the revision and improved interpretation of the reporting” (Stake, 2006, p. 37). In this study, I provided each participant with a summary of her individual case report, rather than the transcripts or full case reports. I decided not to share the actual transcripts due to their length (the transcripts of Amy’s interviews were 49 single-spaced pages and Carina’s were 105 single-spaced pages) and the fact they contained the usual idiosyncrasies of speech which can be off-putting for participants (Seidman, 2006). As previously mentioned, the case reports themselves were also quite lengthy so I felt that a summary was more appropriate. Providing these case summaries allowed me to check some of my interpretations with the participants to ensure that I was presenting their perspectives and experiences as accurately as possible. To share the case summaries, I contacted each participant via email and asked if she would like to receive an electronic copy via email or a hard copy in the mail. When each participant replied, I invited her to make any comments by email, phone or in a face-to-face meeting. Both participants requested to receive their summaries via email and sent their brief comments through email.

CHAPTER 5: THE CASE STUDIES

In this chapter, I present case studies of two Ontario teachers, Amy and Carina. For each of the case studies, I include a description of each school and teacher's background as well as each teacher's view of assessment and her classroom assessment practices. I then discuss the tensions and dilemmas that Amy and Carina described and how each teacher negotiates these in practice. In order to provide context for these case studies I begin with a description of some aspects of curriculum and assessment in the context of Ontario that are contained in various provincial policies and initiatives. This includes an outline of recent revisions in provincial curriculum and assessment policy, some Ontario-specific terminology, and a discussion of some of the issues and challenges in classroom assessment recently acknowledged by the provincial Ministry of Education (OME, 2008).

Curriculum and Assessment in the Context of Ontario

Over the past decade, there have been substantial reforms in both curriculum and assessment in Ontario to reflect current thinking in mathematics education and in classroom assessment (Fullan, 2009; Levin, Glaze & Fullan, 2008). These reforms are closely aligned with the vision for mathematics and assessment proposed by the NCTM (1995, 2000), as well as to broader assessment reforms outside of mathematics education from countries such as Great Britain (Black & Wiliam, 1998a; Leahy et al., 2005; Gipps, 1994, 1999), the United States (Guskey, 2003; Shepard, 2001; Stiggins, 2002, 2005) and Canada (O'Connor, 2002, 2007a). Between 1997 and 2000, new provincial curriculum and assessment policies were released and implemented in Ontario for all grades and subject areas (K-12). These documents described practices aligned with current reforms and represented substantial changes from previous practice. Along with the release of the new curriculum, the OME provided funding for teacher professional development, textbooks and other curriculum

supports, and graphing calculators for each school. The OME also acquired several province-wide software licenses (including dynamic geometry and statistical software, which are referred to in the mathematics curriculum). The mathematics curriculum was further revised between 2005 and 2007 (e.g., OME, 2005, 2007), but these changes were not as extensive.

In Ontario's secondary schools, students are streamed into *Academic* and *Applied* courses for mathematics in Grades 9 and 10⁵, and into *University*, *College* and *Workplace* courses in Grades 11 and 12. The mathematics curriculum lists expectations, or standards, for each course in *curriculum documents*. Many of these curriculum expectations contain verbs such as design, construct, explore, compare and interpret, while others require students to pose and solve their own problems. This active approach to mathematics was emphasized in the most recent curriculum revision (2005-2007) with the inclusion of mathematical process expectations to be integrated throughout each course⁶.

The curriculum documents for all subjects include a section on assessment and evaluation. While the content of this section remains largely unchanged from the previous (1997-2000) version, it is now included in all subject-specific curriculum documents, rather than in a separate document so it is more visible to teachers. *Assessment* is defined as “the process of gathering information from a variety of sources (including assignments, demonstrations, projects, performances, and tests) that accurately reflects how well a

⁵ The Ontario curriculum describes the difference between Academic and Applied courses as follows “*Academic courses* develop students’ knowledge and skills through the study of theory and abstract problems. These courses focus on the essential concepts of a subject and explore related concepts as well. They incorporate practical applications as appropriate. *Applied courses* focus on the essential concepts of a subject, and develop students’ knowledge and skills through practical applications and concrete examples. Familiar situations are used to illustrate ideas, and students are given more opportunities to experience hands-on applications of the concepts and theories they study.” (OME, 2005, p. 6)

⁶ In Ontario the seven mathematical processes are: problem solving, reasoning and proving, reflecting, selecting tools and computational strategies, connecting, representing, and communicating.

student is achieving the curriculum expectations in a course” (OME, 2005, p 17)

Evaluation is used for “the process of judging the quality of student work on the basis of established criteria, and assigning a value to represent that quality” (p 17) The curriculum document states that all curriculum expectations, including the process expectations, must be addressed in instruction, but that course evaluations focus on overall expectations, or the “big ideas” The assessment and evaluation section opens with the statement “The primary purpose of assessment and evaluation is to improve student learning” (p 17) and contains many statements that support teachers’ use of formative assessment A number of specific requirements are set out for classroom assessment practice including that students receive descriptive feedback on their work, that assessment information be used to determine students’ areas of strength and weakness, and that assessment be used to help teachers adapt curriculum and instructional approaches In addition, the curriculum documents indicate that assessment and evaluation must be varied in nature, allow students to demonstrate the full range of their learning and promote students’ ability to set goals and assess their own learning (e g , p 17)

A rather unique aspect of classroom assessment in Ontario is the *Achievement Chart*, which was introduced in 1999 and is included in the curriculum documents for all subject areas, including mathematics The Achievement Chart is defined as a “standard province-wide guide” (OME, 2005, p 18) that describes four levels⁷ of performance standards in four categories of knowledge *knowledge and understanding*, *application*, *communication* and *thinking* The curriculum documents state that the Achievement Chart “enables teachers to make judgements about student work that are based on clear performance standards and on a body of evidence collected over time” (p 18), “guide[s] the

⁷ Level 1 is 50-59%, level 2 is 60-69%, level 3 is 70-79%, level 4 is 80-100%

development of quality assessment tasks and tools (including rubrics)” (p. 18), “assist[s] teachers in providing meaningful feedback to students” and “provide[s] various categories and criteria with which to assess and evaluate student learning” (p. 18). This description has been interpreted differently between school districts in Ontario (OME, 2008). For example, some districts have set out specific guidelines that require teachers to collect and record assessment data for each of the four categories of the Achievement Chart. Some of these districts, including the one where Amy and Carina both work, require teachers to construct their grades using specific mark-weighting guidelines for each category.

Since 1999, Ontario has had a standard province-wide report card that requires teachers to enter a percentage grade for achievement and a separate mark for each of five learning skills⁸. A Ministry of Education document on the provincial report card for Grades 9 to 12 states that “*The assessment and evaluation of learning skills is distinct from and should not influence the determination of percentage grades*. Teachers are encouraged, however, to include comments about the student’s learning skills in the ‘Comments’ section of the report card.” (OME, 1999, p. 10, emphasis in original). This is consistent with recommendations from the measurement community to separate what McMillan (2003) terms “academic enablers” such as effort, homework and motivation from grades that are meant to represent students’ achievement. Many studies on teachers’ grading practices have found that these are often combined in grades (Brookhart, 1994; McMillan, 2001).

The subject-specific curriculum documents in Ontario mandate that 70% of a student’s final grade be determined from regular assessments done throughout the course, including items such as tests, assignments and performance tasks. This portion of the grade

⁸ The five learning skills are *works independently, teamwork, organization, work habits/homework, and initiative*. The marks assigned for these are Excellent, Good, Satisfactory or Needs Improvement

“should reflect the student’s most consistent level of achievement throughout the course, although special consideration should be given to more recent evidence of achievement” (OME, 2005, p. 22). The other 30% of a students’ final grade is determined from one or more evaluations towards the end of the course such as an exam, performance task or essay.

Recently, the Ministry of Education also introduced a *Student Success* initiative that allows students at risk of failing a credit to receive further opportunities to demonstrate their achievement in a particular course. While funded by the province, many details of this initiative are left largely up to individual districts and schools. Through this program, there is funding for a Student Success teacher at every secondary school in Ontario⁹. This teacher works with classroom teachers to monitor and assist students who are at risk, and provides remediation to students who are at risk of not achieving a credit, or have failed a course. These students are given an opportunity to “rescue” a credit by completing work at the end of a course, or to “recover” a credit during a subsequent semester by demonstrating achievement of the overall course expectations without having to repeat the entire course.

Revisiting Ontario’s Assessment Policy

While there were no changes in Ontario’s assessment policy during the 2005-2007 curriculum revisions, a draft document was released by the Ministry of Education in January 2008 (OME, 2008) to start a process of clarifying the provincial assessment policy for implementation in September 2010. In this document, the Ministry of Education acknowledges that concerns have been raised by educators and other stakeholders “about inconsistent and uneven implementation” (p. 1). In particular, the Ministry acknowledges that the assessment policy contains “rich, challenging statements that need to be unpacked so that they can come alive in the classroom” and that the policy statements need to be

⁹ There are approximately 800 secondary schools in Ontario (Levin, Glaze & Fullan, 2008)

understood and implemented in order to ensure that there is “more meaningful data to support student learning.” (p. 2-i). A section at the back of this draft document lists 45 outstanding issues that have “frequently arisen” (p. 10-i) and these include:

- Why is the achievement chart so important in the Ontario Curriculum?
- How is the achievement chart used to develop rubrics and how do the two differ?
- Are the four categories of the achievement chart of equal importance?
- How should the achievement chart categories be balanced to determine marks?
- What is meant by “more recent... most consistent”?
- How should assessment evidence be collected for secondary students for the 70 per cent/30 per cent mark division? Should there be policy related to the timing of assignments?
- What role does professional judgement play in the evaluation of student achievement?
- Should student achievement be reported by achievement chart categories?
- Should students receive a zero for missed assignments?
- Should students be deducted a certain number of marks for late assignments?
- What should the criteria be for deducting marks?
- How should the 0–50 mark range be handled when students’ marks are averaged? (pp. 10i-10iv)

Many of these questions are echoed in the concerns raised and challenges experienced by the two participants in this study.

In this discussion document, the Ministry of Education offers clarification in a number of areas. For example, the Ministry argues that the Achievement Chart categories

help teachers demonstrate the interconnectedness of learning. They focus the teacher’s assessments and feedback on knowledge and skills, rather than the particular tools or methods of assessment, such as tests or labs, which may vary. They assist teachers in focussing assessment on the higher level skills – thinking, communication and application – that allow students to manipulate, organize and use the knowledge and skills they have acquired. (OME, 2008, p. 3-ii).

In addressing the issue of late and missed assignments, the Ministry considers these disciplinary issues best addressed by teachers and school administration, and offers a list of strategies for dealing with late and missed assignments. It emphasizes that the separation of achievement from learning skills and behaviours provides better information to students and parents and stresses that “mark deduction should be used as a last resort” (p. 6-ii).

The revisions to curriculum and assessment in Ontario that have occurred over the past decade are aligned with contemporary reforms (Levin et al., 2008; Suurtamm & Graves, 2007). Ontario's assessment policy places great emphasis on the use of assessment to improve students' learning, use of a variety of tools and strategies, and the integration of assessment and instruction. With respect to evaluation and reporting, teachers are required to separate academic achievement from learning skills or behaviours, and are encouraged to use their professional judgment in considering students' most recent, most consistent work [OME, 2005]. Along with provincially funded professional development initiatives, Ontario's curriculum reforms are considered to have resulted in a relatively successful implementation (Fullan, 2009; Levin, 2009). However, as some research has pointed out (Suurtamm & Graves), and as acknowledged by the OME (2008), some serious issues remain in the area of classroom assessment. The remainder of this chapter presents case studies of two Ontario teachers' experiences with tensions and dilemmas in classroom assessment, and how they negotiate these in practice.

Case Study #1: Amy

This section presents the first case study of Amy. First, I present some contextual information about Amy and the school she teaches at, as well as describe her view of assessment and provide an overview of her classroom assessment practices. I then discuss four tensions and dilemmas that Amy described experiencing in her classroom assessment practice. Finally, I describe how she negotiates these tensions and dilemmas in practice.

Description of the School and Teacher's Background

Amy is currently teaching at Greenwood High School¹⁰, a semestered suburban school in Ontario. The school serves approximately 1100 students from grades 9 through 12 and students take four courses each semester. The main hallways of Greenwood are lined with photos of graduating classes from years past, copies of students' awards, and mounted newspaper clippings recognizing arts productions. The classrooms used for mathematics are all together in a long hallway which is near a department office for the mathematics teachers where they all have space to work.

Amy was in her late-twenties and in her sixth year of teaching during this study, the last five of which were at Greenwood. She has an undergraduate science degree and is qualified to teach both mathematics and science from Grades 7 through 12. In addition, she also holds an Honours Specialist qualification in mathematics which means that she has a strong concentration of undergraduate mathematics courses and an additional qualification in curriculum leadership. While she did teach some science courses in her first few years, Amy has taught mainly mathematics including most of the courses from Grade 9 through 12. Amy is highly involved with extracurricular student activities at her school as well as with several mathematics projects at the district level. These have included a lesson study, two book studies, designing items for district assessment initiatives, participation on committees and chairing a group working on revisions of curriculum support materials. Amy is a very well respected teacher within her district and school and from the lessons I observed appeared to have an excellent rapport with her students.

During the semester of this study, Amy taught Grade 9 and 10 mathematics courses. Her classroom is an average sized room with math posters and samples of student work on the bulletin boards and walls. There are 32 desks in the room and students sit in pairs with

¹⁰ The participants' and schools' names are pseudonyms

the desks aligned in three rows, all facing a wall with an interactive whiteboard mounted on it. At the front of the room, there is also a student-made word wall with math terms and diagrams. On each side of the interactive whiteboard is a large window through which students can see big leafy trees. As students are seated there is a blackboard to their right, a bulletin board to their left, and a blackboard at the back of the room. Mathematics manipulatives and tools are neatly stored along the sides of the classroom, on shelves below the bulletin board and blackboard. There are two large desks at the front of the room on either side of the interactive whiteboard, each with a computer. This semester Amy's Grade 10 class has 22 students, and her Grade 9 class has 26 students which is the maximum allowed under the collective agreement between the district and teachers' federation.

During both of the lessons I observed, one Grade 9 and one Grade 10, the classroom environment was very positive. Students appeared to be engaged in the lessons and most students participated actively in the activities and class discussions. Students knew where to find equipment and tools they needed, such as a small whiteboard and marker, and graphing calculators, and got up freely to get what they needed. In both classes students worked in different groupings, including pairs, small groups and individually, and moved easily into and out of the different groupings. Amy used a large digital timer on the interactive whiteboard during student work periods so that she and the students could keep track of their time. Students in both classes appeared to feel comfortable sharing their thinking, both in volunteering answers, and being called on by the teacher to explain their thinking. Students in both classes also presented written solutions on the interactive whiteboard, which they appeared to be very excited about, and received feedback from both Amy and other students in the class.

View of Classroom Assessment

Amy has experienced a great deal of curriculum and assessment reform in her six years teaching. Revisions to the province-wide mathematics curriculum were phased in over three years, starting in her third year. During this period, district and school assessment policies were also revised to ensure alignment with the provincial assessment policy. Amy views her classroom assessment practice as still evolving. She stated that her “understanding of it just keeps developing as more things come out” (Interview, May 21, 2009), referring to recent provincial and district assessment policy and support documents. Describing aspects of her assessment practice during this study, she used phrases like “always changes” (Interview, May 21, 2009) and “lot of shifting” (Interview, May 21, 2009).

Amy is an early adopter of some of the tools and strategies that the district and her school have encouraged as part of the curriculum and assessment reforms in Ontario, although she has found this can sometimes lead to frustration:

I’m game for trying whatever is coming down the pipelines, but I try it too quickly [laughs]. I’m usually the one that experiences the roadblocks and then says “hold on a second”. Sometimes I get discouraged and go back to what I was doing, but usually, if I’ve tried something new, I’m just going to stick with it (Interview, May 21, 2009)

While recognizing the importance of school-level conversations and experiences, Amy attributes changes in her vision for her own assessment practice as coming largely from central staff at the district level:

So it really is my involvement through Board activities and seeing what’s coming down from above and then getting involved in different opportunities to talk to people that are making those decisions. Then talking to colleagues that are trying to implement those practices. It’s just that opportunity for sharing. It all just sort of comes down. As soon as you hear about an idea you either sort of chuck it or you take it and you try to go with it (Interview, June 18, 2009)

She has

opted to participate in a lot of things because I feel [pause], I feel that people that are making decisions hopefully know a little bit about what they're doing. I'd rather be in the know and I always like to push things forward so I've always taken that leap and wanted to participate in things (Interview, May 21, 2009)

Amy described how she has made changes in her own assessment practice over the past few years:

There was a lot of just building one piece into the next, into the next. And reading more, hearing more, doing more professional development, talking, having staff meetings, talking to my department head. That just puts one piece of the puzzle in [pause]. Every week or two you think about something else or you change or you modify a little bit. That's how I think the changes have happened. That's how my practice has been modified is that I've spent the first while figuring out what it was all about [laughs] and then slowly started to say, "You know what? This part of what I'm doing really probably should be changed. And I'm going to try and change it." If it hasn't worked well I've gone back and then tried to change another piece and some pieces stick and some pieces don't (Interview, May 21, 2009)

The vision of classroom assessment that Amy is trying to enact is "very different" (Interview, May 21, 2009) from what she experienced as a student. She describes her experiences with assessment as a student as being "all about the numbers" (Interview, May 21, 2009). Everything that students handed into the teacher counted and the emphasis was on "the little marks. It was bonus marks and things like that. You were always striving to do more and get more credit" (Interview, May 21, 2009). In one instance, her math teacher was frustrated at poor attendance on the Friday before a holiday so he put six questions up on the blackboard along with the answers and told the students to copy them down to earn ten marks per question. As for the types of assessment she experienced:

It was all tests. That was it. Quizzes and tests and our marks were broken down. A certain percentage for quizzes, a certain percentage for tests and a certain percentage for the exam. And that was it (Interview, May 21, 2009)

Description of Classroom Assessment Practices

This section describes aspects of Amy's classroom assessment practice to help set a context for understanding the challenges she has experienced. Over the past few semesters, Amy has been focusing on developing formative assessment strategies including making her lessons more interactive so she can get a better sense of students' understandings, and increasing the feedback that students receive on their learning. She has also been trying different methods of recording assessment information to consider when determining students' grades. While it is difficult to break classroom assessment practice down into discrete parts, it is useful here to discuss Amy's use of assessment *for* learning and assessment *of* learning.

Assessment for Learning (Formative Assessment)

Amy uses both formal and informal assessments to get a sense of students' learning, inform her teaching and give students feedback. Some of the strategies she uses include circulating and listening while students work, interactive whole class strategies, students' oral presentation of solutions, and formative quizzes.

Circulating and listening to students' discussions.

During both lessons that I observed in Amy's classes, she circulated extensively to listen to students' discussions in pairs or small groups. She views this as an important way to get a sense of students' understanding:

Listening to a lot of what they do in pairs when they work together and they're taking up homework together and they're doing an activity together. Listening to the discussions that are happening to get a sense of which students have a good understanding of what's going on and which ones are still behind. (Interview, May 21, 2009).

Sometimes she posed questions to a group about their strategy or solution, and other times she responded to students' questions. After I observed her Grade 10 class working on a set

of trigonometry problems in groups, she described what she was thinking about as she circulated:

The first thing I was looking for was whether or not they were on the right track as far as the question was going. So, focused on task and also not completely off. But I didn't want to jump in and correct all of their errors. For example, if they were using the wrong ratio or if they'd labelled their diagram incorrectly I let them go with that. I also wanted to see who was taking more of a leadership role in terms of solving the question. And who was more complacent. Just sitting back and trying to go with the rest of the group. There were only two or three that weren't actively participating." (Interview, June 1, 2009).

Interactive whole class strategies.

Amy has been using several interactive strategies that allow her to get a sense of students' understandings that she can use to make quick decisions during a lesson about what activities to do and how long to take with these activities. Sometimes, her students respond to a question on small whiteboards and then hold up their answers for her to see. Each student gets a whiteboard (about 20 cm by 30 cm) with a dry-erase marker and eraser. During the Grade 9 lesson I observed, Amy asked students to respond to eight multiple-choice questions from a previous version of the provincial EQAO assessment¹¹. Students had about one minute to determine their answer, and then all students held up their whiteboards with the letter A, B, C or D written on it. She explained that having the small whiteboards

captures something that you wouldn't normally get ... and it is such a visual. I mean I wish you could have seen it from the front because it's such a visual. Suddenly you see all these 'C's go up or you see 'A, B, C, D, S, T', like when they were doing the points on the graph they all had different and you can just get such a quick idea. "Okay, wait a second. We're not on the same page". (Interview, May 28, 2009)

¹¹ At the end of Grade 9 mathematics, students write a provincial mathematics assessment administered by the provincial Education Quality and Accountability Office (EQAO)

With questions where there was no clear consensus, Amy asked several students to explain their choice and then encouraged other students to respond. Amy described an example of how she has used the whiteboards to get a sense of students' understandings of algebraic expressions from written statements:

I would put something in words up on the board and I had a selection of expressions for them to choose from and they would write on their whiteboard and hold up what they thought the best expression that matched was. It was great because it was such a quick way of seeing where students were at and where the misunderstandings were. Cause that's the biggest issue, I think, for assessment. (Interview, May 21, 2009)

Another quick and interactive strategy that Amy uses with her classes is "agree/disagree". Sometimes students respond to a mathematical statement by writing "A" or "D" on their whiteboard, and other times students respond by a show of hands. Amy insists that all students respond:

Sometimes I'll say "raise your left hand if you think it's A, raise your right hand if you think it's B and raise both hands if you're not sure" so that it's not just people raising their hand. (Interview, May 28, 2009)

Amy described these sorts of quick, informal assessments as allowing her to "change gears" (Interview, May 28, 2009) when she can see that students need more time on a concept. Since she has taught the Grade 9 and 10 courses several times, she plans her lessons with contingencies and options which allows her to use assessment to inform her instructional decisions in a moment-by-moment way. Amy finds that having an interactive whiteboard in her classroom is helpful to her ability to use assessment to inform her instruction since she has quick access to resources such as online curriculum supports, files from her computer, and software programs, without having to make photocopies and transparencies of contingent materials that may not be used.

Students' oral presentation of solutions.

Another way that Amy uses assessment informally is through the presentation of students' solutions. During both lessons I observed, students wrote solutions on the interactive whiteboard and got feedback from both the teacher and their peers. In the Grade 9 class, several students volunteered to put solutions on the whiteboard, and then the class discussed the solutions and made suggestions or corrections. In the Grade 10 class, students were working in groups on a set of trigonometry problems and each group presented a solution to one problem. Amy encouraged classmates to ask questions or challenge parts of the solutions and then offered some feedback to each group focusing on aspects of the solution such as notation, form, possible alternate approaches, checking the reasonableness of answers and common mistakes that she has seen from students' work in the past. After this class, Amy described what she was thinking about when she was giving feedback to each group:

I wanted to try and dispel some of those myths and resolve some of the issues with reading the question and what they're going to draw when they do their evaluations ... I try to not be too nit picky about things because, especially with the focus of today, I didn't want it to be on technicalities. I want it to be on the big picture (Interview, June 1, 2009)

Amy views these presentations as important opportunities for feedback to her students and does not include them as part of students' grades.

Formative quizzes.

In addition to these different types of feedback during class time, Amy sometimes uses quizzes in a formative manner to give students feedback on their learning. The feedback she gives on these types of quizzes tends to be anecdotal. Most of these quizzes are written, but she also does some oral quizzes with her Grade 9 students, to ensure that students get quick, individualized feedback. Amy keeps records of these formative assessments to help her to see if students are missing assignments or have not demonstrated

achievement of particular course expectations in advance of assessments that will count in their grades. These records are often checklists recorded on a hard copy of a class list, along with notes or symbols indicating which students had late or missing work, or work not completed at a satisfactory level.

Assessment of Learning (Summative Assessment)

Students' grades in both of Amy's classes are mainly determined from the results of formal written assessments including tests, quizzes, performance tasks, and a final exam:

In terms of the evaluations I still do quizzes and tests - a lot of pencil and paper tasks. We've introduced performance tasks as sort of the *thinking* portion of their marks. (Interview, May 21, 2009)

Tests.

Tests are a major component of students' grades in Amy's classes. These are administered at the end of each unit of study and she designs them to include all four categories from the provincial Achievement Chart – *knowledge and understanding*, *communication*, *application* and *thinking* (described earlier in this chapter). Amy marks all of her Grade 10 tests using a marking scheme for the *knowledge*, *application* and *thinking* categories. She uses a generic rubric for *communication* and marks this holistically for the whole test. For example, a student in Amy's class could get 14 out of 15 on the *knowledge* component of the test, 7 out of 10 for *application*, 3 out of 5 for *thinking*, and a level 3 for *communication*. She then inputs these marks as four separate entries in her electronic mark book. This year Amy decided to mark her Grade 9 tests in a different way. The Grade 9 tests do not have specific marks for each question, rather, they given an overall level for each category using the Achievement Chart. These levels are entered separately into Amy's electronic mark book. She appeared to believe this more holistic method was more appropriate for students in the Applied level course since she felt these students "tend to be

less focused on the number anyhow” (Interview, May 21, 2009) and was trying this because some other teachers of this course in her district were trying similar things. During the Grade 9 lesson I observed, Amy handed back three pieces of assessment from the previous unit, all marked using levels. Students seemed comfortable with this system and appeared to understand what the levels meant.

In the *thinking* section of a test, Amy often gives students a choice between two questions:

So giving them the option between two very different styles of questions, it releases some of that anxiety. It shortens the test for them, too. So they’re not quite as stressed about timing. And it also gives them a chance to show me what they actually know how to do (Interview, May 21, 2009)

Amy explained further:

I know for myself when I’m stuck on something, I just need to toss it around for however long ... to put a time limit on someone’s thinking I think is tricky, especially when they’re in a pressured situation (Interview, May 21, 2009)

Quizzes.

Amy also uses traditional quizzes as part of students’ grades. These generally cover smaller chunks of material, have a lower weight than tests, and do not assess all four of the Achievement Chart categories. Sometimes Amy only marks one aspect of a quiz, such as *communication*, and gives anecdotal feedback on other parts to help students prepare for an upcoming test or task. The main difference between a quiz included in students’ grades and a formative quiz is that Amy gives her students advance notice so that they are not surprised, and then includes the in her electronic mark book (rather than recording on a checklist).

Performance tasks.

Amy uses performance tasks several times during each course to allow students to demonstrate thinking and problem solving.

