The Influence of Teachers’ Instructional Practices on Student Motivation in Different Assessment Contexts

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Monica J. Kowalski, M.Ed.

Graduate Program in Educational Policy and Leadership

The Ohio State University

2013

Dissertation Committee:

Lynley H. Anderman, Advisor

Eric M. Anderman

Antoinette Errante
Abstract

Current federal educational policy mandates annual standardized achievement testing for students beginning in third grade. Standardized tests are considerably different from regular classroom assessments in format, usage, and meaning. Therefore, teachers are likely to approach the preparation for standardized testing differently than that for regular classroom assessments, with possible implications for students’ motivation in those two contexts. This mixed methods study provides evidence of how five third-grade teachers’ instructional practices differ between classroom assessment and standardized testing preparation contexts. Results from classroom observations and teacher interviews indicate that teachers engage in more mastery goal supportive teaching strategies in the classroom assessment context and more performance goal supportive teaching strategies in the standardized testing context. They also speak differently about the value of the two types of tests and about their expectations for students’ success. The differences are reflective of teachers’ knowledge, beliefs, and perceptions of external pressure related to the different types of assessments. Student survey results provided little evidence to support the hypothesis that they would perceive the classroom motivational climate as being different in the two contexts, however, severe limitations to the survey design are discussed that may have contributed to the lack of significant findings.
Dedication

To Justin, Nathan, and Molly.
May my work help improve your future education.
Acknowledgements

“Holy Spirit, Enlighten my work with your gifts of wisdom and understanding.”

Throughout the process of writing this dissertation, and my graduate school experience in general, I have uttered that prayer countless times. Now, I offer a prayer of gratitude for all of God’s blessings on me that have helped me to complete this project.

I am blessed to have had the help and support of my dissertation committee. I am grateful for my advisor, Dr. Lynley Anderman. Lynley provided me with a tremendous amount of thoughtful feedback on my writing and challenged me in a way that no other teacher ever has. I have become a much stronger writer as a result of being her student. Lynley has also been an invaluable mentor to me outside of the dissertation. I am thankful for her guidance on handling graduate school stress and balancing career ambitions and family priorities. I always appreciated our conversations and her understanding through my occasional tears.

I am grateful for Dr. Eric Anderman and Dr. Antoinette Errante for also serving on my committee. I enjoyed taking classes from Eric and having him as a mentor for my role as a Graduate Teaching Associate. Eric has always been warm and generous with his time, although he is one of the busiest people I know. I appreciate how seriously he takes his role as a mentor, genuinely caring about his students and involving them in his research. I was also happy to have the opportunity to work with Antoinette on a group
research project during my time in graduate school. Her impressive qualitative research skills, as well as her friendliness and approachability, make her an asset to my committee.

There are many other people at Ohio State that I would like to acknowledge for their contributions to this project and/or to my graduate school experience in general. I have had the pleasure of learning from excellent faculty, including Drs. Robert Hite, Ann O’Connell, Bryan Warnick, Ann Allen, Jan Nespor, Richard Lomax, and Bruce Tuckman. I am also grateful for Dr. Anita Woolfolk Hoy and Dr. Belinda Gimbert for serving on my candidacy committee. Deb Zabloudil in Student Services has helped me numerous times with questions big and small and has always been a joy to work with. I would like to thank other OSU students that I have enjoyed working and learning with, such as Dr. Mike Yough, Kathy Kirkpatrick, Dr. Mei-Lin Chang, Dr. Heather Dawson, Ashley Marrietta-Brown, Stephanie Levitt, Dr. David Fultz, and Dr. Paige Shalter Bruening. I especially want to thank students Kate Kovach, Bree Frick, and Megan Sanders for their direct assistance with this dissertation. In addition, I would like to acknowledge the financial support that I have received from the School of Educational Policy and Leadership and from the College of Education and Human Ecology. I am especially grateful for the EHE Dissertation Fellowship Award, which made it possible for me to focus all of my attention on this project during the last year.

There are also many people I would like to acknowledge from my other academic home, Notre Dame. Dr. Julie Turner first introduced me to the field of Educational Psychology and was instrumental in my decision to pursue graduate studies at Ohio State. Her colleague, Dr. Kristen Warzon, provided me with the opportunity to engage in research. Dr. Anre Venter provided strong guidance as my undergraduate advisor.
Finally, many people in the Alliance for Catholic Education at Notre Dame inspired and supported me along my academic journey, including Drs. John Staud, Thomas Doyle, Matt Kloser, Anthony Holter, Jim Frabutt, Christian Dallavis, Ryan Clark, Sean McGraw, CSC, and Sr. Kathleen Carr, Steve Perla, and Lou DelFra, CSC. I am excited to be working with all of them again in my new role as Associate Director of ACE Consulting.

Outside of academia, I am blessed with many other friends and family members who have supported me throughout this project and beyond. I am grateful for friends and neighbors who have provided welcome distractions and helped me keep my sanity during the often stressful time as a graduate student. I am grateful for my brother-in-law, Dr. Brian Kowalski, who truly understood what I was going through, having earned his PhD a few years before me. His wife, Colette, has been a great sister and a constant source of support and friendship. I am thankful for my older brother, Jeff Stone, for being a role model of hard work and efficiency my whole life. I am happy that he married a teacher, Laura, and that their wonderful children know the value of a strong education. I am also thankful for my younger sister, Valerie. In many ways, she was my first student, agreeing to spend hours playing school with me as a child.

I acknowledge my amazing in-laws, Sue and Rich Kowalski. They have always been there to listen and encourage me during stressful times as a student. I am especially grateful for the times they came into town to help out with the kids so that I could write. My own parents, Mary and Joe Stone, have been my first and most important teachers. I owe all my success to them, for the values they taught me and for their unconditional love and support. I appreciate Dad for countless hours of babysitting while I worked and
Mom for always showing such interest in my studies and her willingness to help in anyway, from entering data to proofreading drafts and just being there to listen to me vent. I am truly blessed to be their daughter.

I also wish to acknowledge my own young children, for they have contributed to my success in this endeavor. Justin has colored my entire graduate school experience, since I learned that I was pregnant with him on my first day of graduate school. His presence has taught me to be more efficient than I ever thought possible, and has helped me to balance the priorities in my life. Nathan’s arrival after my candidacy exams further inspired me to focus on completing my studies and his joyful laughter and sweet snuggles were a welcome backdrop to my dissertation work. My new daughter Molly was perhaps the best source of motivation ever, arriving just two days after I turned in a completed dissertation draft to my advisor. While having kids during graduate school has been challenging at times, I am so grateful for the blessings of happy and healthy children and I would not have done it any other way.

Finally, I acknowledge my greatest source of love and support, my husband, Christopher. I would not and could not have done this without him. His unwavering confidence in me gave me the strength to complete this work. Chris, thank you for taking the risk to move here with me five years ago and for holding my hand throughout the ups and downs of this experience. May God continue to bless us as we enter the next adventure together!
Vita

June 1999…………………………………Westerville South High School

2003………………………………………B.A., Psychology and Computer
Applications, University of Notre Dame

2005………………………………………M.Ed., Education, University of Notre Dame

2008-2013………………………………Graduate Associate, School of Educational
Policy and Leadership, The Ohio State University

Fields of Study

Major Field: Educational Policy and Leadership

Specialization: Educational Psychology
# Table of Contents

Abstract........................................................................................................................................... ii

Dedication........................................................................................................................................ iii

Acknowledgements....................................................................................................................... iv

Vita................................................................................................................................................... viii

List of Tables.................................................................................................................................... xiv

Chapter 1: Introduction.................................................................................................................... 1
  Statement of the Problem............................................................................................................... 1
  Overview of Current High-Stakes Testing Policy: NCLB......................................................... 3
  The Current Study.......................................................................................................................... 6

Chapter 2: Literature Review......................................................................................................... 8
  Definitions of Terms...................................................................................................................... 8
  Theoretical Frameworks for Student Motivation................................................................. 12
    Achievement Goal Theory....................................................................................................... 13
    Mastery Goals......................................................................................................................... 15
    Performance Goals.................................................................................................................. 16
    Classroom Goal Structures................................................................................................... 17
  Expectancy-Value Theory........................................................................................................... 18
  Teachers’ Assessment Practices............................................................................................... 21
    Teachers’ Classroom Assessment Practices........................................................................ 21
Teacher Knowledge and Beliefs.................................................................61
Quantitative Measures.............................................................................61
Achievement Goal Structures.................................................................62
Students’ Task Value...............................................................................64
Students’ Expectancy Beliefs.................................................................66
Students’ Prior Achievement.................................................................66
Student Demographics.........................................................................66
Analysis Approach................................................................................67

Chapter 4:
Results..................................................................................................69
Qualitative Results..................................................................................70
  Achievement Goal Theory.................................................................72
  Expectancy-Value Theory.................................................................76
Quantitative Results.............................................................................86
  Reliability............................................................................................87
  Preliminary Analysis.........................................................................90
  Primary Analysis..............................................................................96
Mixed Methods Results.......................................................................98
Summary of Results.............................................................................98

Chapter 5: Discussion..........................................................................100
Instructional Practices Across Assessment Contexts..........................100
Teachers’ Knowledge and Beliefs Across Assessment Contexts...........105
Students’ Motivational Perceptions Across Assessment Contexts.......109
Limitations.................................................................................................................113
Direction for Further Research..................................................................................115
Summary.....................................................................................................................116
References..................................................................................................................118
Appendix A: IRB Approval.......................................................................................136
Appendix B: Email to Recruit Participants...............................................................139
Appendix B: Consent Forms......................................................................................141
Appendix C: Student Survey.....................................................................................151
Appendix D: Teacher Interview Protocol.................................................................154
List of Tables

Table 3.1 Study Participants........................................................................................................53
Table 3.2 Study Timeline............................................................................................................54
Table 4.1 Themes, Subthemes, and Representative Quotes................................................71
Table 4.2 Means and Standard Deviations for Items and Scales and Correlations for First
and Second Time Points........................................................................................................89
Table 4.3 Scale Reliability (Cronbach’s Alpha Coefficients).................................................90
Table 4.4 Motivation Measures for Female and Male Students.............................................91
Table 4.5 Motivation Measures for Student Ethnicity............................................................92
Table 4.6 Motivation Measures at Each Time Point by Teacher........................................93
Table 4.7 Motivation Measures by Terra Nova Score Category............................................95
Table 4.8 Mean, Standard Deviation, t Statistics, and p Values for Measures.........................97
Chapter 1: Introduction

Statement of the Problem

With the current public interest in and federal mandates for increased accountability in our nation’s schools, high-stakes standardized testing has become rampant. The No Child Left Behind Act (No Child Left Behind [NCLB], 2002), Race to the Top (U.S. Department of Education, 2009), and Blueprint for Education (U.S. Department of Education, 2010) federal policies all include extensive standardized testing requirements and subsequent consequences for schools and teachers based on their students’ performance (Ravitch, 2010).

Although few would argue with the aim of holding schools and teachers accountable for gains in student learning, many have criticized the overuse of high-stakes testing as the sole means for evaluating progress of schools and students (Kohn, 2000; Ravitch, 2010). Some of the strongest criticisms center on the argument that an overemphasis on standardized testing lowers the quality of teaching and learning in schools and has a negative influence on student motivation (e.g. Au, 2007; Ryan & Weinstein, 2009). These negative consequences of standardized testing can be better understood by examining how the testing is enacted in classrooms and investigating teachers’ and students’ perceptions of the testing process.

A starting point for understanding the negative consequences of standardized testing is to examine how the standardized testing process is different from regular
classroom assessment activities. After all, for as long as there has been formal education, there has been assessment of student learning (Shepard et al., 2005). Classroom assessment is a natural part of the school experience. Students are subjected to various types of assessments, including quizzes and tests, on a regular basis in their classrooms. These assessments are designed and/or chosen by teachers to evaluate student performance and inform future instruction. Results of these assessments are typically shared only between the teacher, student, and possibly parents.

In contrast, federally-mandated standardized testing is not an everyday occurrence in classrooms. The tests are administered once each year, although usually a substantial amount of classroom time throughout the year, especially in the weeks directly preceding the testing, is devoted to preparing for the tests (Au, 2007; Nichols & Berliner, 2009). Rather than being teacher-created, the tests are written and scored by an external agency approved by the state’s Department of Education. Results are not a private matter between teacher and student, but rather, are used as an evaluation of the school and teacher’s level of success in demonstrating that their students are achieving grade-level proficiency in the tested content areas. The tests are high-stakes for the schools and teachers, as they face consequences such as sanctions if their students do not perform well (NCLB, 2002).

Given the vast differences between regular classroom assessments and standardized testing, teachers’ instructional practices related to these types of assessments are expected to be very different as well (Kowalski & Marietta-Brown, 2011). The ways that teachers approach these assessment activities in their classrooms are likely based on their own knowledge and beliefs about the assessments, as knowledge and beliefs are
known to be related to teachers’ practices (Woolfolk Hoy, Davis, & Pape, 2006). The extent of teacher knowledge of academic content, principles of measurement and evaluation, and educational policy has potential to influence teachers’ instructional practices related to different types of assessments (Brookhart, 1997; Shepard, et al., 2005; Smith, 1991). Teachers’ internal values and beliefs about the nature and purposes of assessments also play a substantial role in determining teachers’ actions related to assessment (McMillan, 2003). In addition, perceptions of external pressure surrounding the high-stakes testing process are expected to influence teachers’ instructional practices (Assaf, 2008; Mathison & Freeman, 2003). Subsequently, teachers’ instructional practices related to assessment and standardized testing have implications for student motivation.

Several motivational theories, including achievement goal theory (e.g. Ames, 1992) and expectancy value theory (e.g. Wigfield & Eccles, 2000), offer explanations for how assessment and testing might influence student motivation. However, empirical research providing evidence for these theoretical assumptions has not been extensive. In addition, the role of the teacher in contributing to students’ experiences with testing has not been specifically addressed. This study will explore the links between assessment, standardized testing, and student motivation and the ways that teachers’ instructional practices and underlying knowledge and beliefs impact students’ perceptions of the classroom motivational climate surrounding testing.

**Overview of Current High-Stakes Testing Policy: NCLB**

Shortly after taking office, President George W. Bush proposed the No Child Left Behind (NCLB) Act of 2001, a reauthorization of the Elementary and Secondary
Education Act of 1965. With enthusiastic bi-partisan support, the Act was signed into law on January 8, 2002. There was wide support for the idea of holding schools and teachers accountable for the learning of all students and standards-based education reform (Linn, 2003; Paige, 2006; Ravitch, 2010). However, the lofty goals of this law are far from being realized and research suggests that the unintended negative consequences of NCLB, particularly with respect to high stakes testing, outweigh the intended benefits envisioned by the writers of the Act (Au, 2007; Kohn, 2000; Nichols & Berliner, 2007).

At the heart of NCLB legislation is a federal mandate that all students in schools receiving public funding be tested annually in mathematics and reading during grades three through eight and once during high school (NCLB, 2002). The individual states are entrusted with the task of designing the standards-based assessments and tracking the results, disaggregated by sub-groups based on socio-economic status, ethnicity, and special needs, with the goal of 100 percent proficiency by the year 2014. Until that year, all schools are expected to make “adequate yearly progress” (AYP) within each sub-group toward perfect proficiency or face a series of increasingly severe sanctions (NCLB, 2002).

The intent of this high stakes testing policy was to improve student achievement and to decrease the achievement gap between white and minority students by holding school districts and schools accountable for effective education of all of our nation’s children (Forte, 2010; Linn, 2003; Paige, 2006; Ravitch, 2010). Policy makers argued that holding schools accountable through attaching consequences to performance of students on achievement tests would motivate schools and districts to improve the quality of education provided, which in turn would result in higher student achievement. While
controversial, some research shows that indeed, test scores, including scores on national tests such as the National Assessment of Educational Progress (NAEP), have risen in many states after implementing high stakes testing policies, signaling an improvement in student achievement (Phelps, 2005). However, other studies have shown the opposite (i.e. Amerein & Berliner, 2002). Discrepancies such as this among studies and threats to validity and reliability of the findings about the effects of testing policies remain troubling for proponents of NCLB (Lee, 2008). The debate over what a rise in test scores really means (i.e. improved learning vs. improved test-taking skills) and the presence of unintended consequences of high stakes testing continues to spark controversy over this law (Harlen, 2005).

Throughout the last decade of NCLB, criticisms of the Act have been widespread. Much attention has been paid to the unintended consequences that have emerged from analyzing the impact of NCLB. Critics of NCLB and high-stakes testing in general argue that implementing this testing system leads to troublesome consequences such as narrowing of curriculum, emphasis on test-taking skills rather than content, corruption by schools and teachers, including cheating and encouraging students not likely to do well on tests to be absent or drop out of school, and decreased student achievement and motivation. (Cimbricz, 2002; Harlen & Deakin-Crick, 2002; Nichols & Berliner, 2007; Ravitch, 2010; Ryan & Brown, 2005; Ryan & Weinstein, 2009). However, empirical evidence of specific ways that these consequences develop at the classroom level has not been well-established. This study will add to the literature by providing more insight into the impact of teachers’ instructional practices related to testing on students’ motivation.
The Current Study

Although testing occurs and has potential to impact students in every grade level, this study will focus on early elementary school, specifically third grade classrooms. According to the current federal policy, NCLB, annual standardized testing is mandated for all students in grades 3-8 in mathematics and reading (NCLB, 2002). This study aims at understanding how initial experiences with this type of standardized testing are associated with teacher actions and student motivational outcomes.

Specific research questions developed for this study are:

1) How do elementary school math teachers’ instructional practices related to assessment differ for summative classroom assessments and preparation for standardized state achievement tests?

2) How do elementary school math teachers’ instructional practices reflect their knowledge, beliefs, and perceptions of external influences related to summative classroom assessments and standardized state achievement tests?

3) How do students perceive the classroom motivational climate surrounding assessment and standardized testing?

These questions will be investigated through a mixed methods research design. Qualitative inquiry will be used to examine the first two questions, comprised of classroom observations and individual interviews with teachers. For the third question, quantitative data will be obtained through a survey of students’ perceptions. The
qualitative and quantitative results will be integrated to answer the research questions and provide a robust understanding of the impact of testing preparation on student motivation.
Chapter 2: Literature Review

The current study examines the impact of teachers’ instructional practices and underlying knowledge and beliefs on student motivation in the contexts of classroom assessments and standardized testing. This chapter begins with a section defining assessment-related terms used in this study. Then, the theoretical frameworks of expectancy-value theory and achievement goal theory are presented. Next, there is a review of the research on teachers’ instructional practices related to classroom assessments and standardized testing followed by review of the literature on teachers’ knowledge and beliefs about the two assessment contexts. Finally, the research questions and expectations and hypotheses of the present study are presented.

Definitions of Terms

This study focuses on assessment contexts in classrooms. Assessment refers to “a formal attempt to determine students’ status with respect to educational variables of interest” (Popham, 1995, pg. 3). This is a broad process that encompasses a wide variety of ways to observe, evaluate, and document student learning. Within the broad area of assessment, there are many ways to classify types of assessments based on purpose and use.
One way to classify assessments is to distinguish between formative and summative assessment. *Formative* assessment refers to assessment for student learning. It is usually informal and occurs before or during instruction (Popham, 1995; Shepard et al., 2005). Formative assessment can be administered through strategies such as oral questioning, quick checks for understanding, practice problems, or discussions. These assessments give feedback to the teacher with regard to how students are progressing in their learning and how effective their teaching has been so that the teachers may continue to refine their lessons and help all students gain better understanding (Black & William, 2009; Shepard et al., 2005).

*Summative* assessment is considered to be assessment of student learning. It takes place after the teaching has occurred and is designed to measure student achievement (Popham, 1995; Shepard et al., 2005). Most summative assessments are associated with grades in typical classroom settings. Summative assessment is generally more formal and evaluative than formative assessment. The traditional form of summative assessment is a paper and pencil test, although many alternative types of assessments have been championed recently including performance assessments, portfolios, and other creative ways for students to demonstrate their learning (Popham, 1995; Wiggins, 1993).

While both formative and summative assessments are essential to effective instruction and can have implications for student motivation, this study will focus on summative assessment contexts. For simplicity, the term “assessment” will be used to refer to “summative assessment” throughout this study unless otherwise specified.

Another important classification regarding assessment is the difference between classroom-based assessment and standardized testing. *Classroom assessment* refers to the
assessment activities that the teacher creates and/or selects and administers in the classroom in order to gather information about his/her students’ learning (Popham, 1995). Summative classroom assessments can include paper and pencil tests, portfolios, projects, exhibits, reports, performances, or other presentations, as well as other forms of demonstration of knowledge and skills (Popham, 1995). For the purposes of this study, paper and pencil math tests will be the classroom assessment activity of interest.

*Standardized testing* refers to a type of test that is given under standardized conditions, such as consistent test directions and time limits, and scored according to uniform procedures (Popham, 1995). Standardized testing has a long history, dating back to the third century B.C. when the Han emperors of China used tests as requirements for government jobs (Eckstein & Noah, 1993). Standardized intelligence tests were developed in the early 1900s, and during World War I, standardized testing was used to assign U.S. military to appropriate duties based on their skills and competencies (Urban & Wagoner, 2009). Standardized testing eventually became more prevalent in educational settings for purposes such as entrance to schools and colleges, identifying students’ strengths and weaknesses, informing instruction, determining which students should be selected for special gifted or remedial education programs, and, more recently, holding schools and individual teachers accountable for students’ learning (Ravitch, 2010).

Standardized tests can be further categorized as either “high stakes” or “low stakes.” A *high stakes standardized test* is a test that is used to make decisions that impact students, teachers, administrators, schools, or districts (Madaus, 1988). Common examples are tests that are used to determine entrance into colleges (i.e. the SAT and
ACT) or graduation tests that require a certain passing score before a student may earn their high school diploma. Tests are also considered high stakes when the results are reported to the public, such as those utilized for current federal accountability policies, discussed below (McNeil, 2000). In contrast, a low stakes standardized test does not have important consequences attached to the results (Wise & DeMars, 2005). These tests may provide useful data to schools and teachers, but they have little or no meaning to students, who do not receive grades nor academic credit for taking them (Wise & DeMars).

The type of standardized test focused on in this study is the state standardized test used to meet No Child Left Behind (NCLB) requirements (NCLB, 2002). The NCLB Act of 2001 requires all public school students in the United States be tested annually in reading and mathematics from third through eighth grade and once again in high school. The individual states were charged with creating or selecting appropriate tests aligned with state standards and determining cut-off scores to indicate proficiency. Results of these federally mandated tests are used to indicate whether or not students, classified into subgroups by categories such as race, gender, and special needs, are making Adequate Yearly Progress (AYP) towards the goal of 100% proficiency by the year 2014 (NCLB, 2002).

These tests are high stakes for schools, teachers, and districts, as results are made public and decisions such as rewards and sanctions are made based on the tests (Ravitch, 2010). However, the tests are low stakes for students, as they personally have no direct consequences attached to their performance on the tests. This issue poses a motivational dilemma regarding the testing: teachers must convince students to take the test seriously even though they may not be motivated to try their best on a low stakes test that holds
little meaning to them (Wise & DeMars, 2005). If students are not motivated to put in effort on the test, the test may not be a valid indicator of what they actually know, which is one major criticism of the accountability movement and the NCLB Act specifically (Nichols & Berliner, 2007; Wise & DeMars, 2005).

Other criticisms of NCLB center on negative consequences associated with increased testing in schools, including narrowing of curriculum, teaching to the test, and increases in unethical behaviors such as cheating (Nichols & Berliner, 2007; Wood, 2004). Research has also investigated the impact of high stakes testing policies on outcomes such as teacher stress and burnout (Dworkin, 2001; Jones & Egley, 2004), student achievement (i.e. Amerein & Berliner, 2002; Lee, 2008; Nichols, Glass, & Berliner, 2006), and student motivation (i.e. Harlen & Deakin-Crick, 2002; Ryan & Weinstein, 2009). The aim of this study is to explore the link between the introduction of standardized testing in third grade classrooms, the first year that federal testing is mandated, and student motivational outcomes, mediated by teachers’ instructional practices and underlying knowledge and beliefs. The following sections will review the literature on these topics, beginning with describing the theoretical frameworks used for student motivation.

**Theoretical Frameworks for Student Motivation**

Many types of motivational beliefs could influence students’ experiences with classroom assessment and standardized testing (Anderman, Anderman, Yough, & Gimbert, 2010). For the purposes of this study, the focus will be on three types of student motivation-related perceptions; specifically, the perception of classroom goal structures, subjective task values, and expectancies for success. In the following sections, the two
main motivational theories related to these perceptions, namely, achievement goal theory and expectancy-value theory, will be reviewed.

**Achievement Goal Theory**

Achievement goal theory emerged through both collaborative and individual work in the 1970s by a number of motivation researchers, including Ames, Dweck, Nicholls, and Maehr (Elliot, 2005). It has now become one of the most prominent and highly researched perspectives in the study of motivation (Maehr, 2002). Achievement goal theory focuses on the goal orientations that students adopt in academic and other contexts (Anderman & Wolters, 2006). The construct of an achievement goal orientation is distinguished from other types of goals, such as objectives or target goals (i.e. Locke & Latham, 2002) that focus only on a task-specific outcome without considering the reason for pursuing the goal (Murphy & Alexander, 2000). Rather, an achievement goal orientation is a pattern of beliefs about goals related to achievement. It includes the reasons or purposes students perceive for engaging in certain tasks as well as the standards or criteria that they use to judge successful performance (Anderman & Wolters, 2006; Pintrich, 2000).

Two primary goal orientations have been identified in the research literature: the goal to learn or achieve mastery at something and the goal to demonstrate competence or look good as compared to others (Ames & Archer, 1988; Dweck & Leggett, 1988; Elliott & Dweck, 1988; Anderman & Wolters, 2006). Students setting and pursuing the first type of goal are focused on improving and learning with little attention to how they compare to others. They evaluate their success in terms of self-improvement. In contrast, students motivated by the second type of goal tend to care more about how they appear in
comparison to others than the extent to which they are learning or improving themselves. They experience a sense of accomplishment when they succeed at outperforming others. Different researchers have used different terminology for the goals, with the first type of goal given names such as a learning goal, a task goal, or a mastery goal, and the second type of goal referred to as an ego goal, an ability goal, or a performance goal (Ames & Archer, 1988; Dweck, 1999; Nicholls, 1984). This study will use the terms mastery and performance goals.

Recently, a further bifurcation between approach and avoidance goals has been proposed for each type of goal (Elliot & Church, 1997; Elliot & Thrash, 2001; Pintrich, 2000). The approach-avoidance distinction has been prevalent in other motivational theories for decades, but had not been applied to achievement goal theory until the late 1990s (Elliot, 2005). At that time, researchers began to question the inconsistency of findings related to the two main goal orientations and theorized that the distinction between approach and avoid may be important (Elliot). Therefore, a separation was made such that students motivated by performance goals focus on demonstrating their competence (performance-approach) or avoiding the demonstration of incompetence (performance-avoid) and students motivated by mastery goals focus on mastering a task (mastery-approach) or avoiding falling short of mastery or losing skills or abilities (mastery-avoid) (Anderman & Wolters, 2006; Elliot & Thrash, 2001; Elliot & McGregor, 2001; Middleton & Midgley, 1997; Pintrich, 2000). While the distinction between approach and avoidance performance goals has been largely supported by empirical and theoretical motivational research (e.g. Harackawicz, Barron, Pintrich, Elliot, & Thrash, 2002; Middleton & Midgley, 1997; Midgley, 2002), there has been less examination of
the approach and avoid constructs within the mastery goal research (Elliot & Thrash, 2001; Meece, Anderman, & Anderman, 2006). Therefore, this dissertation will focus on three primary goal orientations, mastery goals, performance-approach, and performance-avoid goals.

**Mastery goals.** As noted above, students who are motivated by mastery goals focus on developing their understanding and mastering new skills or content knowledge. Mastery goal orientations have been associated with a number of positive student behaviors and outcomes. Students motivated by mastery goals are concerned with self-improvement and will generally persist in their efforts in the face of challenge (Anderman & Wolters, 2006; Wolters, 2004). Mastery goals have been associated with increased effort, intrinsic interest, and deep cognitive processing in students across all grade levels and subject areas (Harackiewicz et al., 2002; Hidi & Harackiewicz, 2000; Midgley, Kaplan, & Middleton, 2001; Wolters, 2004). Given the positive academic behaviors associated with mastery goals, an association with achievement would be expected. However, few studies have shown mastery goals to be positively associated with student achievement (Harackiewicz et al, 2002; Pintrich, 2000). The reasons for the lack of association are not clear, but it may be a reflection of the ways in which achievement is commonly measured. For example, assessments that rely on lower-level thinking skills such as recall and multiple-choice selections may not accurately gage the deep cognitive processing and long-term learning of students pursuing mastery goals (Meece, et al., 2006). More research is needed to examine the mismatch between measures of achievement and mastery goal orientations to better understand the reasons for the apparent contradiction in anticipated and actual empirical findings.
Performance Goals. In contrast to mastery goals, students motivated by performance-approach or performance-avoid goals are concerned with projecting an image of and being recognized for high ability with little effort (Ames, 1992; Blumenfeld, 1992; Middleton & Midgley, 1997; Midgley et al., 2001). Performance goal orientations have generally been considered less adaptive than mastery goals (Ames, 1992; Dweck & Leggett, 1988; Middleton & Midgley, 1997), although some recent research incorporating the approach-avoid distinction shows a more complex range of outcomes (Elliot, 2005). In general, performance goal oriented students may not engage in tasks as deeply as those students with mastery orientations, as they focus more on the outcome (success or failure) rather than the task itself (Hulleman et al., 2008).

Performance-avoid goals in particular have been linked with a number of maladaptive outcomes, including negative associations with academic performance, high anxiety levels, poor study habits, help-avoidance, and lack of interest (Elliot & Church, 1997; Midgley & Urdan, 2001; Wolters, 2004). The research on the negative outcomes of performance-approach goals is less consistent, and in fact, these goals do tend to be associated positively with student achievement and academic interest for some students (Darnon, Butera, Mugny, & Hulleman, 2009; Harackiewicz, et al., 2002; Hulleman, et al., 2008; Senko, Hulleman, & Harackiewicz, 2011). Inconsistency in results may be due to methodological issues in the research. For example, Hulleman and his colleagues (2010) conducted a meta-analysis of 98 studies on performance goals and found that different measures of goals were often measuring slightly different constructs. In particular, there was a difference in studies that defined performance orientations as being focused on demonstrating competence rather than being focused on social comparisons and
outperforming others. Relationships between performance-approach orientation and achievement varied depending on whether measurements reflected the former or latter definition of a performance orientation.

Despite the mixed evidence and methodological concerns, findings about potential positive outcomes associated with performance-approach goals have led to the endorsement of multiple goal perspectives, claiming that combinations of mastery and performance-approach goals may be most adaptive for student achievement (e.g. Senko, Hulleman, & Harackiewicz, 2011). However, most of this research has been done in college settings rather than elementary schools. Developmental considerations for the positive associations with performance-approach goals and student achievement need further research, especially regarding implications for the topic of assessment and standardized testing.

**Classroom goal structures.** In addition to the personal achievement goal orientations that students display, it is also important to consider the goal structures conveyed by the classroom climate. Research shows that teacher practices and features of the classroom environment that emphasize certain achievement goal structures can influence the personal achievement goals that students pursue (Anderman & Maehr, 1994; Blumenfeld 1992; Kaplan, Middleton, Urdan, & Midgley, 2002; Meece, et al., 2006). When students perceive their classroom as emphasizing effort and improvement, characterized as a mastery goal structure, they are more likely to adopt personal mastery goal orientations. When students perceive an emphasis on demonstrating ability and competition, characterized as a performance goal structure, they are more likely to pursue performance goals. Empirical evidence supports this link.
For example, Anderman and Young (1994) showed that students’ endorsement of performance goal orientations was related to teachers’ performance-oriented instructional strategies, such as using the work of the best students as examples for others, in science classrooms. In another study, Turner and her colleagues (2002) found that students were less likely to use avoidance strategies in math classrooms perceived as fostering mastery goal structures than in those perceived as having performance goal structures. Therefore, perceptions of classroom assessment events and performance feedback are likely to have an impact on students’ perceptions of the goal structures communicated by the teacher and subsequent personal goal orientations (Anderman & Maehr, 1994). This study contributes to the research on achievement goal theory by investigating the ways that teachers talk about and prepare students for both classroom and standardized assessments, and potential influences on students’ perceptions of classroom goal structures related to those contexts.

**Expectancy-Value Theory**

Expectancy-value models have been prevalent in psychology and achievement motivation research for many decades. Considered the first expectancy-value theory of achievement motivation, Atkinson’s (1957, 1964) model postulated that achievement behaviors were the result of motives, expectancies, and incentive values (as cited in Wigfield, Tonks, & Lutz Klauda, 2009). Modern expectancy-value models have built on Atkinson’s theory to provide more rich descriptions of the constructs of expectancies and values and their connections to achievement and motivation-related outcomes (Eccles, et al., 1983; Pekrun, 2000; Wigfield & Eccles, 1992; Wigfield, Tonks, & Lutz Klauda, 2009).
The most prevalent expectancy-value model currently used in the field of educational research is that proposed by Eccles and her colleagues (e.g. Eccles et al., 1983, Wigfield & Eccles, 1992). In their model, Eccles and her colleagues proposed that expectancies and values are influenced by many different psychological and situational factors, such as perceptions of competence, task difficulty, previous experiences, and perceptions of social and cultural influences. In turn, expectancies and values are assumed to predict academic choices, persistence, and performance (Wigfield, et al., 2009).

According to expectancy-value theory, the value that students place on tasks, including assessments, is especially relevant to motivation, because value is related to the incentives or reasons for engaging in a task (Eccles & Wigfield, 2002). There are four dimensions of task value, as defined by Eccles and colleagues (see Eccles & Wigfield, 2002). These dimensions are attainment value (importance to the individual of doing well at the task), intrinsic value (interest or enjoyment derived from the task), utility value (usefulness in allowing one to accomplish something else), and cost (potential negative impacts of the task, such as needed effort, anxiety induced by task, or loss of other opportunities that result from choosing one task over others). These components of task values have been associated with a number of motivational outcomes, including effort, persistence, interest, choice of activities, and course enrollment decisions (Hulleman, Durik, Schweigert, & Harackiewicz, 2008).

In addition to subjective task values, students’ expectancies for success on a task also contribute to their motivation. Expectancy beliefs are conceptualized as future-oriented perceptions of the individual’s capacity for success in a particular domain or task
(Wigfield & Eccles, 2000). Expectancy-value theory states that students choose and persist in activities for which they believe they will do well (Wigfield & Eccles, 2000). Expectancy beliefs are also positively related to performance outcomes, as students tend to do well in tasks in which they expect to succeed (Eccles et al., 1989; Wigfield & Eccles, 2000).

Expectancy-value theory has been used widely in educational research in a variety of grade levels and academic subject areas (see Wigfield, et al., 2009). For example, the Eccles et al. model has been used as a framework for understanding the decline in motivation as students progress through school, especially during the transition from elementary to middle schools (Eccles & Midgley, 1989; Eccles, Wigfield, & Midgley, 1993; Wigfield, et al., 1997). Students’ expectancies for success and valuing of certain academic tasks tend to decrease as they enter adolescence. While there may be many reasons for these declines, research suggests that students become more realistic in their perceptions of their own competence as they mature (Stipek & MacIver, 1989) and that the academic and social environment of middle schools is not compatible with the developmental needs of early adolescents (Eccles & Midgley, 1989). This research is important to the current study because it suggests that features of the classroom environment, including teachers’ instructional practices, may influence students’ expectancy and value perceptions. Additional research supporting the link between specific teacher practices and students’ expectancy beliefs and subjective task values is lacking.

In sum, expectancy-value theories of motivation posit that optimal student performance and motivation occur when students are presented with tasks that they
perceive as valuable and for which they have high expectancies for success. This study examined if and how teachers foster these beliefs in students regarding assessments and standardized testing tasks.

The frameworks of achievement goal theory and expectancy-value theory are used in this study to help understand how teachers’ instructional practices and underlying knowledge and beliefs about assessment and standardized testing influence students’ motivation in these two contexts. The following section will review research on teachers’ instructional practices in these two areas and specific links between practices and student motivation, based on the frameworks.

**Teachers’ Assessment Practices**

Assessment has always been important in education, as effective instruction depends on knowing what students know and how they are progressing in their learning (Shepard, et al., 2005). As noted earlier, two main categories of assessment are classroom assessment and standardized testing. This study addresses both, comparing the ways they are carried out in third grade elementary mathematics classrooms. The following sections will present a review of research on teachers’ practices related to classroom assessment and standardized testing. Also, research that specifically links these practices with student motivational perceptions will be reviewed.

**Teachers’ Classroom Assessment Practices**

Research on classroom assessment practices includes studies related to how teachers create, administer, grade, and communicate about formative and summative classroom assessments. In short, this research can include all assessment related
classroom activities, aside from standardized testing, which will be examined in the next section.

In one study, Stiggins and Bridgford (1985) surveyed 228 teachers from second to eleventh grades in a variety of subjects about their day-to-day assessment practices. They found that math and science teachers tended to put more emphasis on paper-and-pencil tests than teachers of other subjects such as speaking and writing. As grade level increased, teachers generally used more self-created objective tests as opposed to published assessments and they expressed more overall concern about the assessment process and specific factors such as quality control and accuracy in judging students’ progress. The authors speculated that the increase in concern about assessment at higher grade levels may be due to the increased importance on grades and student performance as students move through the school system and the impact of grades on future decisions. They concluded that although the teachers in their study expressed concern over assessment and the desire to improve, they also indicated that they lacked the time, opportunities, or means to revise their assessment practices (Stiggins & Bridgeford, 1985). More recently, Zhang and Burry-Stock (2003) investigated classroom assessment practices and teachers’ self-perceived assessment skills and found remarkably similar results. Their 67-item Assessment Practices Inventory (API) of 297 K-12 teacher participants revealed that teachers use objective tests more and show more concern for quality of assessment tools as grade level increases (Zhang & Burry-Stock, 2003). Similarly, a study of assessments and grading in elementary schools showed that grades on homework, extra credit, and constructed-response assessments were more important in higher elementary grades than in lower grades, where effort, improvement, and
behavioral factors tended to be considered more (McMillan, Myran, & Workman, 2002). Taken together, these findings suggest that teachers do not utilize objective assessments to a large extent and do not consider these types of summative paper-and-pencil assessments to be very important in early elementary grade levels, until at least fourth or fifth grade. These findings may have implications for the introduction of standardized testing in early elementary classrooms, as the process and importance of the testing may be very different from what the students typically experience in terms of classroom assessments. As noted earlier, current federal policy mandates standardized testing beginning in third grade (NCLB, 2002). Current research is needed to examine if and how the introduction of testing at this early grade level has changed teachers’ regular classroom assessment practices in order to prepare students for the heightened salience of objective, multiple choice tests as used in standardized testing.