We've introduced performance tasks as sort of the *thinking* portion of their marks. I try to make them more hands-on and more interactive rather than the paper and pencil which we kind of have covered [laughs]. Very, very well [laughs]. (Interview, May 21, 2009)

One example from her Grade 10 class was a “parabola art” project:

It's kind of neat. Some of the things they're coming up with. So there were six parabolas that they had to draw. And then there were six equations in different forms they were given. And they had to do a bunch of analysis. Where is the vertex? Where were the zeros? Etcetera ... They were allowed to use the graphing calculators to do some of the work there. Those six parabolas have to be drawn on and then there were four they were to create on their own to complete some sort of picture of something ... I've got an octopus one. I've got a long-haired one. I've got a ghost one. Yeah. There's a flower and a bunch of honey bees floating around it. Yeah, there's a lot of neat ones. (Interview, June 1, 2009)

Another task that I saw the Grade 10 students preparing for was measurement of the height of an inaccessible object using a clinometer, and Amy spent about 15 minutes during one lesson I observed helping students plan for this task which they were going to do the following day. In the Grade 9 class I saw students receive two small tasks after they had been marked, and Amy described another “Sunflower” task where students get various forms of data (graphs, tables of values and statements) and have to match them up and create any missing pieces. She had students work in a group to come up with a strategy and then complete the rest of the task independently. Performance tasks have been a particular focus in the school district and there have been several projects over the past few years to produce and disseminate quality tasks for teachers to use in the classroom. Amy has been involved in some of these initiatives.

Summative tasks.

For the summative portion of students' grades (described earlier in this chapter), students in Amy's classes write a final exam and do a performance task during the last 2-3 weeks of class. In addition, Grade 9 students also participate in the mandatory province-wide assessment.

Amy's exams are traditional in nature and are written in classrooms during a dedicated exam period at the end of the semester. The exams are common to all students at the school taking the course that semester, and are graded with a marking scheme that is agreed upon and used by all teachers of the course. The summative performance tasks are similar in nature to the performance tasks used through the course, except they generally require students to pull together information from several units. These tasks are marked using a generic rubric, and Amy marked her Grade 10 summative tasks this semester along with her two teaching colleagues to ensure consistency.

Second chances.

In Amy's Grade 9 class, students can receive second chances on assessments during "opportunity club" days. These second chances are encouraged as part of the provincial Student Success initiative through which students can demonstrate curriculum expectations they did not initially achieve. Two or three times a semester Amy prepares individualized folders with assignments for students to do under the supervision of a teacher associated with the provincial Student Success initiative. Students spend half a day out of their regular classes to get caught up on their work that they either initially missed or did not complete to a satisfactory level. This can include re-doing tests and assignments. When Amy assesses these "second chance" assignments, she is looking to see, "okay, has this sunk in anymore? You've seen these questions already once, we've done more stuff - has anything

else changed?’ For some of the students they did do a better job” (Interview, May 21, 2009).

Determining a final grade.

To determine students’ grades, Amy uses commercial grading software. As described earlier in this chapter, term work counts for 70% of the grade, and the summative pieces count for the remaining 30%. The school district requires that teachers enter all marks by Achievement Chart category, so a particular test may require four different entries (one for each category – *knowledge and understanding, application, communication and thinking*), and has set out ranges for the weight of each category.

All of my assignments are weighted depending on their value. I put the highest value on their pencil and paper tests, their five unit tests throughout the semester and then I weight everything else around those unit tests accordingly - whether I feel they’re equal in value or they’re slightly less in value. So, a lot of our *thinking* assignments are usually weighted at about a 3, quizzes are 2, and then smaller assignments go in as 1. (Interview, June 18, 2009)

The software then lets her analyze several different measures of central tendency including average, median, blended mode and blended median, and she also considers students “most recent most consistent [performance] over the semester” (Interview, May 21, 2009). In cases where a student’s achievement has increased over the course of a semester, or if he or she has done particularly well on the course summative assignments (e.g., final exam and performance task), Amy uses the measure of central tendency that she feels best matches the student’s achievement, or simply “bumps” up the final mark by several percent. She has used several different software programs over the past few years as part of different district initiatives.

In a previous semester she used a very different method for determining grades with one of her classes. This involved recording all assessments on a single sheet of paper along

a continuum for each category. She colour coded the notations so that she could tell which ones represented major assessments and which were minor. Before midterm marks were due, Amy conferenced with each student to determine their report card grade. She was very satisfied with this process:

It was really good because it was almost the same percentage to the 'T' as I would have come up with so they saw exactly the same thing that I saw. They didn't think their marks were coming out of nowhere (Interview, May 21, 2009)

Tensions and Dilemmas in Classroom Assessment Practice

Through analysis of all of the data sources, four categories of tensions and dilemmas experienced by Amy were identified and will be described in this section: policy in relation to assignment deadlines, issues in grading and reporting, colleagues resisting current reforms in classroom assessment, and using assessment effectively to improve students' learning.

Policies Related to Late and Missed Assignments

One area that Amy identifies as a particular challenge is interpreting and implementing policies regarding assignment deadlines. She has found that different stakeholders interpret policies at provincial, district and school levels in different ways. The provincial assessment policy found in the curriculum documents does not address this issue directly, but the Ministry of Education position is that late and missed assignments are learning skills that should be reported separately from a student's achievement of curriculum expectations to prevent a distortion of students' academic achievement (OME, 1999, 2008). This is different from traditional practice where marks were routinely deducted for late assignments and zeros given for missed work. Each school district and individual school write more specific assessment policies that are aligned with the

provincial policy, and often include a section for late and missed assignments. Amy finds that these policies are not always clear: “[they’re] not usually laid out how it’s supposed to play out in the classroom” (Interview, June 18, 2009).

In the previous school year, Amy was a member of her school’s assessment committee which updated the school’s assessment policy as required by the district. This committee

came up with a school-wide policy that was an interpretation of the Board-wide policy that was an interpretation of the Ministry policy. So, you’ve got lots of different levels of places to look for support when you’re trying to make a decision and that’s been really hard. The interpretation of the school-wide policy has been taken so broadly by different teachers that everybody’s in a different position (Interview, May 21, 2009)

The policy specifies different steps for teachers to follow for “major” and “minor” assessments that are late or missed, but Amy does not find this particularly helpful in practice:

I think extensions are the biggest thing that are part of our school’s policy that are, according to our Principal very clear, but according to the staff are very unclear. There are a lot of mixed messages being sent. How long do they have to hand things in? ... you always feel like you’re stuck between a rock and a hard place. [As if talking to a student:] “You can’t hand it in. It’s passed the third, fourth, fifth deadline that I’ve now set” (Interview, June 18, 2009)

In the previous semester, Amy experienced a significant dilemma where several students in her Grade 12 class did not submit their summative project on time. She had set a deadline earlier in the semester for the written component of this project, and students were also required to make oral presentations of their work to the class:

[The projects] are not in the folder one week after the deadline. They’re not in the folder two weeks after the deadline. We’re now in the last week of school and it’s their summative mark and they haven’t done a presentation. Now they’re knocking on my door saying “I was sick. I was this. I was that” and “can I just hand it in” and they’re sliding papers under the door saying [to Amy] “I can’t believe you didn’t get it”. (Interview, June 18, 2009)

One student even made a public statement during a school assembly about his missing assignment which upset Amy:

One student stood on stage the day before Christmas holidays, when part of his work was due, and said “oh, by the way, Miss. I just put my work under your door”. Like an hour before Christmas break! It was due far before that and I had said “if I don’t have it for the holidays when I’m marking it you’re out of luck”. And then it was up to me to scramble to find it [the student’s work]. It wasn’t under my door so I didn’t know where it had been put ... And I check with my department head and I can’t get a hold of him because now everybody’s gone for the holidays. So I’m thinking [the student] made a public announcement to try to make a point so that he could say he wasn’t lying about it. If I call him on it now, he’s going to say “well, like I would have gotten up on stage and told you it was there if it wasn’t”. So, yeah. The extensions thing - the whole deadline thing is a huge pet peeve. (Interview, June 18, 2009)

During the semester of this study, Amy faced a similar problem with her Grade 10 class when several students were over a month late in handing a performance task. Students were supposed to complete all of their work in class and were asked to stay after school if the task was unfinished. Amy was frustrated with several students who handed in their tasks on the day of their final exam:

I have kids that never stayed after school. They just brought it to me the morning of their exam. And it wasn’t originally supposed to go home ... now I have zeros in these marks and I’m thinking “well, okay they’re half finished it. What choices do I have?” I could fight the battle. And that’s the thing - it comes down to how much support I’m going to get if I put a zero here and a parent complained. What’s in my best interest as well as the student’s? And you know what? Just take it home and finish it. It’s not worth the fight. Because all I get is resistance from everywhere. (Interview, June 18, 2009)

Amy does not believe that teachers at her school are dealing with this issue in a consistent manner: “Everybody’s kind of doing their own thing. Which is working for some, but it’s certainly not consistent” (Interview, June 18, 2009). She finds the school, district and provincial assessment policies are not clear enough on this issue, which leaves teachers with difficult decisions about how to deal with late and missing assignments, and she feels that some students are taking advantage of the policy.

The issue became quite contentious at Amy's school towards the end of the semester. Several staff members posted comments on the school's internal electronic conference describing incidents where significant numbers of students did not submit assignments on time and the impact that this is having on teachers' workload and students' learning. Many teachers feel that students do not perceive any consequence for late or missed assignments. Amy stated:

It's a very emotional situation including the A and E [assessment and evaluation] policy and what we're expected to do and what people are trying to sort of push us into without realizing the practicality of how to make that happen on a day to day basis ... Everybody is ready to say, "you know what? This is not working. If we're going to make this happen it's going to have to be done differently because it's just not working" and I think that people are just kind of tired of being jerked around (Interview, June 18, 2009)

Issues in Grading and Reporting

Amy has also had trouble with several aspects of grading and reporting. These include interpreting and implementing the provincial Achievement Chart, dealing with limitations of grading software and one incident where her principal overturned a final grade.

Interpreting and implementing the provincial Achievement Chart.

Amy has experienced difficulties interpreting and translating the provincial Achievement Chart (described earlier in this chapter) into practice. The district requires students' grades to be broken down into four categories: *knowledge and understanding*, *application*, *communication* and *thinking*. Since Amy did not experience this as a student - her marks were broken down into categories such as tests, quizzes, and class participation - she was initially unsure how to connect the Achievement Chart to her classroom assessment practice:

For the first three years of teaching I was just trying to figure out what was happening with those categories and levels of achievement. What a rubric was and how do I even make one and how am I supposed to use it and I'd rather just do it by marks. (Interview, May 21, 2009)

Even though she has participated in several district assessment initiatives that have involved some consideration of the Achievement Chart, Amy is still struggling with how to make sense of it in her day-to-day work with her students. She has had many discussions with other teachers at her school, during professional development sessions and with district-level consultants and finds that people have very different interpretations:

The thing I've had the hardest time working with is the Achievement Chart - trying to figure out how that translates into what I'm actually seeing in front of me on paper. Choosing what types of questions to ask to fit the Achievement Chart so that I know that I'm evaluating *application* in an *application* question. And there's been so much discussion. [Name of colleague] has a very different interpretation of *application* than I do and so we've had a lot of discussions that way ... I do see a lot of the descriptions overlapping but in terms of what I see looking at a piece of work, how I interpret that question has been a huge challenge and it has been within our department as well. Just not knowing when you're writing a test or when you're writing something that you're getting the students to complete - which category do you put it in? What are you looking for? So I think PD [professional development] around that would be really useful, but at the same time with reference to real student work because that's where the challenge comes in. Theoretically, it's great. In words they can make it sound good, but when we're seeing a question, actually interpreting that and putting it into somewhere, [more PD] is needed (Interview, May 21, 2009)

Amy is not sure about her district's decision to set specific mark weighting guidelines by category because this does not come directly from the provincial assessment policy:

Really, the curriculum documents don't say that any one [category] should be worth more or less than the other. So it's very arbitrary that that's been decided by our Board. It's their interpretation of the curriculum documents. Yet, some people say that their interpretation is the way it's written and it's really. Well, we're sort of not reading them all the same (Interview, May 21, 2009)

Amy also expressed uncertainty over combining marks in different forms to report a student's achievement within a category. Some assessments are graded with a marking

scheme while others are graded with rubrics so teachers need to combine different types of scores (such as 'level 3' and '9 out of 10') to determine one overall mark. This mark must be expressed as a percentage for provincial report cards. Recent district assessment initiatives have used different assessment tools (such as marking schemes and rubrics), but there has been a recent push to mark exams more holistically according to the categories of the Achievement Chart. Amy is not clear about what direction to take in grading her own classroom assessments:

We're evaluating those categories, within those categories, and we are levelling or marking or whatever we're doing. We're coming up with a sense of the student's achievement in that category. And that's whether you do it with levels or not. I'm not getting clear direction or a clear sense of "this is where we want to go" (Interview, June 18, 2009)

This semester, Amy designed her Grade 9 tests with different sections for each category and then included a photocopy of the relevant part of the Achievement Chart at the end of each section. She was not sure that the open and generic nature of these rubrics was helpful for her assessment: "I just gave [students] a level based on how well they met that criteria. So I didn't make [the rubric] particular to the test, which is good in a way, but also not good because it's not specific enough" (Interview, May 21, 2009).

Limitations of commercial software for grading and reporting.

In Amy's six years of teaching, she has used three different commercial software programs to analyze and determine grades, as well as a more holistic approach (described above). The program that Amy is using this semester is the one promoted by the school district and it is licensed for all teachers to use. The program allows teachers to analyze trends in students' marks and to calculate a grade using four different methods: average, median, blended modes and blended median. Amy wants to consider each student's "most recent, most consistent [performance] over the semester" when she is determining final

grades. This has been encouraged during district professional development sessions, but Amy is having difficulty figuring out how to use the software efficiently to help her do this:

I've tried using [name of software] to come up with my final grades and using the four different outputs and trying to determine which one is best and which one doesn't fit. And looking at the trends. But I find that within distinct units, looking at trends isn't really all that useful because you see a dip in their mark, but it might be a realistic dip. In the quadratics unit they maybe didn't do as well ... You can't just sort of discount that work that they didn't quite grasp in quadratics. So, I'm still struggling with that. (Interview, May 21, 2009)

One aspect of this particular software program that she is frustrated with is the recording and management of assessments that are not included in the final grade. Items entered as "formative" are not displayed on the summary report of a student's achievement. Amy would like to be able to see all of the assessment data that she has for each student in order to make better decisions and to communicate more effectively with students. She would find it useful to keep track of student completion of work as well as overall trends that include pieces of work she is not going to include in the final grade:

Every time I made something in [name of software] out of nothing, it eliminates it from your list of things ... What I was told by the Board was to set up a separate file for assessment that you could use to look at, but it's not as useful because you're really looking at two distinct things that overlap. And so it's hard to overlap them because you see them as separate sets, ... [it] is useful to see the trends and what's, what's not coming in and what am I not seeing evidence of at different points in the semester. (Interview, May 21, 2009)

Amy also worries about some of her colleagues' use of the grading software. At the beginning of each semester, individual teachers need to configure the program according to the district's mark weighting guidelines. In addition, teachers need to decide how to use the software to weight different assignments. Amy is concerned that some teachers do not fully understand this process and make errors that can result in minor assessments counting more than major assessments. She is also concerned that some teachers do not fully understand

the implications of using the different measures of central tendency when determining grades, which is a relatively new feature of the software:

That really frightens me. Cause at least I have a concept of how [name of program] is coming up with their marks. There are so many factors. Depending on what averaging method you use. Whether you weight out of the number of marks it's worth. Whether you set up your bins properly. Those marks are coming out and people are taking them at face value. I don't even think they're considering how that mark is being developed and that's very frightening ... You have to be a math major to figure it out. So, you know, a large percentage of our school is stuck. (Interview, June 18, 2009)

Principal overturning a student's final grade.

At the end of the previous school year, Amy's principal responded to a parent's appeal of a final grade by raising the mark five percent. This incident deeply affected her and she mentioned it several times during the interviews. Rather than using grading software to determine a final grade for the course, Amy had assessed all of the course assignments in levels and then she holistically determined a mark for each category. Amy learned about this method from several teachers at another school and felt it was consistent with current assessment initiatives at the provincial and district levels. The students were aware of this grading method and Amy conferenced with her students to determine the midterm grades.

Upon the appeal, Amy's principal did not support this grading method: "it's not really a solid mark. We don't really have numbers here" (Interview, May 21, 2009). This frustrated Amy since she believed she had enough data to make a valid decision and she felt "more confident about that [grade] coming out of that rubric than I did coming out of a mark program" (Interview, May 21, 2009). She felt that her "professional opinion wasn't valued" (Interview, May 21, 2009) and that her principal did not understand the direction that teachers had been getting from the district's own professional development initiatives:

my administration was not understanding that process which was more frustrating than anything. When that move was sort of coming down from above - to move in this direction and then try stuff like this. We're really going towards levels and we're [supposed to be] looking at how well we're meeting the categories of achievement (Interview, May 21, 2009)

Colleagues Resisting Current Reforms in Classroom Assessment

A significant tension that Amy experiences is the reaction of some of her colleagues to the vision of assessment reform put forth in provincial curriculum documents and the subsequent interpretations by local districts and schools. She feels that some of her colleagues have not moved away from traditional approaches to assessment and are resisting making reforms in their own practice.

One particular situation arose during a guest speaker's presentation to the staff at Amy's school. The workshop focused on the implications of recent changes for grading practices, and Amy felt as though the speaker received an unreasonably hostile reception:

There was a lot of rebellion. It was at the end of the year. It was presented in a way that I don't think people were ready to buy in. They're never going to be ready so I mean at some point you have to kind of say this is the way that things are going and we need to make it happen so you're going to have to figure it out ... It was sort of attack, attack, attack, from the staff. And peoples minds started to close up. (Interview, June 18, 2009)

She felt that the reaction of her colleagues in the mathematics department was "Half and half. Some people took it okay and some people really just went in not planning on receiving it well." (Interview, June 18, 2009).

At another session, mathematics teachers from across the district examined and graded samples of student work from a common assessment. Amy had been part of a committee that designed and proofread the assessment items. During this workshop, teachers worked in small discussion groups and Amy was frustrated with one of the teachers in her group:

I'm sitting across from someone that really didn't like really anything that was done and made it very well known. Which I just [pause] I said nothing, you know. That's her opinion and that's fine. So [the colleague said] "I'm not marking this question properly because I don't agree with it and it's a bad question and it shouldn't be there and this is ridiculous" and "blah, blah, blah, blah". Okay, bite your tongue, bite your tongue. And then at one point she goes, "Well nobody here was involved with this, right?" [laughs]. It's like, "Yep. I was." I said, "it's really hard to come up with questions like this" so I said "I'm sure they'd welcome your opinion when they ask for volunteers. They would be more that happy to have people volunteer for that". (Interview, June 18, 2009)

A further issue with this particular session was that Amy felt it was poorly facilitated (she was not involved in either the organization or running of the session). She found that the session leader did not give clear instructions, did not have a good sense of the audience and kept interrupting the small group activities. At the same time Amy was annoyed with her colleagues' behaviour and felt that the teachers did not listen well or follow the facilitator's instructions. The whole experience exasperated Amy: "my whole body was tense because it was just driving me batty" (Interview, June 18, 2009). Amy felt that the session was an important opportunity for teachers to help each other make sense of some aspects of assessment through dialogue in their small groups, but that the occasion was largely wasted. Although Amy acknowledges that her involvement with district assessment initiatives has helped her develop her understanding of classroom assessment, she finds that these opportunities can sometimes be frustrating and even "detrimental" (Interview, June 18, 2009).

Issues in Using Assessment Effectively to Improve Learning

A final area in which Amy has experienced tensions and dilemmas is in using assessment information effectively to improve students' learning. While she is quite proud of many of the developments she has made in her assessment practice, she remains unsatisfied with several aspects.

Surprise/disappointment in students' performance.

Even though Amy uses a variety of different formative assessment strategies within her classroom, she still finds that she is sometimes “disheartened” (Interview, May 21, 2009) by some students’ performances on assessments. There is sometimes a gap between her sense of a student’s learning and his or her performance on a major assessment:

when I get their evaluations, for about 20% of my class it’s just sort of that “ooh - wait a second. That just doesn’t match up with what I was anticipating”. Especially with my Grade 9s. There’s about four kids that’s happened with on the most recent unit test, and I thought “something’s not working for them, something’s changed in their position from the beginning of the semester to now, or I had a very preconceived notion about how they were doing in this unit because they’d done well on previous evaluations”. I was listening to their answers and observing them through that lens. I don’t think I was focusing as much on the areas they were struggling in ... even when I’ve identified two of them to say “yeah, they’re not quite getting this as well” but then I look at their evaluation and it’s like “holy smokes! They were way further behind than I thought they were!” (Interview, May 21, 2009)

This suggests that the increased focus on teachers’ professional judgement is an area that needs to be addressed through professional development. Amy believes that “the biggest issue” for assessment is being able to see “where students are at and where the misunderstandings are” (Interview, May 21, 2009). She tries to use this information to inform her teaching which means that she often finds herself “changing gears” (Interview, May 28, 2009).

Providing quality feedback.

Amy recognizes the importance of quality feedback to students’ learning, but she has found it difficult to provide specific, anecdotal feedback on assignments in a timely fashion. She has been trying to focus on

lots of smaller things in the classroom, quicker turn-around time and one or two comments. Not a lot of feedback because that was killing me and that just wasn’t working (Interview, May 21, 2009)

She found this a particular concern for some of the formative items that she gave in advance of a major assessment to help students prepare. She felt that sometimes she was

not getting it back fast enough and so the whole purpose of immediate feedback and the whole idea behind it was completely blown because I was taking just as long to mark those as I was their evaluations. (Interview, May 21, 2009)

During class time, Amy circulates to listen to small group discussions so that she can provide oral feedback to her students. Sometimes she finds it difficult to decide how long to let students struggle with a concept and when to intervene. She feels that sometimes she intervenes too quickly “in situations where they should do that by themselves a little more” (Interview, May 28, 2009).

In one class, I observed Amy’s Grade 10 students working on a set of trigonometry problems in small groups and then each group presented a solution. After the class, Amy stated that she found it challenging to balance students’ need for sufficient time in groups to discuss strategies and solutions with time for oral presentation of students’ solutions. She viewed the presentations as an important way for students to get feedback on their own work, as well as on common errors she has encountered in the past. Due to time constraints, only four of the seven groups were able to present their solutions during the class. Even though only a few groups had presented, she felt that students were “disengaged” (Interview, June 1, 2009) by the end of the period. She does not find that students recognize or value this sort of opportunity for feedback in the same way that she does.

Amy would like to begin evaluating these types of oral presentations as a part of students’ grades to broaden the range of assessments beyond pencil-and-paper items:

It’s something I would like to be doing if I could get them up there individually more. I am just not sure with how to coordinate that with a large class. Just because of the time restrictions. So it’s something I would love to be doing ... because some students are so much better orally than they are in writing. But I

just haven't been able to figure out a way for that to work, yet. (Interview, June 1, 2009)

Negotiating these Tensions and Dilemmas in Practice

This section describes the ways that Amy negotiates these different tensions and dilemmas in her assessment practice, addressing each of the four areas.

Policies Related to Late and Missed Assignments

Amy finds it difficult to reconcile different interpretations of various levels of policy for late and missed assignments. Although she does not find it to be clear enough, she looks primarily to school policy when making her decisions:

[School policy] affects us most directly because if a parent has an issue the first person they go to is the principal and the principal refers to the school policy. Hypothetically, the school policy is built on the Board policy, which is built on the provincial policy. So it's everybody's interpretation of a document ... it's really more the school policy that dictates where you go with it (Interview, June 18, 2009)

Amy appeared to feel that her concerns were validated by recent comments posted by her colleagues to the school's online electronic conference. Describing her reaction to this sharing of other teachers' experiences with late and missed assignments, Amy said she felt "sort of empowered. In fact, I think I wasn't the only one because people are writing back 'hear, hear' and 'send more stories' and it's sort opened the floodgates" (Interview, June 18, 2009).

Another way that Amy negotiates this issue is by handling late and missed assignments differently between grade levels. Amy's current Grade 9 is an Applied¹² course, and she is more comfortable letting students in this course have a second opportunity to demonstrate achievement of curriculum expectations:

¹² As described earlier in this chapter, Grade 9 courses in Ontario are offered at the Applied and Academic levels

I feel like they're [Grade 9 students] not going out there looking for answers to a test before they come into a Student Success day. So I don't mind giving them a test and redoing it and trying it again. With Grade 10 Academic students, I would find that I little bit challenging because there is a lot more conversation going on. If I were to redo evaluations, which I haven't done yet with the [Grade 10s], it would be a different evaluation. (Interview, May 21, 2009)

One student in Amy's Grade 9 class was repeating the course and had skipped several mathematics classes where there had been assessments. The young woman told Amy that her mother knew about the absences but said that she would have to "deal with the consequences" (Interview, May 28, 2009). Amy did not feel that the student sensed any consequences for skipping school so she decided to tell the girl that she would be receiving zeros for missed assignments

Because as soon as she said there were no consequences from her mom, I can't go and tell her that there's an opportunity for her to go and make up all the tests she's missed. So I said, "Yeah. As of right now they're at a zero". (Interview, May 28, 2009)

Amy then encouraged the student to come to all the remaining classes (about three weeks) and try her best. She felt confident that the young woman could demonstrate a sufficient level of achievement to earn her credit.

Issues in Grading and Reporting

Amy negotiates her challenges with grading in reporting in different ways. When struggling to interpret and implement the provincial Achievement Chart as part of her classroom assessment, Amy looks to school and district policy for more specific guidance:

... the weighting of the Achievement categories is completely discretionary according to the Board. So, really, trying to follow what the province says about that is sort of futile because the Board has already decided how they want to approach it. If the school then subsequently changes that approach you're kind of stuck to that. So, although we're following Ministry guidelines, the Board and the school are really holding you more to what they've put in print (Interview, June 18, 2009)

Following the mandate of her school district, Amy breaks all of her tests into four separate sections, one for each of the Achievement Chart categories (*knowledge and understanding, application, communication and thinking*). Amy believes that she has

a sense of what each of the categories embodies in what it's trying to evaluate. And then, when I'm designing my evaluations I design them with that in mind. But it's sort of ambiguous because questions could go in different categories, depending on what you're evaluating - what's really *knowledge* versus *thinking* versus *application*? I do basically design my evaluations with the categories separated out. Sometimes it's just by question and sometimes it's by section. So I'll have a *knowledge* section, an *application* section. Because the Board policy has those categories sort of decided in terms of their weighting factor, they go in based on how the evaluation was created. So I stick to those percentages. I try to evaluation *knowledge* and *application* more so than *communication* and *thinking* because they're worth more in their overall grade. That's sort of how I do it. (Interview, June 18, 2009).