Other research on classroom assessment has looked at individual variation in teachers’ instructional practices. Research has shown high use and variety of both formative and summative assessment methods at all levels (Black et al., 2004; Volante, et al., 2010). In addition, there are significant individual differences among teachers’ assessment and grading practices. Across all grade levels and subject areas, individual variation in selection, administration, and evaluation of assessments has been found and differences in assessment practices tend to vary more between individual teachers than between schools (Brookhart, 1994; Frary, Cross, & Weber, 1993; Shepard et al., 2005; Stiggins & Conklin, 1992). This is evidence that individual teachers largely determine their own classroom assessment practices, with personal preferences and professional judgments contributing to their decisions. While it is established that teachers have
autonomy to make decisions about classroom assessments and that general patterns of assessment practices exist, as described in the studies above, research depicting what the actual process of assessment looks like in the classroom is lacking. Therefore, this study will contribute to research on assessment practices by investigating teachers’ instructional practices related to assessment through classroom observations.

**Teachers’ Standardized Testing Practices**

In contrast to classroom assessment, teachers do not have many choices about whether or how to conduct state-mandated standardized achievement testing. Nevertheless, their instructional practices related to how they prepare students for the tests are likely to be more individually varied and influenced by teachers’ knowledge and beliefs about testing as well as contextual variables of their schools and districts (Dooley & Assaf, 2009; Mathison & Freeman, 2003). The links between knowledge, beliefs, and practices will be explored later. This section will review research on instructional practices related to testing.

Much research has established that the implementation of standardized testing policies results in narrowing of curriculum and increases in test-focused teaching practices, often considered “teaching to the test” (Au, 2007; Cimbricz, 2002; Nichols & Berliner, 2007; Smith, 1991; Wood, 2004). When teachers and administrators feel pressured to ensure that students achieve high scores on specific tests, the subjects of those tests become the basis of the curriculum taught. Other subjects, like social studies or the arts, that are not typically included in state testing are squeezed out of the curriculum to enable teachers to spend more time on math and reading instruction (Au, 2007). In a qualitative metasynthesis of 49 studies investigating the result of high stakes
testing on curriculum, Au found that in the large majority of cases, the implementation of testing policies did lead to narrowing of curriculum to tested subjects. Moreover, content was increasingly taught only within the context of the tests themselves (i.e. multiple-choice teaching strategies) and with greater teacher-centered rather than student-centered teaching methods.

The emphasis on increasing teaching time for tested subjects has also led to the loss of non-academic activity time in schools. When administrators and teachers feel the pressure to find more time in the day for instruction, they may resort to taking time from elective classes, eliminating assemblies, field trips, and even recess. For example, Nichols and Berliner (2007) report that elementary schools in Waltham, Massachusetts have eliminated recess on all but one day per week and have shortened lunch periods to less than 15 minutes on many days in order to make time for testing and to meet NCLB requirements. In 2004, in Clark County, Nevada, which includes the city of Las Vegas, the school district also decided to eliminate recess in the elementary schools in order to handle the pressures of NCLB (Richmond, 2004). Similar actions are being taken all across the country with ramifications for students including decreased motivation and enjoyment of school, fewer opportunities to cultivate personal interests, and even rising rates of childhood obesity due to inactivity (Winter, 2009). Although these practices may result in students scoring higher on tests, it is widely argued that overall student learning and well-being suffers in the process of focusing on preparing for standardized tests (Kohn, 2000; Nichols & Berliner, 2007; Ravitch, 2010; Wood, 2004).

While the general consequences of narrowing curriculum and increased teaching to the test at the school level are well-established, less research has examined individual
teachers’ classroom behaviors related to these issues. There are several factors that might contribute to variability between teachers in practices surrounding preparing students for standardized testing. A cross-case analysis of two teachers in Texas revealed that pressures to increase test scores had a stronger impact on teacher practices in an urban school than a higher-income suburban school (Dooley & Assaf, 2009). While teachers in both school contexts expressed negative attitudes towards the demands of testing policies, the urban school teacher changed her practices much more drastically as a result of testing pressures, engaging in far more teaching to the test. For example, she used practice test passages rather than authentic texts for reading instruction and she spent more time on test-taking strategies than on deciphering the meaning of texts. This reflects that in school contexts in which students are largely minorities or come from backgrounds with few financial resources, schools are more likely to show achievement gaps and be subject to sanctions due to low performance on state tests (McCaslin, 2006; Nichols & Berliner, 2007; Novak & Fuller, 2003; Ravitch, 2010). Therefore, teachers may feel more pressure and, in turn, engage in more instructional behaviors such as narrowing curriculum and teaching to the test in higher poverty schools than in more affluent schools. More empirical research showing how teachers respond to testing pressure in different socio-economic contexts is needed to support these conclusions.

Level of teaching experience is another factor that influences how teachers respond to testing pressure. Teachers with more experience tend to exhibit less concern over the danger of being reprimanded or dismissed due to testing results. In a year-long study of fourth-grade teachers in two New York schools, Mathison and Freeman (2003) showed that some experienced teachers were willing to exercise their own professional
judgment, “almost with an air of defiance” (11), and make their own decisions about curriculum and instructional practices even against the district’s mandates about test preparation. Similarly, Bolden and Newton (2008) interviewed teachers and found that more experienced elementary school teachers worried less than new teachers about justifying their instructional decisions and accounting for student activities in a highly monitored and restrictive school setting.

**Summary of Teachers’ Assessment Practices**

In summary, research on classroom assessment practices reveals that teachers at all grade levels and across subject areas tend to utilize a wide variety of assessment tools and practices in their classrooms. Patterns of use indicate that self-made objective tests are increasingly used and summative assessments have more importance in higher grades levels as compared to lower grade levels. In contrast, early elementary grade teachers show less concern over the assessment process and tend to utilize summative paper-and-pencil tests less frequently. However, assessment practices tend to vary greatly across individual teachers, as teachers generally retain a great deal of autonomy over classroom assessment activities.

In contrast, research on standardized testing practices shows that teachers generally engage in narrowing the curriculum and teaching to the test in order to prepare students for standardized tests. Individual variability in test preparation practices depends on several factors, including the context of the school, experience level of the teacher, and knowledge and beliefs related to testing (discussed later). The next section will review research on assessment and standardized testing practices specifically connected
to the motivational frameworks of this study: achievement goal theory and expectancy-value theory.

**Assessment Practices and Student Motivation**

Motivation researchers have suggested that theoretical links exist between teachers’ assessment practices and student motivational outcomes (i.e. Anderman, et al., 2010; Ryan & Brown, 2005; Wood, 2004). However, empirical evidence linking specific assessment practices with students’ motivational perceptions, such as goal orientations, subjective task values and expectations for success, is limited. In one study, Brookhart and Bronowicz (2003) interviewed 161 students of different grade levels about their perceptions of specific classroom assessments. Their analysis suggested developmental differences in students’ perceptions of intrinsic and importance values of assessment tasks. In elementary schools, students relied on their perceptions of their own abilities to determine how interested they were in tasks and thus how valuable the tasks were to them. That is, students expressed more interest in an assessment when they expected to do well on the task, and those were the tasks they characterized as being most valuable. In contrast, high school students tended to emphasize their perception of the importance of the task in gauging the value. They were able to make more sophisticated judgments than the younger students about how important a task was in terms of factoring into their course grades or enabling them to succeed on other tasks in the future. High school students also thought that assignments were important if they found them to be personally intrinsically interesting. These results suggest that as students progress through school, they begin to value assessment tasks less based on their own abilities and expectations and more based on their perceptions of subjective task interest or importance. These
findings would be strengthened by evidence of how students come to perceive tasks as being interesting and of importance to them, and in particular, how teachers contribute to students’ perceptions. Further research may explore how teachers communicate these messages to students in different grade levels.

Another study by Anderman and his colleagues (2001) used hierarchical linear modeling to explore elementary and middle school students’ valuing of mathematics and reading as related to their perceptions of classroom goal structures over the course of a year. They found that valuing of math and reading were lower in classes in which the teacher reported emphasizing performance-oriented instructional practices, such as working for top grades, achieving higher test scores, and knowing who is doing the best and striving to do as well, as compared to mastery-oriented classrooms. These performance-oriented teacher assessment practices are therefore maladaptive for student motivation because they lead students to decrease valuing of learning those subjects. Interestingly, the authors found that gender was unrelated to changes in valuing of math and reading although prior research suggests that there are significant gender differences that emerge over time in the value of these subjects, with females tending to value English and reading more and mathematics less than males (Eccles, Wigfield, Flanagan, Miller, Reuman, & Yee, 1989). Gender may have been insignificant in this study because the students were too young for value differences to be prominent or because, as the authors suggest, instructional practices are more influential in changing value beliefs than the effects of gender (Anderman et al., 2001).

Feedback practices also have an impact on student valuing of academic tasks. According to the research literature on teacher feedback, the nature and informational
quality of written and verbal feedback are more important than grades or generic praise in
motivating students and increasing achievement (Black & Wiliam, 1998; Butler & Nisan,
1986; Hattie & Timperley, 2007; Narciss, 1999). Feedback is most effective when given
without evaluative grades, as students tend to ignore comments if grades are also given
(Black et al., 2004). An experimental study of 200 fifth- and sixth-grade students by
Butler (1987) supports these findings and links feedback with valuing of academic
subjects. Students were given no feedback or feedback consisting of standardized praise,
umerical scores, or individual comments after performing interesting divergent thinking
tasks. Students receiving only individual comments as feedback found divergent thinking
tasks to be more interesting, or intrinsically valuable, than students in the other condition
groups. In a similar study, Elawar and Corno (1985) found that achievement, enjoyment,
and attitudes towards mathematics increased for students receiving informational
feedback rather than information on correctness of their work only. In the experimental
study, informational feedback included at least one positive remark about something done
correctly and then specific comments on errors with suggestions on how to improve.
Students in the treatment group performed significantly better on a post-test and reported
significantly more positive attitudes about math and about school in general, higher self-
concept, and lower test anxiety than students in the control group. This evidence suggests
that teachers can influence students’ valuing of assessment tasks, and towards subjects
and even school in general, by providing effective informational feedback on their
performance.

In addition to traditional classroom assessment events, task value is also an
important consideration with regard to standardized testing. High-stakes tests are
designed to be motivating to students because of the value of consequences attached to them, such as passing a course or graduating from high school. In contrast, low-stakes tests do not have direct consequences for students, although they often have consequences for teachers, schools, and/or districts (Nichols & Berliner, 2007; Ryan & Brown, 2005). Particularly in low-stakes testing conditions, students are apt to ascribe low personal value to the exams and thus are less likely to exert much effort on the task (Wise & DeMars, 2005). This is troubling because the combination of low value and low effort is likely to diminish the validity of the test results, because scores are not an accurate reflection of what students are capable of if they are not trying their best to demonstrate their ability (Nichols & Berliner, 2007; Paris, Lawton, Turner, & Roth, 1991; Wise & DeMars, 2005). Therefore, understanding students’ perceptions of the value of standardized testing is important to increasing motivation and maintaining validity of such tests.

To explore this issue, Ryan and her colleagues (Ryan, Ryan, Arbuthnot, & Samuels, 2007) conducted interviews with eighth-grade students about their motivation for standardized mathematics exams. They found that students expressed a variety of reasons for valuing math and math exams, including both intrinsic interest, and importance or utility value. For example, some students intrinsically valued math because it was their favorite subject or they just loved math. Others valued math, and/or the math exam, for its importance to them for future careers or to do well in school to graduate or be eligible for sports. However, the authors did not provide any evidence of students valuing the test itself for intrinsic reasons, nor did they address value for low-stakes exams. In particular, the cases of students intrinsically valuing the subject but not the...
testing poses an interesting pattern for future research to consider. It may be that the students see the standardized testing process as being very different from their regular experiences in that subject and therefore have different perceptions of the subjective value of the tasks.

In another investigation of students’ perceptions of value, Paris and his colleagues (2000) conducted a series of studies examining students’ perceptions of standardized tests and found significant developmental differences. Survey data from students in grades 2-11 revealed that younger students tended to consider tests useful whereas older students were more skeptical about the value of tests and therefore were less likely to give their full effort. In addition to age differences, the authors also found that achievement level is associated with students’ perceptions. Lower-achieving students reported more disillusionment with the value of testing than high-achieving students (Paris, Roth, & Turner, 2000). This is especially troubling considering current policies of implementing testing to try to increase achievement, especially for low-performing students (Nicholls & Berliner, 2007). It is unlikely that testing will be effective for those most in need of improvement if they are not motivated by the value of the task.

In summary, theoretical associations between perceptions of the value of assessment tasks and student motivation and subsequent achievement have been supported through empirical studies with students at elementary and middle-school levels. More research is needed to explore how students form beliefs about the value of assessment and standardized testing and the teacher’s role in communicating such messages to the students. For both assessment contexts (classroom and standardized testing), in-depth, classroom-based examinations of individual teacher’s instructional
practices are lacking in the literature. There is a need for more research that shows how teachers’ instructional methods and communication about assessment vary with regard to classroom and standardized testing to help understand how motivational consequences of testing policies emerge. The goal of this study is to investigate the specific differences in individual teachers’ practices between the two contexts, as well as connections between these practices and teachers’ underlying knowledge and beliefs. The next section will review the research about teachers’ knowledge and beliefs related to assessment and testing.

**Teacher Assessment Knowledge and Beliefs**

The findings that teachers’ classroom assessment and standardized testing preparation practices are highly individually varied (Brookhart, 1994; Dooley & Assaf, 2009) suggest that many factors may contribute to teachers’ differing practices. Among them, the influence of teachers underlying knowledge and beliefs about assessment and testing may play an important role in the manifestation of assessment and testing practice in classrooms.

The knowledge and beliefs of teachers has been an area of interest to educational researchers for many decades (Woolfolk Hoy, Davis, & Pape, 2006). What teachers know and believe about educational topics, such as assessment, strongly influences their instructional practices, which, in turn, have implications for student learning and achievement (Brown, 2004; Pajares, 1992).

In the research literature, distinguishing between knowledge and belief can be difficult, as there is much overlap between the two constructs (Southerland, Sinatra, & Matthews, 2001; Woolfolk Hoy et al., 2006). Most people think that what they “believe”
they also “know” and vice versa, so it is not always easy to draw the line between knowledge and belief. Some researchers have made the distinction that knowledge is based on evidence of truth whereas beliefs can be held regardless of evidence (e.g. Richardson, 1996). Others have argued that beliefs are a type of knowledge (Nisbett & Ross, 1980) or that knowledge is a type of belief (Rokeach, 1968). Despite the philosophical and conceptual disagreements over the definitions of these terms, it is widely acknowledged that teachers’ knowledge and beliefs interact with one another to inform and influence their decision making and instructional practices in the classroom (Brown, 2004; Nisbett & Ross, 1980; Pajares, 1992; Southerland et al., 2001; Woolfolk Hoy et al., 2006).

Southerland and colleagues (2001) argued that while theoretical distinctions between the constructs of knowledge and beliefs certainly exist, they become “hopelessly blurred at the empirical level” (p. 348). In the absence of very clear definitions for these terms, it often makes more sense to treat them as overlapping constructs for the sake of empirical research on teachers’ knowledge and beliefs. Therefore, many studies do not explicitly differentiate between knowledge and beliefs when collecting and analyzing data (Southerland et al., 2001; Woolfolk Hoy et al., 2006). Following that precedent, in this study, teacher knowledge and beliefs will be treated as one complex construct, unless specifically noted otherwise. The following sections will review the research linking teachers’ knowledge and beliefs to instructional practices in the two assessment contexts: classroom assessment and standardized testing.

Teacher Knowledge and Beliefs about Classroom Assessment
It is widely argued in the assessment field that teachers do not receive adequate training in assessment and educational measurement (i.e., Cizek et al., 1995/1996; Impara, Plake, & Fager, 1993; Shepard et al, 2005; Volante et al, 2010). Some of the knowledge teachers should be gaining from pre-service and in-service training includes knowledge of formative and summative assessment methods, knowledge about reliability and validity of assessments, criterion vs. norm-referenced grading practices, and aligning assessments with learning objectives and state standards (Shepard et al, 2005). However, many pre-service and in-service teacher education programs do not even require coursework on assessment (Impara, Plake, & Fager, 1993). Therefore the amount and types of assessment knowledge that teachers hold vary greatly based on their preparation and experience.

Regardless of preparation and experience, however, research has shown that teachers’ assessment practices rarely conform to what measurement specialists would consider to be sound testing and grading practices (Bonner & Chen, 2009; Brookhart, 1994; Frary et al, 1993; Karp & Woods, 2008). This mismatch may be due to a lack of knowledge and/or training, or it may be the result of other factors, such as resource limitations, contextual variables of the school or classroom, or teachers’ beliefs and value judgments related to assessment. For example, Zhang and Burry-Stock (2003) showed that there were some differences in the assessment practices of teachers with and without specific training in educational measurement. Those with measurement training reported higher levels of self-perceived skills in a number of assessment-related activities, such as using performance measures, revising tests and instruction, and communicating assessment results. This would suggest that knowledge about assessment does inform
practice. Similarly, Bonner and Chen (2009) developed and administered a Survey of Assessment Beliefs (SAB) to 143 pre-service teachers to measure their perceptions about assessment and grading practices both at the beginning and at the end of a course on educational assessment. They found that measurement training had a significant impact on teachers’ perceptions. Specifically, after the assessment training course, teachers were less likely to endorse grading solely based on alternative assessments, using grades to motivate student effort, and using grades to manage student behavior (Bonner & Chen, 2009). However, this study only considered pre-service teachers’ perceptions and did not include examination of teachers’ actual assessment and grading practices.

In contrast, Brookhart (1993) found that measurement training made little difference in teachers’ grading practice. In assigning grades to students, teachers with measurement training thought differently about the meaning of grades in terms of their validity than their peers without training, but they did not think differently in terms of value implications and social consequences of grades. That is, they had similar responses to scenarios about the meaning of the assigned grade to the student and about what will happen to the student because of the grade. Hence, knowledge of assessment practices may not be sufficient to inform practice because beliefs about the function of assessment also play a role in determining action. Teachers may choose not to follow recommended assessment and grading practices because they are concerned about the social consequences that the grades have for students, even if they have sufficient knowledge about educational measurement (Brookhart, 1993).

Other research supports the role that values and beliefs play in determining student grades, in contrast to what measurement specialists would advise (Bonner &
Measurement principles state that grades should be determined by achievement on academic tasks and should not include factors such as effort or improvement as those are difficult to measure accurately and can cause confusion over the meaning of grades (Shepard et al., 2005). However, in addition to academic achievement, a “hodgepodge” of non-academic factors has been found to play a part in teachers’ determinations of grades, including effort, ability, completed or missing work, and improvement (Brookhart, 1993; 1994).

Frary, Cross, and Weber (1993) surveyed 536 secondary teachers in Virginia about their practices and opinions of classroom testing and grading. They found that tests and quizzes were the most dominant determinants of grades. However, large majorities of teachers agreed with the use of gain, ability, and effort in determining grades and 31% even agreed with including student behavior as a factor in grading. Similarly, McMillan (2001) surveyed 213 high school science teachers about their grading practices and found great variety in how grades were determined. Consistent with Frary et al. (1993), teachers did hold academic achievement as the most important determinant of grades, but they also included factors such as effort, ability, and improvement.

Through cases studies of interviews and observations of 15 high school teachers, Stiggins, Frisbie, and Griswold (1989) also determined that teachers’ practices differed from the recommendations of measurement principles. They found that teachers did believe that achievement should be the primary consideration in grading, but they also valued student motivation and effort in grading and many considered student ability level as a factor. About half of the teachers had different assessment and grading procedures for students with different levels of ability, such as considering only achievement in
determining grades for high ability students but factoring in both achievement and effort for lower ability students (Stiggins et al., 1989). The small sample size limits the generalizability of the findings, but the results are consistent with the previously reviewed studies and contribute to the understanding of the ambiguity of grading practices due to teachers’ knowledge and beliefs.

Classroom assessment practices are also influenced by teachers’ knowledge and beliefs about the general purposes of assessment and their views on learning. Brown (2004) found evidence for four main conceptions about assessment purposes: assessment improves teaching and learning, assessment makes students accountable for learning, schools and/or teachers are held accountable for learning through assessment, or that assessment is irrelevant to the life and work of teachers. Through surveys of 525 New Zealand primary school teachers, Brown used structural equation modeling (SEM) to show support for these four distinct conceptions of assessment held by teachers and for interrelations among them. In his sample, teachers tended to agree with the Improvement and School Accountability conceptions and disagreed with the Student Accountability and Irrelevance conceptions. There were several significant inter-correlations of conceptions as well, indicating that teachers’ conceptions of assessment are multi-dimensional. For example, teachers who believed that assessment is about School Accountability were highly likely to also hold conceptions of assessment as Student Accountability and Improvement. He argued that these four conceptions serve as filters through which teachers view their roles and make decisions about instruction and assessment. However, the study only considered teachers’ self-reported beliefs, so direct implications for instructional practice are not warranted. Also, it is important to note that
this study was conducted in the New Zealand context, which had no externally mandated standardized testing and accountability program, so generalizability to the United States context may be limited.

Another perspective on the general purposes of assessment is the distinction between the use of assessment as formative or summative. Formative assessment is assessment for learning that provides information for teachers and students to use as feedback and to help improve instruction and student learning (Black et al, 2004). Through interviews with 20 Canadian teachers on their views of conducting formative assessment, Volante and his colleagues (2010) found that most teachers highly valued formative assessment and believed that emphasis should be primarily on the learning process rather than evaluative grades. Unfortunately, even teachers who held this belief about the purpose of assessment expressed conflict with external pressures to use assessment for other purposes, such as ranking students or demonstrating accountability (Volante, Beckett, Reid, & Drake, 2010). Thus, the implication of their beliefs for actual practice is not clear. Further research is needed to explore the links between teachers’ espoused beliefs about the purposes of assessment and actual teaching practices, especially with regard to the external pressures for accountability.

In summary, the knowledge and beliefs that teachers hold clearly have implications for classroom assessment practices. Knowledge acquired from measurement training may impact instructional practices to some extent, but is certainly moderated by teachers’ beliefs and perceptions of external pressure related to assessment and grading. Beliefs about the purposes of assessment and grades may also influence instructional practices. However, empirical evidence for
these claims is lacking in the extant literature. This study will contribute to the research in this area by investigating whether teachers' practices related to classroom assessment reflect their knowledge and beliefs as revealed through observations and interviews.

**Teacher Knowledge and Beliefs about Standardized Testing**

Recent research examining teachers’ knowledge and beliefs about assessment practices has largely focused on high-stakes testing, due to the increased presence and controversy surrounding such tests (Cimbricz, 2002; Delandshere & Jones, 1999; Nichols & Berliner, 2007). Teachers need to have knowledge of federal and state accountability policies and standardized testing and research on the effects of such testing on students (Shepard, et al., 2005). Understanding issues of reliability and validity with respect to large-scale tests and knowing the impact of testing on students can help teachers prepare students appropriately, intellectually and emotionally, to take these tests. Teachers also need to understand how to interpret and communicate testing results and to use results to inform their instructional practices (Shepard, et al., 2005). Empirical research is needed in this area to understand how knowledge about tests impacts teachers’ instructional practices.

Subject matter knowledge may also have an important impact on teachers’ instructional practices. Research has shown that teachers who are less comfortable with knowledge of their content area may have trouble preparing students for standardized testing (Cimbricz, 2002; Wilburne & Long, 2010). For example, in a study of pre-service teachers’ content knowledge and confidence in teaching, Wilburne and Long (2010) showed that teachers with weak content knowledge in mathematics were less confident in
their abilities to prepare their students for state standardized testing. Teachers answered questions taken from state standardized tests and then rated their confidence both in their answers and in their ability to teach that particular question to their students. An alarming number of teachers (as many as 40% of the 70 participants, for some problems) reported that they could not solve the problems, and there was a strong correlation between ability to solve the problem and confidence to teach it to students. The authors concluded that pre-service math teachers need more content knowledge related to the types of math problems that their students will be asked to learn, in order to be able to prepare them for the standardized tests (Wilburne & Long, 2010). However, because the sample was comprised of pre-service teachers, this study did not look at the implications of a lack of knowledge and confidence for teachers’ actual instructional practices or subsequent student performance. Future research on the impact of teachers’ level of content knowledge for preparing students for standardized testing would be beneficial.

In spite of the importance of the types of knowledge discussed above, teachers’ internal beliefs and values have been cited as the most important influence on assessment and testing decisions (McMillan, 2003). However, the increase of high stakes standardized testing represents an area of tension between teachers’ values and external demands on their practices. The nature of the relationship between beliefs and practices, especially concerning state-mandated accountability testing, is complex and in need of further research (Cimbricz, 2002).

While many teachers do support the notion of holding schools and teachers accountable for student learning, the value of using large-scale standardized tests for this purpose is much more controversial (Cimbricz, 2002; Dooley & Assaf, 2009; Hill, 2004;
Leighton, Gokiert, Cor, & Heffernan, 2010; Smith, 1991). Leighton and colleagues (2010) surveyed 265 teachers in Canadian secondary schools and found that teachers believed that large-scale tests were less useful than classroom assessments for providing meaningful information about student learning and achievement. Likewise, a survey of 1,280 fourth through eleventh grade teachers in Missouri revealed that teachers assumed responsibility for student achievement and were willing to be held accountable, but they did not believe that large-scale standardized tests appropriately measure student learning and progress for this purpose (Bryant et al., 2009). The teachers cited reasons such as student background factors, including socioeconomic status, school attendance, cognitive ability, and parental support, as well as lack of student effort and test-taking skills. In addition, they expressed the belief that a single test was not sufficient to demonstrate student progress throughout the year or across multiple school years. Teachers in the study reported considerable pressure, stress, and frustration about the testing process that was characterized as hostility towards the testing policy (Bryant et al., 2009).

Similar findings were reported based on classroom observations and interviews with elementary teachers in Arizona (Smith, 1991). This study revealed that teachers experienced negative emotions due to the emphasis on test scores, especially because they believed the tests were invalid or incomplete measures of students’ educational achievement. The teachers resented the demands they felt from administrators and parents to raise students’ test scores when the teachers believed the scores were more or less worthless as indicators of achievement. As a result, teachers reported feelings of dissonance and alienation as well as anxiety because they felt they could not directly control how their students would do on the tests (Smith, 1991).
Another common belief expressed by teachers is that standardized testing diminishes their professional identity and autonomy as teachers. This belief is frequently linked to the issue of pressure to narrow the curriculum and teach to the tests. For example, Bolden and Newton (2008) investigated three elementary school teachers’ beliefs through interviews and observations and found that they believed that standardized tests were a barrier to their professional identity because they prevented them from teaching in their preferred ways. Specifically, they felt that they had to focus too much on coverage of content needed on the test and could not allow their students to take an investigative or constructivist approach to their learning. Likewise, Mathison and Freeman (2003) reported that teachers in New York were losing autonomy over choosing textbooks and making other curricular decisions due to pressure to teach to the test. These teachers expressed frustration at having to gear their instructional practices towards lower-level skills tested in multiple-choice state tests at the expense of teaching the content in more depth. These studies show that often, teachers feel torn between what they feel is best for their students’ learning and what will help them pass the tests (Assaf, 2008).

Teachers in Smith’s (1991) study also expressed concern over the impact of testing on students’ emotional and mental well-being. Teachers reported feelings of anxiety and guilt when watching their students complete standardized testing because they believed that it would have a negative emotional impact on the students. These teachers frequently claimed that they considered the tests to be “cruel and unusual punishment” for young elementary school students (Smith, 1991).
These beliefs may be particularly salient for teachers working with under-performing student populations. For example, a cross-case analysis of fourth-grade teachers in Texas revealed that the teacher in an under-achieving urban school context reported more feelings of concern over her students’ reactions to testing pressure than the teacher in a suburban context (Dooley & Assaf, 2009). Although both teachers expressed beliefs that multiple-choice tests are not a valid measure of students’ ability, the teacher in the urban school context felt more pressure to avoid sanctions and engaged in more instructional behaviors such as teaching to the test. She expressed frustration about the situation and concern over the impact that the emphasis on testing would have for her students. Although it was only a case study of two teachers, Dooley and Assaf demonstrated that perceptions of external pressure as well as teachers’ knowledge and beliefs about testing influenced instructional practices related to standardized testing.

In summary, various types of knowledge and beliefs influence teachers’ instructional practices related to preparing students for standardized testing, including knowledge and beliefs about the tests and testing policies, subject matter, and consequences of testing policies for teachers and students. Teachers’ views on effective teaching and learning are often in conflict with external pressures of testing policies. This study explores the potential mismatch that this influence creates between teachers’ knowledge and beliefs and their actual classroom practices.

**The Present Study**

Given all of the research reviewed in this chapter, the present study is guided by three main research questions. The questions and expectations of findings are provided below.
**Question One**

*How do elementary school math teachers’ instructional practices related to assessment differ for preparation for summative classroom assessments and preparation for standardized state achievement tests?*

Based on the research reviewed in this chapter, it is expected that there will be notable differences in the instructional practices of teachers between the two assessment contexts. The research suggests that teachers might communicate differently about the task value of classroom assessment activities as opposed to standardized testing and that teacher practices may align more with those known to foster a performance goal structure in the standardized testing context than in the classroom assessment context (Anderman et al., 2010; Kowalski & Marrietta-Brown, 2011).

**Question Two**

*How do elementary school math teachers’ instructional practices reflect their knowledge, beliefs, and perceptions of external influences related to summative classroom assessments and standardized state achievement tests?*

Research indicates that teacher knowledge and beliefs are important influences on their instructional practices (Brown, 2004; Pajares, 1992; Woolfolk, Davis, & Pape, 2006). However, external accountability policies create pressure to demonstrate student achievement that may cause conflict between what teachers know and believe to be effective teaching and what they feel they have to do in their own classrooms in response to the policies (Assaf, 2008; Cimbricz, 2002). Therefore, it is expected that teachers’ instructional practices will reflect their reported knowledge and beliefs about the types of testing, as well as their perceptions of external pressure related to testing policies.
**Question Three**

*How do students perceive the classroom motivational climate surrounding assessment and standardized testing?*

Because teacher practices are known to be associated with student motivation (Ames, 1992; Brophy, 2004), it is hypothesized that students will perceive the motivational climates of their classrooms differently in the two contexts, to the extent that teacher practices differ. Therefore, results are expected to align with the results from the first research question.
Chapter 3: Methods

This study examined the influence of teachers’ instructional practices on students’ motivation in different assessment contexts. In this chapter, a description of the participants and study procedures is provided. The methodology for investigating the research questions is described, including the instruments and measures used and the analysis approach.

The present study addressed the following research questions:

1. How do elementary school math teachers’ instructional practices related to assessment differ for preparation for summative classroom assessments and preparation for standardized state achievement tests?

2. How do elementary school math teachers’ instructional practices reflect their knowledge, beliefs, and perceptions of external influences related to summative classroom assessments and standardized state achievement tests?

3. How do students perceive the classroom motivational climate surrounding assessment and standardized testing?

In order to investigate these questions, the study utilized a mixed methods design. Mixed methods refers to research that incorporates both qualitative and quantitative approaches in order to broaden understanding and strengthen the research to provide more results than either qualitative or quantitative research alone (Creswell, 2009).
Mixed methods approaches have become more widely used in social science research since Campbell and Fiske introduced the term “multimethod” in 1959 to refer to the strategy of mixing different modes of data (Tashakkori & Teddlie, 2003). There are many different mixed methods designs incorporating different ways of using various forms of data, from seeking convergence across different data sources (known as data triangulation) to using one method to identify participants for research using another method, to merging data from multiple sources into a single, integrated analysis (Creswell, 2009; Tashakkori & Teddle, 2003). Different forms of data can be collected simultaneously (concurrent design) or one form can be collected at a later time in order to expand or elaborate on the findings of the first form (sequential design). The forms of data can be weighted equally in terms of their importance in the findings or one form can be considered primary, with the secondary form of data being embedded within the study to provide support for or answer a different research question from the predominant form of data (Creswell, 2009).

In this study, both quantitative and qualitative data were collected at the same time and the analysis integrated the information to interpret the overall results in what is known as a concurrent embedded strategy design (Creswell, 2009). Qualitative inquiry was primary in this study, comprised of classroom observations and interviews with teachers, and quantitative surveys of student perceptions were embedded to explore the problem from multiple perspectives and address each research question. More details about the study design can be found under the procedures section of this chapter.
This dissertation was completed with the approval of the Institutional Review Board of the Ohio State University. IRB approval documentation can be found in Appendix A.

Participants

Schools. The focus of this study was third grade classrooms in public elementary schools. Third grade was chosen because it is the first year in which students take state standardized tests, as mandated by federal policy (NCLB, 2002). Contextual variability was intentionally limited by confining the investigation to two demographically similar schools within a single school district in the state of Ohio. The district, Bellehaven City Schools (all names of places and people are pseudonyms), was chosen because personal experience in the district helped to facilitate access to the schools. In addition, the district is known for relatively high performance on standardized tests. This record of high test performance was desirable for this study because it indicated that the schools are successful in preparing students academically for the tests. The motivational impact of this preparation was the focus of this study.

According to the district website, Bellehaven City Schools serves approximately 15,000 students. The district includes 16 elementary schools, four middle schools, and three high schools enrolling students from a 52-square-mile area. Students are culturally and economically diverse, with 64% White, 21% Black, 5% Hispanic, 3% Asian, 7% Multiracial or Other, and 22% classified as economically disadvantaged for the 2011-2012 school year. In addition, 13% of students are identified as students with disabilities and 8% as having limited English proficiency, with approximately 1200 enrolled in English as a Second Language programming. Average per-pupil expenditure for the year
2011-2012 was $10,890, the graduation rate was 93%, and more than 80% of graduates each year pursue post-secondary education at colleges or universities. Bellehaven City School district achieved a rating of “Excellence with Distinction” from the State Department of Education on the 2010-2011 State Report Card, a designation that reflects its performance on state assessments as well as attendance and graduation rates.

Principals of all sixteen elementary schools were contacted by email in early fall of 2011 asking them to consider allowing research to be conducted in their schools (see Appendix B for a copy of the communication). Four principals agreed to forward the invitation email to their third grade teachers. Teachers in one of the schools never responded to the email nor to a reminder sent two weeks later. One teacher responded from the second school, but she was considered ineligible because she was going to be out on maternity leave during the spring and would therefore not be able to participate for the duration of the research. Positive responses were received from all third grade teachers from the other two schools, three from Martin Grove and two from Fairview. These two schools are located on the same road only a mile and a half away from each other. The proximity of the schools within similar neighborhoods provided a desirable degree of homogeneity in terms of demographic and contextual variables. Therefore, the two schools were selected as the research sites and no further attempts were made to recruit additional participants.

State Report Cards were accessed from the Ohio Department of Education website (http://www.ode.state.oh.us) to provide more detailed performance information about the participating schools. The State Report Cards are issued annually for each school and include students’ performance data based on state standardized test scores,
rates of improvement in student scores, attendance information, and graduation rates.

More specifically, there are eight state performance indicators that are assessed for each school: students’ state standardized test scores in third, fourth, and fifth grade mathematics and reading, state standardized test scores in fifth grade science, and a measure of average daily attendance. To meet a state indicator, at least 75% of the students must have scored proficient on the test corresponding to the indicator. For the student attendance measure, 93% is the required average daily attendance to meet that indicator.

Rates of improvement are analyzed based on the school’s progress toward Adequate Yearly Progress (AYP) for each subgroup of students, according to federal requirements of NCLB (NCLB, 2002). AYP goals are set by the state and applied to ten student groups: all students, economically disadvantaged, Asian, Black, American Indian, Hispanic, Multi-Racial, White, students with disabilities, and students with limited English proficiency. If any of the groups does not meet AYP in reading proficiency, math proficiency, or attendance, then the school overall fails to meet AYP and federal and state consequences occur after consecutive years of not meeting AYP (NCLB, 2002). For the state report cards, schools are assigned one of six designations based on state performance indicators met and AYP status: Academic Emergency, Academic Watch, Continuous Improvement, Effective, Excellent, or Excellent with Distinction.

Martin Grove Elementary School earned a designation of Excellent on the State Report Card for the 2009-2010 school year, having met seven of the eight state performance indicators. Martin Grove met the attendance indicator with 96% and met six of the seven test indicators. The only indicator not met in 2009-2010 was fifth grade
science, on which 74.7% scored proficient, just 0.3% away from the 75% cut-off. The school also failed to meet Adequate Yearly Progress (AYP) designation for the third year in a row. This determination was based on the scores of the Limited English Proficient students on the Reading assessment. All other subgroups made AYP in all other areas.

In 2009-2010, the Martin Grove student population was 68% White, 18% Black, 4% Asian, 4% Hispanic, and 6% Other or Multi-Racial. The percentage of students classified as economically disadvantaged was 23%, and 14% of students were identified as students with disabilities. In addition, 12% of students at this school were identified as Limited English Proficient. All teachers held appropriate certification and Bachelor’s degree and 86% held Master’s degrees.

The population of Fairview Elementary School for the 2009-2010 school year was 64% White, 16% Black, 8% Hispanic, 5% Asian, and 7% Other or Multi-Racial. The population included 38% classified as economically disadvantaged, 15% students with disabilities, and 16% limited English proficient. The school was also designated as Excellent on the State Report Card for 2009-2010, having met all eight of eight state indicators. The school also met AYP in all areas for all subgroups. All teachers held appropriate certification and Bachelor’s degree and 71% held Master’s degrees.

**Teachers.** Five teachers responded to email invitations to participate in this research, with the support of their building principals, as described above. Martin Grove teachers included Ms. Simons, Ms. Taylor, and Ms. Orville, with 13, 13, and 15 years of teaching experience, respectively. Fairview teachers were Ms. Nelson and Ms. Edwards, who had both been teaching for seven years. All teachers were White and female. Every
teacher except Ms. Nelson had a Master’s degree. See Table 3.1 for details on the teachers.