In her negotiation of the challenges in recording and analyzing assessment data, Amy speaks to colleagues. When she was frustrated with the way that formative assessments disappeared from her student reports, Amy contacted the district's software support staff and then did implement their suggestion of trying to keep a separate file for these assessments, even though she was not satisfied with this. She has also looked to other colleagues for what they are doing

One of my colleagues uses Excel. That's all he does. He doesn't even use a grading program. Actually two of my colleagues do that. And it works fine. I think it's just finding something that works for me. (Interview, May 21, 2009).

While she has found it frustrating that the district has changed their recommended grading software several times, this willingness to be part of a group that pilots software appears to be part of her way of trying to find software that she likes.

With regard to the incident where her principal changed a mark, Amy felt confident that her mark was an accurate representation of the student's achievement. She said that she would use this holistic/rubric method again in the future: "I'd do it again, for sure. I just

[pause] not right now” (Interview, May 21, 2009). When introducing this grading method to her class early in the semester, Amy felt confident, despite some initial scepticism on the part of her students. When a student asked her why the assessment was changing from the previous semester, Amy replied, “we’re in a different classroom” (Interview, May 21, 2009). She found that the students bought in as the semester progressed and they participated in a conference about their grade before the midterm report card. Amy’s department head, who she described as “forward thinking” (Interview, May 21, 2009), was aware of her grading practice for that class and had copies of the rubrics which showed how the final mark was determined. While Amy made no reference to any conflict between her principal and department head, she appeared to feel that messages she was receiving from her department head and district staff were coherent, and that she did not agree with her principal’s perspective on this particular occasion.

Colleagues Resisting Current Reforms in Classroom Assessment

While Amy described feeling frustrated with some of her colleagues, she acknowledged that teachers find assessment a particularly hot button topic: “it’s a struggle because it’s a passionate issue. Something that we feel is a part of our day-to-day operations and what we need to do with the students in the classroom” (Interview, May 21, 2009).

Amy tends to negotiate this tension by aligning herself with other teachers who are supportive of the current assessment reforms in Ontario. She largely develops this sense of “what’s coming down from above” (Interview, June 18, 2009) through her participation in district activities. Amy enjoys making positive connections with other teachers who are also trying to develop their assessment practice:

Discussions with colleagues are really beneficial just to find out what other people are doing. I've tried to support other colleagues as they've tried new things so that gives me an opportunity to think more about it because then they say "why are you doing this?" and "why are you doing it that way?" when they're trying to figure it out. It really forces you to reflect on your own practice (Interview, May 21, 2009)

A few years ago, Amy was able to participate in an extended professional development opportunity with a group of teachers in her district. The group had extended discussions about the "big ideas" in several of the secondary school mathematics courses and began considering how to use assessment to develop these understandings in their students. Amy described this experience as "great in terms of what I brought back to the classroom" (Interview, May 21, 2009).

While Amy finds that some centrally-led professional development experiences can be "detrimental" (Interview, June 18, 2009), she tends to seek out supportive colleagues to talk through issues in her assessment practice. For example, during the semester of this study, she and two colleagues decided to collaboratively mark their Grade 10 summative performance tasks. Amy found this experience to be "very, very valuable" (Interview, June 18, 2009). She appears to have a sense that the assessment reforms have value for students and that she will seek out people who can help her interpret and implement a coherent vision for her classroom.

Issues in Using Assessment Effectively to Improve Learning

Amy is quite reflective about the effectiveness of her use of assessment to improve students' learning. After one lesson where I observed students in her Grade 10 class working on a trigonometry problem set, she said, "Given that I tried that [lesson] a couple of times now, seven is too long [referring to the 7 questions on the sheet]. There's not enough time for the presentations. So I would cut it back to five, I think" (Interview, June

I, 2009). Her dissatisfaction with some elements of her practice motivates her to continue developing her understanding and participating in more learning opportunities. This participation helps her to “really find out what the latest buzz is and try something else” (Interview, June 18, 2009).

Amy has focused recently on more informal and interactive assessment in her classes, both to improve students’ learning as well as to inform her own teaching. She has been pleased with a provincial curriculum resource for the Grade 9 course that contains quality activities that help her establish a “completely open, low risk environment” (Interview, May 28, 2009) in her classroom where students are comfortable contributing to discussions, collaborating in small groups, and sharing their thinking. Amy views these elements as essential to classroom assessment.

Amy has found that her (recently acquired) interactive whiteboard allows her more flexibility to adapt her lessons since she has access to her own files, resources from publishers and the Ministry of Education, and the internet. This allows her greater flexibility to adapt her lessons on-the-fly and be more responsive to students’ needs.

Amy described that one of the elements of her assessment practice that she was trying to develop was getting more regular and timely feedback to students. She is trying to accomplish this by using “lots of smaller things in the classroom and quicker turn around time and one or two comments.” (Interview, May 21, 2009).

Summary

A summary of the tensions and dilemmas that Amy described in her assessment practice, as well as how she negotiates these is included in Table 5.1. Over the past year, Amy has been focusing on three main goals in her assessment practice: grading fewer assignments to focus more on assessment-for-learning, making her lessons more interactive

so she can get a better sense of students' understandings, and students presenting solutions to mathematics problems. Amy has grown more comfortable with more informal assessment strategies such as observation, quick conferences with students, and quizzes that do not count for grade. She tries to give students quick feedback in order for them to use this information to further their learning. The interactive strategies help inform her teaching including student use of individual whiteboards and student displays of hands. She has also had students present their solutions, which she uses as an opportunity for students to get feedback on.

There are several other aspects of her assessment practice that have been in flux. She has used several different tools for determining students' grades, including different commercial grading software programs and a holistic chart method where she conferenced with students. There has been significant turnover in her department heads and district staff, as well as a fairly recent change in Principal which can contribute to uncertainty in the school and district vision. Amy is very involved in several district initiatives and has volunteered several times to be part of pilot projects. She tends to be an early adopter of change and admits this can sometimes lead to her feeling frustrated, but she likes to be in the leading edge of teachers adopting practices encouraged by the school district.

Amy described feeling dissatisfied with some elements of her practice, including being surprised by students' performance on tests and ensuring students received timely feedback. These elements of dissatisfaction appear to be a major influence in motivating her to continue her participation in professional learning activities. She appeared to feel a constant sense of strain in interpreting policies from different levels and noted that school policy was an interpretation of district policy which was an interpretation of provincial assessment policy.

In negotiating these challenges, Amy looks to school and district policies for more specific guidance and seeks out input from colleagues at her school and the district, particularly with what they are doing. She felt that her concerns over late work were validated by staff comments to the school's online electronic conference where several teachers shared examples of students submitting late work, along with a list of steps that they had taken to get students to submit the work. Amy appeared to feel more strain around late summative assignments and performance tasks in Academic courses than other types of assessments, due to concerns over academic integrity and lack of student respect for her time. She is reflective about what she feels is effective in her classroom and has been focused on developing more informal and interactive assessment strategies to ensure more regular and timely feedback to her students, as well as to inform her own teaching. She identified a curriculum support developed by the Ministry of Education as being key to helping her establish a classroom culture where students are comfortable sharing ideas and discussing, and suggested that she was adapting examples from this resource into other courses. She also expressed that the interactive whiteboard in her classroom allowed her more flexibility in her lessons to adapt and be more responsive to students' needs.

Amy's response to a summary of her case, including description of classroom practices, challenges in classroom assessment and ways these challenges are negotiated was:

Thanks for the summary. Interesting to see how it was interpreted. I feel it is quite accurate and have no concerns to report. If you are looking for any specific feedback, please feel free to let me know and I can send some your way. Thanks again for sharing. (Personal Communication, November 2009)

Table 5.1: Summary of tensions and dilemmas Amy has experienced in her classroom assessment practice and how she negotiates them

Tensions and dilemmas experienced in classroom assessment practice	How these tensions and dilemmas are negotiated
1. Policies related to late and missed assignments a) Interpreting policies from different levels b) Students submitting work late	Compliance with policies and directives at school, district and provincial levels (e.g., Principal's perspective on late and missed assignments; mark weighting according to the categories of the Achievement Chart) Confidence in her own approach to teaching, learning and assessment
2. Issues in grading and reporting a) Interpreting and implementing the provincial Achievement Chart b) Limitations of commercial software for grading and reporting c) A student's final grade being overturned	Belief that current assessment reforms are beneficial to students' learning Participation with mathematics teacher community within and outside her school, in both formal and informal activities – particularly conversations with department heads, district staff, and participation in formal district PD initiatives
3. Colleagues resisting current reforms in classroom assessment a) At the school level b) At the district level	Reflective about changes she can make to improve students' learning (e.g., still dissatisfied with elements of her practice and seeks out new ideas to help her resolve these concerns/gaps)
4. Issues in using assessment effectively to improve learning a) Surprise/disappointment in students' performance b) Providing quality feedback	Use and knowledge of a wide variety of resources

Case Study #2: Carina

Following the structure of Amy's case, this section also has two main parts. Again, I provide some relevant information about the context of the case including Carina's school and her professional background, as well as her views on assessment and a description of

her classroom assessment practices. I then discuss four tensions and dilemmas that Carina described in her classroom assessment practice, and how she negotiates these.

Description of the School and Teacher's Background

Carina is a teacher at Blueridge Secondary School, a semestered suburban school in Ontario. Like Amy's school, this school also serves students from Grades 9 through 12 and has approximately 1100 students. While in different parts of the city, the two schools are quite similar in terms of the socioeconomic, cultural and ethnic composition of the student body, both being located in fairly diverse, middle class neighbourhoods. Both schools have a history of above average scores on provincial and district tests, and strength in both athletic and arts programs. Blueridge is one of the newer schools in the district and many of the brightly lit halls are lined with students' visual artwork – paintings, collages and graphic design projects.

Carina was in her mid-forties and in her twentieth year of teaching during this study, all of which were at Blueridge. Carina's undergraduate degree is in science and mathematics and she is qualified to teach in both these areas from grades 7-12. Like Amy, she also holds an Honours Specialist qualification in mathematics. Earlier in her career, Carina taught courses in both mathematics and science but she has been teaching mathematics exclusively for the last four years. During this time, she has had the opportunity to focus on developing three mathematics courses since she has had a very stable timetable with the same courses for several semesters.

Carina has been actively involved in developing curriculum resource materials that align with the provincial curriculum, including performance tasks, and she has shared some of these with teachers at her school and others in her district. She is part of a district-wide group of mathematics teachers focused on sharing resources, staying current on new

initiatives and planning professional development for other teachers. In addition, she has participated in several assessment initiatives organized by mathematics consultants at the school district. Over the past few years, she has made presentations to her own school's mathematics department and at the district level. Like Amy, Carina is also a very well respected teacher within her school and district and appeared to have an excellent rapport with her students.

During the semester of this study, Carina taught Grade 10 and 12 mathematics courses. Carina's classroom is bright and tidy with thirty desks, arranged in five single rows, facing a blackboard with an interactive whiteboard mounted in the middle of it. Students are seated so that the empty desks are at the back of the room. From the view of students, there is a blackboard on the left wall, a bulletin board on the right wall, and two big windows behind them at the back of the room. There is an empty teacher's desk on the left side of the room and students sometimes work in groups in this space. Carina uses her own laptop on a small desk at the front of the room in the right corner. The department has a large collection of mathematics manipulatives, tools and resources that are stored in a room near the department office, and teachers sign out the equipment they need for their lessons. Carina brought graphing calculators to all three of the classes I observed, which some students used.

During all three of the lessons I observed, one Grade 10 and two Grade 12 classes, students were quiet, but appeared to be engaged in the lessons and were very involved in their class work. After a 10-15 minute teacher-directed lesson, students generally worked in pairs or small groups for the remainder of the class. While students were working, Carina circulated extensively to focus on individual student or groups concerns, stopping occasionally to address the entire class.

View of Classroom Assessment

Carina has experienced a great deal of change since she started teaching. She describes, “when I first started teaching, math was not very creative then. It was pretty much drill and skill and that’s all the emphasis was. It was very procedural without a lot of connections to anything else” (Interview, February 20, 2009). Now she feels that “math has evolved cause it’s not just right or wrong which there used to be a lot more of” (Interview, February 20, 2009). Over the past eight years, Carina has been gradually developing her classroom assessment practice to align with these changes. She now spends “a lot of time thinking about overall expectations. Thinking about how kids are going to get there and how am I going to help them get there” (Interview, February 20, 2009). She has introduced performance tasks and investigative activities and making an effort to increase her use of formative assessment. She believes this approach to assessment requires a shift in teachers’ grading practices, from averaging and a reliance on commercial grading software to a more holistic view of students’ achievement through professional judgment: “it is two different ways of thinking about it: the mark is just sort of a summary at the end of what you think.” (Interview, June 19, 2009). Carina attributed many of these changes to the recent revisions in provincial curriculum and assessment policy, as well as district and school professional development opportunities, collaborations with colleagues at her school, and her strong professional drive to continually develop and improve her practice. Referring to the curriculum and assessment reforms that have been occurring in Ontario, she declared, “I’ve totally bought into what’s going on” (Interview, February 20, 2009). When I asked Carina how she felt her classroom assessment practice had evolved since she started teaching, she replied:

I think it's more relaxed. The kids are more relaxed. I'm more relaxed. Because it's not like everything counts... It's more like, "You know this is an opportunity to find out how you're doing". And they buy into it... It seems to be more in a friendlier way. It seems to be more like "I really want to help you. Here's how to do it." But if I give 2 out of 10, that's all they see. (Interview, February 20, 2009).

The changes that Carina has made have been very gradual. Her approach was to listen to colleagues' interpretations and then to make incremental changes in her own practice with new goals for each semester. She believes that the sorts of fundamental changes called for by the current curriculum and assessment reforms in Ontario must be taken "at your own pace. You have to introduce it at a level you're comfortable with" (Interview, June 19, 2009). For example, Carina used to assign scores to all of the quizzes she gave her students and include these in their final grades. She reconsidered this practice several years ago when other teachers in her department started using quizzes exclusively to give students feedback on their learning. She was initially worried that her students would not take quizzes seriously if they did not count in their grade, so she reduced the weight on the quizzes to a quarter of their previous value:

I remember the first time they [colleagues in her mathematics department] started with we're not going to mark quizzes, it's only going to be tests and I thought "Oh my god! How are they going to do it?" I thought "they [students] won't study". I think it's easy to react to an idea right away. But I'm the type of person that will think about it and try it. (Interview, February 20, 2009)

Carina now gives students feedback on their quizzes in the form of comments rather than scores. This approach to assessment is very different from what she experienced as a student, where almost everything that teachers collected was scored and counted in the final grade.

As she participated in more activities and discussions, she began to feel more confident with this new approach to assessment. She feels that it is important for her to

participate in on-going learning opportunities: “I think the more involved you are the more opportunity you have to hear about different opinions” (Interview, February 20, 2009).

Description of Classroom Assessment Practices

This section describes key aspects of Carina’s classroom assessment practice to provide a context for the subsequent discussion of the tensions and dilemmas that she experiences and how she negotiates these in practice. While it is difficult to separate classroom assessment practices neatly into distinct categories, it is useful here to focus on several aspects of Carina’s use of assessment *for* learning, and assessment *of* learning.

Assessment for Learning (Formative Assessment)

In recent years, Carina has focused on developing her use of assessment *for* and *as* learning. This section will describe Carina’s Assessment Checklist, and how she uses quizzes and observational assessment to improve students’ learning.

Assessment Checklist.

An important aspect of Carina’s approach to assessment is a recording system that she calls an Assessment Checklist. She has developed this over the past few semesters to help her keep track of how students are doing on the course expectations in advance of major assessments that count in students’ grades. This double-sided “checkbric”¹³ has a section for each unit of the course (see Appendices G & H for copies of the teacher and student versions of these checkbrics for Carina’s Grade 10 class). Along the left side there is a list of the overall course expectations written with more detail than in the provincial curriculum document. Along the top there are four columns with headings that describe the level of achievement that the student has demonstrated: exceeds expectation, meets expectation, meets with help, and does not meet.

¹³ A “checkbric” is similar to a rubric, but has less detail.

For each of her classes, Carina keeps an “assessment binder” with checkbrics for each student (filed alphabetically). She periodically updates these records based on class activities including formative quizzes, observations, and interactions that she has with her students. She distinguishes between three levels of achievement: “meets expectations”, “meets with help” or “did not meet”. Carina described these distinctions:

“Meets with help” means you can do it, but you’re not there yet unless you can do it independently. But there’s some kids, even if you help them... they still can’t do it. That’s “does not meet”. So I wanted to make that distinction ‘cause I was doing a lot of my assessment with allowing them to ask some questions ... I don’t mind them asking questions if it’s an assessment, but I will make a note of that to myself so that I know it’s not independently done. (Interview, February 20, 2009)

The regularly updated written record allows her

to get a really good sense and to get it down on paper... It’s not just a gut feeling. You can actually say “well, this is the feedback we had.” This is the opportunity I had to talk to them (Interview, March 2, 2009)

Carina updates these notes daily as part of her end-of-day routine along with entering attendance electronically and updating the class website. She described, “as I’m recording them I can see right away which expectations - if we’re good or not” (Interview, February 20, 2009). She regularly looks over these data to ensure that students are prepared for upcoming major course assessments, and if they are not, to help them take steps to ensure they are prepared.

Carina calls her copy of the Assessment Checklist the “evidence of expectation” or “teacher checklist”. Students receive a blank copy for their own reference that she calls the “student checklist”. During each unit, they are prompted several times to fill in their own checklist as a form of self-assessment. Carina wants students to “come up with their own impressions of what they thought they knew” (Interview, February 20, 2009). Before major assessments, Carina conferences with her students to compare the student and teacher

checklists and discuss any differences. If the two checklists do not match, students can show that they have done additional work or re-done their quizzes and Carina will update her teacher checklist. She asks her students, “what have you done since my ‘does not meet’ to your ‘can do it’? And [for] some of them it is a good discussion that way” (Interview, March 2, 2009). The student version of the checkbric also has a space for students to write a few sentences in self-reflection towards the end of each unit. She wanted to provide

an opportunity for them to say “okay, I really struggled with this the whole unit. I better get help before the exam”. I’m hoping they do it before the unit test. And hopefully the day before the day before the unit test cause then they can still have some action with it.” (Interview, February 20, 2009).

Carina believes this process clarifies the learning goals for the course and helps students begin to reflect on their own learning. She described some of her rationale behind developing and using the checkbric and the distinction between some of the different levels of meeting course expectations:

I want them to be a bit more accountable. I said, “if you check off ‘always’ on every single thing you should be fine. At the end, when you get to the exam you should have them all checked off if you want to be really successful (Interview, February 20, 2009)

Carina decided to bring this recording system into her Grade 12 class because in previous semesters she had been frustrated with students leaving test questions blank or writing that they did not know how to do a particular question. With the Assessment Checklist, Carina feels that students “cannot say [they] don’t remember how to do that. That’s like a basic given that it’s there. And I see more and more of that” (Interview, February 20, 2009). This recording system means that Carina is rarely surprised by poor performances on evaluations: “I have a sense as to where they’re at. So that tomorrow [a test day] it’s not [pause] it shouldn’t be a huge surprise. I’ve already contacted parents” (Interview, March 2, 2009). In the assessment binder Carina also keeps notes about parent

contact, as well as other information about her students. Carina has been using a version of the teacher checklist for several semesters with her Grade 10 class, but this is her first time trying the student checklist and self-assessment. This is the first semester that she is trying this recording system with her Grade 12 class.

Quizzes.

Carina uses formative quizzes as the major source of evidence for her Assessment Checklist in both her classes. At the beginning of each semester, she establishes with her students that in her courses quizzes are assessments that will NOT count in students' grades. Carina views these as an opportunity for her to get a sense of her students' understandings so she can help them individually. The quizzes also inform her teaching since she has "a chance to instruct again and clarify some things" (Interview, March 2, 2009). She believes the quizzes also help students to identify the big ideas in a course because the quizzes can "reinforce things they might not have thought were important" (Interview, February 20, 2009). She works to develop a classroom culture where students understand that quizzes are

an opportunity to see the kinds of questions that they'll get, my thinking around certain things because, really, teachers do take different interpretations on how to write evaluations. That's their opportunity to check that they know what they think they know (Interview, February 20, 2009).

The quizzes also help students to see how well they really know the material without the benefit of having immediate access to the answers, as with questions in a textbook:

I think the important thing about these little quiz assessments is to give them an opportunity to do something that doesn't have an answer that they can go back to. ... Can they actually put it together? How do they confront a new question? (Interview, March 2, 2009).

These quizzes are not given traditional scores because Carina believes that if her students "saw a mark [they] wouldn't read any of the comments that I tried to write"

(Interview, February 20, 2009). Instead, students receive anecdotal feedback in a variety of ways. Sometimes Carina collects quizzes and writes comments on them. Other times she displays the solution on the interactive whiteboard and discusses different elements with students, who then check their own work. Most of the time, Carina circulates and checks quizzes during the period to give oral feedback to each student individually: “I find that they really do listen way better when you do it one-on-one like this. With *their* questions.”

(Interview, March 2, 2009). She believes this helps to motivate and encourage students:

I find that when you check their quizzes like I did today...to have that vocal and verbal acknowledgement of their work. You could just see them straightening up. And that’s an opportunity you don’t have if you’re marking for marks (Interview, February 20, 2009)

When Carina is circulating during a quiz, she assists students as needed. In one Grade 12 lesson I observed she said to her students, “if you need help during the quiz, you can ask because it’s not for marks” (Field Notes, March 2, 2009). She explained her philosophy on assisting students during quizzes:

... this is assessment so why not ask questions? Because it might be the first time they’ve actually thought about [it] ... it’s an opportunity to do one-on-one work with them that they will then remember because it was their question you were answering (Interview, February 20, 2009).

Towards the beginning of a unit, Carina tends to use quizzes as a way to help students see how well they understand a particular topic. She tries to make this a very low risk activity and often does not collect these early quizzes: “So there’s no pressure at the beginning. And they have a chance to keep getting better and better.” (Interview, February 20, 2009). She hopes that these quizzes foster reflection: “I say ‘I’m not even going to look at it. All you should think about is are you more successful as the time goes on?’” (Interview, February 20, 2009).

She calls some quizzes “skills drills” which are very short and meant to reinforce foundational knowledge and skills that students need to know “just like you breathe”

(Interview, February 20, 2009). These quizzes

start with the basics but then you repeat the theme... it takes them five minutes or so and then I either have them put it up or I take it up and they just check it. I don't even look at it. So then the next day they come in and I greet them with this so they know. It's expected. (Interview, April 23, 2009).

Other shorter quizzes can involve skills on a graphing calculator:

I will go round and say “I need to see that screen” and I'll go. And that's very quick. Cause anyone who doesn't have it, like it's easy to see. Then you help them and then you say “meets with help”. Like really nice and easy (Interview, April 23, 2009).

Towards the end of a unit, Carina uses “challenger quizzes” as an opportunity for students to demonstrate their conceptual understanding in advance of a major assessment.

With these quizzes, she wants to see:

Can they start putting their minds in a place where they can handle something that's new? I think it's important to do that. I don't mean to do that before a test because I want to scare them, but I want them to be prepared that this is the kind of thing they should be ready for. Now, remember, this is the third assessment quiz we've had (Interview, March 2, 2009)

These quizzes help students with strategies for problem solving and allow Carina to emphasize areas of difficulty that she has recognized from previous years, such as students getting lost in the algebra while solving a problem:

[Sometimes] they almost lose sight of what the question is and a lot of them for inequalities won't even have a statement at the end. Which is really the answer to the question. This [quiz] gives you an opportunity to do stuff like that. (Interview, March 2, 2009).

Near the end of a unit of study, Carina also likes to give students feedback on the written presentation of their solutions: “Just before a test I think it's important to address form ... what can we do to tidy it up?” (Interview, March 2, 2009).

Carina believes that quizzes help both her and her students ensure they have met and demonstrated course expectations. Describing a class earlier in the day where she had given a quiz, Carina said, “I know that everyone met the expectations today. A lot of them needed help to get there but they all ended up with my initial on their paper” (Interview, February 20, 2009). Sometimes, if a few students are struggling with a particular quiz, she will allow them to take it home and then hand it during the next class. She tells her students, “look, if you’re not ready to do this today, take it home. Give it to me tomorrow” (Interview, March 2, 2009).

This approach to quizzes is very different from the approach that Carina took earlier in her career when all quizzes counted in students’ grades. She used to find that students “never wanted a quiz before” (Interview, February 20, 2009), but now she finds they are “totally positive on this” (Interview, February 20, 2009). She described:

I’ve had Grade 12 kids ask me “So miss, I really think we should have a quiz soon” whereas they never wanted a quiz before... I have had a few kids say, “Oh, I think we better have a quiz before the test”, you know “we haven’t had a quiz”. So they’re getting antsy because they haven’t had a quiz. Which I think is a great response (Interview, February 20, 2009)

Observational assessment.

In addition to the observation that Carina does as part of quizzes, she also uses several other observational methods to get a sense of her students’ understandings. For example, when studying transformations in her Grade 10 and 12 classes, she has students do “functions aerobics” where they stand up and move their arms to indicate transformations of various functions. She values this type of assessment “because I can see right away: do they get it or do they not?” (Interview, February 20, 2009). Another example is students discussing “thinking” and “reflecting” questions from the textbook. Carina prefers to hear student discussions of these questions because she can listen and respond “if

they're talking. Whereas, if they're writing, how do I know? They'll check the back of the book or something which is not very meaningful" (Interview, February 20, 2009). She also listens to students discussions during classroom activities and performance tasks and gives students feedback. Carina does not always record this information, primarily using it to inform her own instruction and address students who are struggling.

Assessment of Learning (Summative Assessment)

In this section I describe Carina's use of assessment *of* learning including how she uses tests and performance tasks, summative tasks, when and how students can get second chances on major assessments, and how she uses her professional judgment in finalizing students' grades.

As previously described, provincial assessment policy requires that students' grades be comprised of term work, which counts for 70% of the final grade, and summative work, which counts for the remaining 30%. For the term work portion of the grade, Carina uses tests and performance tasks. In the Grade 10 course, Carina plans five units of study with five tests and two or three performance tasks. She has organized the Grade 12 course into six units with six tests and three performance tasks. The summative portion of the grade for both courses is comprised of a final exam and a final performance task.

Tests.

Carina grades all of her tests using a marking scheme. The tests usually address three of the four categories from the Achievement Chart: *knowledge and understanding*, *communication* and *application*. Occasionally she will include a short *thinking* question, but believes this is assessed more effectively through performance tasks. In addition to written *communication* questions such as explanations or comparisons, each test is assessed holistically for form using a generic *communication* rubric. This rubric score is converted

to a mark out of 10 for recording purposes so that these marks can be easily compared when Carina needs to determine a grade.

Carina believes that her students have been well prepared for tests and have had many opportunities for practice. Through her use of the Assessment Checklist, Carina feels that she has a good sense of students' understandings prior to major assessments and has had sufficient time to help students prepare for tests, by either encouraging them to complete missing work or seek extra help. Before tests, Carina refers students to their own Checklist:

[as if speaking to her students] "If you're not checking these things off as 'always met' there's work to be done and you've only got one week now 'cause we will have the test the week after and that's when the game is on. There isn't 'can we have a replay?' The game's on. That's it. You've had all this time to practice" (Interview, February 20, 2009)

As opposed to quizzes, which she does not change much from semester to semester, Carina spends a lot of time writing her tests each semester. She explained:

I do spend a lot of time making that [test] for each class because that's the one that has to be new and different. It has to also 'jive' with the way the class has been going. Did we spend a lot of time on this cause we took up this homework and now they should be able to make connections based on that? (Interview, March 2, 2009)

She takes pride in having well written tests and stated it is very rare that she needs to either make a change to a question during a test, or not count a question due to poor wording or an error in the question. She also believes that her students find her tests fair, even if they did not do well: "I've never had any comments about 'oh, the test wasn't fair'. But I do spend a lot of time making my tests ... [students] don't come out feeling frustrated. They might think "oh, I should have studied harder" and that's all I get. (Interview, February 20, 2009).

While Carina believes that her students have every opportunity to prepare for tests, she sometimes offers additional flexibility. For example, she offers students some choice in their test date by finalizing a date through a class vote. During one lesson, I observed Carina gave students the option of two test dates and was pleased that they chose the later one because she “knew some kids weren’t ready. If I didn’t see all the quizzes within 15 minutes then not everybody’s ready” (Interview, March 2, 2009).

Because of the large size of Carina’s school, there is generally more than one section of a course running at a time. Teachers in the department generally share tests, but usually administer substantially different versions for the first few units to discourage students from a morning class sharing test questions with students in an afternoon class.

Performance tasks.

Carina has a broad repertoire of performance tasks for the Grade 10 course because of collaborative efforts within her department, including several tasks that she created, as well as an ongoing board assessment initiative where several sample tasks are created and disseminated to teachers. Some of these tasks involve students working together in groups on the first day to plan a strategy for the task, even though they only receive partial information on the first day. Students then receive the rest of the information on the second day and work individually to finish solving the problem. Only the individual work from the second day is graded.

Carina marks all her performance tasks using a generic rubric for the category of *thinking* (on the provincial Achievement Chart). Some tasks are also assessed for *communication* using a rubric. The levels from the rubric are converted to a mark out of 10. Within these categories of the Achievement Chart, some aspects of the task can be more heavily weighted than others due to their length or the complexity of the question. Carina

explains: “you make a judgment based on what you think are the most important aspects” (Interview, June 19, 2009). The tasks usually take two days to complete with students completing a pre-task activity on the first day and completing the actual performance task on the second day. In her planning Carina tries to make sure there is time for three performance tasks for each course to ensure she has sufficient data upon which to make a judgment about the grade for the *thinking* category.

For the Grade 12 course, Carina has been creating many of her own tasks, either based on ideas from textbooks, resources produced by professional organizations including the Ontario Association for Mathematics Education (OAME) and the NCTM, or other sources on the internet. In this course, there is a strong focus on modelling so most of the tasks require students to analyze and model data. Sometimes students collect data themselves using technology, while other times it is provided to them. Most tasks require students to use a graphing calculator for the analysis. Carina tries to connect these tasks to real applications of mathematics, such as linking the length of daylight over a year at different latitudes to seasonal affective disorder, and comparing the trajectory of the AIDS epidemic in Sub-Saharan Africa and in the United States.

Summative tasks.

As described earlier in this chapter, Carina’s school board requires two distinct components to be included in the summative portion of the mark. For both the Grade 10 and 12 courses, these take the form of an exam and a final performance task.

The exams are traditional in nature and are written in the school’s gym during a two-hour block at the end of the semester. The exams are common to all students taking the course that semester, and are graded with a marking scheme that is agreed upon and used by all teachers of the course. Carina provides her students with a practice exam several

weeks in advance, and posts the solutions to this practice exam in the classroom the week before the exam.

The summative performance tasks are similar in nature to the performance tasks used through the course, except they generally require students to pull together information from several units. For example, this semester's Grade 12 task required students to analyze and model three different sets of data from a biology experiment, and write a short report. These tasks are graded using a generic rubric, but Carina makes notes on the rubric so that the mark can be later justified. She explained: "[Students] are not going to see it but I have enough notes. I can certainly account for why their mark was such" (Interview, June 19, 2009).

Second chances on major assessments.

Students who have not done well on some assessments have a few opportunities for a second chance to demonstrate what they know and can do. Some of these are department practices, Carina's own, or mandated by the provincial policy on Student Success (described earlier in this chapter).

Two years ago, the mathematics department at Carina's school began a formal remediation program for students in Grades 9 and 10. In several units, students have the option of participating in a remediation program that allows them a second opportunity on a test. This program involves students attending two lunch hour sessions where material is reviewed, students work on a remediation assignment, and receive assistance from a teacher. After they have completed the assignment, students write a shorter version of the test. This re-test is marked holistically, by category, up to a maximum of level 3, which is considered the provincial standard. Carina is supportive of this program and speaking of a class from a previous semester she stated:

It just kept giving them hope so they wouldn't give up. That was really important, though. Because otherwise they could turn into really negative people and yet they just thought it was great that they were getting these opportunities so "I'm going to try again" and "okay, you know, keep the optimism up. We're going to try" (Interview, February 20, 2009)

Occasionally, Carina gives students remediation opportunities beyond this formal program on a case-by-case basis. Speaking specifically about one student who had been doing very well, but missed a large number of classes over a short period due to extracurricular activities and illness:

I said "You've done so well during the year that if you want help re-learning that unit I'm willing to help you through it. And then I'm also willing to give you a new evaluation" (Interview, June 19, 2009)

In addition, if she has to create another test due to an unexpected student absence, she will sometimes allow students to write a second version test if they feel they have not done well the first time. She stated: "[As if speaking to students] I have to make up a new one anyway so if you want to write another one I won't look at the first one" (Interview, June 19, 2009). Carina does not do this for every unit, but does it more at the beginning of a course to help students feel successful and to encourage students to revisit material.

Finally, students who have not demonstrated achievement of course expectations at a sufficient level to get their credit have an opportunity to "rescue" their credit. As previously described, provincial policy requires each school to run a Credit Rescue program at the end of each semester that allows students with marks in the 40% to 50% range to receive another opportunity to demonstrate course expectations. The subject teacher creates an individualized package of work for students to complete with the Student Success teacher during the exam period. The subject teacher then marks this material and determines if the student has demonstrated the course expectations at a sufficient level to grant the credit. This option is open to all students in the school in all courses.

Determining a final grade.

Carina uses grading software to determine her students' grades from the major course assessments. After all the assessments are complete, she looks at all of the evidence by category and compares with the exam:

If they're fairly consistent in their marks then it's pretty well what they have. But, if there's an anomaly that's when I look [further]. With things like *communication* ...I just really eyeball that. (Interview, February 20, 2009)

With the *communication* mark generated from the generic rubric used on tests and tasks Carina uses her grading software to select all of these assessments to analyze them for each student. Since they are all entered out of 10 marks, she is able to holistically determine a final grade out of 10 considering each student's most recent, most consistent performance over the semester. She does not average these marks, rather she decides by inspection after looking at all of the evidence: "I'm like, 'it's going to be closer to an 8' but I don't take out a calculator" (Interview, February 20, 2009).

Carina likes to make sure that there are at least three tasks from which to determine a grade for the *thinking* category. She believes this allows her to make valid professional judgments about grades without using an arithmetic average, which may be distorted by one poor performance. She explained:

With two [tasks] it's kind of hard not to give them the mark that they got. But if there's three, I really do look at the three...The more evidence you have [over a course], the easier it is to make a decision for the kid. I try and have at least three. At least three. It would be nicer to have more, but if they take two days each, that's a lot of time. (Interview, February 20, 2009)

Carina will adjust a final grade "if the kid does better on the final exam than they did on the term because they met the overall expectations in another way so I'll up it one mark or like whatever the difference is" (Interview, February 20, 2009). Before she submits her final

grades she then looks them over one more time, focussing on students with marks close to the next grade level:

I'm not proud enough of my own ability that I can say that I can zero it in to exactly one mark. I think about what it means to go from a 69 to a 70. For a kid it means a lot... It doesn't affect their average, really, [but] it does affect their perception of the course. Now let's say they bombed the exam then that 69 should stay a 69. (Interview, June 19, 2009).

Carina also described how she sometimes incorporates observations and evidence from her Assessment Checklist into students' grades. She recognizes that some students can demonstrate their understanding of overall curriculum expectations in class, but sometimes "fall apart on a test" (Interview, February 20, 2009). She indicated that she is comfortable using her professional judgment, based on her observations and the Assessment Checklist, in these cases to determine a student's final grade.

Tensions and Dilemmas in Classroom Assessment Practice

This section describes four areas where Carina experiences tensions and dilemmas in her classroom assessment practice: students' poor work habits and study skills, issues in the design and use of performance tasks, issues in collecting, analyzing and recording assessment data and colleagues resisting current reforms in classroom assessment.

Students' Poor Work Habits and Study Skills

Carina described several tensions and dilemmas that she experienced with students' poor work habits and study skills. These include students not doing homework, not taking responsibility for their learning and progress, and taking remediation opportunities for granted. Carina connects her concerns about poor work habits to issues in classroom assessment because she believes that many students are missing important opportunities for feedback.

Students not completing homework.

One area that Carina finds particularly unsatisfying is the number of students in her classes that do not complete assigned work. She wonders, “How do we get kids to do their homework and to realize that it is part of your commitment to the course? The math is about doing. It’s not about listening to someone else doing” (Interview, June 19, 2009). She feels that she puts a lot of effort into presenting material clearly, incorporating fun and worthwhile activities into her lessons, planning opportunities for students to get feedback on their learning, regularly updating the class website with handouts and homework, but is frustrated that this does not always translate into a strong commitment to the course from all her students. She explained:

Problem is they like to listen to it. They like to participate - they think the class is great. But then they don’t do the follow up. They tell me how they really enjoy my class - they actually tell me that. But they don’t do the follow up and then they wonder why their marks aren’t what they should be. When it comes down to the mark. Well, they’ve enjoyed the class, but they haven’t committed fully to the process (Interview, June 19, 2009)

Carina asks her Grade 12 students to commit to a half-hour per night of math homework, which she believes is reasonable for a senior Academic class, but she is frustrated by how few students actually do this. She has tried to explain to students that if they fall behind on their schoolwork then the subsequent lessons will not be as meaningful and their difficulties will build exponentially. Carina emphasizes to her students that cramming for a course just before a major assessment is not “as effective as just doing it bit by bit. And they don’t seem to get that and it’s really frustrating cause I’m seeing that more and more” (Interview, April 23, 2009).

Sometimes when several students are struggling with a formative quiz, Carina allows them to take it home to complete. She gets very frustrated when students do not hand these in. I observed this during one lesson where the Grade 12 class had done a quiz

during the previous class, and several students had not handed it in. Carina voiced her frustration to me, and explained to the students that they had missed an opportunity for feedback, and still offered to give feedback when students completed them.

In light of feeling unsatisfied with some of her students' commitment, Carina sometimes feels that she might give students more time in class to complete the work, but feels very conflicted about this:

So [pause] So I don't know. I [pause] have to think of how do I [pause] am I going to enable them to not do homework by giving them more time in class? Maybe I'd have to water some things down to be able to do it. So I felt a little pressured in trying to [pause] finish the course, get those assessments in (Interview, June 19, 2009)

Students not taking responsibility for their learning and success.

At the end of the Grade 12 course, before the exam, Carina was very frustrated that “nobody came for help. Nobody came to look at the solutions for the sample exam. You know. So that's telling” (Interview, June 19, 2009).

In previous semesters, Carina has been frustrated with students writing desperate notes on their tests and exams such as “I don't remember” or “I've never seen this question before”. Through her use of the Assessment Checklist and formative quizzes she believes that the main topics should be clear to students. She explains to students:

If you check off 'always' on every single thing [on the Assessment Checklist] you should be fine and at the end when you get to the exam you should have them all checked off if you want to be really successful” (Interview, February 20, 2009)

Carina explained how she feels when she gets these sorts of notes from her students: “Don't give me those notes. Where were you the whole semester when I was offering help? Where were you?” (Interview, June 19, 2009).

While she supports the goals of assessment reform, Carina worries about students with poor work habits not being prepared for post-secondary education. She believes that students will not receive the same sorts of extensions and second opportunities for assessments, but hopes “we’ve created enough in them that they realize that it’s in their best interest to get things done in a timely fashion (Interview, February 20, 2009). She related several stories she had heard through friends and the media where parents of university students had phoned professors or employers to help their children, and expressed concern that students are not developing life skills such as independence, persistence, resilience and self-advocacy. Unlike some of her colleagues, she does not believe there is strong relationship between current assessment reforms and the perception that young adults are not as responsible as in previous generations, but at the same time she does appear conflicted: “the only thing I worry about [with current assessment reforms]: is this going to perpetuate that even more? How accountable are we letting them be?” (Interview, February 20, 2009).

Students taking remediation opportunities for granted.

In Carina’s Grade 10 class during the previous year, she felt that a few students were taking the remediation program for granted and did not appreciate the extra work that teachers had to put in to the program. Teachers in the mathematics department at her school had decided that for students to be eligible for the remediation assignment (re-test), students needed to complete an assignment to demonstrate that they had taken steps to learn the material. Carina, along with other mathematics teachers, found that some students would not complete this new assignment, or did not attend the two mandatory support sessions (during lunch), yet still expected to be allowed to write the re-test. This sometimes resulted in conflict (Carina used the word “confrontations”) when a teacher would not

allow these students to write. Carina and her colleagues felt upset by these situations because they believed the remediation program was a positive program that they volunteered their time for, yet they did not feel this work was appreciated by all students.

Issues in the Design and Use of Performance Tasks

Carina feels that she has had “a mixed success rate with tasks” (Interview, February 20, 2009). Some of the issues she has faced are: difficulties in designing quality tasks for senior mathematics courses, issues with the academic integrity of performance tasks, and students struggling to use graphing calculators effectively as a tool to help their thinking.

Difficult to design quality tasks for senior mathematics courses.

Carina is not satisfied with her use of performance tasks in either of her courses: “I’m still struggling with that... I haven’t had a lot of success with tasks” (Interview, June 19, 2009). Although the tasks that Carina has designed are highly regarded by her colleagues, she is not satisfied with them:

I don’t know if they’re that great [laughs]. I know they’re more procedural that open ended, but I think you tie the course in a different way. So, it’s still that sort of [pause] I think we don’t have enough exemplars to know exactly what we should be doing with tasks, especially at the higher levels (Interview, February 20, 2009)

While her school district has had an initiative to design and assess performance tasks in the Grade 10 course, she believes these examples are not open enough, and in support materials released for the Grade 12 course Carina found “there really wasn’t much there for tasks” (Interview, February 20, 2009). She is currently trying to make her own performance tasks more open-ended: “One of the things that I’m trying right now is to try to get away from one answer” (Interview, February 20, 2009).

Carina finds one of the most difficult aspects of task design is setting the task at an appropriate level of challenge for students. She finds it hard to write problems in senior

mathematics that challenging yet accessible. She felt that one of the tasks she had recently designed and used in her Grade 12 course “was too easy” (Interview, June 19, 2009), but when she actually administered the task, many students had trouble with an aspect that she thought was common knowledge and had even briefly explained in the instructions to the task. Sometimes she believes that she simplifies a problem too much by scaffolding the question with steps. She wants the tasks to assess students’ *thinking*, not whether they can follow directions.

Academic integrity of tasks.

Carina is very concerned with issues of academic integrity and this is demonstrated in a variety of ways. For instance, due to the difficulty of finding or designing quality performance tasks, Carina struggles to find new tasks each semester to ensure the academic integrity of the assessments. While she often creates her own tasks, she finds them “really hard to come up with” (Interview, February 20, 2009). Carina does not believe it is fair to give the same task two semesters in a row since she often has at least one student repeating the course from a previous semester, and believes that students do share information about tasks from one semester to the next:

Some kids have wicked memories for these things. They might not remember to do their homework or whatever, but they’ll remember the third question on the fourth page of the exam from last June. Because it’s something that’s stuck with them. So they have two days to work over these tasks. It’s not that hard to replicate it on a piece of paper. (Interview, June 19, 2009)

One particular dilemma that Carina faced during this study was with a district-wide summative performance task in her Grade 10 class where six out of the 14 students who wrote it got a perfect mark. While Carina said this was a strong class, she was skeptical that so many students would write a perfect task without some aspect of the task having been compromised. She spoke to her class after having marked the tasks:

I said “this will not affect your mark but could you on this piece of paper. Don’t put your names. I don’t care. I just want to know. I know some of you have used resources at hand. All I want to do is accumulate some information to be able to say I don’t think the way that we’re doing the summative task is working. And so if you heard about the solution from anyone just tell me ‘yes’ and how you got the information. No names, please - I don’t care about that. And I’m not going to use that against you”. And six were up front that had ... not the six necessarily that had got perfect, either. (Interview, June 19, 2009)

Students told her that they had received information about the task through text messages from friends in the morning section of the same course (with a different teacher), while two others had received information from friends at another school. The students then discussed strategies and a solution during the lunch hour. When Carina spoke to teachers from other schools about their experiences, a colleague shared that when teachers at that school compared “how many perfects you had in period one, two, three and four. By period four they had about 75% perfect. Like that’s huge!” (Interview, June 19, 2009).

Graphing calculators.

Another issue that Carina has faced with performance tasks is that students struggle to use graphing calculators as tools to help their thinking. Carina believes that graphing calculators can be a valuable tool for problem solving and often incorporates them in instructional activities. She finds that many students are reluctant to use them, and during problem solving activities, students often choose paper-and-pencil strategies that take much longer. Many of her students need a great deal of assistance to perform what she feels are fairly basic operations on the calculator. While there are enough graphing calculators in Carina’s school so that each student can use one daily in class, students cannot take them home to use. Very few students own their own calculators. Despite her encouragement to use the graphing calculators and her use of them during instructional tasks, Carina finds most of her students are not fluent enough to use the calculator as a tool in problem solving.

She finds this frustrating because she believes that graphing calculators are an important problem solving tool: “it’s the actual thinking that you want to bring out, not the procedural stuff. That’s why you can do it on the graphing calculator” (Interview, June 19, 2009). She believes that this lack of fluency limits the types of problems she can pose during assessments.

Issues in Collecting, Recording and Analyzing Assessment Data

Carina has experienced several challenges related to collecting, analyzing and recording assessment data, both for formative and summative purposes. This section will describe some of these challenges related to her design and use of the Assessment Checklist, and in determining students’ grades.

Assessment Checklist.

As previously described, Carina has used an Assessment Checklist for the past few semesters in her Grade 10 classes, but this semester she decided to use it with both her Grade 10 and Grade 12 classes. She uses this Checklist to keep track of her formative assessment and modified it from the previous version to focus on the overall expectations of each course. The Checklist is regularly updated for each student.

The evidence that Carina uses from the Checklist comes from either written quizzes or from observations during formative assessment activities. In order for Carina to use a piece of evidence from a class activity to record on the Checklist, she likes to have all students complete it. For a written quiz, it is fairly straightforward, since she collects the quizzes and has physical evidence from which to update the Checklist, rather than having to remember how each student did. Often, however, Carina treats these formative quizzes more informally, circulating to see how students are doing and gives oral feedback. If they complete the activity to her satisfaction, Carina will initial the quiz, and then record this on

the Checklist. Carina did this during one lesson I observed where she declared, “It’s madness trying to get around to everyone!” (Field Notes, March 2, 2009). Since Carina views the major purpose of these formative quizzes to be individualized teaching she has a very high number of interactions with individual students or small groups during these informal assessment activities which she finds worthwhile, but exhausting:

It’s a lot of running around ... Keeping [pause] the recording. You have a pretty good sense, but the physical recording of it is what’s hard. Just to make sure. Because I will spend about fifteen, twenty minutes today. Because it has to be today. Before you forget ... But this is still the follow up. Now, this might take, you know, a minute each. But I’m still looking about fifteen, twenty minutes. (Interview, March 2, 2009)

Sometimes she feels that it is hard to keep up with because she often has students waiting for her attention and she points out that “it does take a long time to do all this” (Interview, March 2, 2009). Carina finds that the students that benefit most from the close monitoring are the most difficult to keep track of since they frequently defer assessments until a later date, take quizzes home and then do not submit them, or are absent for observational assessment. Carina would like to include more evidence on this Checklist, such as student presentation of solutions, but she has not determined a satisfactory way to record this evidence when not everyone has had the same opportunity.

Carina finds the Assessment Checklist much more difficult to keep up with her Grade 12 class, compared with her Grade 10 class. One reason for this is that the material is more complex so she often helps students to a greater extent than in her Grade 10 class. Also, more students take their formative quizzes home to finish and not all students hand them in. Near the end of an instructional unit, Carina often uses “challenger” quizzes to help students make connections between concepts and takes up the problem in class, which means that she is not always as sure how each individual student did with the problem,

compared to a more traditional written quiz. Also, with Grade 12 students Carina does not want to do as many small quizzes due to time constraints (covering the curriculum) and because she wants these senior students to be more responsible for their own learning:

It's just because of the nature of how much more [pause] um [pause] you know [pause] full the course is in terms of the expectations. You don't have as much time to do all these little quizzes. Or, I don't want to, also. I don't want to give them these little piecemeal quizzes. So it's okay to have them assess themselves with little skills drills, because that is like a mini-quiz. But I think it's more to put it on them. I think that's how it's working out to be. Instead of me taking in everything." (Interview, April 23, 2009)

She has also found the Grade 12s are reluctant to complete their own student Checklist as self-assessment. During one lesson I observed, Carina prompted the students three times during one period to fill in their own Checklist so that they could compare it with Carina's. While noting it is the first time that she is using the Checklist for self-assessment, Carina finds "that it's hard to get them to buy into it. That there is a purpose why they do it" (Interview, March 2, 2009)

Grading.

Carina uses both percentages and levels for marking, depending on the task. For instance, when using performance tasks she marks in levels, but uses percentages for tests. However, Carina is not satisfied with the school district's directive to convert levels from a rubric into percentage grades. The district has provided teachers with a conversion chart between levels and percents (e.g., a level 3- is 72%, a level 3 is 75%, a level 3+ is 78%, a level 4- is 84%, and a level 4 is 89%). She finds this "a huge problem. I think that it's sort of contradictory... How do we peg a mark at an 89? And does it have any meaning?" (Interview, June 19, 2009). She explained further:

If there's anything that I struggle with in this whole A and E [Assessment and Evaluation] is really about [pause] You do a task, you do it on a rubric. You're looking at overall. We don't mark point by point, right? You just sort of look at

the overall “how are they coping with that problem” and then you have to convert it to a mark. So it’s almost like, you know, sort of the fizz is out of the pop. So you’re like. Aw. Yeah. Flat. There - now it’s a mark. How do we get around that? (Interview, February 20, 2009)

Another challenge that Carina has is including evidence from classroom observations as part of students’ grades: “I do find that very hard ... I’m not there yet. I’m still not there” (Interview, June 19, 2009).

Colleagues Resisting Current Reforms in Classroom Assessment

A final area where Carina experiences tensions and dilemmas is with colleagues resisting reforms in classroom assessment. She has encountered this both at her own school and during district-wide professional learning opportunities.

In own school.

While very respectful of her teaching colleagues, Carina does not feel that regular department office talk is sufficient to move teachers towards the vision of assessment as put forth in the Ontario curriculum documents. She said, “the current thinking is not necessarily coming from the math teacher community” (Interview, February 20, 2009). She does value the learning opportunities presented at professional activity days and stated that those “are already trying to go beyond. I’m saying just in terms of math teachers getting together ... It is not going to happen” (Interview, February 20, 2009).

She finds that there is a critical mass of teachers at her school who are reluctant to reconsider their assessment practice:

I think a lot of people are not willing to look at an idea and a view because they’re so ingrained with what they’ve always done that they can’t step outside and say “why is it that I’m doing this?” ... And it’s easier to keep doing the same things over and over. (Interview, February 20, 2009)

One assessment issue that has emerged as particularly controversial at Carina’s school is the debate over whether zeros should be assigned for missing work. The district

has released a new policy that states that every effort should be made to avoid giving a student zero for missing work because it skews a mark if a grade is being determined by averaging. She feels impatient that several recent meetings she has attended have been consumed with teachers expressing anger over this new policy. She believes that many of her colleagues do not understand the rationale underlying this approach and often use it as a means of wielding power to try to control students' behaviour. She thinks many teachers need to reflect on this practice:

[as if speaking to a colleague] "Why is it so important that you have that zero there to hang over their heads?" Let's look at the language. Because, really, that's what you're doing. You're hanging it over their heads. So is it a punishment? And is it a punishment because they don't have the concepts ... Or is it because you want to have control? ... So what if the kid is a real jerk, but man can he do math! Does he deserve anything less than what he can give you on the evaluations because maybe he didn't show up for your classes three or four times?
(Interview, February 20, 2009)

Another area where Carina has encountered some resistance to current assessment reforms is with the remediation program. She referred to the implementation of this program as a "political issue" (Interview, February 20, 2009). Several teachers in the mathematics department were concerned about an increased workload since they had to help supervise extra help sessions and then mark the re-tests. Other teachers expressed discomfort with the holistic grading that was to be used (rather than a marking scheme).

Carina explained:

A lot of people had problems with the whole idea of eyeballing. That's just still not [pause]. You know, math teachers are [pause] they trust numbers. So they're good at numbers. They get to plug it in and see what the number comes out. So what is a problem is being a little more creative in identifying what the overall expectations are and how well they've reached them, right? (Interview, February 20, 2009)

The school administration supported this initiative and even reduced the supervision duties for the teachers who organized and ran the program (two teachers in the department). This

caused friction with some other teachers outside the mathematics department who took issue with the supervision credit, which Carina described as a “huge, huge kafuffle” (Interview, February 20, 2009).