**Students.** All students in each of the five third-grade classrooms were invited to participate in this study. Students were given two weeks to return parental consent forms prior to the first data collection. An incentive of a $25 gift card to a pastry shop for a class breakfast was used to encourage students to return the forms. Across the five classrooms, 70% of students were given parental consent and agreed to participate for a total of 93 out of 132 students, 50 from Martin Grove and 43 from Fairview. The sample was nearly evenly split by gender, with 46 females and 47 males, and racially diverse with 65% White, 17% Black, 3% Hispanic, 2% Asian, and 4% Other or Multi-racial. Fifteen percent of the students were economically disadvantaged, as evidenced by their eligibility for federally funded free or reduced lunch. Generally, the demographic makeup of the study sample was highly similar to the populations of the schools, indicating that the sample is a good representation of the overall demographics of the target population. Table 3.1 provides a summary of the participants.

Table 3.1. *Study Participants.*

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Experience</th>
<th>Degree</th>
<th>Class Size</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martin Grove Simons</td>
<td>13</td>
<td>Masters</td>
<td>27</td>
<td>12 (7 M, 5 F)</td>
</tr>
<tr>
<td>Taylor</td>
<td>13</td>
<td>Masters</td>
<td>24</td>
<td>21 (9 M, 12 F)</td>
</tr>
<tr>
<td>Orville</td>
<td>15</td>
<td>Masters</td>
<td>24</td>
<td>17 (8 M, 9 F)</td>
</tr>
<tr>
<td>Fairview Nelson</td>
<td>7</td>
<td>Bachelors</td>
<td>29</td>
<td>22 (10 M, 12 F)</td>
</tr>
<tr>
<td>Edwards</td>
<td>7</td>
<td>Masters</td>
<td>28</td>
<td>21 (13 M, 8 F)</td>
</tr>
</tbody>
</table>
Overview of Study Design

This study included three different sources of data: classroom observations and teacher interviews, which were analyzed qualitatively, and student surveys, which were analyzed quantitatively. Classroom observations and student survey data were collected simultaneously at three time points and teacher interviews were conducted subsequently. The observations and surveys were scheduled so that the first two time points were during review lessons for classroom summative assessments and the final time point was during a lesson focused on preparing for the state standardized test. See table 3.2 for an overview of the timeline for the study procedures.

Table 3.2

*Study Timeline.*

<table>
<thead>
<tr>
<th></th>
<th>Introductory Visit/ Consent</th>
<th>First Observation/ Survey</th>
<th>Second Observation/ Survey</th>
<th>Third Observation/ Survey</th>
<th>Teacher Interviews</th>
</tr>
</thead>
</table>

Procedures

**Communication with teachers.** As previously described, school principals were initially contacted to gain permission to conduct the research in their buildings. A sample contact email is included in Appendix B. Principals identified the third grade teachers who were then invited to participate. Through email, we scheduled times to meet to
discuss the project further and obtain consent. I visited Martin Grove Elementary during lunchtime one day in October to meet with the three teachers. I explained the study design and had them sign consent forms. I also gave them parental consent forms to be sent home with the students and returned within a two-week period. Copies of consent forms can be found in Appendix C. Incentives were offered, including a $100 honorarium to be paid to each teacher at the completion of the data collection for their time and a $25 gift card for the class to a pastry shop as an incentive to return the parental consent forms. The class incentive was offered to encourage students to show the consent forms to their parents and bring them back signed, whether or not they agreed to let their children participate. After meeting with the teachers during lunch, I stayed to visit each classroom for about 20 minutes. The purpose of these classroom visits was to familiarize myself with the classrooms and to allow the students to be aware of the presence of an observer.

Martin Grove teachers practice flexible ability grouping for math. Prior to every math unit, students are given a pre-test in their homerooms. Based on the results of the pre-tests, students are assigned to one of the three teachers for their math section for that unit. The teachers rotate among high, middle, and low groups so that they all get the chance to teach all levels of ability and to reduce students’ awareness of group differences. Upon completion of the unit, students return to their homerooms to take the post-test, which is the same as the original pre-test. They are scored and encouraged to pride themselves on growth from the pre-test to the post-test. As the teachers described, they staple the post-test on top of the pre-test and have students flip through to compare their scores and acknowledge their growth.
In December, I visited Fairview Elementary one day to observe in each third grade class for about 30 minutes before meeting with the teachers after dismissal. Again, the study was explained and teachers signed consent forms. The same procedures for parental consent and incentives were explained. None of the teachers expressed any concerns or had any questions about the study, other than logistics of scheduling the visits.

Fairview classrooms were completely self-contained and teachers did not switch classes. They actually were looping teachers with second grade, meaning that they had taught their current third graders students the previous year, in second grade. The following year, they would return to second grade to start with a new group of students and stay with them two years as well. Unlike the teachers at Martin Grove, they did not regularly pre-test their students. Instead, they taught a unit of instruction based on their textbook chapters and then gave chapter or unit tests at the end of the period of instruction. According to the teachers, the tests were mainly drawn from the textbook and publisher resource materials, although they did occasionally write their own tests or make modifications to the tests to align with their teaching.

One participating teacher from each of the schools served as a contact person for me during the data collection period. I communicated with the teachers via email to schedule the data collection times. First I returned to each of the school two weeks after the initial meetings to collect parental consent forms and class rosters. To ensure confidentiality, I assigned ID numbers to the students based on their teacher (homeroom teacher at Martin Grove) and alphabetical order in the class, so the first student in the first
class was assigned ID number 101, and so on. We then continued to set up scheduled visits through email.

**Classroom observations.** A previous study (Kowalski & Marietta-Brown, 2011) served as a pilot for the observation component of this research. In that exploratory case study, one third-grade math teacher was observed during three lessons to investigate differences between typical math lessons and standardized testing preparation lessons. Through this pilot work, I developed a process for observing and taking running field notes focused on the motivational climate surrounding assessment and testing. The general approach was to capture as much evidence as possible, including dialogue and descriptions of student activities and engagement, of teachers’ instructional practices related to preparing students for classroom and standardized testing.

For this study, each of the five classes was observed during three separate math lessons throughout the year, for a total of 15 classroom observations. Math lessons lasted for one hour each at Martin Grove and 55 minutes each at Fairview, for a total of 14.5 hours of classroom observations (three hours for each Martin Grove teacher and two hours and 45 minutes for each Fairview teacher).

Three advanced graduate students assisted with the classroom observations for this study. One of the assistants completed only one observation while the other two completed four observations each. I completed the remainder of the observations myself.

I had previous experience working with two of the assistants on a different qualitative study involving classroom observations that utilized a similar observation protocol, so training for them was minimal. The third assistant had not had any prior experience with classroom observations, so she was given examples to study and discuss.
before completing the first observation. I observed with her the first time to check for reliability of her observation field notes and was satisfied with the consistency of our observations.

The study, research questions, and theoretical framework were explained to the assistants and they were instructed to take running field notes of as much dialogue and content as they could during the class, especially focusing on anything related to assessment and testing. All observers sat near the back of the classrooms in order to view as much classroom activity and interaction as possible and to be less of a disturbance in the room. At no time during the observations did any observers interact with the students or disrupt the flow of the lessons.

I accompanied each of the assistants who did multiple observations once each in order to have two sets of field notes to compare for reliability. We sat apart from one another during the lesson and met briefly afterwards to compare notes. There were no major discrepancies between the notes for either observation and minor differences consisted of omission of small amounts of content that were judged to not be significantly related to the research topic. There were no incidences of contradictory accounts in the sets of field notes.

In keeping with the research questions and study design, the first two observations for each classroom took place on the day before a unit test. The final observation for each classroom occurred during a lesson specifically focused on preparation for the mathematics portion of the Ohio Achievement Assessment (OAA) test in the Spring.

Student surveys. In order to investigate students’ perceptions of the motivational climate surrounding assessments and testing, short surveys were administered to those
students with parental consent immediately following the observed lessons. The assistants conducting the observations also conducted the survey administration. They were given scripts and had been trained through discussion and role playing on how to introduce and administer the surveys to the students.

The researchers distributed surveys by calling out students’ names and giving them the surveys with their ID number already written on it. Then a verbal script for participant assent was read aloud (see Appendix C). The script emphasized that participation was voluntary and that honest responses were very important to the researchers in order to help us better understand how students feel about math tests. Students were instructed to follow along with the researchers as the survey was read aloud and to think about their upcoming math test while answering the questions.

Due to the young age of the participants, it was assumed that they did not have much experience completing opinion surveys. To compensate for this lack of experience, several measures were taken. First, the survey was kept to an appropriately short length and each question was written in clear, concise language. Secondly, the agreement scale for the questions was designed to be easy for the students to understand. Rather than using a numbered or worded scale, a pictorial scale of “smiley” faces was used. See Appendix D for the complete survey. Thirdly, in order to familiarize students with the response scale, example questions were read aloud and demonstrated on the board. An example question may have been, “Fairview is the best school in the world,” and students were engaged in a discussion about which smiley face picture they would circle to answer that question based on their level of agreement. Finally, the entire survey was slowly read aloud to the students as the researcher circulated the room to ensure that students
followed along and to minimize confusion or challenges due to difficulty with reading. Students without parental consent or personal assent were instructed to sit quietly and/or work on reading or other independent work during the administration of the survey.

The survey administration for each observation took approximately 12-15 minutes. Researchers then collected the completed surveys and left the classroom. Survey responses were later entered into SPSS software for analysis.

**Teacher interviews.** After all observations and surveys were completed, teachers were contacted to schedule one-on-one interviews. The interviews occurred last so that questions about the observations could be included in the semi-structured, open-ended interview protocol. Interviews were scheduled during the school day at the request of the teachers. For the teachers at Martin Grove, the interviews were completed during lunch and recess periods. At Fairview, interviews occurred while students were at art or music classes during an afternoon. The scheduling limitations meant that the interviews could not be lengthy and in fact, ranged from 21-32 minutes long. With participant permission, interviews were audio recorded and later transcribed for coding and analysis.

**Qualitative Design**

**Teacher instructional practices.** Classroom observations were conducted to examine teacher instructional practices related to assessment and preparation for standardized testing. As previously described, each of the five participating teachers was observed during three lessons, two focused on preparing for a classroom summative assessment (unit or chapter test) and one lesson specifically focused on preparing for the math portion of the Ohio Achievement Assessment (OAA) state standardized test. Altogether, data included a total of 14.5 hours of observations. Running field notes were
taken during the lessons and were coded for evidence of promotion of certain classroom goal structures, task value, and expectancy messages. Open coding (Strauss & Corbin, 1998) techniques were also used to identify other emerging themes from the data.

**Teacher knowledge and beliefs.** Qualitative inquiry was also used to examine teacher knowledge and beliefs about assessment and testing through open-ended, semi-structured interviews. A copy of the interview protocol can be found in Appendix E. Sample questions included: “Where have you gained most of your knowledge about assessment” and “What do you believe is the main purpose of assessment/standardized testing?” Teachers were also asked about how they tried to talk to their students about testing and what messages they believe students receive. Interviews took place after all other data were collected so that questions related to the observations could also be included in the interviews. For example, after observing a teacher praise students for making “great mistakes” on a sample problem, I was able to ask about her beliefs about the role of mistakes in the learning and assessment process, referring to her words. In this way, I was able to better understand the complexity of how the teachers’ knowledge and beliefs related to their specific instructional practices during the observations.

Interviews were audio recorded, transcribed, and coded for evidence of promotion of certain classroom goal structures, task value, and expectancy beliefs. Open coding (Strauss & Corbin, 1998) techniques were also used to identify other emerging themes from the data.

**Quantitative Measures**

A short survey was administered to students to assess their perceptions of classroom goal structures, subjective task values, and expectations for success for each of
the assessment contexts. Questions were read aloud and students responded by circling the smiley face that corresponded to their agreement on a five-point pictorial scale. A copy of the survey can be found in Appendix D.

**Achievement goal structure.** Students’ perceptions of classroom achievement goal structures for each assessment context (classroom assessment and standardized testing) were measured using items chosen from the Patterns of Adaptive Learning Study (PALS) scales (Midgley et al., 2000). The PALS scales were developed by a team of researchers at the University of Michigan for use in a large-scale, longitudinal study of young adolescents (Midgley et al., 1998). The PALS manual includes a number of student and teacher scales measuring personal goal orientations, goal structures, efficacy, and other academic-related perceptions, beliefs, and strategies (Midgley et al., 2000).

For this study, the PALS scales assessing students’ perceptions of classroom goal structures were used. These scales have been used widely in motivation research. The original PALS researchers published extensively on their projects using these scales with a range of elementary and middle school student samples (e.g., L. Anderman, 1999; L. Anderman & E. Anderman, 1999; Midgley & Urdan, 2001, Urdan, Midgley, & E. Anderman, 1998). Other researchers have used the classroom goal structure scales in diverse studies including a hierarchical linear modeling analysis of person-context interactions in fifth-grade classrooms in Singapore (Lau & Nie, 2008), a study of eighth-grade student achievement and goal structures in a standards-based educational system in the United States (Ballard, 2010), and an investigation of the association between goal structures and intrinsic motivation and self-concept in Japanese adolescents (Murayama & Elliot, 2009), to name a few.
There is substantial evidence of the reliability and validity of the PALS scales (Anderman & Midgley, 2002; Midgley et al., 1998; Midgley et al., 2000). The Personal Goal Orientation scales in particular have been analyzed for internal consistency, convergent validity, construct validity, and discriminant validity (Midgley et al., 1998). The Goal Structure scales, focused on in this study, have also been established as valid in the plethora of studies utilizing them (Anderman & Midgley, 2002). According to the PALS Manual (Midgley et al., 2000), the Perceptions of Classroom Mastery Goal Structure scale has an established internal consistency (Cronbach’s alpha) of 0.76. The PALS Performance-Approach Goal Structure scale has an established alpha of 0.70 and the Performance-Avoid Goal Structure scale has an established alpha of 0.83.

For this study, two measures of perceptions of achievement goal structures are used: Mastery Goal Structure and Performance Goal Structure. While the PALS authors separated the Performance Goal Structure scales into approach and avoid scales, consistent with theoretical perspectives of achievement goal theory and research on personal achievement goals, there is little empirical evidence to support the distinction at the classroom goal structure level (Linnenbrink & Pintrich, 2002). Rather, current conceptions of goal structures tend to focus only on the distinction between mastery and performance goals without the differentiation of approach and avoidance (Linnenbrink & Pintrich). Therefore, in this study, items from the two Performance Goal Structure scales from the PALS study were combined to create a single measure of perceptions of performance goal structure, described below.

Three items from the PALS scale measuring perceptions of classroom mastery goal structure were adapted for the Mastery Goal Structure measure. Items were “In our
class: 1) trying hard is very important, 2) really understanding the material is the main goal, and 3) it’s ok to make mistakes as long as you are learning.” For this study, the items were adapted to reflect the task-specific context rather than overall classroom goal structure by changing “in our class” to “for this test” for each item.

Student perceptions of classroom performance goal structures for each assessment context (classroom and standardized testing) were measured using two items from the three PALS items measuring performance-approach goal structure and two items from the five PALS items measuring performance-avoid goal structure. Performance-approach items chosen were, “In our class, getting the right answers is very important,” and “In our class, it’s important to get a high score.” Performance-avoid items chosen were, “In our class, it’s important not to do worse than other students,” and “In our class, it’s important not to look dumb.” The items were adapted to reflect the task-specific testing context rather than overall classroom goal structure. For example, performance-avoid item “In our class, it’s important not to do worse than other students” was changed to, “For this test, it’s important not to do worse than other students.”

**Students’ task value.** Perceived task value was operationalized according to expectancy-value theory as being comprised of importance value, utility value, and interest value (Eccles, 1983; Wigfield et al, 1997). Sample items from the work of Wigfield, et al (1997) were used to measure students’ valuing of the assessment task for each of those three dimensions of value. In their study, they developed scales for expectancy beliefs and subjective task values for math and reading for use in a three-year longitudinal study of first through sixth grade elementary students. For subjective task value, two separate scales were created: interest, and importance and usefulness. Across
different domains and times throughout the study, internal consistency for the interest scale was high, with Cronbach’s alphas ranging from .73-.92. Internal consistency reliability for the importance and usefulness scale was low in the first year of data collection ($\alpha = .36-.57$), but was higher across the second and third years ($\alpha = .61-.88$). This indicates that students as young as first and second grade, who were included in the first year, may have a difficult time determining the usefulness and importance of tasks or using scales to respond to survey questions in general. However, reliability of the scale was acceptable when the participants were all above second grade, as in the subsequent two years of the study (Wigfield et al, 1997).

Items from these scales have also been used extensively in the research literature, including a study of patterns of motivational beliefs in middle school students (Conley, 2012) and a longitudinal study of growth trajectories with students as young as first grade (Fredricks & Eccles, 2002). These and other studies provide ample evidence of the reliability and validity of the scales.

For this study, items from the subjective task value scales were modified in two ways. First, they were presented on a five-point pictorial scale rather than the seven-point numbered Likert-type scale used by Wigfield et al. (1997). This modification was made to be consistent with the other items on the survey and because of the young age of the participants. Second, the items were adapted to be more context-specific, referring to the assessment task rather than the academic domain. An item reading “For me, being good at math is important” was changed to, “For me, doing well on this math test is important,” “In general, how useful do you find math for your future?” became, “Taking
this math test will be useful for me in the future,” and an item reading, “How much do you like doing math?” became, “How much do you like taking this math test?”

**Students’ expectancy beliefs.** A single item was used to measure students’ expectancy for success in each assessment context. The item was adapted from Wigfield et al (1997) to be context-specific. “How well do you expect to do in math this year?” was changed to, “How well do you expect to do on this math test?” The item was also adapted to be on a five-point pictorial scale rather than a seven-point numbered Likert scale in order to be consistent with the rest of the survey items.

**Students’ prior achievement.** Student records were accessed to determine prior standardized testing achievement based on second grade Terra Nova mathematics test scores. The Terra Nova is a widely-used national low-stakes standardized test taken by students in the Bellehaven district in the fall starting in second grade. The test is published by CTB/McGraw Hill and measures achievement in math, reading, language arts, vocabulary, science, social studies, and spelling. A norm-referenced composite score is given ranging from 0-100. For this study, composite scores were translated into categorical variables, with a score of 0-19 coded as “1”, 20-39 coded as “2”, 40-59 coded as “3”, 60-79 coded as “4”, and 80-100 coded as “5”.

**Student demographics.** Gender, ethnicity, and socio-economic status indices were obtained through student records. Gender was coded “0” for female and “1” for male. Ethnicity was labeled according to the categories used on the State Report Card from the Department of Education and was coded “1” for White, “2” for Black, “3” for Hispanic, “4” for Asian, and “5” for Other or Multi-racial. Socio-economic status was determined based on eligibility for federally supported free or reduced lunch and was
coded “0” for yes and “1” for no. Those students eligible for free or reduced lunch were considered economically disadvantaged.

**Analysis Approach**

This mixed methods study included two phases of data analysis. For the qualitative data, transcriptions of observation field notes and teacher interviews were coded based on the *a priori* theoretical framework of the study in order to most directly answer the research questions (Saldana, 2009). To this end, a provisional list of codes was developed for teacher communication, expressions of beliefs, and knowledge about achievement goal structures, expectancies for success, and task values. In addition, open coding (Strauss & Corbin, 1998) was used to identify other themes and subthemes that emerged from the data that might have an influence on the student experience and motivation associated with assessment and testing. One of the research assistants who participated in the data collection also coded a subset of the data, comprised of three classroom observations and one teacher interview transcript, to check for inter-rater reliability. There were only minor discrepancies in coding that were resolved through discussion. Patterns of codes were identified and analyzed to respond to the research questions.