Another more recent issue arose when a colleague who has been very vocal in his criticism of assessment reforms reported results of an internal teachers’ federation poll about classroom assessment. Carina felt that he presented a skewed version of the data that undermined many of the district and school initiatives around assessment, and capitalized on some teachers’ perceptions that current reforms are not good for students’ learning. She noted that he rolled his eyes in some parts of the presentation and felt that the presentation was an example of “how to lie with statistics” (Interview, June 19, 2009). She feels particularly torn because she believes this colleague is “an amazing teacher” and actually does practice many aspects of the current assessment reforms. She thinks “he just resents being told that he has to do it... He’s dangerous because he doesn’t actually practice that what he’s saying” (Interview, June 19, 2009).

Some of Carina’s other colleagues worry that parents will misuse a written assessment policy that allows multiple opportunities and does not allow teachers to assign zeros for missed assignments. Carina has suggested that teachers should “explain to people what the intent is. And I have tried to argue in favour of it” (Interview, February 20, 2009). While Carina is supportive of this policy, she is not comfortable defending it in public forums like staff meetings because she finds the conversations get quite ugly and some colleagues at her school can be aggressive.

At the district level.

At the district level, Carina is supportive of most of the assessment initiatives, but she does have concerns with how some of the staff are promoting the changes. In some

aspects, she feels that some of the changes are being pushed forward without an explanation to teachers who have not yet bought in. She feels that some of the assessment reforms could be promoted in a more appealing manner. Carina perceives that some of the recent changes have been introduced as

“This is the way it’s going to be and that’s it” whereas [teachers] just want to know why is this better? Like why am I going to invest a whole bunch of learning here in how to mark this and all that? Because it’s going to be agonizing for a lot of people who haven’t done much with rubrics. Which is unforgivable at this point. But you know what I mean, right? That whole slow change thing. (Interview, June 19, 2009)

She is worried that some of the recent changes may not last when the consultants’ terms end and

I’ve already spent [pause] I’ve invested a lot of energy and a lot of time in making sure I’m careful in how I mark the way I mark. So why am I now going to give that up? And all that investment and time and energy to take this new thing on when maybe next year they’ll decide that didn’t work. Let’s just go back to the way things were. I need some assurances.” (Interview, June 19, 2009)

Negotiating Tensions and Dilemmas in Practice

In this section I will describe some of the ways that Carina negotiates these tensions and dilemmas in practice.

Students’ Poor Work Habits and Study Skills

Carina has several ways in which she deals with students’ poor work habits and study skills. At the beginning of each course she gives her students a handout called “How to Get an ‘A’ in Math!” (Weiner, 2004, p. 1) and takes time to discuss it in class. It outlines specific study strategies for math as well as how to effectively ask questions of your teacher. During the course, Carina also monitors students’ progress using the Assessment Checklist, which, in addition to the learning function, provides an accountability check for her since she has a record of any notes or contacts she has made. She does not expect to be

surprised very often by students' performance on major assessments "at least I have a sense as to where they're at. So that tomorrow it's not [pause] it shouldn't be a huge surprise. I've already contacted parents" (Interview, March 2, 2009). Carina can follow up with individual students who were absent or missed work when she reviews her assessment binder. Throughout a unit, the formative quizzes get progressively more complex as she generally moves from "skills checks" to "challenger" quizzes. Carina uses the challenger quizzes to integrate ideas from the unit and talk specifically about strategies for problem solving:

So what do you write down? These are things you can write down, and you know what? By the time you write them down you could have some strategy of what to do. So, that's sort of what you try and teach them in Grade 10 and you can do it through the assessment (Interview, March 2, 2009)

She feels that students appreciate these learning opportunities and, referring to one formative quiz described:

This was a challenger and I wanted to see what they could do with it. Really, as you saw, most of them had at least a strategy of how to do that and might have gotten hung up on the algebra cause it was a little bit more complicated than some things. But, really, I think by the end of it, I know [name of student] said "oh, I'm really glad we did that." They usually feel that that was a good exercise (Interview, March 2, 2009)

Circulating during these quizzes allows her to see every student individually with their own questions. Sometimes Carina asks students to review reflection or *thinking* questions in small groups so she that students are forced to engage with the material orally and she can hear their thinking.

With her Grade 12 classes, she does not monitor students' class work as closely, because she believes that they need to learn to be more responsible for their work. She does not want to do daily homework checks in Grade 12, and finds that she does not have as much time as in earlier grades for lots of small quizzes to keep students on track:

You don't have as much time to do all these little quizzes. Or, I don't want to, also. I don't want to give them these little piecemeal quizzes. It's okay to have them assess themselves with these little skills drills, because that is like a mini-quiz. But I think it's more to put it on them. I think that's how it's working out to be. Instead of me taking in everything. This one quiz I wanted to do because it's a big chunk of it [the unit of study]. I will provide feedback. And I will jot it down [record it on the Assessment Checklist] (Interview, April 23, 2009)

While frustrated with students' work habits, Carina describes that she feels more "relaxed" with the current approach to classroom assessment. She feels that she now has more flexibility in how she administers assessments. She delays tests when she feels students are not prepared or allows students to take formative quizzes home to complete. On one particular quiz

I had about half do it on the day of. The other half weren't comfortable so they didn't do it. Then I said "well, why aren't you comfortable? Cause this is not new material today. It means you're not caught up. But I haven't seen you for help, so why is that?" I'm hoping to generate a bit more accountability (Interview, March 2, 2009)

While she admits that with the formative quizzes "it's true that a lot of times they don't study, but I think they're still valid in that 'this is what you know without studying - Is this good enough for you?'" (Interview, February 20, 2009).

She believes that the new approach to assessment has changed her relationship with her students for the better. She feels that she and the students are

in it together. It just feels that way. Like, you know. And it's frustrating then when they don't do anything about it, but they at least come out of it knowing they had the opportunity to. And you're never there just sitting back. (Interview, February 20, 2009)

Even with students who have ended up not doing well in her courses, Carina believes that they think she is a fair teacher who did give them the opportunities to succeed, but that they did not pull their own weight. Referring to a conversation she had some students after an exam she asked the students: "so how was the exam" and they said, 'yeah. It was a good

exam' and then that 'hmmm. Wish I had studied a bit more'." (Interview, June 19, 2009).

She feels that students find her very supportive: "they know that I am their biggest cheerleader. And I'm there for them" (Interview, February 20, 2009).

Carina also tries hard to use engaging activities in class to motivate her students. She regularly combines humorous graphics on the interactive whiteboard with stories or music in class activities. Describing one activity with her Grade 10 class where she used two cartoon characters and a story to discuss the number of solutions to a system of linear equations, she found her students "were so excited ... even now they're still working on it because it's a little more fun than just 30 questions from the textbook" (Interview, February 20, 2009). Another time she did a treasure hunt with her class, where students had to solve math problems to find clues and then a small prize. Describing her approach, Carina said, "I try and grab them right away with word problems and make things kind of neat. I do have a lot of neat stuff in the first unit because I do think it's important to get them" (Interview, February 20, 2009).

Carina is a very reflective teacher and she is thinking about making changes to the Grade 12 course, which she has now taught four semesters in a row and will teach again next year. She is specifically thinking about changing her approach to homework. She currently does not give a lot of time in class for homework:

I don't give them a lot of time in class to do their work. I know there's two different camps. One is to show them an example and then have them work through the textbook and then they'll encounter other problems and maybe they'll ask. But at least they'll get a chance to practice some stuff. Mine is, I try to explain the background and how it all hooks together, you know, so they have a conceptual framework. And then I figure if they do their homework they'll know how to do it all (Interview, June 19, 2009)

But she is now thinking that she may "rethink the Grade 12 course because I'm finding that for those kids who do all their work and that it's great. I have it set up perfectly for

success” (Interview, June 19, 2009), but she is concerned that students who are not doing their homework are not being successful. She felt that many of the quizzes ended up being homework practice, instead of being treated by students as opportunities for feedback into their learning. Doing the quizzes

was sort of like giving them homework practice instead of actually checking to see what they knew. So I’d give them [quizzes] towards the end then there wouldn’t be enough time for me to take them in, look through them and give it back with enough time for the feedback to be useful. Do you know what I mean? And yet, I know that they would tell me “yeah, I don’t know how to do that” already. So that’s [pause] “okay, that’s your assessment at this point” (Interview, June 19, 2009)

Carina is also interested in reconsidering the remediation program and stated that she “wouldn’t mind looking at some other options because I was feeling like it was taken almost as a given” (Interview, June 19, 2009).

Issues in the Design and Use of Performance Tasks

The primary way that Carina negotiates tensions and dilemmas in the design and use of performance tasks is through conversations and collaborations with colleagues. She looks carefully at any samples of tasks that she finds, either from colleagues, online, or in course resources and discusses these with other teachers who are currently teaching the same course. She has gained confidence over the past two or three years to share some of her tasks beyond her own school with other teachers in the district. In recent years the Grade 10 course, which she has taught many times, has been the focus of a district assessment initiative to design and disseminate quality tasks. In addition to the actual resources, teaching this course has given Carina the opportunity to attend professional development sessions on tasks and to hear perspectives of district consultants, mathematics department heads from other schools and other teachers. She has brought some of these ideas into her own task design in her other courses. In addition, by listening to other

teachers' difficulties, she feels validated in her struggles to create tasks that are open-ended and at an appropriate level for her students.

The issue of academic integrity has been raised several times at these meetings, and Carina can see that there is not an easy solution to this problem. Many of the district's Grade 10 sample tasks require two class periods to complete, where students get partial information on the first day and work in groups, while on the second day they receive the complete information and complete the task independently. Two veteran teachers who have been leaders at the district level on tasks have different approaches to the problem:

"[Colleague A]'s solution is that they have to be open-ended. [Colleague B] says you can't have them over two days." (Interview, June 19, 2009). When Carina experienced the issue of academic integrity on her Grade 10 summative performance task during this study, she sought these two colleagues and others out to discuss solutions for future semesters. For district wide performance assessments, she believes that they must either be set up so that all students write them at the same time, or that there needs to be "more than one where the twist is significant. You can have a similar situation but the twist is not just change a number or two" (Interview, June 19, 2009).

Although she finds it challenging to set tasks at a level that is not too prescriptive, yet not too difficult, Carina enjoys the creative aspect of the design. Since she has taught several of the senior mathematics courses many times, she has had an opportunity to share her tasks with several different teachers at her school and to refine them with each use. On one particular task that she thought was too easy, she has since found a web-based simulation of one aspect of the problem that she would now show her class as part of the pre-task activity to clear up confusion.

Since it is difficult to come up with new tasks each semester Carina takes advantage of the few quality *thinking* questions in the course textbook to help students learn how to approach the questions. She sometimes puts questions from the textbook together to use as “mini-tasks” (Interview, April 23, 2009) to help students prepare for a major performance task.

Carina prepares her students for the performance assessments using pre-task activities that are “directly connected to the task” (Interview, February 20, 2009). She describes one example from her Grade 12 class where the task involved using a pendulum to investigate a periodic function:

We did have a pendulum out as a pre-task thing where they had to with the CBR [Calculator Based Ranger] look at how the [graphing] calculator takes the analysis and does the regression and what each part means. And what does it mean to swing a pendulum. So they had one day to do that in groups like in a physics lab. And the next day I had data because I wanted to make sure it all worked. (Interview, February 20, 2009)

Since she finds that students frequently struggle with using the graphing calculators, Carina also reviews any functions on the graphing calculator that students may need during the pre-task activities: “I said ‘This is what you have to be able to do with a calculator so is there anything you don’t know how to do?’ or went through it” (Interview, February 20, 2009). She is also incorporating the graphing calculators into more class activities. She recently acquired software for her interactive whiteboard that emulates the graphing calculator. This has a display that allows students to see several screens at once so they can follow the keystrokes without getting lost: “Part of the problem is that with tasks they can use the graphing calculators and they’re not good with the graphing calculators so that’s what I’m hoping this emulator will allow me to do” (Interview, February 20, 2009).

Issues in Collecting, Recording and Analyzing Assessment Data

Carina finds it challenging to ensure that she gives students individualized feedback on their formative assessments during class time and then to accurately update her Assessment Checklist. She finds this easiest when students hand in a written quiz so that she has physical evidence by which to update the Checklist:

So it's easier when I have it on a piece of paper instead of trying to remember. I sit afterwards when I do their attendance. I update [the class website] everyday with the stuff they have to do. So I do have a routine ... It's just a matter of making sure I do it right away. Cause you know, 25 kids in there and try to remember who needed help and who didn't (Interview, February 20, 2009)

There are several instances where Carina refers to the Assessment Checklist as not just being important for keeping students' learning on track, but also to hold students accountable for their own efforts and learning. With the Checklist and the accompanying notes, Carina can demonstrate all of the efforts that she has made to help students and she feels that this puts some of the onus back on her students:

I am not the great evaluator who gets to decide. They get to decide and that's what all this [assessment reform] has given us. "So what have *you* done about it?" Like "You haven't taken the time to fill it in [self-assessment]. So does that mean you haven't thought about the course at all? I have!" (Interview, February 20, 2009)

Over the semester she has found that students are quite reluctant to complete the student version of the Checklist and accompanying self-assessment and believes:

They're not used to it, partly. They don't see the relevance yet. I think that they see the relevance when it comes to the final exam. That's how I'm selling it to them. I said to them, "for the final exam, you will know exactly what the expectations are" (Interview, March 2, 2009)

She has decided that she will "emphasize it a lot more next time I do it. I did it in the first unit where I did compare mine to theirs" (Interview, April 23, 2009).

While it is exhausting for Carina to collect the individualized assessment data through quizzes and classroom interactions, and maintain the Assessment Checklist, she

feels the effort is worth it. Describing how she felt after a lesson where I observed a class doing a formative quiz that she recorded on the Checklist, Carina said, “I was pretty happy with that. It was crazy and I was tired but I initialled everybody’s paper at the end of it” (Interview, February 20, 2009). Referring to the one-on-one work that she does with students during the quizzes:

Well, today it was nuts, but honestly... They were totally involved. They were totally engaged and they even said “I learned a lot today. I thought I knew it” and then it came to doing an exercise on the graphing calculator and it’s like “I didn’t know how to move my window” so that was a question that didn’t come up yesterday (Interview, February 20, 2009)

Carina felt very proud that her own principal, who happened to be in the photocopy room as she was copying it for her classes, has photocopied her Assessment Checklist. A department head from another school also asked her for it, which made her very pleased.

In another area related to recording assessment data, as previously described, Carina has been unsatisfied the district’s “pegging” system where levels on a four-point rubric are converted into percents. She finds this unreasonably precise, but does follow the district procedure. For each task, she converts the level into a percent and then in her grading software enters it out of 10 so that all the tasks are equally weighted and can easily compared when the final grade is being determined.

With regard to Carina’s struggle to incorporate observational data into students’ grades, she finds it difficult to include data for activities that not all students completed, but is willing to consider it where a student’s performance on a major assessment does not match prior performances. If Carina is surprised by a student’s low mark on a test, she is willing to refer back to the Assessment Checklist and uses her judgment to either allow the student to re-do a different version of the assessment, or to use data from the Checklist in place of the low mark, where she is certain that the student has met the particular course

expectation. She described: “If they’ve always shown in everything they’ve done that they have been meeting the expectations and they fall apart on a test - with a kid like that you have to take [the Assessment Checklist] into consideration. That’s when your judgment comes in” (Interview, February 20, 2009).

Colleagues Resisting Current Reforms in Classroom Assessment

Carina deals with tensions and dilemmas in relation to her colleagues by talking out the issue privately with other colleagues that she feels share similar views on assessment. She does not choose to respond to comments that display resistance to assessment reforms during staff meetings, but discusses these later with trusted colleagues to see if they also disagreed and to then pick apart the opposing argument. Carina seeks out professional relationships with like-minded teachers who she sees as also having a good sense of the “new” approach to assessment. This has taken the form of non-district initiated collaborations on test, task and exam design with teachers from other schools. Sometimes Carina will respond to resistance by seeking conversations with district consultants so that they are aware of dissonance at the grass roots level and she hopes will address these sorts of concerns in future professional development sessions.

She is very aware of the political nature of current assessment discourse and that words must be chosen very carefully. Carina cited an example where a colleague in her school made a worthwhile suggestion that was lost due to the blunt nature of the delivery. Carina believes that communication skills are very important for changes in assessment to take broader hold: “that’s why sometimes you don’t get a message across. Because you don’t say it quite right that people will take it and think about it rather than react to it” (Interview, June 19, 2009).

Carina is quite confident in her own assessment practice. She is recognized within her own department as having well thought out assessments, and takes pride in sharing these within her school's mathematics department.

Carina recognizes that broad changes to classroom assessment will only happen as individual teachers make sense of the reforms and then share points of view or collaborate with other colleagues:

It can't be just from the top - it's got to be directed as a peer. I think that's the only way to succeed. I think [the problem with current district assessment initiatives] is they've been trying to do it by having an expert in each school and that's not working. (Interview, June 19, 2009)

She believes that the most effective learning happens when sharing courses. In these situations, teachers have to negotiate common assessment practices and often share examples of what they are doing.

Summary

A summary of the tensions and dilemmas that Carina described in her assessment practice, as well as how she negotiates these is included in Table 5.2. Over the past year, Carina has focused on developing her Assessment Checklist and trying to encourage students to engage in self-assessment. While she is happy with the Assessment Checklist she has designed and uses, she finds it difficult to maintain. She does feel that the effort is worthwhile, particularly early in the semester, to serve as an early warning system for students who are struggling. She also believes that the Assessment Checklist serves to hold students accountable for their own efforts and learning since she can demonstrate all of the efforts she has taken to support their learning. She stated that next semester she plans to emphasize the self-assessment portion of the Checklist more by having students compare their own copy to hers for each unit. Her confidence in the value of this tool was reinforced

by several colleagues who have seen it and asked for copies, including her own Principal who wanted to share it at a PD session. She is cautious about making change in her own practice and carefully considers anything before adopting it. For example, several years ago she decided that her quizzes would no longer count in students' grades after several other teachers in her department did this. She describes needing to understand the rationale behind change and why this is better for her students' learning before she is willing to adopt a new strategy or tool. She also prefers to make these changes at the beginning of a semester, rather than making changes as she goes, to present a coherent assessment plan to her students. She believes that the "new" approach to assessment has improved her relationship with students and feels that even students who do not commit to the coursework think she is a fair teacher. Carina is very reflective about her teaching and students' engagement in her classes. She believes they enjoy the class and lessons, particularly some of the rich activities, investigations and simulations done as instructional activities.

Carina's response to a summary of her case, including description of classroom practices, challenges in classroom assessment and ways these challenges are negotiated was: "I think it's pretty accurate" (Personal Communication, November 2009).

Table 5.2: Summary of tensions and dilemmas Carina has experienced in her classroom assessment practice and how she negotiates them

Tensions and dilemmas experienced in classroom assessment practice	How these tensions and dilemmas are negotiated
1. Students' poor work habits and study skills a) Students not completing homework b) Students not taking responsibility for their learning and success c) Students taking second chance opportunities for granted	Confidence in her own approach to teaching, learning and assessment Belief that current assessment reforms are beneficial to students' learning Participation with mathematics teacher community within and outside her school, in both formal and informal activities – particularly informal conversations with colleagues she trusts and collaboration in course planning and assessment design
2. Issues in the design and use of performance tasks a) Difficult to design quality tasks for senior mathematics courses b) Academic integrity of tasks c) Graphing calculators	Close monitoring of individual students' achievement of overall course expectations (both formative and summative evidence) with follow up (assessment checklist and individual conferencing, parent contact)
3. Issues in collecting, analyzing and recording assessment data a) Assessment Checklist b) Grading - converting levels (e.g., from a performance task marked with a rubric) to percentage	Enjoyment of the creative aspect of designing instructional and assessment activities (e.g., performance tasks, integrating stories or humorous graphics with simulations on interactive whiteboard)
4. Colleagues resisting current reforms in classroom assessment a) In own school b) At the district level	Use and knowledge of a wide variety of resources Reflective about changes she can make to improve students' learning (e.g., reconsidering use of class time for independent work, plans to emphasize self-assessment in future) Compliance with policies and directives at school, district and provincial levels (e.g., pegging marks, mark weighting)

CHAPTER 6: DISCUSSION OF FINDINGS

The previous chapter presented case studies of Amy and Carina individually with a description of each teacher's assessment practices and goals, tensions and dilemmas and how she negotiates these tensions and dilemmas in practice. This chapter now discusses the findings from both cases considered together. First, I address the two research questions that have guided this study, using the conceptual framework to both organize the discussion and point to some of the causes for some of these tensions and dilemmas. Finally, I put forward three assertions based on cross-case analysis.

Addressing Research Question #1: What Tensions and Dilemmas do Two Secondary Mathematics Teachers Experience as they try to Implement Current Assessment Reforms?

In this section, I address the first research question using the conceptual framework (Figure 3.1) to organize the discussion. First, I discuss assessment decision making (McMillan, 2003) and compare the tensions and dilemmas Amy and Carina experience to the four components of assessment (purpose, eliciting, interpretation and use) proposed by McMillan (2004). This framework is useful for identifying common areas where both Amy and Carina experience struggles in their assessment practice. Second, I consider the external and internal influences on Amy's and Carina's assessment decision making. While there are tensions between these various influences (McMillan, 2003; Saxe et al., 1999), teachers' awareness and experience of potential or actual conflicts between these influences that need to be taken into account in classroom assessment practice is what I refer to as *tensions* in this study. Next, I highlight some of the assessment *dilemmas* that arose during the course of the study where the teacher was required to make some difficult decisions. These dilemmas often brought out strong feelings of uncertainty and frustration in both Amy and Carina.

Assessment Decision Making

As described in Chapter 3, along with McMillan (2003), I view decision making as a central aspect of teachers' work, particularly in classroom assessment. Teachers have a huge number of choices they need to make as they plan for lessons, during lessons and after lessons as they consider students' work and plan for future learning. These decisions include instructional decisions such as what activities would work best for a particular group of students, how to sequence topics effectively, and what materials to use. They also include assessment decisions such as when to give tests, quizzes or performance tasks and what items to include on these formal assessments. Many other decisions need to be made in areas such as administration, management and motivation. The nature of teaching and learning in classrooms means that many of these choices need to be considered and acted upon simultaneously (McMillan, 2004). As McMillan (2003, 2004) and others point out (e.g., Brookhart, 2004; Even, 2005; Suurtamm, 2004), decisions about assessment and instruction are highly interrelated. Teachers are constantly making decisions as they observe their students to decide when and what questions to ask, when they should "switch gears", which students they need to attend to, how long to spend with each student or group, when to probe deeper into a students' thinking, and even what location to be in the classroom to give themselves the most valuable information.

McMillan (2003) defines assessment decision making as "a process in which teachers balance the demands of external factors and constraints with their own beliefs and values to determine classroom assessment practices" (p. 42). I see this as strongly connected to tensions and dilemmas in teaching. As previously discussed, Talanquer et al. (2007) define teaching dilemmas as "problem spaces created in the minds of teachers as they engage in the practice of teaching" (p. 401) and Woods et al. (1997) suggest that

teaching dilemmas occur when teachers are “confronted by choice, which they are free to make, although at times it may be difficult to do” (p. 19). In the cross case analysis presented in this chapter, I focused on the choices and decisions that both Amy and Carina had to negotiate, and how the influences on these were often in conflict, resulting in tensions and dilemmas. In the analysis here, I use McMillan’s (2004) outline of four components of classroom assessment that point to different types of assessment decisions that must be made by teachers: the purpose of the assessment, the means for eliciting data, interpretation of the data, and the use of the assessment data. Here, I use the components of this framework to highlight particular areas where Amy and Carina experienced tensions and dilemmas in their practice, and areas where they did not.

Both Amy and Carina appeared quite comfortable with two of the components of assessment: decisions concerning the purpose of various classroom assessments, and how they choose to elicit assessment data. Over the past few years, both teachers have increased their use of formative assessment including more attention to oral and written feedback to students, use of assessment to inform instruction, and recognition of the value of informal assessments including observation. Both teachers purposefully select assessment evidence when determining students’ grades, rather than using all assessment evidence. Generally, grades are determined from unit tests, performance tasks, and a final exam, but both teachers use a much wider range of strategies to get a sense of their students’ learning. They appear to view the performance tasks as addressing *thinking* and *communicating* more effectively than tests and are generally satisfied with this. Both teachers tend to see formative and summative assessment as separate, but sometimes offer students further opportunities to demonstrate their achievement when scores on summative assessments were lower than expected (based on formative assessment). Both Amy and Carina

mentioned that they would like to incorporate some oral assessments (e.g., from oral presentations) to expand the variety of assessment modes, but neither is sure how to manage this in terms of class time or issues of fairness (consistency). This did not appear to be a major source of tension for either teacher.

On the other hand, Amy and Carina both experienced significant tensions and dilemmas around the other two components of assessment: interpretation and use of assessment data. In particular, both teachers expressed uncertainty in decisions related to the analysis of assessment data for reporting purposes. For instance, Carina takes into consideration students' most recent, most consistent work in the Achievement Categories of *communication* and *thinking*, but is dissatisfied with how few examples there are to give teachers direction in this analysis. Amy expressed concern with the lack of consistency she sees between teachers at her school where some teachers purposefully select assessment items while constructing grades, while others strongly supporting averaging all available data. Both teachers also indicated they were not confident in combining different types of assessment data (such as scores from marking schemes, percentages and levels from rubrics) to determine a course grade for reporting. In addition, Amy stated she is not confident that she interprets data from formative assessments effectively since she still finds herself surprised by some students' performances on tests and performance tasks. Finally, both teachers expressed some concern over their *students'* use of the assessment data, particularly from quizzes and tests, to improve their learning.

McMillan's (2004) framework for the four components of assessment is helpful here because it points to two main areas of tensions and dilemmas: the interpretation and use of assessment data. Now, I turn to a discussion of tensions between internal and external influences on classroom assessment practice.

Tensions Between Internal and External Influences on Classroom Assessment Practices

McMillan's (2003) framework for assessment decision making highlights tensions between internal and external influences on teachers' assessment practice. Saxe et al. (1999) also point to this idea, and in their discussion they distinguish between two types of external "presses" on practice (institutional and key stakeholder) and internal "presses" on assessment practice. Both McMillan and Saxe et al. suggest that some of these influences or presses act to move practice in the direction of current reforms while others act to inhibit such changes. In this section, I discuss the different influences on Amy and Carina's classroom assessment practice beginning with the external influences from institutional and key stakeholder presses. I then address internal influence on teachers' assessment practice.