For the quantitative phase of this study, statistical analysis of student survey data was conducted using SPSS software. Reliability of the measures was examined through calculations of Cronbach’s alphas. Analysis of variance (ANOVA) was used to investigate the effects of demographic variables on each measure. In order to answer the research question, correlations of values between time points were analyzed and paired t-
tests were conducted on each measure to determine differences in student motivation between the two contexts.

Findings from the two phases of analysis were integrated in the interpretation of the results to answer the research questions. Qualitative data provided the teachers’ and researchers’ perspectives on classroom motivational climates for the two contexts, while the quantitative data represented the perspective of the students. Taken together, a robust understanding of the classroom experiences was achieved.
Chapter 4: Results

This chapter describes the data analysis approach and the resulting findings for this study. The aim of the study was to address the following research questions:

1. How do elementary school math teacher instructional practices related to assessment differ for summative classroom assessments and preparation for standardized state achievement tests?

2. How do elementary school math teacher instructional practices reflect their knowledge, beliefs, and perceptions of external influences related to summative classroom assessments and standardized state achievement tests?

3. How do students perceive the classroom motivational climate surrounding assessment and standardized testing?

These research questions were addressed using a mixed methods research design. The results will be presented in three sections. The first section provides a description of the analysis approach and results for the first two research questions, which were investigated using a qualitative approach. The second section describes the analysis approach and results for the third research question, which was addressed using a quantitative approach. Finally, the third section provides an integration of the qualitative and quantitative results.
Qualitative Results

Observation field notes and audio-recorded interviews were transcribed and coded by the researcher. A provisional list of codes was devised, based on the *a priori* theoretical framework of this study, in order to help directly answer the specific research questions (Saldana, 2009). An initial pass through the data employed descriptive coding methods (Saldana; Wolcott, 1994), after which the coding scheme was revised to include additional subthemes emerging from the data. At that point, a second pass through the data served to finalize the coding scheme. The codes and sub-codes were reorganized through researcher reflection on the theoretical frameworks. The list of codes, sub-codes, and representative quotations is included in Table 4.1. One of the research assistants who participated in the data collection also read through a subset of the data, comprised of three classroom observations and one teacher interview transcript, to check for inter-rater reliability. She was given my coding scheme but also encouraged to add additional codes or suggest other changes. There were only minor discrepancies in coding that were resolved through discussion. Finally, data were organized into matrices by teacher and context (assessment vs. standardized testing) (Miles & Huberman, 1994). This allowed me to more easily see patterns of similarity and differences for each of the codes. Although there was some minor variability in patterns between teachers, all of the codes varied more by context than by teacher. The purpose of this study is to investigate differences between the contexts, however, individual teacher differences, though not pronounced, will be described when they are evident.
Table 4.1

Themes, Sub-Themes, and Representative Quotes

<table>
<thead>
<tr>
<th>Theme</th>
<th>Sub-Theme</th>
<th>Representative Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus on learning</td>
<td></td>
<td>“This is the chance to see not just that you got the right answer…but to really make sure you know what you are doing.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I’m going to give you a bonus problem that’s really hard and what that means is if you get it wrong, it will not count against you. I’ll just be really proud of you for trying.”</td>
</tr>
<tr>
<td>Mastery</td>
<td>Effort over performance</td>
<td>“That’s a really great mistake, isn’t that a great mistake?!”</td>
</tr>
<tr>
<td></td>
<td>Mistakes are okay</td>
<td>“If you got the answer, that’s two points. If you got the explanation then it’s four points.”</td>
</tr>
<tr>
<td>Performance</td>
<td>“They” for testmakers</td>
<td>“Do you think the person grading is going to know you? Or care about you? Nope.”</td>
</tr>
<tr>
<td></td>
<td>Mistakes not okay</td>
<td>“If you missed #1, you made a reading error…you should not be making reading errors!”</td>
</tr>
<tr>
<td>Expectancy</td>
<td>Communication of expectancy beliefs</td>
<td>“Every single one of you is capable of getting that one right. You are all smart enough.”</td>
</tr>
<tr>
<td>Emphasis on preparation</td>
<td></td>
<td>“You see this all the time on morning work, you guys are totally prepared for patterns and rules!”</td>
</tr>
<tr>
<td>Importance</td>
<td></td>
<td>“This is on the test, so it’s really important that you pay attention.”</td>
</tr>
<tr>
<td>Utility</td>
<td></td>
<td>“This is an activity that has both math and social studies.”</td>
</tr>
<tr>
<td>Value</td>
<td>Interest/Fun</td>
<td>“In school you have to do things you don’t want to do. You will learn, in life, you have to do that too.”</td>
</tr>
<tr>
<td></td>
<td>Rewards</td>
<td>“Your group is eligible to earn a third-grade dollar for every problem out of the five that you get right.”</td>
</tr>
</tbody>
</table>
Analysis of classroom observation field notes and interview transcripts revealed that there were distinct differences in teachers’ instructional practices, knowledge, and beliefs related to classroom assessments and standardized testing preparation that point toward motivational implications for students. These results will be described within the frameworks of achievement goal theory and expectancy-value theory.

**Achievement Goal Theory**

Classroom observations and teacher interviews revealed that teachers were more likely to engage in practices associated with classroom mastery goal structures in the classroom assessment context than in the context of standardized testing preparation. During classroom assessment review lessons, teachers consistently stressed learning over performance, dissuaded social comparisons, and embraced mistakes as opportunities for learning. For example, Ms. Orville began going over answers on a review worksheet by saying, “Now this is a chance to see not just that you got the right answer- so don’t just be busy erasing and writing down the right answer, but to really figure out if you got it right and make sure you know what you are doing.” While handing back papers, Ms. Simons frequently reminded the class, “Whatever anyone else got doesn’t matter, you just worry about what you got.” During a review game, Ms. Edwards reminded the class that she does not keep track of points because the competition was not important; instead, reviewing the material was the point of the activity. When a student shared a mistake on a problem, Ms. Simons replied, “I’m so glad you taught us that, that’s so interesting. I don’t think anyone has ever told me that mistake before and now nobody in here will make that mistake again!” Nearly all classroom assessment review lessons involved students working cooperatively and helping each other learn. For example, five lessons
included review games based on the television show *Jeopardy* where students were trying to solve problems and answer questions in teams. Three other lessons included math centers in which students worked in groups and rotated through a series of collaborative activities, such as using blocks as manipulatives to solve volume, area, and perimeter problems. Ms. Orville praised her class for their collaboration after a review lesson including group problem solving challenges, saying, “I loved that! I saw so many people explaining how to get the answer to other people. Sometimes you will do that to help other people understand.” Throughout the classroom assessment lessons, there was a clear emphasis on learning and understanding, consistent with a mastery goal structure (Blumenfeld 1992; Kaplan, Middleton, Urdan, & Midgley, 2002; Meece, Anderman, & Anderman, 2006).

In contrast, performance goal structures were more highly endorsed during standardized testing lessons. All of the standardized testing lessons were individual work periods where students were reminded to “keep your eyes on your own paper. I want to know what you know, not your neighbor” (Ms. Edwards). In these lessons, teachers generally distributed worksheets or math practice test packets and students worked individually on problems before going over answers together. There was a strong emphasis on getting answers correct and earning points, with teachers constantly reminding students how many points a particular question would be worth on the test. For example, Ms. Orville explained, “If you got the answer, that’s two points. If you got the explanation then it’s four points. How many got four points? Good! Anyone get two points? Ok, let’s make sure we all get all four points with good explanations.” This focus
on earning points is consistent with a performance oriented goal structure (Kaplan et al., 2002; Meece et al., 2006).

One of the most interesting findings with regard to emphasizing performance was the prevalence of the pronoun “they” in teachers’ references to the test makers and/or graders. All teachers talked frequently about the test makers as external, nebulous characters to whom the students were charged with demonstrating their competence. Ms. Simons chastised her students, “It’s wrong if they can’t read it! Do you think the person grading is going to know you? Or care about you? Nope. That’s why I want you to always do it right so you won’t miss it.” When a student tried to explain her graph to Ms. Taylor before turning it in, she scolded her, “You can’t tell me anything. Because are you going to be able to tell the grader anything? NO! So you can’t tell me what you mean.” Similarly, Ms. Orville told a student that her answer was correct, but “you know what they’re gonna get you on? In that box, they aren’t going to understand what those numbers mean…we have to figure out how to write the division so they can read it in the box.” The test makers and graders were considered foreign to the classroom with no existing relationship with the students or teachers.

These examples of the use of the term “they” are also indicative of the emphasis on teaching strategies to maximize the points earned, not necessarily the content learned. Again and again teachers told students to work only inside the box and to write neatly so that the external graders would not penalize them. The number of points that students would receive through focusing on test-taking strategies such as writing in the box was emphasized much more than the knowledge of the content. In this way, performance goal structures were fostered during standardized test preparation lessons.
The observed differences in goal structures for the two assessment contexts were reflective of the teachers’ knowledge and beliefs about the testing circumstances. When asked about the different ways they approach the types of tests with their students, teachers were aware of the different messages they were sending. For example, with regard to the role of mistakes, Ms. Simons described her view of the classroom assessment process by saying, “It’s a continuum, you are given an idea and you need time to learn. You need a grace period...you need to be able to feel that it is okay to make mistakes.” However, in describing standardized testing, she stated, “It’s like a competition and we’re gonna win...You can’t make mistakes on the OAA...it’s the end, it’s the performance, it’s like the Olympics. Your mistakes are all out, you are ready to be perfect.” Similarly, Ms. Orville explained how she talks about mistakes in the two contexts, saying, “We really go with the concept of, ‘Thank you for letting us learn from your mistake, thank you for your effort,’ but with the standardized tests, it’s more of, ‘Here are the mistakes that people make. I don’t want to see you making this mistake’...yeah, it’s a definite difference.”

The belief that it was okay to make mistakes on regular classroom assessments but not on the standardized tests is an indication of the external pressure that teachers feel to have their students perform well on the tests. Ms. Edwards explained, “There’s a lot of focus that is put on it, whether it is by the district or the school or the state. I mean, we start talking about it right from the beginning of the year at teachers’ meetings, and it’s like, really? Is that all there is to this? I mean, they are kids, they’re not just scores.” Ms. Orville reiterated this feeling of pressure:
I think we approach it so differently about what we expect from the kids versus the testing that someone else is going to come back to us on... With (classroom assessment) you’re really all about how they’re doing, like did they make growth, did they show progress...I think that we approach (the standardized test) so seriously when I know my boss is going to have my name with my score beside it on a chalkboard on the first day of school.

Ms. Orville’s sentiments are indicative not only of an emphasis on performance goal structure for her students during standardized testing, but also a school-wide performance goal structure supported by a principal who practices social comparison of scores among teachers. There is evidence that goal structures are present at the school level and that they can influence classroom goal structures (Maehr & Midgley, 1991). The top-down pressure for performance goals is apparent in the context of standardized testing.

In sum, the pressure to demonstrate high student achievement leads teachers to approach standardized testing differently than regular classroom assessment activities. Teachers engage in practices more likely to support mastery goal structures during classroom assessments and performance goal structures during standardized testing preparation.

Expectancy-Value Theory

Differences in classroom observations and teacher interviews between the two assessment contexts were also apparent from the theoretical perspective of expectancy-value theory. Teachers frequently communicated messages of expectancy for success to their students in both classroom assessment and standardized test preparation contexts.
While reviewing for a classroom assessment, Ms. Taylor began the lesson by saying, “This is the last day, everyone in this math group has been working hard, I know you can get it.” She concluded the same lesson with, “Any questions about the post test tomorrow? You should be thinking ‘I’ve got this, I know this.’” In a similar classroom assessment review lesson, Ms. Simons demonstrated solving a problem on the board and stated, “Every single one of you is capable of getting that one right. You are all smart enough.” After a particularly challenging enrichment lesson on long division, Ms. Orville helped her students prepare for a post test by telling them:

_I think a few of you started to doubt whether you were really good at multiplication and division, right? Today I want to restore your confidence that you are good at multiplication and division, it’s just that long division was really difficult and was making you stressed. By the time you leave today you are going to feel so much better. Today we are going to review the third grade stuff that you need to know for your test tomorrow, because that long division was really fourth and fifth grade math, and I want you to remember that you are good at the math you need for third grade._

Messages of expectancy for success on standardized testing were generally attributed to students being “smart” or “capable” enough to succeed and more about them being adequately prepared to succeed. Teachers made references to preparing students for the tests throughout the year, including test-like questions in regular classroom lessons and frequent reminders of how to succeed on the test. As early as November, five months before the standardized test, Ms. Orville presented the class with a review problem for their upcoming classroom assessment by saying, “I’m going to be honest with you, this
question has been on every OAA for the last four years. They always ask this! So let’s look at it…” When preparing for the standardized test, Ms. Taylor told her students excitedly, “The next part is patterns and rules, you see that all the time in your morning work, you guys are totally prepared for patterns and rules!” In this way, teachers were helping students see the similarity of test items to previous successful experiences, therefore building their self-efficacy, a distinct but related construct to expectancy for success (Bandura, 1997; Wigfield & Eccles, 2000). Students were reassured time and again that they were well prepared to succeed on the test and should not feel stressed about it.

In spite of the positive expectation messages that students heard verbally from their teachers, there was also evidence of teachers possibly inducing stress in students through their cautionary tone with regard to the standardized tests. For example, Ms. Nelson referred to a practice problem saying, “I guarantee you that several of you, if this were on the OAA, you wouldn’t know what to do, because there is no fraction here.” Ms. Simons chastised her class, saying, “If you missed #1, you made a reading error. Boys and girls, you are in the third grade now, you should not be making reading errors!” Ms. Edwards showed students a problem in their practice test packet and stated, “You would be surprised how many kids would just move on to the next question at this point and just skip the second part…Everyone should be able to get two points on this question, but you won’t get it if you forget to answer the question at the bottom!” Similarly, every teacher warned students multiple times about remembering to write their answers within the designated response boxes on the test. These examples were all intended to increase the students’ preparedness for the test by emphasizing the strategies and test-taking skills that
they had learned, but it is possible that students interpreted the message as lowering their expectations for success on the standardized test due to the cautionary tone.

The unintended consequence of inducing stress and possibly lowering students’ expectancy for success is not surprising given the pressure that teachers reported feeling to have students do well and their concern about their students’ ability to succeed. In interviews, all teachers expressed positive views of their expectations for their students to succeed on their classroom assessments. As Ms. Taylor summarized, “I think we have high standards and I think that 95% of the time the kids are going to show growth and prove that they are meeting or achieving above the standards.” While teachers also believed their students were capable and prepared for the content of the standardized tests, they indicated concern over students performing well under stressful testing conditions. Ms. Simons predicted that her students would do really well:

....if they really show what they know. But here’s the thing, all kinds of things get in the way. They make mistakes and you can’t go back and fix it...They skip questions, they lose their endurance, they’re just tired. Our post tests are maybe like 25 questions. They are taking a post test for all of third grade that is like 47 questions long. Nobody thinks about their endurance. They are only eight and nine years old. That’s a lot to ask of them. I mean, it’s like telling a kid to clean a house. Are you kidding me? They’re responsible for their room, but maybe just one part of their room. And the (standardized test) just feels like they are responsible for the whole house. You cannot possibly do all that yet.

All teachers echoed this sentiment about the test demanding a large degree of responsibility from and imparting pressure on students. They realized that this belief
contributed to a different message being communicated to the students regarding the expectations for the test. Ms. Orville summarized this, saying:

> In (classroom assessments) situations, we allow them to feel comfortable and confident about it even if they’re not getting everything, but I think with the OAA, I feel like we’re all very, you know, ‘you’ve got to do your best, you’ve got to work hard, you’ve got to focus,’ so it’s different.

Clearly, the pressure that teachers feel to have students succeed on the standardized tests had an impact on the type of messages they sent related to expectancy for success.

While messages about expectancies for success were common, messages about the subjective task value of the tests were less apparent in both assessment contexts. For example, there was no evidence during observations of teachers helping students see the attainment value of either type of assessment. That is, there was no communication related to the personal importance of either testing task.

There were some instances of teachers mentioning the importance of doing well on tests, alluding to utility value. For example, Ms. Nelson reviewed for a classroom assessment, warning students, “Tomorrow you take a test. This is on the test, so it’s really important that you pay attention.” Ms. Taylor told her students, “Tomorrow is your post test. It counts for a real grade for this math group.” Presumably, students were led to believe that grades were important and therefore classroom tests were a serious matter. It was important to pay attention and work hard on the test because test performance would lead to grades.
Students were also given the clear message that the standardized tests were important. This was apparent in the amount of time, both throughout the year and in the intense review period in the weeks leading up to the test, spent on preparation. They were told that testing strategies were important, such as writing only in the boxes, taking their time, and checking their work. For example, Ms. Taylor made a point of explicitly reminding students to “NEVER leave a question blank,” telling them, “When you have two and a half hours you don’t have to worry about leaving one blank because you can go back and check and recheck!” Similarly, Ms. Edwards warned her students, “It’s frustrating when I see kids who are done like that (snaps fingers) and I know they have made a lot of mistakes because they went too fast...so take your time and check your work and watch out for those little mistakes.”

While students were told that the test and these strategies were important, they were not really told why they were important or what utility value they held. During observed lessons, no teacher ever talked about what students’ performance on the test meant for the students or the school or why they were taking the tests to begin with. Rather, when talking about the standardized tests, teachers would stress that the tests were imposed by an external source, often referred to as “they” as discussed earlier, and offered no further explanation. In fact, at times the teachers actually made comments to devalue the testing process. The strategy of writing inside the box in particular seemed to be problematic in terms of value. For example, Ms. Edwards reminded students to write answers only in the boxes. When a student asked why they had to do that, she shrugged and responded, “That’s just the way it has to be. Follow the directions.” Ms. Nelson
went so far as to remark, “Remember you can’t do anything outside this box…it’s a dumb rule, but that’s the facts.”

This issue reflects teachers’ beliefs about the lack of value in certain test taking procedures. In their interviews, all teachers expressed frustration with aspects of the testing policy, including the exorbitant two and a half hour length of the test, the “work in the box” rule, and the lack of individualization in the test. Ms. Orville explained that she believed it would be more valuable if it were, “more ongoing and not just one snapshot of time.” She advocated using classroom assessments instead of standardized testing, because: “to look at all the growth would be more effective than to say, ‘here’s two and a half hours and that kid may have had 18,000 things go wrong that morning, but hey, that’s your problem.’”

The rule about working inside the box was a frequent point of complaint for the teachers. As Ms. Nelson explained in her interview:

*On the OAA, their work has to be shown within that certain box, and if it shows outside that box, they don’t count that work, which I think is utterly ridiculous because it’s right there on the paper. And that’s what I mean by teaching to the test, because you’re teaching you have to work inside this box and don’t go out of the box.*

Teachers found that having to teach students to work inside the box was difficult because it was not something they felt was meaningful. They talked about other ways that the test was not formatted or graded in the same way that they would typically choose to teach and assess their students. Ms. Simons explained that she tried to be honest with her students about this:
I have to say to them, ‘this is not what I would do but this is how you have to be graded...I don’t agree with this, but I can’t change. I can’t do anything about it.’

It’s almost like good cop bad cop.

Ms. Simons’ sentiments clearly indicate a lack of value and meaningfulness of the testing policies. Messages such as this have implications for students’ understanding of the testing process, potentially decreasing their valuing of the task and also further contributing to the promotion of a performance goal structure, as previously described. Again, the message is that grades and performance are most important and students should just do whatever they need to do to get the points.