External Influence #1: Institutional Presses on Classroom Assessment Practice

The changes that Amy and Carina have made in their assessment practice over the past few years have coincided with recent revisions to Ontario's curriculum (in all subjects) which reflect the goals of current educational reforms (Fullan, 2009; Levin, Glaze & Fullan, 2008; Levin, 2009; Suurtamm & Graves, 2007). The institutional influences on teachers' assessment practice discussed in this section include provincial curriculum documents, district and school assessment policies, professional development, curriculum supports (such as textbooks and provincially funded resources), and large-scale assessment.

As described in Chapter 5, Ontario has a reform-oriented province-wide mathematics curriculum that was first implemented in secondary schools in 1999 and then revised in 2005. The curriculum documents include a section on assessment, which describes strategies and tools aligned with current assessment reforms. Among other considerations, teachers are directed to use assessment strategies that are varied in nature, allow students to demonstrate the full range of their learning, promote students' ability to

assess their learning, and ensure students are given directions for improvement (OME, 2005). Along with these changes, the province provided funding for new textbooks and supports aligned with the curriculum, as well as professional development. In addition, a new provincial Student Success initiative requires formal supports for students at risk of failing courses through Credit Recovery and Credit Rescue opportunities (described in Chapter 5). More recently, there has been a focus on reviewing assessment policies at the provincial, district and school levels (OME, 2008). All of these initiatives were introduced to encourage Ontario teachers to adopt reform-oriented practices, including classroom assessment practices.

In Ontario, messages from these provincial policies and initiatives are translated into more specific policies at the district and school levels. The province requires each school district and school to write their own assessment policies that are aligned with provincial directives. Amy and Carina both experience difficulties in interpreting and implementing assessment policies from different levels and do not believe there has been sufficient clarification offered with respect to issues in grading and reporting. For example, both teachers have recently shifted their methods for recording assessment evidence for both major assessments as well as formative assessments. While each teacher believes her current methods are an improvement on past practices, neither is satisfied. Both Amy and Carina have received the message that they should be considering students' most recent and most consistent work when determining course grades, but neither has seen satisfactory examples of what this looks like in practice. Both teachers also feel they should be doing more to individualize students' assessments but they are not sure how to reconcile this with their own ideas about fairness and reliability. In determining grades for reporting, both teachers expressed uncertainty in interpreting assessment policies from their own

mathematics department, school, district, and provincial Ministry of Education in addition to messages from professional development and professional reading and integrating them into a coherent practice.

While there has been provincial and district funding for professional development to support Ontario's curriculum and assessment reforms, both Amy and Carina share the concern that these sessions have not been effective in raising teachers' understanding of the foundations and rationale for these reforms. When some of their colleagues publicly resist messages supporting current reforms, Amy and Carina are frustrated that these issues are not sufficiently addressed. Both teachers would like to see school administration and district staff taking a more direct approach and being more proactive to countering this resistance.

While McMillan (2003) and Saxe et al. (1999) identify curriculum supports as a source of influence on teachers' assessment practices, both Amy and Carina expressed concerns about the lack of specific direction from curriculum resources (such as textbooks and Ministry-funded curriculum supports) in classroom assessment. Amy and Carina both stated that they had benefited from examples of performance tasks in curriculum resources, particularly at the Grade 9 and 10 levels, but felt they provided little guidance on scoring. In particular, Carina was frustrated that there were not more examples of quality performance tasks for Grades 11 and 12 mathematics, and Amy would like more assessment examples to support her use of the provincially funded TIPS4RM resource for Grade 9. Neither teacher uses commercially produced tests from publishers.

McMillan (2003) cited large-scale accountability testing as an important external influence on assessment practice, but this does not appear to be a major influence on Amy or Carina's practice. Since Ontario's province-wide assessment occurs in Grade 9, Carina

has never experienced it since she has never taught Grade 9 mathematics. During this study, Amy was teaching Grade 9 mathematics and I observed her using some of the publicly available release items from the assessment during one lesson (all multiple choice). She appeared to believe that the sample items were helpful to students, and used them for both instructional and formative assessment purposes. In describing the changes she has made in her own assessment practices, she never mentioned the provincial assessment as a significant source of influence, nor did she appear to use the format of the assessment (mainly multiple choice) in her major assessments. This may be because the content of Ontario's provincial mathematics assessment is quite closely aligned with the mathematics curriculum. Writing specifically about this assessment, Suurtamm, Lawson and Koch (2008) argue that it is "aligned with many tenets of reform oriented curricula" (p. 42) but note that "there is still more work to be done to have full coherence and we are concerned with the impact on classroom practice." (p. 42).

Amy and Carina appear to see a great deal of alignment in messages about assessment between many of these formal institutional presses including curriculum documents, policies at different levels and messages from professional development sessions, which may be why they have been able to adopt many strategies aligned with current assessment reforms. The tensions that they experience mainly arise in areas where they perceive a lack of clarity, which was a point acknowledged by Shepard (2000) in her writing about the implementation of a reformed vision of classroom assessment. She stated that while there was a strong theoretical basis for the emergent view of classroom assessment, it was not clear how this would work in practice.

External Influence #2: Stakeholder Presses on Classroom Assessment Practice

Another significant external influence on teachers' classroom assessment practices identified by both McMillan (2003) and Saxe et al. (1999) is key stakeholders such as colleagues, administration, students and parents. In this section, I examine how each of these affects Amy and Carina's assessment practice.

Amy and Carina both emphasized the importance of working with colleagues, in both formal and informal settings, at their own school and beyond to develop their classroom assessment practice. They both have developed networks of teachers with whom they feel comfortable sharing issues and questions. Both teachers are highly involved in professional activities at their school and in the district, and are comfortable discussing issues of classroom assessment with teachers in their own mathematics departments and district staff (instructional coaches and curriculum consultants). Informal conversations with district staff appear to have a great influence on both teachers since they feel these teachers are well informed and speak with authority and insight about interpretations of curriculum documents and district policies. For Amy and Carina, the most productive activities for developing assessment practices aligned with current reforms appear to be opportunities to collaborate in the creation or scoring of assessments, as well as discussions and sharing examples of constructing grades.

While acknowledging the beneficial nature of working with supportive colleagues to develop assessment practices, both Amy and Carina expressed concerns about other colleagues who are resisting the current reforms. Both teachers described feeling uncomfortable during staff meetings at their schools, or at district professional development sessions where some colleagues were openly hostile towards current reforms and district initiatives. They both believe that district staff and school principals need to be more active

in helping teachers understand the foundations of current reforms, and to work with teachers who are not supportive. Amy and Carina are both frustrated by the facilitation of meetings where negative comments were unchallenged. Neither teacher feels comfortable sharing her own views in these large-group settings for fear of provoking angry discussions that may not be handled well by the session leaders (in professional development) or school administration (in staff meetings).

Both teachers support the aims of current assessment reforms, yet have experienced significant tensions and dilemmas in the implementation. Their experiences in these sessions led to a feeling of guardedness in whom they shared their struggles with. Carina stated that she believes that the implementation of assessment reforms needs to be considered a political process and is quite intentional with her comments and aware of her audience when discussing her experiences with assessment. With people that she knows are supportive, she is willing to share her experiences of tensions and dilemmas, while with other who are resistant, she is very careful and aims to present a largely positive picture of assessment reform. This resistance from teaching colleagues complicates the already difficult task of re-negotiating classroom assessment practice. The dual nature of some colleagues helping to develop their understanding of classroom assessment, while other resisting these changes will be explored in more depth later in this chapter.

Amy and Carina generally found their school administrators supportive of current assessment reforms. Carina was pleased when her principal asked for a copy of her Assessment Checklist to share with a professional development group. Amy described her principal's leadership in drafting the school assessment policy, and in inviting a guest speaker to address the staff about reform-oriented views on grading. However, she also felt that her principal was not enforcing consequences for late and missed work as laid out in

the school's assessment policy, and was not proactive enough in the school at addressing teachers actively resisting the assessment reforms. One exception to this was Amy's experience where her principal increased a final course mark by five percent in response to a parent's appeal. In this case, she felt her principal was out of touch with the district's own messages from professional development sessions.

Both teachers also described their students as being generally supportive of their classroom assessment practices. Carina has been pleased that students in her classes have begun to internalize her use of formative quizzes and have even asked her for quizzes before major assessments in order to check their understandings. Amy has found students cooperative with her assessment practices, even when they recognize some aspects as different from other teachers'. In a previous year, students in a Grade 12 class asked about her use of a holistic assessment tool, but accepted her rationale. Amy believes they found it valuable since it was accompanied by individual conferencing and several students told her that they really understood where the mark came from. Both teachers indicated that they felt their students believed them to be fair in their approach to assessment.

Finally, parents appeared to have an indirect but significant influence on both Amy and Carina's assessment practices. While both teachers support a greater use of professional judgement in determining grades, they both expressed concerns over the defensibility of grades. One of Carina's reasons for developing and using her Assessment Checklist was so that there was documentation and evidence to support her judgments. Amy also used a combination of paper and electronic documentation, for this same reason. While Amy did have an experience where a final grade was overturned, neither teacher described ever having to justify marks directly to parents, but both appeared to feel that it

was an ever-present possibility that informed their record keeping and assessment decisions.

Internal Presses on Classroom Assessment Practice

In addition to external influences presented by institutional and stakeholder presses, there are internal factors that influence teachers' assessment practices. As previously mentioned, McMillan (2003) found that teachers' beliefs and values are the greatest influence on their assessment decision making, while other research points to significant roles for teachers' emotions and identity (Geijsel & Meijers, 2005; Hodgen & Askew, 2007). In this section, I discuss some of the internal presses that appear to influence Amy and Carina's assessment practices, including the teachers' beliefs and values, identity, and emotions.

Both Amy and Carina support current assessment reforms and believe that they are beneficial for students' learning. Both teachers indicated that their assessment practice is quite different from what they experienced as students, and believe their relatively recent emphasis on formative assessment is a big improvement on what they experienced. While Amy and Carina have both made similar changes in their classroom practice, their approach to change is quite different. Amy described herself as an early adopter of many aspects of assessment reform while Carina stated that she was quite cautious about changing aspects of her practice. This difference in Amy and Carina's approach to change may be a function of these teachers' years of experience: during this study, Amy was in her sixth year of teaching while Carina was in her twentieth. This is in line with several studies that have found that as teachers become more experienced, they become more cautious in their approach to change (Hargreaves, 2005; Huberman, 1993).

Amy's identity as an early adopter appeared to influence her assessment practice, and her experience of tensions within her practice. She volunteered for several different pilot projects at the district level for commercial grading software, which resulted in her using four different grading programs over the past few years. Through her involvement in professional activities, she often hears of new assessment tools and strategies that she frequently tries out shortly after learning of them. She views some of these as successful and incorporates them more widely in her practice, while other ideas are tried and discarded. In contrast, Carina characterized her own approach to change as very gradual. While she is open to innovation, she is quite confident in her own approach to instruction and assessment and is quite cautious about making any changes until she understands the benefit for students and the implications of these changes for other aspects of her teaching practice. Before adopting new assessment tools and strategies, Carina actively seeks out the opinions of other trusted colleagues to debate the possible benefits and costs of the change. She tends to wait until the start of the next semester before making any changes so that she incorporate them as part of a carefully considered plan for teaching, learning and assessment. She also consults extensively with colleagues at her school who are teaching the same courses. For example, when she began giving comments on quizzes instead of grades she did this in a shared course along with a colleague at her school that was already doing this. Her identity of being a highly competent teacher who carefully considers the implications of any changes appeared to be a significant influence on her adoption of new assessment practices.

Amy and Carina both described tensions between their support for current assessment reforms and their perception that many of these are resulting in a greater workload. This was often an emotional issue for both teachers. For example, both teachers

expressed support for the principle of allowing students multiple opportunities to demonstrate learning and re-do some assessments, but find this to be very time consuming. Both teachers also shared concerns about fairness and academic integrity in relation to allowing students to repeat assignments or complete them at a later date than their classmates. For example, Amy said that she was comfortable allowing students in her Grade 9 Applied course to re-do assignments and tests since she felt that students would not make the effort to get other students' completed assignments to copy. She is NOT currently allowing students in her Grade 10 Academic class to re-do assessments due to concerns over academic integrity. Although she supports the idea that students should have multiple opportunities to demonstrate their achievement of curriculum expectations, she feels that if these students were to re-write the same version of a test or task they would have an unfair advantage. She believes that students repeating an assessment in Academic courses need to complete a different but equivalent version of the test or task, and she stated that she is not currently able to provide this opportunity due to the substantial amount of time it takes her to write assessments. Carina does sometimes allow students in her Academic courses to have a second opportunity at a test, but she does re-write a new version of the test to avoid issues of academic integrity. She also stated that this new version is usually slightly harder than the first opportunity. The tensions between their support for current reforms and the increased workload they were experiencing was evidently an emotional issue for both teachers. As they described these sorts of issues, both Amy and Carina took long pauses in their speech and became more animated in their tone and body language.

Dilemmas in Classroom Assessment

In the past few sections, I discussed many of the tensions that Amy and Carina experience in their assessment practice and I have characterized a *tension* as “the emotional

awareness of a possible conflict” (Tillema and Kremer-Haydon, 2005, p. 204). In this section, I now discuss several specific assessment *dilemmas* that Amy and Carina experienced during the study. The distinction I am making here is that dilemmas are “more narrowly focused on actual situations of teaching” (p. 205) where the teachers had to respond to situations where they were confronted by difficult choices. These “lived experiences” (Woods, et al., 1997, p. 19) were centred around interpretations of assessment policies, issues of fairness and student responsibility.

Both Amy and Carina encountered assessment dilemmas that required them to make difficult decisions involving their interpretation of assessment policies and their sense of fairness. Amy had several experiences where students handed in work past deadlines (despite several opportunities), and she had to make a decision on whether she should accept the work, or refuse to mark it. These were assessment dilemmas for Amy because she felt these students had the advantage of extra time that their classmates did not receive, but also felt there would be significant resistance to a refusal to accept the assignments from students, parents and administration. In addition, since the assignments were handed in during the exam period (a very busy time for teachers), she found this to be an unfair burden to her own workload. While she believed that her own school’s assessment policy allowed her to give a zero for work that was submitted past a well-established deadline, Amy did not think her principal would support this, and felt resigned to accepting and marking the late work. She expressed her frustration at this situation: “The extensions thing – the whole deadline thing is a huge pet peeve” (Amy, Interview, June 18, 2009).

In a different issue also related to the interpretation of assessment policies and fairness, Carina encountered a situation where several students in her Grade 10 class admitted to sharing or receiving information about a district-wide performance task. This

was an assessment dilemma for Carina because the task was supposed to have been done independently, but several students admitted to consulting with their friends from a class earlier in the day, or from other schools. Her response indicated that she felt there were structural problems with the assignment, rather than being angry with the students. She marked the assignments as they were written and did not ask students to repeat the assignment. She did, however, bring the issue up with other mathematics teachers at her school, at a subsequent professional development meeting for teachers of that course, and with a district mathematics consultant.

Both teachers also experienced assessment dilemmas where they felt their support for allowing students to have multiple opportunities to demonstrate their learning came into conflict with their belief that students need to be responsible for their own learning. Again, when several of Amy's students handed in late assignments she felt that she had to accept their work since they had demonstrated their learning, but she believed that the school's assessment policy allowed these students to get away with procrastinating on the assignment. Carina experienced a similar feeling with the work habits and study skills of students in her Grade 12 class. During the final interview, which took place in June, Carina was reflecting on changes that she felt she might make in that course for the following year. During the past three semesters that she taught that course, she used most of the class time for instructional tasks, which students engaged in quite willingly during class time, and assigned daily follow up for students to do at home. She felt very frustrated that students were not doing this homework regularly, particularly because they were Grade 12 students in a course that is a prerequisite for mathematics at the university level. She stated several times that she felt that students needed to take more responsibility for their learning and success. She felt this was a dilemma related to classroom assessment because so few

students were able to complete the formative quizzes that she intended as checks for students' understanding during the intended class period. Many of the students already recognized they could not solve the problems she presented, so the quizzes became a source of motivation for students to ask questions and finish some homework. Carina's intention was to use these quizzes to help give more detailed feedback. She felt that she needed to rethink her approach, but felt that in a Grade 12 class that it was quite reasonable to ask students to spend 30 minutes outside of class each day completing class assignments. She felt that she would be compromising on some of the content if she assigned less work.

Summary of Tensions and Dilemmas in Classroom Assessment

It is evident that Amy and Carina balance many different influences or “presses” on their assessment practice. Conflict between some of these various influences results in tensions and dilemmas in classroom practice. Some of these experiences are more an “awareness” of the possibility of conflict (termed *tensions* here), while others result in actual lived situations where difficult decisions are required (termed *dilemmas* here).

Through the data analysis for this study, McMillan's (2003) framework for assessment decision making, along with work by Saxe et al. (1999), was helpful in pointing to external and internal influences on teachers' assessment practices. External influences included institutional “presses” such as provincial curriculum documents, district and school assessment policies, professional development and curriculum supports such as textbooks and provincially funded resources); as well as key stakeholder “presses” from colleagues, administrators, students and parents. Internal influences included the teachers' beliefs and values, identity, and emotions. McMillan's (2004) work on the four components of assessment was also helpful in identifying that many of the tensions and dilemmas are related to the interpretation and use of assessment data, rather than on the purpose for

assessments and means for eliciting assessment data. These tensions and dilemmas included the use of assessment data for reporting purposes; issues of fairness, reliability and consistency between teachers; interpretation and implementation of policies from different levels; use of formative assessment data by both teachers and students to improve learning; and the defensibility of their students' grades (to parents and administration). There were also significant tensions and dilemmas related to the implementation of current assessment reforms given resistance from colleagues, and with the perception that some aspects of current assessment reforms are resulting in an increased workload. While these do not fit directly into McMillan's four components of assessment, these concerns are clearly related to assessment decision making, since the decisions that teachers need to make about their assessment practice cannot occur in isolation. Teachers' assessment decisions are clearly influenced by the context in which they work.

Addressing Research Question #2: How are these Tensions and Dilemmas Negotiated in Classroom Practice?

In this section, the second research question, "How are these tensions and dilemmas negotiated in classroom practice?" is addressed and discussed in relation to the two case studies. Amy and Carina have very similar ways of negotiating these tensions and dilemmas. These include participation in their local mathematics communities, confidence in their approach to classroom assessment along with the belief that current assessment reforms are valuable for students' learning, use of a variety of resources, a differentiated approach based on the grade level, compliance with official policies and directives and reflection on their own practice.

For both Amy and Carina, the primary way in which they negotiate many of the tensions and dilemmas they experience is through participation in their various

mathematics communities. Amy indicated that her involvement on committees for district projects allows her to build relationships with other teachers who were keenly interested in developing their understanding of current assessment reforms. This has helped her develop a network of trusted colleagues with whom she can discuss issues and problems in her practice, and she is confident speaking directly to district staff to seek clarification when needed. This participation also means that she is among the first to grapple with new initiatives and messages being promoted by the district, so she has time to try them and discuss the implications with other teachers. She also finds time for informal discussions at workshops to be very valuable in learning about new assessment tools and strategies that other teachers are trying in their classrooms and what their experiences are with assessment reforms. In particular, she has valued conversations around different methods of recording marks and determining grades. This participation helps her to “find out what the latest buzz is and [laughs] try something else” (Amy, Interview, June 18, 2009). In addition, she stated that discussions with teachers at her own school, especially her department head, were important to her developing understanding of current assessment reforms. Similarly, Carina described discussions and collaboration with colleagues in her own school and beyond as being essential aspects of her adoption of current assessment reforms. She has cultivated a small group of trusted colleagues from several schools and she actively seeks out collaborative efforts in writing tests, performance tasks and exams. She also enjoys working with teachers she shares courses with at her own school. Carina also participates in district professional development initiatives and, like Amy, feels comfortable advocating her perspectives and concerns directly to district consultants. During this study, she specifically mentioned discussing issues of academic integrity of performance tasks with district staff. Both Amy and Carina have found that their participation has resulted in a

close circle of colleagues with whom they share ideas to make sense of different perspectives on contentious aspects of assessment reforms.

Another important way that Amy and Carina negotiate their tensions and dilemmas is through their confidence in both their own approach to classroom assessment, as well as their belief in the value of current assessment reforms. Both Amy and Carina are comfortable making changes in their practice in response to ideas from colleagues, or after reflection on what has worked or needs improvement for future semesters. Even though Amy had a mark overturned by a Principal in a previous semester, she felt confident that the holistic assessment tool she was using was aligned with provincial and district assessment policy. She felt that her Principal during that semester did not really understand how the tool was being used and stated that she would use it again in the future. Carina described feeling confident in her understanding of current assessment reforms and believes that the “new” approach to assessment has changed her relationship with her students for the better. She also believes that her students believe she is a fair teacher who gives them every opportunity to succeed. Her approach to assessment, including the use of her Assessment Checklist, has been validated by colleagues at her own school, including her Principal who asked for a copy of the Checklist to share at a summer course he was teaching. Over the past few semesters, Carina has shared some of her tasks beyond her school. This confidence appears to sustain their efforts to implement current assessment reforms, even in the face of tensions and dilemmas they experience.

A third support for Amy and Carina is their knowledge of and use of a variety of resources in their classroom instruction and assessment practice. These resources, many of which were developed by the provincial Ministry of Education or shared by colleagues, have helped to provide Amy and Carina with exemplars of rich tasks and sample

assessments. These artefacts are helpful as anchors for discussions with other teachers of what current assessment reforms look like in practice. For example, during the lessons observed during this study, Amy used several provincially funded curriculum supports, including the TIPS4RM¹⁴ program for Grade 9 Applied math and some of the publicly available sample items for the provincial assessment (EQAO). She described the TIPS4RM resource as particularly helpful since it contains quality activities that help her establish a “completely open, low risk environment” (Amy, Interview, May 28, 2009) in her classroom where students are comfortable contributing to discussions, collaborating in small groups, and sharing their thinking, all of which she views as essential to classroom assessment. During the lessons observed during this study, Carina used activities, tasks and assessments she created herself. She also described using performance tasks produced as part of a district initiative, resources shared electronically by colleagues in the district, sample performance tasks from many different textbooks, and internet resources such as the NCTM Illuminations website. Both teachers also credit technology they have available in their classrooms, including interactive whiteboards and graphing calculators, to be helpful to making their lessons more interactive. This variety of resources has helped Amy and Carina to develop a sense that there is coherence between the provincial assessment policy, district initiatives, and messages from a broader mathematics education community beyond Ontario. Both teachers describe this as being critical in their determination to pursue further learning, to support district initiatives, and to maintain their efforts to implement current assessment reforms.

¹⁴ The TIPS4RM resource is a publicly available complete course guide including full lesson plans for each day along with blackline masters. It was funded by the Ontario Ministry of Education and written by teachers (not a commercially produced program)

Another way that Amy and Carina negotiate tensions and dilemmas in their classroom assessment practice is to use different approaches based on the grade level (Grades 9-12), and course (e.g., Applied or Academic). For example, Amy felt more comfortable with second opportunities for students in her Grade 9 Applied math course than in her Grade 10 Academic course where she expressed concerns over academic integrity. She also indicated that she was more comfortable using levels on assessments in her Applied course compared to her Academic courses where she still used traditional marking schemes on her tests. She stated

I think about them [assessments in Applied versus Academic courses] very differently as a teacher, which I shouldn't. But I really do focus on that [use of levels on assessments] more with the Applied classes. I feel like I'm further ahead with the Applied classes than I am with the Academics (Interview, June 18, 2009)

Carina felt her Grade 12 students should demonstrate more independence and responsibility than Grade 10 students so she did more follow up with Grade 10 students with the Assessment Checklist. Using different approaches depending on the grade and course level has helped both teachers make sense of some new assessment tools and strategies, and to see how and if some of these could be useful in other courses.

A fifth support to both Amy and Carina was to look to official policies and directives, as well as statements from their principals, department heads and district staff, to help them negotiate tensions and dilemmas in their assessment practice. Sometimes, these are not satisfactory, such as Carina's discomfort with using her district's "pegging" system for converting rubric levels to percentage grades, but both teachers comply with required policies.

Finally, both teachers stated that it was important to be reflective about their own current practices and to stay open to ideas from other colleagues or professional

development. Amy was particularly reflective about the effectiveness of her use of assessment to improve students' learning. She stated that her own dissatisfaction with some elements of her practice motivates her to continue developing her understanding and participating in more learning opportunities. During the final interview (in June) in particular, Carina was reflective about changes she could make to give students more time to work in class, particularly in the Grade 12 course. She also plans to keep using her Assessment Checklist, but plans to focus more on using this tool to foster self-assessment next semester. This reflective approach helps Amy and Carina negotiate tensions and dilemmas since they are constantly trying to improve their assessment practice.

Summary of Negotiating Tensions and Dilemmas in Practice

As Enyedy et al. (2006) suggest, some of these tensions and dilemmas are managed rather than solved. For both teachers, the greatest source for solutions to tensions and dilemmas is conversations with other teacher who also support current assessment reforms and are working on implementing new assessment practices. Amy and Carina both enjoy learning about new ways to improve their assessment practice. This has included shared strategies for formative assessment, new methods for recording and analyzing assessment data, and discussions about using performance tasks. When Amy and Carina cannot find satisfactory ways to solve their tensions and dilemmas, they find ways to manage them. Often, this is by complying with official policies and directives, or with common practice within their mathematics department.

Key Findings

In this section, I put forth three assertions based on cross-case analysis. First, determining grades for reporting purposes is a major source of dilemmas. Next, collegial relationships can both help and hinder the implementation of assessment reforms. Finally,

that Amy and Carina appear caught in a paradigm shift. For each assertion, I make connections between the findings of this study and the literature.

Assertion #1: Determining Grades for Reporting Purposes is a Major Source of Tensions and Dilemmas

Many of the tensions and dilemmas that both Amy and Carina described centred on the task of determining grades for reporting purposes, rather than other aspects of classroom assessment. These two teachers do not appear to be struggling as much with new assessment *forms* as they are with integrating various pieces of assessment evidence to construct a final grade. In principle, Amy and Carina both support notions of reform assessment related to grading such as giving students multiple opportunities to demonstrate course expectations, using a variety of strategies and tools, and using the most consistent, most recent evidence. However, they are not sure how to integrate these with their own understandings of reliability, fairness and defensibility, especially given that the dominant practice in their schools is to average the results of all major assessments. Both teachers expressed discomfort with their current methods for constructing grades using (different) commercial grading software. While they are unsatisfied with their current methods for determining grades for reporting, neither Amy nor Carina believes she has seen examples of any suitable alternatives.

Amy and Carina's experiences and concerns with grading and reporting are reflected in the literature. Over twenty years ago, Stiggins, Frisbie & Griswold (1989) characterized high school teachers' grading practices as a "complex array of myth, tradition, uncertainty, and procedures" (p. 14). Even recently, O'Connor (2007a) contends that grading is still a largely "unexamined and private practice" (p. 117) that is "clothed in myth, mystery and magic" (p. xiii). This is troubling since high school grades are linked to

many aspects of student's learning, success and future including motivation and applications to post-secondary institutions. Grading is an emotional issue for students, teachers and parents, and grades are one of the most common public uses of educational assessment (Brookhart, 1994). It has been argued that there is a gap between teachers' grading practices and measurement theory (Guskey, 2002; McMillan, 2001, 2003; Stiggins et al.). Stiggins et al. suggest three reasons for this: best practices for determining grading may be a matter of opinion, recommended practices do not take some of the practical aspects of teaching into account, and/or teachers lack training or expertise in sound practices.

While the calls to reform grading practices are relatively consistent (e.g., NCTM, 1995, 2000; O'Connor, 2002, 2007a, 2007b; Wormeli, 2006), Guskey (2007) points out that there is no consensus "about what specific measures should be used, the most appropriate methods for synthesizing that evidence, or the most effective procedures for communicating that information" (p. 25). He argues that teachers "have to face the difficult challenge of deciding what evidence or combination of evidence represents the truest and most appropriate summary of students' achievement and performance" (2002, p. 778). Stiggins, Frisbie & Griswold (1989) use words like "uncertainty", "frustration" and "ambiguity" to describe teachers' experiences with grading processes, which reflect Amy and Carina's experiences over two decades later.

Much of the reform-oriented literature on grading argues that teachers' professional judgement is a key determining factor in assigning grades since students should have multiple opportunities to demonstrate their learning and that not all of these trials should be included in students grades (NCTM, 1995, 2000; O'Connor, 2007a). There is a strong case made in this literature that averaging is no longer considered a satisfactory method of

determining grades because past assessment results may no longer accurately reflect their learning (Guskey, 2002; O'Connor, 2002, 2007a). These are all tenets that Amy and Carina agree with, but they are uncertain about how to determine grades in a fair, reliable, transparent and publicly understood way without the use of averaging. Furthermore, their school district requires teachers to calculate grades using published mark weighting guidelines, and has licensed commercial grading software for teachers to use. Shepard (2001) argues that “good assessment practice should include a combination of both locally negotiated scoring routines and clinical, or interpretivist, approaches to data synthesis” (p.1089), but recognizes that teachers need further training in qualitative methods. In particular, she suggests that teachers need to be adept at ensuring the dependability and confirmability of their interpretations and conclusions, through such methods as triangulation, purposeful sampling and audit trails. These are areas that both Amy and Carina indicated they would like opportunities for further professional development.

Even though the grading and reporting practices of teachers in Ontario must be informed by policies from different levels, Amy and Carina still feel a sense of dissatisfaction and uncertainty in this important professional duty. In Ontario, the provincial guidelines on grading are currently very general and state only three main points. First, students' grades must be based on the Achievement Chart and reported as percentages. Second, that 70% of a students' mark come from work during the term taking in consideration the “most consistent level of achievement throughout the course, although special consideration should be given to more recent evidence of achievement” (OME, 2005, p. 22). Finally, the remaining 30% of the grade must come from a final evaluation “administered towards the end of the course” (p. 22). Assessment policies at the district and school levels are more specific and include directives indicating that teachers should avoid

using zeros, how teachers should handle late and missed assignments as well as plagiarized work, and instructions on weighting assessment evidence according to the categories of the Achievement Chart. While these policies represent a development from two decades ago where such policies were rare (Brookhart, 1994; Stiggins et al., 1989), Amy and Carina both expressed uncertainty in how to interpret and integrate these policies and neither has experienced significant professional development in how these policies should inform the determination of grades.

Assertion #2: Teacher Community is a “Double-Edged Sword” in the Implementation of Assessment Reforms

Both Amy and Carina found some colleagues to be a great asset to their developing understanding of assessment reform while other colleagues resisted these reforms and contributed to Amy’s and Carina’s feelings of tensions and dilemmas. In this section, I describe both the positive and negative impact of teacher community as evidenced in this study, and how this connects to the literature.

For the teachers in this study, the primary way that they both develop their understanding of assessment reform, as well as negotiate tensions and dilemmas, is through conversation and collaboration with colleagues. Both Amy and Carina described the most fruitful relationships as ones where there was mutual trust so they could be open and honest about elements of their practice they were dissatisfied with. This allowed them a venue to share their struggles and challenges in a low-risk, supportive environment. Most of the valuable interactions that the teachers described were opportunities they had initiated themselves, away from formal structures such as staff meetings and professional development sessions. These included private conversations in department offices with teachers they shared courses with, collaborative efforts they pursued on their own time with

teachers beyond their own schools, and informal conversations during breaks at formal professional development sessions. For example, Amy described two contrasting experiences in collaboratively marking summative performance tasks for her Grade 10 course: she found it very beneficial to work with two colleagues in her department office compared to a similar activity done during a workshop sponsored by the school district which she found to be a fairly negative experience. In particular, both teachers described sharing of assessment tools and strategies as helpful, designed either by themselves or other teachers, where there was an element of critical conversation about what this looks like in a classroom, stumbling blocks they experienced, and ways that they could be improved in the future. Sharing challenges helped them to feel validation in the efforts since they knew others were struggling with similar issues. For example, Carina described feeling validated in her efforts to design quality performance tasks since she knows others are struggling to find or create tasks that are open-ended but at an appropriate level of difficulty for their students. Both teachers felt confident, through their involvement with activities within their district mathematics community, to initiate conversations with district consultants and share their concerns with the hope that this can inform future professional development sessions.

These experiences are reflected in the literature on teachers' professional development. Amy and Carina appear to be members of several professional learning communities. These include both formal and informal groups, both inside and outside their schools. These connections reflect the important social dimension of teachers' learning which is widely identified in the literature (Harris, 2003; Spillane, 1999). Relationships that Amy and Carina described as being particularly helpful to their thinking included discussions that probed below surface features of teaching to discuss the implications of reform on their teaching. As Little & Horn (2007) note, "talk in teacher communities is

likely to be generative of profession learning and instructional improvement to the extent that it invites disclosure of and reflection on problems of practice” (p. 91). Amy and Carina appear to have been supported in their efforts towards reform by their participation in these sorts of communities.

While acknowledging that their colleagues were essential resources for negotiating their understanding of assessment reforms, both Amy and Carina described feeling a negative influence from colleagues resisting reforms in classroom assessment. As previously described, both teachers have experienced situations where meetings or professional development sessions focused on assessment took on ugly tones. Both teachers expressed discomfort with the political nature of current assessment discourse and are careful in their approach with colleagues who are resistant to contemporary reforms. When encountering colleagues resisting assessment reforms at staff meetings or professional development sessions, Carina does not speak out publicly, but rather engages in discussion of these ideas and view later with other colleagues who support the “new” assessment and dissects the opposing arguments (to solidify her own sense of assessment).

While there is a great deal of support for professional learning communities in the literature on educational change, there is also evidence of the “dark side” of community (Noddings, 1996, p. 245). As McLaughlin and Talbert (2001) and Rousseau (2004) point out, strong communities not always associated with innovative practice. Gellert (2008) explains:

an over-confirmed framework of collective orientations can prevent professional development. Mutual validation can turn experience into law. In that sense, mathematics teachers do not inevitably learn from experience. The contrary may hold: Collective teaching experience can be blind to alternative conceptions of teaching (p. 105)

All reform efforts in education imply a “loss of competence from that teacher’s perspective” (Carlgren, 1999, p. 45) and the implementation of these reforms requires teachers to renegotiate their practices and routines which results in discomfort (Gellert, 2008). Research into teacher community has repeatedly shown that problems expressed by teachers met with normalizing responses to reassure and express solidarity that result in turning the discussion away from the teaching, rather than towards the issue as an object of collective attention (Little & Horn, 2007). In this study, Amy and Carina both experienced this “dark side” of community. Both teachers felt that many of their colleagues would prefer to side-step difficult conversations about changing assessment practice, and believed that some of their teacher colleagues felt that their competence was being questioned by current assessment reforms. This was particularly true in Amy’s experience where some teachers planned to not receive a guest speaker well, and had made up their minds even before he spoke.

Amy and Carina’s experiences with the “double-edged sword” (Gellert, 2008, p. 98) of teacher community is reflected in some of the literature on teacher professional communities. As Gellert points out “On the one hand, routines and collective orientation might serve as backing and support; on the other hand, both might result in conservative forces impeding or even foiling the necessary change process” (p. 98). While it has been suggested that subject departments in secondary schools can serve as an important locus for teachers’ work (Little, 2002), it is rare to find professional practices that allow deep and sustained collaboration to support teachers’ on-going learning (Little & Horn, 2007; McLaughlin and Talbert, 2007). Harris (2003) warns “unless careful attention is paid to *how* collaborative processes are fostered and developed in schools and classrooms the

potential for well intentioned initiatives to impact upon teaching and learning will remain severely restricted” (p. 380, emphasis in original).

Assertion #3: These Teachers are Caught in a Paradigm Shift

As stated earlier, in the review of literature on classroom assessment (Chapter 2), Shepard (2000, 2001) has suggested that some of the difficulties that are currently being experienced by teachers arise due to the concurrent use of instructional practices that are based in the emergent paradigm and assessment practices based in the old paradigm. In her discussion of the emergent paradigm she suggests that “it is not fully developed theoretically and, surely, not adopted in practice” (p. 1073). In the intervening decade, there has been a great of literature written about reform-oriented instructional and assessment practices, and in Ontario, the province-wide curriculum for all subject areas, including mathematics, was revised to be consistent with these principles.

As evidenced in this study, Amy and Carina have adopted many aspects of current instructional and assessment reforms in their classroom practice, yet are still experiencing significant tensions and dilemmas in their practice. These experiences go beyond what Shepard (2000, 2001) identified as problems from using practices from two different paradigms. Some of these issues arise due to the “uncertain” and “underdetermined” (Ball, 1996, p. 506) nature of the emergent paradigm, but many arise because teachers do not work in isolation and many of the stakeholders in education still value the traditional paradigm, particularly for classroom assessment. This creates difficulties for individual teachers who have views about teaching, learning and assessment that align with the emergent paradigm.

There were several examples in this study where the two participants described encounters with colleagues who were still operating within the traditional paradigm. In

particular, the explanation Amy's principal gave for overturning a final grade – “it's not really a solid mark. We don't really have numbers here” (Amy, Interview, May 21, 2009) – demonstrated the difficulties teachers have in adopting new assessment practices in isolation. Amy and Carina both described several experiences in staff meetings and professional development sessions where they were frustrated by the resistance of some of their teaching colleagues to current reforms. Both teachers interpreted messages from policy (at provincial, district and school levels), professional development, and district staff to be supportive of assessment reforms and the changes they were making in their practice, but found that some of their colleagues either had different interpretations or disagreed with the reforms.

Many of the dilemmas that Amy and Carina describe point to issues where there is institutional and stakeholder support for the traditional paradigm. While both of these teachers have adopted many aspects of the emergent paradigm, they are still working in a system that has not yet fully embraced reform, despite messages in curriculum documents, policies and professional development. School principals in particular, appear to have a pivotal role to play in encouraging wider adoption of current reforms since they are often the intermediary between the implications of this new vision for instruction and assessment (e.g., new policies) and the public (e.g., explaining changes in grading practices to parents). Amy and Carina also felt that their school principals needed to be more direct in addressing resistance from teachers at their schools.

Summary

In addition to descriptions of the tensions and dilemmas that Amy and Carina experience in their assessment practice, and the ways they negotiate these tensions and dilemmas, I have put forth three major findings from this study. First, that grading is a

major source of tensions and dilemmas in classroom assessment. Second, that professional communities can be both a source of support and resistance to current reforms. Finally, I suggest that Amy and Carina appear caught in a paradigm shift between traditional and reform-oriented practices.

CHAPTER 7: CONCLUSIONS

There have been calls for two decades now for reforms in curriculum, teaching and assessment to improve learning (e.g., Black & Wiliam, 1998a; Gipps, 1994; NCTM, 1989, 1991, 1995, 2000; Shepard, 2000), yet these changes, in particular in classroom assessment, have been slow to emerge (Delandshere & Jones, 2001; Lock & Munby, 2000; Ohlsen, 2007). In Ontario, there have been province-wide curriculum and assessment policies aligned with current assessment reforms for a decade, yet teachers are still experiencing difficulties fully embracing and implementing reform-oriented assessment practices (OME, 2008; Suurtamm & Graves, 2007).

This research involved two case studies of secondary mathematics teachers in Ontario implementing reform-oriented assessment, and described the tensions and dilemmas they experience as well as how they negotiate these tensions and dilemmas in practice. Amy and Carina described significant tensions and dilemmas in several areas including the recording and analysis of assessment data for the purpose of reporting, resistance from some colleagues to the adoption of current assessment reforms, difficulty interpreting and integrating policies from different levels (school, district and province), and an increased workload caused by provincial and district initiatives intended to support the implementation of these reforms (such as Credit Rescue). The ways that Amy and Carina negotiated these tensions and dilemmas was remarkably similar and included participation in their local mathematics communities, confidence in their approach to classroom assessment, belief that current assessment reforms are valuable for students' learning, use of a variety of resources, a differentiated approach based on the grade level, compliance with official policies and directives, and reflection on their own practice.

Three key assertions were put forth based on cross-case analysis. First, determining grades for reporting purposes is a major source of tensions and dilemmas. Both Amy and Carina described feeling comfortable with different forms of assessment and considering evidence from multiple and varied sources, but both teachers expressed uncertainty over how to integrate this information into a grade for reporting purposes. Second, teacher community plays a complex role in the implementation of assessment reforms. While Amy and Carina both characterized having a network of trusted and respected colleagues as a crucial element in developing (and continuing to develop) their understanding of current assessment reforms, they both also shared experiences where some of their teacher colleagues actively resisted assessment reforms. This resulted in some uncomfortable situations and frustrations that they were not sure how to manage. Finally, even though both these teachers appear to be using reform-oriented instructional and assessment practices based in what Shepard (2000, 2001) calls the *emergent paradigm* they are still caught in a paradigm shift.

Contributions and Implications of this Study

This study contributes to research in the fields of classroom assessment and professional development in several ways. First, it provides insight into the experiences of two secondary mathematics teachers' as they implement classroom assessment practices that are aligned with current reforms. Descriptions of the tensions and dilemmas teachers experience can be helpful for understanding the complexities of "doing" classroom assessment in a manner that is aligned with the current reforms. These teachers are not struggling as much with new assessment forms as they are integrating various pieces of assessment evidence for reporting purposes. Teachers need opportunities to consider examples of how to do this, and could benefit from professional development involving

case studies of analysis and reporting of assessment data. Examination of tensions and dilemmas can be a productive way to “press deeply into the work of reform” (Ball, 1996, p. 505) for both pre-service and in-service teachers. Literature on teachers’ learning suggests that the emphasis needs to shift from dissemination and implementation of programs to posing questions and critically interrogating one’s practice and the practice of others (Cochran-Smith & Lytle, 1999; Talanquer et al., 2007; Windschitl, 2002). As Ball argues, “Given the uncertainties and underdeterminedness of the reform visions, local interpretation and invention are both inevitable and desirable” (p. 506). Identifying and studying particular dilemmas of practice may lead to strategies to cope with these dilemmas (Tilemma & Kremer-Haydon, 2005). Case studies of teaching dilemmas have been used in mathematics education for both pre-service and in-service teachers (Barnett, 1998) and can provide portraits of practice to stimulate important discussions. As Windschitl (2002) argues, teachers “could benefit greatly from the case histories of other teachers who have attempted pedagogical reforms” (p. 162).

In addition, this study supports other literature that calls attention to the “dark” side of community in teaching (Noddings, 1996; Westheimer, 1998). Calls for the encouragement and development of professional learning communities often do not address the possibility that collective orientations can reinforce traditional practices and stand in the way of innovation (Gellert, 2008; McLaughlin & Talbot, 2001). While it is certainly difficult to mitigate this within informal communities, it might be beneficial to offer outside perspectives to teachers during more formal professional development opportunities (such as district staff, researchers or teachers from other schools or districts that support the reforms). Common dilemmas of reform-oriented teaching and assessment could be

presented and discussed, modelling and encouraging inquiry-oriented approaches to examining practice.

To support teachers using reform-oriented assessment practices, there needs to be coherence in policies on grading and reporting from different levels. This study is relevant to policymakers who may also be interested in using dilemmas to identify areas where there may be institutional or organizational barriers to reform initiatives. In addition, there needs to be recognition that school administrators and district staff, along with classroom teachers, need opportunities to develop their understanding of the rationale behind policy changes so they can communicate effectively and persuasively with teachers, students and parents about the value of these reforms. As Spillane et al. (2002) point out, making sense of these changes is a very complex process and there is often a tendency to focus on superficial similarities between new messages and old practices rather than critically important differences. It may help ease some of the tensions teachers experience in the analysis and reporting of assessment data if efforts were made, possibly at the school or district levels, to communicate to parents and other stakeholders about how this process is different in the emergent paradigm compared to the traditional paradigm that many experienced as students themselves.

Limitations of this Study

This study looked at the experiences of two confident, experienced, highly qualified teachers who were deliberately trying to improve their classroom assessment practice in ways consistent with current reforms. The experiences discussed here may not reflect those of teachers who are less confident. Amy and Carina volunteered to participate in this study and were willing to discuss tensions and dilemmas in their classroom assessment practice, which is a very personal aspect of their teaching, and other teachers who are less confident

may have different experiences of tensions and dilemmas in their assessment practice. Amy and Carina are both experienced teachers, with six and 20 years of teaching experience respectively. Their experiences may not reflect those of less experienced teachers, and the implications may be more useful for in-service professional development than pre-service teacher education. Both Amy and Carina are highly qualified teachers, both holding specialist qualifications in mathematics. There was no evidence during the classroom observations nor the interviews that either teacher struggled with the mathematical content they were teaching, so these teachers' experiences of tensions and dilemmas may be different than teachers who are not as highly qualified. Both teachers were making deliberate efforts to improve their classroom assessment practice so their experiences are likely different from teachers who are either satisfied with their current practice, or teachers who are actively resisting current reforms. Finally, Amy and Carina teach in a jurisdiction where there has been a province-wide reform-oriented curriculum and assessment policy for over a decade. The school district in which they both work in has substantial resources directed at professional development. The findings in this study may be different in jurisdictions with different organizational and policy environments.

Directions for Future Research

As I conclude this research, several questions and directions for new research are emerging. The data for this study involved interviews and lesson observations. As Carlgren (1999) notes, an increasing amount of teachers' work is done *outside* of the classroom and this is where a great deal of assessment decision making occurs. In future studies it would be interesting to include observation of teachers engaged in their work outside of the classroom, in activities such as planning for assessment (individually and with colleagues), giving written feedback to students on work, scoring tests, analyzing assessment data and

determining grades for reporting. Also, it would be helpful to videotape interviews and classroom observations for a more complete record, although careful consideration should be made in how this could affect possible participation since some teachers may be uncomfortable with video recordings of their teaching. Video recordings of lessons would allow for another layer of exploration of tensions and dilemmas as suggested by Enyedy et al. (2006). In their work, they recorded lessons and later showed clips individually to the teachers, which prompted rich discussions of specific dilemmas of practice.

As I reflect on this research, several new questions have emerged. There appears to be a significant gap in the literature regarding the analysis and reporting of assessment data. Several studies have used predominantly quantitative methods to examine teachers' grading practices (e.g., Duncan & Noonan, 2007; McMillan, 2001, 2003; Ohlsen, 2007), but there are few studies that examine teachers' experiences constructing grades for reporting purposes. I am interested in how teachers decide what assessment evidence to include for reporting, how analysis may differ between classes or even between students in the same class, teachers' notions of validity and reliability and how these affect choices they make in analysis and reporting, how teachers use commercial grading software as a tool for analysis, and how teachers work together and negotiate these issues while teaching the same course (e.g., a Grade 9 team within a school or district). There is also a need to further explore the characteristics of teacher community that encourage innovation and help sustain reform efforts – particularly for new teachers who often revert quickly to traditional practices. Finally, I am interested in further exploring the use of tensions and dilemmas in professional development as a way to examine and focus teachers' practice. As Ball (1993) argues, "How teachers learn to frame and manage the dilemmas of 'intellectually honest'

practice in ways that do indeed benefit all students is crucial to the promise of such work” (p. 395).

Concluding Thoughts

As I conclude this study, I reflect back to Clark and Lampert’s (1986) quote that I cited at the start: “the teacher is in the difficult position of doing a cognitively and conceptually complex job while hiding the complexity and uncertainty – making it look easy to teach” (p. 28). I believe that future work, by both researchers and teachers, needs to focus on exploring the complexities and uncertainties of teaching, and particularly in teaching and assessing in ways aligned with current reforms. It is only by bringing these out into the open that we can work towards improvements in teaching and learning, and understand how to better support teachers as they make changes in their practices. As evidenced by the findings in this study, the examination of tensions and dilemmas in classroom assessment is one way to approach this, but there is a great deal more work that needs to be done if we wish to realize the vision of current reforms.

REFERENCES

- Adler, J. (1995). Dilemmas and a paradox – secondary mathematics teachers' knowledge of their teaching in multilingual classrooms. *Teaching and Teacher Education*, 11(3), 263-274.
- Adler, J. (1997). A participatory-inquiry approach and the mediation of mathematical knowledge in a multilingual classroom. *Educational Studies in Mathematics*, 33, 235-258.
- Adler, J. (1999). The dilemma of transparency: Seeing and seeing through talk in the mathematics classroom. *Journal for Research in Mathematics Education*, 30(1), 47-64.
- Ball, D. L. (1990). Reflections and deflections of policy: The case of Carol Turner. *Educational Evaluation and Policy Analysis*, 12(3), 247-259.
- Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal*, 93(4), 373-397.
- Ball, D. L. (1996). Teacher learning and the mathematics reforms: What we think we know and what we need to learn. *Phi Delta Kappan*, 77(7), 500-508.
- Barnes, M., Clarke, D., & Stephens, M. (2000). Assessment: the engine of systemic curricular reform?. *Journal of Curriculum Studies*, 32(5), 623-650.
- Barnett, C. (1998). Mathematics teaching cases as a catalyst for informed strategic inquiry. *Teaching and Teacher Education*, 14(1), 81-93.
- Beijaard, D., Meijer, P. C., & Verloop, N. (2004). Reconsidering research on teachers' professional identity. *Teaching and Teacher Education*, 20(2), 107-128.
- Berlak, A., & Berlak, H. (1981). *Dilemmas of schooling: Teaching and social change*. London: Methuen, 1981.

- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Buckingham, UK: Open University Press.
- Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education*, 40(2), 119-156.
- Briscoe, C. (1993). Using cognitive referents in making sense of teaching: A chemistry teacher's struggle to change assessment practices. *Journal of Research in Science Teaching*, 30(8), 971-87.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and practice*, 22(4), 5-12.
- Brookhart, S. M. (2004). Classroom assessment: tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429-458.
- Brown, G. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science Publishers.
- Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Carlgrén, I. (1999). Professionalism and teachers as designers. *Journal of Curriculum Studies*, 31(1), 43-56.

- Clark, C., & Lampert, M. (1986). The study of teacher thinking: Implications for teacher education. *Journal of Teacher Education*, 37(5), 27-31.
- Cochran-Smith, M., & Lytle, S. L. (1999). Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education*, 24, 249-305.
- Cohen, D. (1990). A revolution in one classroom. The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12(3), 327-345.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Crockett, M. D. (2002). Inquiry as professional development: creating dilemmas through teachers' work. *Teaching and Teacher Education*, 18, 609-624.
- Davis, B., & Sumara, D. (2007). Complexity science and education: Reconceptualizing the teacher's role in learning. *Interchange*, 38(1), 53-67.
- Day, C., & Leitch, R. (2001). Teachers' and teacher educators' lives: the role of emotion. *Teaching and Teacher Education*, 17, 403-415.
- Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 8(2), 113-133.
- Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record*, 104(7), 1461-1484.
- Delandshere, G., & Jones, J. H. (1999). Elementary teachers' beliefs about assessment in mathematics: a case of assessment paralysis. *Journal of Curriculum and Supervision*, 14(3), 216-240.
- Delandshere, G., & Petrosky, A. R. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher*, 27(2), 14-24.

- Denzin, N. K., & Lincoln, Y. S. (1998). *Collecting and interpreting qualitative materials*. Thousand Oaks, CA: Sage.
- Dilley, P. (2000). Conducting successful interviews: Tips for intrepid research. *Theory into Practice*, 39(3), 131-137.
- Drake, C., Spillane, J. P., & Hufferd-Ackles, K. (2001). Storied identities: teacher learning and subject-matter content. *Journal of Curriculum Studies*, 33(1), 1-23.
- Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *The Alberta Journal of Educational Research*, 53(1), 1-21.
- Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin.
- Enyedy, N., Goldberg, J., & Welsh, K. M. (2006). Complex dilemmas of identity and practice. *Science Education*, 90(1), 68-93.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching*, (pp. 119-161). New York: Macmillan.
- Erickson, F. (2004). Demystifying data construction and analysis. *Anthropology & Education Quarterly*, 35(4), 486-493.
- Ernest, P. (1991). *The philosophy of mathematics education*. London: Falmer Press.
- Even, R. (2005). Using assessment to inform instructional decisions: How hard can it be?. *Mathematics Education Research Journal*, 17(3), 45-61.
- Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers College.
- Fullan, M. (2009). Large-scale reform comes of age. *Journal of Educational Change*, 10, 101-113.

- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915-945.
- Geijsel, F., & Meijers, F. (2005). Identity learning: the core process of educational change. *Educational Studies*, 31(4), 419-430.
- Gellert, U. (2008). Routines and collective orientations in mathematics teachers' professional development. *Educational Studies in Mathematics*, 67, 93-110.
- Gerring, J. (2007). *Case study research: Principles and practices*. Cambridge, UK: Cambridge University Press.
- Gill, M. G., Ashton, P. T., & Algina, J. (2004). Changing pre-service teachers' epistemological beliefs about teaching and learning in mathematics: An intervention study. *Contemporary Educational Psychology*, 29, 164-185.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24(1), 355-392.
- Guskey, T. R. (2002). Computerized gradebooks and the myth of objectivity. *Phi Delta Kappan*, 83(10), 775-780.
- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6-11.
- Guskey, T. R. (2007). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues & Practice*, 26(1), 19-27.