Although teachers were quick to express their disapproval of the testing process, they were also aware of the messages they sent to students affirming the importance of the test. As Ms. Edwards explained: “There is a little bit of a different message, like you REALLY need to show what you know on this, because, you know, it’s looked at as super important and they HAVE to do well.” Ms. Taylor agreed, saying:

They’re going to say that it is so much more important because they just hear about it over and over and over again from their parents and from the school and the fliers that go home and from the teachers. So they are going to feel the importance of that.

Ms. Nelson also felt that the importance of the test was over-emphasized to students, explaining, “they feel the pressure of these tests. The third grade test, if they don’t pass they can be held back, and parents know that so I think they pressure their children.” Overall, regardless of their personal views about the value of the test, teachers
knowingly fostered a heightened sense of importance value in their students with regard to the standardized tests.

Finally in terms of value, there was also a clear difference between the levels of intrinsic value communicated to students in reviewing for classroom assessments as compared to standardized testing. Every single classroom assessment review lesson observed included a fun, engaging activity for the students. Ms. Simons told students, “We are going to do something really fun today! It’s marshmallow math day!” They proceeded to use marshmallows as manipulatives to review their addition and subtraction for their upcoming test. Other lessons included Jeopardy-style review games, math centers with various group activities, and challenging group problem solving enrichment problems for which students earned small rewards. For example, Ms. Orville introduced a review activity saying, “This is an activity that has both math and social studies. Your group is eligible to earn a third grade dollar for every problem out of the five that you get right.” While these activities did not add to the intrinsic value of the classroom tests themselves, they at least made the review experience positive and enjoyable for the students. Students were generally engaged and enthusiastic during these lessons.

In contrast, every single standardized testing preparation lesson included teacher-directed lessons and individual work on review sheets or practice test packets. Rather than collaboration, students were told, “Math has started. We are not talking, we are working in our mind” (Ms. Taylor). There were no messages of intrinsic value related to the review or to the test itself. Students did not exhibit the same level of engagement in these lessons that they had in the more fun classroom assessment review lessons. At one point, a student in Ms. Nelson’s class complained of being tired and bored. She replied,
“You know what, I’m tired. In school, you have to do things you don’t want to do. You will learn in life, that you have to do that too. Ok, next page...”

In their interviews, teachers explained their decision-making with regard to planning review lessons and they acknowledged the vast difference between the two contexts. Ms. Nelson succinctly stated: “In all honesty, I think when we do class testing, it tends to be more fun, right?” She further explained:

When you have a whole class lesson (for standardized preparation), you can keep an eye on much more of them than you can when they’re in groups or centers. And that’s more important when you are preparing for standardized testing than regular testing...when we do games, those kind of things, we know that if they don’t pass, we could re-teach and re-assess. And the OAA is just a one time shot. You know, you take it and if you fail, you fail.

Ms. Orville reiterated this difference, explaining:

We had drilled and drilled and we were doing questions on the overhead and we were just working. They were doing practice and practice. There was no fun, like, happy-go-lucky...whereas when it’s just a typical unit we do a lot more laid-back where everyday they have some group time or partner time. But the weeks before the OAA, it’s just ‘you do this’ because you have to do it by yourself on the test.

Another reason that teachers gave for the difference in the lessons was the different format and wording of the tests. Ms. Edwards explained that standardized test preparation is “less of the game-type” fun and engaging lessons because:

We really want them to look at samples and become familiar with how the text is written, because sometimes just the wording of the questions might be tricky...it’s
just to get them familiar with how it's going to look like, while reviewing the concepts as well. And I feel like with the classroom assessments, they are so used to the wording and how they go and so we feel like we can branch out a little more and find creative ways to review.

The teachers’ comments in their interviews demonstrated that they intentionally designed review lessons that were less intrinsically enjoyable for students in the standardized testing context due to their beliefs about the importance and external nature of the test.

In sum, there were differences in messages that students received regarding expectancy for success and the subjective task value of classroom assessments and standardized testing. Teachers’ knowledge and beliefs about the types of tests were consistent with the observed differences in teaching practices between the two contexts. Motivational implications of these differences will be examined further in Chapter 5.

**Quantitative Results**

The quantitative phase of this study was designed to answer the third research question: How do students perceive the classroom motivational climate surrounding assessment and standardized testing? The hypothesis was that there would be differences in students’ perceptions of motivational climates between the different contexts. A survey was developed and administered on three occasions to examine students’ perceptions of achievement goal structures and messages about expectancy and value related to the tests. Each student was assigned a unique ID number in order to match data across collection waves and to student records. Survey responses were coded and entered by ID number
into a data spreadsheet. The statistical software package SPSS version 19 was used to conduct all analyses described below.

**Reliability**

The survey consisted of three scales: perceived mastery goal structure (3 items), performance goal structure (4 items), and task value (3 items). Scales were constructed using the arithmetic means of included items. In addition, a single item was used to measure students’ expectancy for success. Prior to creating the scales for analysis, Cronbach’s alphas were calculated for each of the measures, separately at each of the three survey administrations, to examine their internal consistency. In general, internal reliability values were lower than desired for each of the measures. Alphas were especially low for the mastery goal structure scale at each time point (T1=.24, T2=.51, T3=.46). The scale was analyzed to see if internal consistency would be increased if any one of the items were dropped, but there was no improvement. Therefore, the mastery goal structure items were analyzed individually instead of as a scale. The items were: Mastery 1 = “For this test, trying hard is very important.” Mastery 2 = “For this test, really understanding the material is the main goal.” Mastery 3 = “For this test, it’s ok to make mistakes as long as you are learning.”

For the performance goal structure and task value scales, alphas ranged from .407-.675. Though still lower than desired, given the relatively small sample size and small number of items in each scale, the low values are understandable and were considered acceptable for the remaining analyses.

Though there were three survey administrations, the first two both occurred in the same context (classroom assessment). Therefore, correlations between the first and
second administrations were examined, revealing significant correlations for each scale and item at the 0.01 level (see Table 4.2). Subsequently, alpha values were constructed using arithmetic means for combined Assessment Context measures including both time one and time two data. Data from time three is considered the Standardized Testing Context. Table 4.2 shows means and standard deviations for items measuring expectancy for success and perceived mastery goal structure and scales for performance goal structure and task value at each time point as well as Pearson’s correlations for the first two time points for each measure. Table 4.3 shows alpha coefficients for performance goal structure and task value measures at each time point.
Table 4.2

*Means and Standard Deviations for Items and Scales and Correlations for First and Second Time Points*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Correlation (Time 1 and Time 2)</th>
<th>Assessment Context (T1 and T2 combined)</th>
<th>Standardized Context (T3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expectancy</td>
<td>4.26</td>
<td>4.34</td>
<td>.521*</td>
<td>4.32 (.75)</td>
<td>4.39 (.91)</td>
</tr>
<tr>
<td></td>
<td>(.91)</td>
<td>(.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery 1</td>
<td>4.62</td>
<td>4.66</td>
<td>.296*</td>
<td>4.64 (.61)</td>
<td>4.87 (.37)</td>
</tr>
<tr>
<td></td>
<td>(.77)</td>
<td>(.72)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery 2</td>
<td>4.20</td>
<td>4.27</td>
<td>.347*</td>
<td>4.25 (.66)</td>
<td>4.15 (.98)</td>
</tr>
<tr>
<td></td>
<td>(.79)</td>
<td>(.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery 3</td>
<td>4.53</td>
<td>4.56</td>
<td>.509*</td>
<td>4.54 (.71)</td>
<td>4.23 (1.13)</td>
</tr>
<tr>
<td></td>
<td>(.91)</td>
<td>(.72)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Scale</td>
<td>3.61</td>
<td>3.71</td>
<td>.535*</td>
<td>3.68 (.84)</td>
<td>3.64 (.97)</td>
</tr>
<tr>
<td></td>
<td>(.99)</td>
<td>(.92)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value Scale</td>
<td>4.08</td>
<td>4.09</td>
<td>.592*</td>
<td>4.08 (.73)</td>
<td>4.05 (.93)</td>
</tr>
<tr>
<td></td>
<td>(.86)</td>
<td>(.75)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Standard deviations appear in parentheses below means. *p < .01*
Table 4.3

Scale Reliability (Cronbach’s Alpha Coefficients)

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
<th>Assessment Context (T1 and T2 combined)</th>
<th>Standardized Context (T3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>.675</td>
<td>.673</td>
<td>.674</td>
<td>.662</td>
</tr>
<tr>
<td>Value</td>
<td>.514</td>
<td>.407</td>
<td>.461</td>
<td>.631</td>
</tr>
</tbody>
</table>

*Note. *p < .01

Preliminary Analysis

Prior to engaging in the primary analysis of the data, a series of one-way ANOVAs were conducted to examine any potential differences by gender (Table 4.4), ethnicity (Table 4.5), or teacher (Table 4.6) for each measure in each assessment context. For the ethnicity groups, it should be noted that the uneven and, in some cases, very small cell sizes make the interpretation of any differences limited. Overall, there were only two significant findings from the ANOVA analysis. Gender was a significant factor on students’ perceptions of performance goals in the classroom assessment context \( (F(1, 83) = 4.81, p = .031) \), however, the effect size was small \( (\eta^2=.055) \). Means for the groups were 3.87 (SD = .78) for females and 3.48 (SD = .86) for males, indicating that female students perceived higher performance goal structure than males during classroom assessment lessons. Gender was also a significant factor for the second item measuring students’ perceptions of mastery goals in the standardized testing context \( (F(1,84) = 5.50, p = .021) \), though the effect size was also small \( (\eta^2=.052) \). Group means were 4.40 (SD = .73) for females and 3.85 (SD = 1.14) for males. In other words, females reported higher
agreement than males to the item “Really understanding is the main goal” in the standardized testing review lessons. Because of the small effect sizes and since no other significant differences were found, data were collapsed across all demographic variables for the remaining analyses.

Table 4.4

Motivation Measures for Female and Male Students

<table>
<thead>
<tr>
<th>Measure</th>
<th>Gender</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>F</td>
<td>η²</td>
</tr>
<tr>
<td>Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery 1</td>
<td>4.68</td>
<td>4.58</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.52)</td>
<td>(.71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery 2</td>
<td>4.31</td>
<td>4.22</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.67)</td>
<td>(.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery 3</td>
<td>4.61</td>
<td>4.50</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.74)</td>
<td>(.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>3.87</td>
<td>3.48</td>
<td>4.81*</td>
<td>.055</td>
</tr>
<tr>
<td></td>
<td>(.77)</td>
<td>(.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>4.19</td>
<td>3.98</td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.76)</td>
<td>(.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectancy</td>
<td>4.35</td>
<td>4.28</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.92)</td>
<td>(.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardized</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery 1</td>
<td>4.90</td>
<td>4.85</td>
<td>.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.37)</td>
<td>(.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery 2</td>
<td>4.40</td>
<td>3.85</td>
<td>5.50*</td>
<td>.052</td>
</tr>
<tr>
<td></td>
<td>(.73)</td>
<td>(1.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery 3</td>
<td>4.36</td>
<td>4.05</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.98)</td>
<td>(1.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>3.80</td>
<td>3.48</td>
<td>2.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.94)</td>
<td>(.99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>4.17</td>
<td>4.28</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.87)</td>
<td>(.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectancy</td>
<td>4.41</td>
<td>4.37</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.95)</td>
<td>(.89)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Standard deviations appear in parentheses below means. *p < .05.
Table 4.5

Motivation Measures for Student Ethnicity

<table>
<thead>
<tr>
<th>Measure</th>
<th>Ethnicity</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White n=65</td>
<td>Black n=17</td>
<td>Hispanic n=3</td>
<td>Asian n=2</td>
<td>Other n=4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment</td>
<td>Mastery 1</td>
<td>4.69</td>
<td>4.43</td>
<td>4.83</td>
<td>4.00</td>
<td>4.75</td>
<td>1.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.56)</td>
<td>(.75)</td>
<td>(.29)</td>
<td>(1.41)</td>
<td>(.50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mastery 2</td>
<td>4.30</td>
<td>4.20</td>
<td>4.17</td>
<td>4.75</td>
<td>4.00</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.68)</td>
<td>(.56)</td>
<td>(.29)</td>
<td>(.35)</td>
<td>(.82)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mastery 3</td>
<td>4.66</td>
<td>4.27</td>
<td>3.70</td>
<td>4.75</td>
<td>4.25</td>
<td>2.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.53)</td>
<td>(.94)</td>
<td>(1.04)</td>
<td>(.35)</td>
<td>(.72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>Performance</td>
<td>3.61</td>
<td>3.81</td>
<td>4.04</td>
<td>3.31</td>
<td>4.44</td>
<td>1.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.84)</td>
<td>(.95)</td>
<td>(.19)</td>
<td>(.44)</td>
<td>(.46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>Value</td>
<td>4.12</td>
<td>4.06</td>
<td>3.67</td>
<td>4.58</td>
<td>4.00</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.73)</td>
<td>(.78)</td>
<td>(.17)</td>
<td>(.59)</td>
<td>(.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectancy</td>
<td>Expectancy</td>
<td>4.26</td>
<td>4.57</td>
<td>3.83</td>
<td>4.00</td>
<td>4.50</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.78)</td>
<td>(.58)</td>
<td>(1.26)</td>
<td>(.71)</td>
<td>(.41)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardized</td>
<td>Mastery 1</td>
<td>4.89</td>
<td>5.00</td>
<td>4.00</td>
<td>5.00</td>
<td>4.75</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.32)</td>
<td>(.00)</td>
<td>(1.00)</td>
<td>(0.00)</td>
<td>(.50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mastery 2</td>
<td>4.19</td>
<td>4.27</td>
<td>3.33</td>
<td>4.00</td>
<td>3.75</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.94)</td>
<td>(1.16)</td>
<td>(.57)</td>
<td>(1.41)</td>
<td>(.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mastery 3</td>
<td>4.37</td>
<td>3.93</td>
<td>4.33</td>
<td>4.50</td>
<td>3.00</td>
<td>1.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.99)</td>
<td>(1.28)</td>
<td>(1.15)</td>
<td>(.71)</td>
<td>(1.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>Performance</td>
<td>3.56</td>
<td>3.63</td>
<td>4.50</td>
<td>3.00</td>
<td>4.49</td>
<td>2.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.94)</td>
<td>(1.11)</td>
<td>(.66)</td>
<td>(0.00)</td>
<td>(.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>Value</td>
<td>4.10</td>
<td>4.04</td>
<td>3.78</td>
<td>4.17</td>
<td>3.58</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.85)</td>
<td>(1.10)</td>
<td>(1.07)</td>
<td>(1.18)</td>
<td>(1.57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectancy</td>
<td>Expectancy</td>
<td>4.44</td>
<td>4.40</td>
<td>3.00</td>
<td>4.00</td>
<td>4.75</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.83)</td>
<td>(.91)</td>
<td>(2.00)</td>
<td>(1.41)</td>
<td>(.50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Standard deviations appear in parentheses below means.
Table 4.6

Motivation Measures at Each Time Point by Teacher

<table>
<thead>
<tr>
<th>Measure</th>
<th>Teacher</th>
<th>Teacher</th>
<th>Teacher</th>
<th>Teacher</th>
<th>Teacher</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simons</td>
<td>Taylor</td>
<td>Orville</td>
<td>Nelson</td>
<td>Edwards</td>
<td></td>
</tr>
<tr>
<td>T1 Mastery 1</td>
<td>4.83</td>
<td>4.94</td>
<td>4.64</td>
<td>4.26</td>
<td>4.60</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td>(.41)</td>
<td>(.25)</td>
<td>(.70)</td>
<td>(1.15)</td>
<td>(.68)</td>
<td></td>
</tr>
<tr>
<td>T2 Mastery 1</td>
<td>4.33</td>
<td>4.60</td>
<td>4.65</td>
<td>4.77</td>
<td>4.65</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(.58)</td>
<td>(.91)</td>
<td>(.67)</td>
<td>(.53)</td>
<td>(.75)</td>
<td></td>
</tr>
<tr>
<td>T3 Mastery 1</td>
<td>4.82</td>
<td>5.00</td>
<td>5.00</td>
<td>4.74</td>
<td>4.86</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>(.39)</td>
<td>(.00)</td>
<td>(.00)</td>
<td>(.56)</td>
<td>(.36)</td>
<td></td>
</tr>
<tr>
<td>T1 Mastery 2</td>
<td>4.33</td>
<td>4.31</td>
<td>4.08</td>
<td>4.26</td>
<td>4.15</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>(.82)</td>
<td>(.60)</td>
<td>(.70)</td>
<td>(.87)</td>
<td>(.99)</td>
<td></td>
</tr>
<tr>
<td>T2 Mastery 2</td>
<td>4.00</td>
<td>4.04</td>
<td>4.65</td>
<td>4.09</td>
<td>4.40</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(.94)</td>
<td>(.49)</td>
<td>(.92)</td>
<td>(.75)</td>
<td></td>
</tr>
<tr>
<td>T3 Mastery 2</td>
<td>4.35</td>
<td>4.05</td>
<td>4.29</td>
<td>3.68</td>
<td>4.45</td>
<td>1.88</td>
</tr>
<tr>
<td></td>
<td>(.86)</td>
<td>(.92)</td>
<td>(.95)</td>
<td>(1.25)</td>
<td>(.76)</td>
<td></td>
</tr>
<tr>
<td>T1 Mastery 3</td>
<td>4.67</td>
<td>4.88</td>
<td>4.48</td>
<td>4.26</td>
<td>4.50</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>(.82)</td>
<td>(.34)</td>
<td>(1.16)</td>
<td>(1.05)</td>
<td>(.76)</td>
<td></td>
</tr>
<tr>
<td>T2 Mastery 3</td>
<td>4.67</td>
<td>4.36</td>
<td>4.70</td>
<td>4.55</td>
<td>4.65</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>(.58)</td>
<td>(.81)</td>
<td>(.57)</td>
<td>(.80)</td>
<td>(.67)</td>
<td></td>
</tr>
<tr>
<td>T3 Mastery 3</td>
<td>3.88</td>
<td>4.57</td>
<td>4.71</td>
<td>3.89</td>
<td>4.50</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td>(1.27)</td>
<td>(.68)</td>
<td>(.49)</td>
<td>(1.45)</td>
<td>(.76)</td>
<td></td>
</tr>
<tr>
<td>T1 Performance</td>
<td>3.38</td>
<td>3.38</td>
<td>3.75</td>
<td>3.96</td>
<td>3.43</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>(1.13)</td>
<td>(1.02)</td>
<td>(.87)</td>
<td>(1.05)</td>
<td>(.95)</td>
<td></td>
</tr>
<tr>
<td>T2 Performance</td>
<td>4.17</td>
<td>3.60</td>
<td>3.45</td>
<td>3.95</td>
<td>3.75</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>(.38)</td>
<td>(1.09)</td>
<td>(.78)</td>
<td>(.86)</td>
<td>(.90)</td>
<td></td>
</tr>
<tr>
<td>T3 Performance</td>
<td>3.63</td>
<td>3.42</td>
<td>3.68</td>
<td>4.11</td>
<td>3.36</td>
<td>1.84</td>
</tr>
<tr>
<td></td>
<td>(.91)</td>
<td>(.85)</td>
<td>(.89)</td>
<td>(.95)</td>
<td>(1.11)</td>
<td></td>
</tr>
<tr>
<td>T1 Value</td>
<td>3.67</td>
<td>4.31</td>
<td>4.05</td>
<td>3.88</td>
<td>4.30</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>(.87)</td>
<td>(.46)</td>
<td>(.77)</td>
<td>(1.14)</td>
<td>(.86)</td>
<td></td>
</tr>
<tr>
<td>T2 Value</td>
<td>4.11</td>
<td>4.08</td>
<td>4.00</td>
<td>4.02</td>
<td>4.28</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>(.51)</td>
<td>(.91)</td>
<td>(.61)</td>
<td>(.71)</td>
<td>(.78)</td>
<td></td>
</tr>
<tr>
<td>T3 Value</td>
<td>4.10</td>
<td>4.02</td>
<td>3.43</td>
<td>4.07</td>
<td>4.18</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>(.88)</td>
<td>(.70)</td>
<td>(1.56)</td>
<td>(.93)</td>
<td>(.96)</td>
<td></td>
</tr>
<tr>
<td>T1 Expectancy</td>
<td>4.17</td>
<td>4.19</td>
<td>4.20</td>
<td>4.47</td>
<td>4.25</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(1.60)</td>
<td>(.54)</td>
<td>(.87)</td>
<td>(.84)</td>
<td>(1.07)</td>
<td></td>
</tr>
<tr>
<td>T2 Expectancy</td>
<td>5.00</td>
<td>4.32</td>
<td>4.45</td>
<td>4.27</td>
<td>4.21</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>(.00)</td>
<td>(.85)</td>
<td>(.51)</td>
<td>(.83)</td>
<td>(1.08)</td>
<td></td>
</tr>
<tr>
<td>T3 Expectancy</td>
<td>4.41</td>
<td>4.52</td>
<td>4.14</td>
<td>4.58</td>
<td>4.05</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>(.80)</td>
<td>(.68)</td>
<td>(1.22)</td>
<td>(.61)</td>
<td>(1.31)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Standard deviations appear in parentheses below means.
A one-way ANOVA was also conducted to examine differences for each measure at each time by prior achievement. Results are shown in Table 4.7. Prior achievement was measured by a composite score on a previously taken standardized test (Terra Nova) converted into a categorical variable with five groups as described in the methods chapter. The only significant finding was that students’ expectancy for success in the third survey administration, the standardized testing context, varied significantly by prior achievement, $F(4, 73) = 4.13, p = .005$; the effect size was large ($\eta^2 = .184$). The means were 3.25 (SD = 1.71) for the first group (lowest test scores), 3.80 (SD = 1.79) for the second group, 4.14 (SD = .69) for the third group, 4.50 (SD = .67) for the fourth group, and 4.65 (SD = .53) for the fifth group (highest test scores). Thus, the means for expectancy for success on the standardized test increased with higher prior achievement scores.
Table 4.7