- Hancock, D. R., & Algozzine, B. (2006). *Doing case study research: A practical guide for beginning researchers*. New York: Teachers College Press.
- Hargreaves, A. (1998). The emotional practice of teaching. *Teaching and Teacher Education*, 14(8), 835-854.
- Hargreaves, A. (2005). Educational change takes ages: Life, career and generational factors in teachers' emotional responses to educational change. *Teaching and Teacher Education*, 21(8), 967-983.
- Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, 39(1), 69-95.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies. *The Curriculum Journal*, 16(2), 207-223.
- Harlen, W. (2007). *Assessment of learning*. London: Sage.
- Harris, A. (2003). Behind the classroom door: The challenge of organizational and pedagogical change. *Journal of Educational Change*, 4(4), 369-382.
- Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom factors that support and inhibit high-level mathematical thinking and reasoning. *Journal of Research in Mathematics Education*, 28(5), 524-549.
- Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the message across: The problem of communicating assessment feedback. *Teaching in Higher Education*, 6(2), 269-274.
- Hodgen, J., & Askew, M. (2007). Emotion, identity and teacher learning: Becoming and primary mathematics teacher. *Oxford Review of Education*, 33(4), 469-487.
- Huberman, M. (1993). *The lives of teachers* (J. Neufeld, Trans.). New York: Teachers College Press. (Original work published 1989).

- Joughin, G. (2009). Refocusing Assessment. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education*. Palo Alto, CA: Springer.
- Katz, S., Earl, L., & Olson, D. (2001). The paradox of classroom assessment. *McGill Journal of Education*, 36(1), 13-26.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Lampert, M. (1985). How do teachers manage to teach?. *Harvard Educational Review*, 55(2), 178-194.
- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, 27(1), 29-63.
- Lampert (2001). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: minute by minute, day by day. *Educational Leadership*, 63(3), 18-24.
- Levin, B. (2009). Reform without (much) rancor. In A. Hargreaves & M. Fullan (Eds.), *Change Wars*, (pp. 259-272). Bloomington, IN: Solution Tree.
- Levin, B., Glaze, A., & Fullan, M. (2008). Results without rancor or ranking: Ontario's success story. *Phi Delta Kappan*, 90(4), 273-280.
- Little, J. W. (2002). Professional community and the problem of high school reform. *International Journal of Educational Research*, 37, 693-714.
- Little, J. W., & Horn, I. S., (2007). 'Normalizing' problems of practice: converting routine conversations into a resource for learning in professional communities. In L. Stoll &

- K. S. Louis (Eds.), *Professional learning communities: Divergence, depth and dilemmas*, (pp. 79-92). New York: Open University Press.
- Lock, C., & Munby, H. (2000). Changing assessment practices in the classroom: A study of one teacher's challenge. *Alberta Journal of Educational Research*, 46(3), 267-79.
- Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. W. (2003). *Designing professional development for teachers of science and mathematics* (2nd ed.). Thousand Oaks, CA: Corwin.
- McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago: University of Chicago Press.
- McLaughlin, M. W., & Talbert, J. E. (2007). Building professional learning communities in high schools: challenges and promising practices. In L. Stoll & K. S. Louis (Eds.), *Professional learning communities: Divergence, depth and dilemmas*, (pp. 151-165). New York: Open University Press.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43.
- McMillan, J. H. (2004). *Classroom assessment: Principles and practice for effective instruction* (3rd ed.). Boston: Pearson.
- Milne, J., & Oberle, K. (2005). Enhancing rigor in qualitative description: a case study. *Journal of Wound, Ostomy and Continence Nursing*, 32(6), 413-40.
- National Council of Teachers of Mathematics [NCTM] (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

- National Council of Teachers of Mathematics [NCTM] (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics [NCTM] (1991). *Professional standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics [NCTM] (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics [NCTM] (n.d.). Write for Mathematics Teacher. Retrieved June 28, 2008 from <http://www.nctm.org/publications/content.aspx?id=10452>
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149-170.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Noddings, N. (1996). On community. *Educational Theory*, 46(3), 245-267.
- O'Connor, K. (2002). *How to grade for learning: Linking grades to standards*. Glenview, IL: LessonLab.
- O'Connor, K. (2007a). *A repair kit for grading: 15 fixes for broken grades*. Portland, OR: Educational Testing Service.
- O'Connor, K. (2007b). The last frontier: Tackling the grading dilemma. In D. Reeves (Ed.) *Ahead of the curve: The power of assessment to transform teaching and learning*, (pp. 127-145). Bloomington, IN: Solution Tree.
- Ohlsen, M. T. (2007). Classroom assessment practices of secondary school members of NCTM. *American Secondary Education*, 36(1), 4-14.

- Ontario College of Teachers [OCT] (n.d.). One-session Honour Specialist Additional Qualifications. Retrieved July 12, 2009 from http://www.oct.ca/additional_qualifications/schedule_e/?lang=en-CA
- Ontario Ministry of Education (1999). *Guide to the provincial report card, Grades 9-12*. Toronto: Queen's Printer for Ontario.
- Ontario Ministry of Education (2000). *Program planning and assessment*. Toronto: Queen's Printer for Ontario.
- Ontario Ministry of Education (2005). *The Ontario curriculum: Grades 9 and 10 mathematics*. Toronto: Queen's Printer for Ontario.
- Ontario Ministry of Education (2007). *The Ontario curriculum: Grades 11 and 12 mathematics*. Toronto: Queen's Printer for Ontario.
- Ontario Ministry of Education (2008). *Growing success: Assessment, evaluation and reporting*. Toronto: Queen's Printer for Ontario.
- Pellegrino, J. W., & Goldman, S. R. (2008). Beyond rhetoric: Realities and complexities of integrating assessment into classroom teaching and learning. In C. W. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning*, (pp. 7-52). New York: Lawrence Erlbaum Associates.
- Pope, N., Green, S. K., Johnson, R. L., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education*, 25(5), 778–782.
- Romano, M. E. (2006). “Bumpy moments” in teaching: Reflections from practicing teachers. *Teaching and Teacher Education*, 22(8), 973-985.
- Rousseau, C. K. (2004). Shared beliefs, conflict, and a retreat from reform: the story of a professional community of high school mathematics teachers. *Teaching and Teacher Education*, 20(8), 783-796.

- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Saxe, G. B., Gearhart, M., Franke, M. L., Howard, S., & Crockett, M. (1999). Teachers' shifting assessment practices in the context of education reform in mathematics. *Teaching and Teacher Education*, 15(1), 85-105.
- Schram, T. H. (2006). *Conceptualizing and proposing qualitative research*. Upper Saddle River, NJ: Pearson.
- Seidman, I. (2006). Interviewing as qualitative research: A guide for researchers in education and the social sciences (3rd ed.). New York: Teachers College Press.
- Senk, S. L., Beckmann, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education*, 28(2), 187-215.
- Sfard, A. (2003). Balancing the unbalanceable: The NCTM Standards in light of theories of learning mathematics. In J. Hiebert (Ed.), *A research companion to the Principles and Standards*, (pp. 353-392). Reston, VA: NCTM.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 1-14.
- Shepard, L. A. (2001). The role of classroom assessment teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching*, (pp. 1066-1101). Washington, DC: AERA.
- Sherin, M. G., Mendez, E. P., & Louis, D. A. (2004). A discipline apart: the challenges of 'fostering a community of learners' in a mathematics classroom. *Journal of Curriculum Studies*, 36(2), 207-232.

- Sipe, L. R., & Ghiso, M. P. (2004). Developing conceptual categories in classroom descriptive research: some problems and possibilities. *Anthropology & Education Quarterly*, 35(4), 472-85.
- Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: the mediating role of teachers' zones of enactment. *Journal of Curriculum Studies*, 31(2), 143-175.
- Spillane, J., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387-431.
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Stake, R. E. (2006). *Multiple case study analysis*. New York: Guilford.
- Stiggins, R. J. (2002). Assessment crisis: the absence of assessment *FOR* learning. *Phi Delta Kappan*, 83(10), 758-765.
- Stiggins, R. J. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86(1), 22-27.
- Stiggins, R. J. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5-14.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning*, (pp. 133-145). London: Sage.
- Stobart, G. (2008). *Testing times: the uses and abuses of assessment*. New York: Routledge.

- Suurtamm, C. (2004). Developing authentic assessment: Case studies of secondary mathematics teachers' experiences. *Canadian Journal of Science, Mathematics and Technology Education*, 4(4), 497-513.
- Suurtamm, C., & Graves, B. (2007). *Curriculum Implementation in Intermediate Mathematics (CIIM): Research Report*. Retrieved October 24, 2007, from <http://www.edu.gov.on.ca/eng/studentsuccess/lms/>
- Suurtamm, C., Koch, M., & Arden, A. (in press). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice*.
- Suurtamm, C., Lawson, A., & Koch, M. (2008). The challenge of maintaining the integrity of reform mathematics in large-scale assessment. *Studies in Educational Evaluation*, 34, 31-43.
- Talanquer, V., Tomanek, D., & Novodvorsky, I. (2007). Revealing student teachers' thinking through dilemma analysis. *Journal of Science Teacher Education*, 18, 399-421.
- Tierney, R. D. (2006). Changing practices: Influences on classroom assessment. *Assessment in Education: Principles, Policy & Practice*, 13(3), 239-264.
- Tillema, H., & Kremer-Hayon, L. (2005). Facing dilemmas: teacher-educators' ways of constructing a pedagogy of teacher education. *Teaching in Higher Education*, 10(2), 203-217.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, 14(3), 281-294.

- van Veen, K., Slegers, P., Bergen, T., & Klassen, C. (2001). Professional orientations of secondary school teachers towards their work. *Teaching and Teacher Education*, 17, 175-194.
- Watson, A. (2006). Some difficulties in informal assessment in mathematics. *Assessment in Education: Principles, Policy & Practice*, 13(3), 289-303.
- Webb, M., & Jones, J. (2009). Exploring tensions in developing assessment for learning, *Assessment in Education: Principles, Policy & Practice*, 16(2), 165-184.
- Weiner, J. (2004). *The Mathematics Survival Kit*. Toronto: Nelson.
- Wenger, E. (1998). *Communities of practice: Learning, meaning & identity*. Cambridge: Cambridge University Press.
- Westheimer, J. (1998). *Among schoolteachers: Community, autonomy and ideology in teachers' work*. New York: Teachers College.
- Whittemore, R., Chase, S.K., & Mandle, C.L. (2001). Validity in qualitative research. *Qualitative Health Research*, 11(2), 522-537.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design*. Alexandria, VA: ASCD.
- William, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning*, (pp. 1051-1098). Reston, VA: NCTM.
- Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural and political challenges facing teachers. *Review of Educational Research*, 72(2), 131-175.
- Wood, T., Williams, G., & McNeal, B. (2006). Children's mathematical thinking in different classroom cultures. *Journal for Research in Mathematics Education*, 37(3), 222-255.

- Woods, P., Jeffrey, B., Troman, G., & Boyle, M. (1997). *Restructuring schools, reconstructing teachers*. London: Falmer Press.
- Wormeli, R. (2006). *Fair isn't always equal: Assessing & grading in the differentiated classroom*. Portland, ME: Stenhouse.
- Yin, R. K. (2003). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.
- Zeibarth, S. W. (2003). Classroom assessment issues related to teaching mathematics through problem solving. In H. L. Schoen (Ed.). *Teaching Mathematics through Problem Solving* (pp. 177-189). Reston, VA: NCTM.

APPENDICES

Appendix A: Interview Guide for Introductory Meeting

Note: This interview was combined with the First Interview for both participants

Introductory Meeting

[written notes only – no audio recording]

Anticipated length: 15 minutes

Date:

Time:

Location:

Interviewer: Ann Arden

Participant:

Position:

Email Address:

Found out about study:

[Thank potential participant for agreeing to meet]

[Describe research project and go over informed consent]

Questions:

1. a) How long have you been teaching?
b) How long have you been teaching mathematics at the secondary school level?
2. a) What courses are you teaching this semester?
b) What courses are you teaching next semester?
c) Which other secondary mathematics courses have you taught?
3. Have you had any other teaching experiences, or taught any subjects other than secondary mathematics?

[Set up first formal interview; ask teacher to think about which class they would like me to observe; ask about when would be a good time for classroom observations]

Appendix B: Interview Guide for First Interview

First Interview (Semi-Structured) [to be audio recorded]

Anticipated length: 60 minutes

Date/Time:

Participant:

Location:

Interviewer: Ann Arden

[Thank participant for agreeing to participate; confirm it is okay to audiotape the interview]

Questions:

A) Teacher's background

1. Can you tell me how you came to be a secondary mathematics teacher?
2. What were your experiences with classroom assessment when you were a student?

B) Experiences with curriculum and assessment revisions

3. In relation to assessment, what sorts of things have you tried in your classroom over the past few years? *(follow up with questions asking participants to describe their experiences – both positive and negative)*
4. Can you describe for me how you get a sense of your students' understandings and learning in mathematics?
5. How do you determine your students' grades for mathematics? *(in follow up questions ask about grading tools for particular assignments as well as determination of course grades)*
6. As you know, there have been revisions to both curriculum and assessment policy in Ontario over the past decade. How have these changes impacted the way you conduct assessment and evaluation in your classroom?

C) Challenges and supports

7. What challenges have you experienced in trying to implement these changes in assessment? *(ask about strategies for coping with these challenges in follow up questions)*
8. What supports have you experienced in trying to implement these changes in assessment?

[Ask participant if there is anything else they would like to add. Thank participant]

Appendix C: Classroom Observation Protocol**Classroom Observation Protocol****Date/Time:****Class:****Participant:****Observer:** Ann Arden

Descriptive Notes	Reflective Notes

Appendix D: Interview Guide for Post-Observation Debriefings

Post-Observation Debriefings (Informal/Conversational) [to be audio recorded]

Anticipated length: 30 minutes

This interview will be an informal conversation about the teacher's perspective on the lesson. While I expect that questions will emerge during the observation, I have included several possible questions.

Date/Time:
Location:

Participant:
Interviewer: Ann Arden

Questions:

1. How did you feel about the lesson today?
2. In relation to assessment, can you describe for me some of the things you were thinking about in preparation for and during today's lesson?

(in follow up questions ask about how the teacher got a sense of students' understandings and learning during the lesson; ask about any "changes of plan" that the teacher made during the lesson; ask about any upcoming evaluations)

3. Did you experience anything today that you were particularly proud of or found challenging?

I also anticipate asking questions to clarify contextual issues.

[Ask participant if there is anything else they would like to add. Thank participant]

Appendix E: Interview Guide for Final Interview (Amy)

Final Interview (Semi-Structured) **[to be audio recorded]**

Anticipated length: 60 minutes

Date/Time:
Location:

Participant:
Interviewer: Ann Arden

[Thank participant for allowing me to observe classes and agreeing to meet today]

1. How did the summative assessments go in your courses this semester?
2. How have you been developing your understanding of new assessment practices over the past few years? What sorts of experiences and resources have you found helpful?
3. What sorts of things have you struggled with this semester? What have you learned this semester?
4. With regard to your assessment practice this semester, what will you continue to do? What do you plan change? Why?
 - plans for recording (mark book) for next year?
 - interpreting achievement chart?
5. Given that there are provincial, school board and school policies for assessment and evaluation, how do these influence your decisions in your classroom?

[Ask participant if there is anything else they would like to add. Thank participant]

Appendix F: Interview Guide for Final Interview (Carina)

Final Interview (Semi-Structured) **[to be audio recorded]**

Anticipated length: 60 minutes

Date/Time:
Location:

Participant:
Interviewer: Ann Arden

[Thank participant for allowing me to observe classes and agreeing to meet today]

1. How did the summative assessments go in your courses this semester?
2. How has your assessment checklist worked for you and your students this semester?
How did the student self-assessment checklist go?
3. With regard to your assessment practice this semester, what will you continue to do?
What do you plan change? Why?
 - plans for recording assessments for next year?
 - interpreting achievement chart?
 - remediation program next year?
4. What sorts of things have you struggled with this semester? What have you learned this semester?
5. Given that there are provincial, school board and school policies for assessment and evaluation, how do these influence your decisions in your classroom?
 - whole staff sense?
 - math department sense?
 - using levels vs. percents?
6. How have you been developing your understanding of new assessment practices over the past few years? What sorts of experiences and resources have you found helpful?

[Ask participant if there is anything else they would like to add. Thank participant]

Appendix G: Carina's "Assessment Checklist" – Student Version

Excerpt from Student Version of Carina's Assessment Checklist for Grade 10

	Course Expectations	Ability to meet Expectation			
		Always	Mostly	At times	Rarely
Unit 1: Linear Systems	Writing systems of equations (writing equations from word problems)				
	Determine the number of solutions of a system				
	Solving systems of equations:				
	- graphing by hand				
	- using graphing technology				
	- by method of substitution				
	- by method of elimination				
Self-reflection:					
Unit 2: Analytic Geometry	Properties of line segments (slope midpoint & length)				
	Equation of a circle				
	Verifying geometric properties				
	- equation of median line, perpendicular bisectors, altitudes				
	- centroid				
	- shortest distance between a line and a point				
	- proving triangles & quadrilaterals				
Self-reflection:					

Appendix H: Carina's "Assessment Checklist" – Teacher Version

Excerpt from Teacher Version of Carina's Assessment Checklist for Grade 10

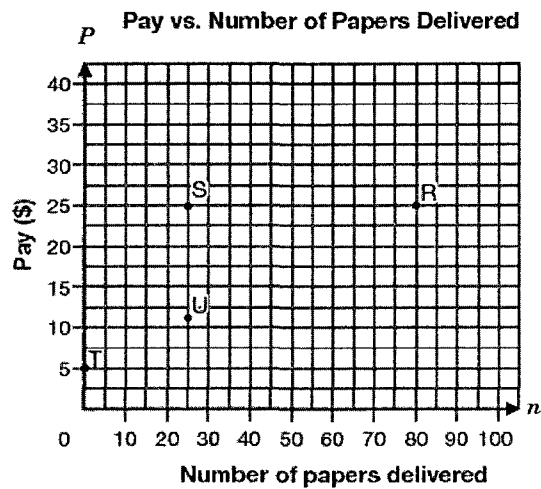
	Course Expectations	Evidence of Expectation								
		Exceeds	Meets	Meets with Help	Does not meet					
Unit 1: Linear Systems	Writing systems of equations (writing equations from word problems)									
	Determine the number of solutions of a system									
	Solving systems of equations:									
	- graphing by hand									
	- using graphing technology									
	- by method of substitution									
	- by method of elimination									
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%;">Notebook Check:</td> <td style="width: 20%;">Notes:</td> <td style="width: 20%;">Homework:</td> <td style="width: 20%;">Review:</td> <td style="width: 20%;">Organization:</td> </tr> </table> Comments:						Notebook Check:	Notes:	Homework:	Review:	Organization:
Notebook Check:	Notes:	Homework:	Review:	Organization:						
Unit 2: Analytic Geometry	Properties of line segments (slope midpoint & length)									
	Equation of a circle									
	Verifying geometric properties									
	- equation of median line, perpendicular bisectors, altitudes									
	- centroid									
	- shortest distance between a line and a point									
	- proving triangles & quadrilaterals									
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%;">Notebook Check:</td> <td style="width: 20%;">Notes:</td> <td style="width: 20%;">Homework:</td> <td style="width: 20%;">Review:</td> <td style="width: 20%;">Organization:</td> </tr> </table> Remediation: Y/N Days Attended: Work submitted: Marks on original test: $\frac{\quad}{K} + \frac{\quad}{A}$ New Marks: $\frac{\quad}{K} + \frac{\quad}{A}$ Comments:						Notebook Check:	Notes:	Homework:	Review:	Organization:
Notebook Check:	Notes:	Homework:	Review:	Organization:						

Appendix I: EQAO Items used in Amy's class

Publicly available at www.eqao.com

Two sample questions are included here:

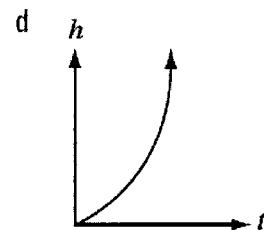
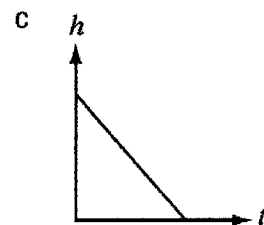
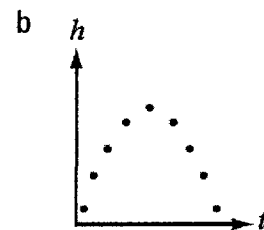
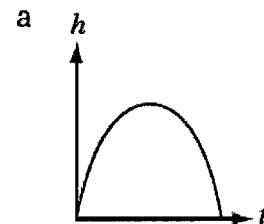
Mia delivers the local newspaper. Her base pay is \$5 per week, and she gets \$0.25 per paper.



Which of the points on the graph represents Mia's pay for delivering 25 newspapers in a week?

- a Point R
- b Point S
- c Point T
- d Point U

Which of the following graphs represents a linear relation?



Appendix J: Sample of Worksheets from Amy's class

Publicly available at www.edu.gov.on.ca/eng/studentsuccess/lms/tips4rm.html

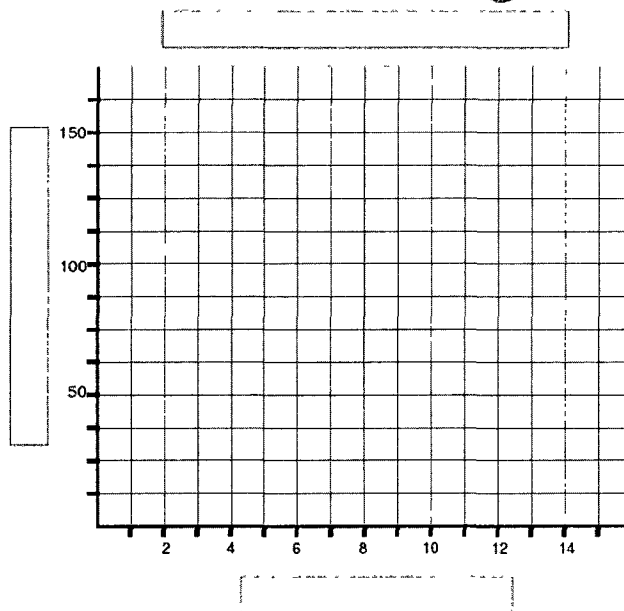
Two sample items included here:

5.6.2: Descriptions, Tables of Values, Equations, Graphs (continued)

3. Nisha is just learning how to snowboard. White Mountain charges \$10/hour for lessons and \$40 for the lift ticket and snowboard rental.

- Make a table of values for up to 6 hours.
- Graph the relationship.





- Write an equation to model the relationship.

_____ = _____

- Does this relation represent a partial or direct variation? Explain.

- How much will it cost in total for Nisha to take 2.5 hours of lessons?
Show your work.

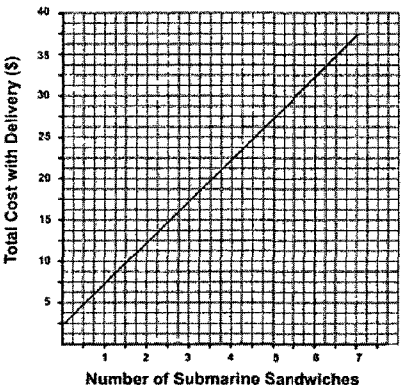
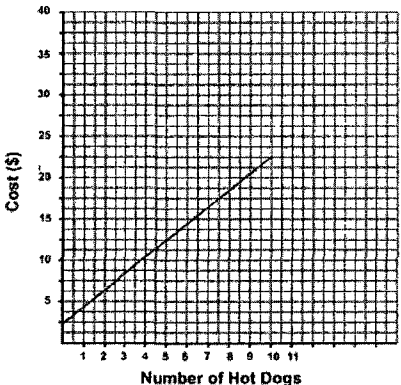
- If Nisha paid \$75, how long was she at the White Mountain getting lessons?
Show your work.

5.8.1: Modelling Linear Relations with Equations

Food Frenzy

Partner A: _____ Partner B: _____

Write the equation for each relationship in the space provided. Show any calculations you made. Indicate if the relation is a partial or direct variation and whether the line modelling the relationship is solid or dashed.

A coaches B	B coaches A																								
1. A family meal deal at Chicken Deluxe costs \$26, plus \$1.50 for every extra piece of chicken added to the bucket.	2. A Chinese food restaurant has a special price for groups. Dinner for two costs \$24 plus \$11 for each additional person.																								
3. Total Cost of Submarine Sandwiches 	4. Total Cost of Hot Dogs at the Baseball Game 																								
5. <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>Number of Toppings</th> <th>Cost of a Large Pizza (\$)</th> </tr> </thead> <tbody> <tr><td>0</td><td>9.40</td></tr> <tr><td>1</td><td>11.50</td></tr> <tr><td>2</td><td>13.60</td></tr> <tr><td>3</td><td>15.70</td></tr> <tr><td>4</td><td>17.80</td></tr> </tbody> </table>	Number of Toppings	Cost of a Large Pizza (\$)	0	9.40	1	11.50	2	13.60	3	15.70	4	17.80	6. <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>Number of Scoops</th> <th>Cost of Ice Cream with Sugar Cone (\$)</th> </tr> </thead> <tbody> <tr><td>0</td><td>1.25</td></tr> <tr><td>1</td><td>2.00</td></tr> <tr><td>2</td><td>2.75</td></tr> <tr><td>3</td><td>3.50</td></tr> <tr><td>4</td><td>4.25</td></tr> </tbody> </table>	Number of Scoops	Cost of Ice Cream with Sugar Cone (\$)	0	1.25	1	2.00	2	2.75	3	3.50	4	4.25
Number of Toppings	Cost of a Large Pizza (\$)																								
0	9.40																								
1	11.50																								
2	13.60																								
3	15.70																								
4	17.80																								
Number of Scoops	Cost of Ice Cream with Sugar Cone (\$)																								
0	1.25																								
1	2.00																								
2	2.75																								
3	3.50																								
4	4.25																								

Appendix K: Editing quotes from transcripts

Here is an example of a quote that was edited to ensure readability while maintaining the meaning:

Original quote from transcript:

“If you’re not checking these things off as ‘always met’ there’s work to be done and you’ve only got one week now cause we will have the test the week after and that’s when the game is on, right? There isn’t like ‘can we have a replay?’ The game’s on. That’s it. You’ve had all this time to practice.” (P2Int1, p. 34)

Edited quote:

If you’re not checking these things off as ‘always met’ there’s work to be done and you’ve only got one week now cause we will have the test the week after and that’s when the game is on. There isn’t ‘can we have a replay?’ The game’s on. That’s it. You’ve had all this time to practice” (dissertation, p. 112)