Motivation Measures by Terra Nova Score Category

| Measure | Terra Nova Score | | | | |
|---------|-----------------|---|---|---|---|---|---|---|---|---|---|
|         | 0-20 | 21-40 | 41-60 | 61-80 | 81-100 | F |
| Assessment | | | | | | |
| Mastery 1 | 4.13 | 4.13 | 4.58 | 4.64 | 4.70 | 1.43 |
| | (.85) | (.75) | (.80) | (.54) | (.61) | |
| Mastery 2 | 3.75 | 4.50 | 4.67 | 4.18 | 4.33 | 1.49 |
| | (.50) | (.58) | (.41) | (.59) | (.70) | |
| Mastery 3 | 4.38 | 4.25 | 4.67 | 4.54 | 4.58 | 0.26 |
| | (.48) | (.50) | (.26) | (.86) | (.65) | |
| Performance | 4.03 | 3.88 | 3.31 | 3.85 | 3.50 | 1.16 |
| | (.19) | (.74) | (.83) | (.71) | (.92) | |
| Value | 3.88 | 3.83 | 4.33 | 4.09 | 4.17 | 0.50 |
| | (.28) | (.43) | (.52) | (.74) | (.72) | |
| Expectancy | 4.75 | 3.63 | 4.25 | 4.35 | 4.40 | 1.70 |
| | (.50) | (1.75) | (.52) | (.62) | (.53) | |
| Standardized | | | | | | |
| Mastery 1 | 4.50 | 5.00 | 5.00 | 4.78 | 4.93 | 2.04 |
| | (.58) | (.00) | (.00) | (.52) | (.26) | |
| Mastery 2 | 3.25 | 4.80 | 4.43 | 4.35 | 4.00 | 2.08 |
| | (.50) | (.45) | (.78) | (.78) | (1.13) | |
| Mastery 3 | 4.75 | 4.80 | 4.43 | 4.09 | 4.13 | 0.77 |
| | (.50) | (.45) | (.54) | (1.16) | (1.24) | |
| Performance | 3.81 | 3.90 | 3.32 | 3.84 | 3.46 | 0.86 |
| | (.66) | (1.07) | (.86) | (.93) | (1.00) | |
| Value | 3.42 | 4.33 | 4.14 | 3.98 | 4.14 | 0.83 |
| | (.63) | (1.13) | (.69) | (.98) | (.80) | |
| Expectancy | 3.25 | 3.80 | 4.14 | 4.50 | 4.65 | 4.13* |
| | (1.71) | (1.79) | (.69) | (.67) | (.53) | |

Note: Standard deviations appear below means. *p < .05.

Simple linear regression was then performed to further analyze the predictive value of prior testing achievement on expectancy for success in standardized testing.

Results indicated a significant association, $b = 3.16$, $t(75) = 9.58$, $p < .001$, with 17.4% of the variance in expectancy for success accounted for by prior achievement, $R^2 = .174$, $F(1, 76) = 16.00$, $p < .001$. That is, when thinking about their performance on an
upcoming standardized test, students’ expectancy for success on that test was
significantly related to their performance on a previous standardized test. However, there
was no significant association between prior test performance and expectancy for success
in the classroom assessment context. This difference in expectancy for success will be
discussed in Chapter 5.

Primary Analysis

The primary analysis for this study comprised of paired sample t-tests to
determine whether or not there were significant differences between the two assessment
contexts in each measure of student motivation. Means, standard deviations, t statistics,
and p values are shown in Table 4.8 for each measure in both the Classroom Assessment
(C.A.) and Standardized Testing (S.T.) contexts. T-tests revealed only two significant
differences in students’ perceptions of the classroom motivational measures between the
two contexts, both pertaining to mastery goal structure items. Responses to items number
one, “For this test, trying hard is very important,” and three, “For this test, it’s ok to
make mistakes as long as you are learning,” differed significantly between the classroom
assessment and standardized testing contexts. For item number one, students reported a
significantly lower level of agreement in the classroom assessment (M=4.63, SD=.61)
than the standardized testing (M=4.88, SD=.37) contexts; t(81) = -3.48, p = .001. For
item number three, students reported a significantly higher level of agreement in the
classroom assessment (M=4.56, SD=.72) than the standardized testing (M=4.21,
SD=1.14) contexts; t(80) = 3.00, p = .004. T-tests for all other items and measures
revealed insignificant differences between the two contexts.
### Table 4.8

Mean, Standard Deviation, t Statistics, and p Values for Measures

<table>
<thead>
<tr>
<th></th>
<th>Mean (C.A.)</th>
<th>SD (S.T.)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mastery 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.A. (n=82)</td>
<td>4.63</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.T. (n=82)</td>
<td>4.88</td>
<td>0.37</td>
<td>-3.48*</td>
<td>.001</td>
</tr>
<tr>
<td>Mastery 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.A. (n=81)</td>
<td>4.27</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.T. (n=81)</td>
<td>4.14</td>
<td>0.98</td>
<td>1.20</td>
<td>.233</td>
</tr>
<tr>
<td>Mastery 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.A. (n=81)</td>
<td>4.56</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.T. (n=81)</td>
<td>4.21</td>
<td>1.14</td>
<td>3.00*</td>
<td>.004</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.A. (n=80)</td>
<td>3.68</td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.T. (n=80)</td>
<td>3.68</td>
<td>0.97</td>
<td>-0.04</td>
<td>.972</td>
</tr>
<tr>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.A. (n=79)</td>
<td>4.11</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.T. (n=79)</td>
<td>4.10</td>
<td>0.90</td>
<td>0.37</td>
<td>.713</td>
</tr>
<tr>
<td>Expectancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.A. (n=79)</td>
<td>4.31</td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.T. (n=79)</td>
<td>4.42</td>
<td>0.91</td>
<td>-1.14</td>
<td>.260</td>
</tr>
</tbody>
</table>

*Note.* *p < .01.*
Mixed Methods Results

A mixed methods approach was chosen for this study in order to examine the research questions from different perspectives and to integrate findings from both the qualitative and quantitative phases in the interpretation of the results. To this end, the quantitative survey data were collected to explore the degree to which students’ perceptions reflected the findings of the qualitative inquiry in this study. The presumption was that students would perceive the motivational climates in the classroom contexts in a similar way as the observers and interviewers interpreted. As the results show, there was little evidence to support this presumption for most measures. Survey results revealed that students perceived the classroom motivational climates to be the same for all but two items during regular classroom assessments and standardized testing preparation, whereas the researcher interpreted the climates to be drastically different.

Having the perspective of the students from the quantitative survey data presents a challenge in the interpretation of the results. However, the combined qualitative and quantitative results provide a more complete portrayal of the classroom environment than that which would have been available had only one or the other of the methods been used. There are many potential reasons why the results were not consistent between the students’ and researcher’s perspectives, including severe limitations of the survey instrument: these will be explored further in Chapter 5.

Summary of Results

This study included classroom observations and teacher interviews, analyzed qualitatively, and student surveys, analyzed quantitatively. Qualitative results revealed
strong differences in teachers’ instructional practices and underlying knowledge and beliefs between the classroom assessments context and the standardized testing context that were consistent with the theoretical frameworks of achievement goal theory and expectancy-value theory.

Quantitative results indicated that students did not perceive the classroom motivational climates as different between the two contexts, with two exceptions: They responded with higher agreement to the statement “Trying hard is really important” in the standardized testing context and to the statement “It’s ok to make mistakes as long as you are learning” in the classroom assessment context. The latter finding is consistent with the qualitative results while the former is in opposition to the interpretation of classroom observations and the literature on mastery goal structures. Those two items were the only evidence in support of the hypothesis that students would perceive the classroom motivational climate to be different in the two contexts. Students did not report significant differences in perceptions of any of the other measures.

Taken together, the data suggest that there is little evidence to support the presumption that the students and observers perceived the classroom environment in a similar way, however, limitations in the survey instrument should be noted that may make interpretation of the results somewhat inconclusive. Further description will be provided in the next chapter.
Chapter 5: Discussion

This chapter presents a discussion of the results of the three research questions guiding this study:

1. How do elementary school math teachers’ instructional practices related to assessment differ for summative classroom assessments and preparation for standardized state achievement tests?

2. How do elementary school math teachers’ instructional practices reflect their knowledge, beliefs, and perceptions of external influences related to summative classroom assessments and standardized state achievement tests?

3. How do students perceive the classroom motivational climate surrounding assessment and standardized testing?

The findings related to each question will be discussed in turn in the subsequent sections of this chapter. Finally, sections will detail the limitations of this study and suggestions for future related research.

**Instructional Practices Across Assessment Contexts**

The first research question concerned how teachers’ instructional practices differed when preparing students for classroom math assessments as opposed to preparing for standardized tests. This question was investigated qualitatively through
classroom observations and teacher interviews. Two observations for each teacher occurred the days before classroom assessments and one observation occurred during the weeks leading up to the state standardized test, in which lessons were specifically focused on the preparation for that test. Teachers were interviewed after all observations took place to share their perspectives on the differences in their teaching between the two assessment contexts. The findings revealed that there were drastic differences in teachers’ instructional practices between the two contexts that point toward motivational implications for students.

From the theoretical framework of achievement goal theory, teachers generally engaged in practices known to be supportive of a mastery goal structure during their classroom assessment lessons. Teachers focused on student understanding and effort, dissuaded competition and social comparison, and emphasized that mistakes were opportunities for learning, all consistent with a mastery goal structure (Kaplan, Middleton, Urdan, & Midgley, 2002; Meece, Anderman, & Anderman, 2006). However, standardized testing preparation lessons were more indicative of a performance goal structure in each classroom. Teachers emphasized using test-taking strategies to earn as many points as possible, demonstrating their ability to the test makers and/or graders, and avoiding mistakes that might decrease their scores. This focus on the score and importance of demonstrating competence is consistent with theoretical descriptions of a performance goal structure (Kaplan, Middleton, Urdan, & Midgley, 2002; Meece, Anderman, & Anderman, 2006).

There were also clear differences in teachers’ instructional practices between the two contexts from the theoretical framework of expectancy-value theory. While teachers
communicated positive expectancy messages to students in both contexts, there were some differences in the tone of the messages sent. For example, many of the classroom assessment lessons included teacher language making attributions to students being “smart” and “capable” of doing well on their tests. In the context of standardized testing, the message was more about being well-prepared to succeed. The focus of teachers’ comments was not as much on knowing the content to be tested as it was on being knowledgeable about the format and structure of the standardized test. Students were told that they could expect to succeed on the test if they remembered their test-taking strategies and focused for the two and a half hour length of the test. There was also a cautionary tone, however, warning students not to get “tricked” by the test-makers and to avoid making mistakes in order to do well. In general, the focus of the teachers’ instruction, and subsequent messages to students about how to succeed, was on knowledge of mathematical content during classroom assessment lessons and on knowledge related to test-taking skills during standardized testing lessons.

Teacher instructional practices related to the subjective task value of the tests were less common than expectancy messages in both contexts. During preparation for classroom assessments, teachers occasionally talked about the importance of “showing what you know” on the test and working hard to get a good grade, which reflects utility value of the assessments (Wigfield & Eccles, 1992). They also created fun and engaging collaborative review lessons, such as Jeopardy-style games and centers with manipulatives, for their students that served to increase the intrinsic task value of reviewing for the test, if not for taking the test itself.
The value of the standardized tests was obviously communicated to the students through the teachers’ instructional practices. They stressed the importance of the test throughout the year through consistent reminders to students and as evidenced by the intensity of the review period in the weeks leading up to the test. Teachers used the term “important” frequently in reference to the test and testing strategies, but their comments did not reflect the expectancy-value theory definition of importance in terms of attainment value (Eccles & Wigfield, 2002). They were clearly told that their performance on the test was important, but at no time during the observations was there evidence of a rationale given for the tests. Teachers did not talk about the meaning or purpose of the tests during the observed lessons, nor did they tell students what types of consequences were attached to the tests.

One of the most striking differences in the two different contexts was related to intrinsic value. As mentioned above, classroom assessment review lessons were enjoyable and consisted of games or other group activities. In contrast, standardized test preparation lessons were all teacher-directed individual student work periods with worksheets or practices tests. Students complained of being “bored” during these lessons and were far less engaged than in the classroom assessment review lessons. It seemed that teachers were sending the message that the standardized test was a serious independent performance activity and not a very interesting or exciting event.

Finally, the recurrent use of the pronoun “they” to refer to standardized test makers and/or graders was another notable difference between the two contexts. Teachers consistently highlighted that the standardized tests were different from regular classroom tests because students had to follow different rules and strategies, such as writing only
inside the designated answer box, using neat handwriting, and not expecting help from
the teacher, due to the fact that the tests were imposed and graded by an external source.
This finding can be understood both by achievement goal theory and expectancy-value
theory. From an achievement goal perspective, the focus in the standardized testing
context was on demonstrating competence to the test graders (“them”), which is clearly
indicative of a performance goal structure (Midgley, 2002). In addition, teachers
frequently explained that they did not agree with the way the tests were constructed or
scored, but that was just the way that it had to be, communicating to students both a
devaluing of the task according to expectancy-value theory (Wigfield & Eccles, 1992)
and diminishing mastery-supportive student and teacher autonomy and meaningfulness of
the task (Midgley, 2002).

The differences in teachers’ instructional practices between the two contexts are
not surprising, given the literature and theoretical assumptions about the effects of high
stakes testing in schools. A vast body of literature explains that standardized testing
results in narrowing of curriculum and “teaching to the test” or focusing on test-taking
strategies rather than academic content (e.g. Au, 2007; Nichols & Berliner, 2007; Wood,
2004). The observations in this study revealed that a focus on preparing students for the
standardized tests was present throughout the year, with frequent reminders about
standardized test-type questions even during regular classroom assessment reviews.
During the weeks immediately preceding the standardized tests, instruction was explicitly
focused on teaching to the test and preparing students for testing conditions. Therefore,
this study supports the research on the curricular impact of standardized testing.
In addition, motivation research posits that the inherently performance-laden nature of standardized testing leads to more emphasis on teaching practices that foster a performance goal structure rather than a mastery goal structure (Anderman, Anderman, Yough, & Gimbert, 2010; Kowalski & Marietta-Brown, 2011). This study provides further empirical evidence to support that claim, showing that teachers altered their practices in the standardized testing context to emphasize scoring points, demonstrating competence, and avoiding mistakes while diminishing student autonomy and collaborative learning efforts. This study also contributes to the literature on expectancy-value theory, demonstrating how teachers communicate different messages to students in the two assessment contexts related to students’ likelihood of success and the value of the assessment tasks. Expectancy-value research predicts that students’ perceptions of these constructs can be influenced by teachers and can have an impact on students’ effort and achievement on standardized tests (e.g. Wigfield & Eccles, 2000; Wise & DeMars, 2005). This study provides empirical evidence of how teachers’ instructional practices communicate messages to students about expectancy and value, potentially impacting students’ perceptions of, and subsequent achievement on, assessment tasks.

**Teachers’ Knowledge and Beliefs Across Assessment Contexts**

The second research question of this study was focused on the teachers’ knowledge, beliefs, and perceptions of external pressure surrounding classroom assessment and standardized testing. Open-ended, semi-structured interviews were used along with classroom observational data to investigate this question. Interpretation of results indicated that teachers’ instructional practices were largely reflective of their
knowledge, beliefs, and perceptions of external pressure related to assessment and
standardized testing.

Consistent with research on teachers’ assessment-related knowledge (e.g., Impara,
Plake, & Fager, 1993; Volante, Beckett, Reid, & Drake, 2010), teachers reported having
little to no formal training in assessment and standardized testing during their teacher
preparation experiences. Rather, they explained that they learned about assessment and
testing through experience, publisher-provided materials, and consulting with peer
teachers. I did not employ any measures to formally assess their knowledge, as I wanted
the interviews to remain comfortable and non-threatening to the teachers, but I did try to
gauge a general sense of their knowledge about assessment principles from their
responses to the questions about their beliefs and practices. My contention is that teachers
seemed to have a better understanding of the classroom assessment process than
standardized testing. For example, teachers at Martin Grove spoke articulately about their
flexible ability grouping and use of pre- and post-tests to measure student growth and
track mastery of grade-level standards, indicating a solid knowledge base of assessment
principles (Shepard et al., 2005). Teachers also made references to their use of both
formative and summative assessment techniques and the use of assessment results to
inform future instruction, consistent with assessment literature (Black & William, 1998;
Shepard et al., 2005; Stiggins & Bridgeford, 1985). However, when asked about
standardized testing, they made vague comments or indicated incomplete understanding,
such as simply stating that the purpose of testing is “the government” (Ms. Taylor). Ms.
Orville summed up her vague knowledge of the testing policy and its purpose by stating,

“The has to be some kind of data that they can give you...like some kind of bar that you
want your teachers to jump to.” Though not specifically questioned about their knowledge of federal or state standardized testing policies or best practices related to testing and the use of performance data, the lack of substantive conversation about these topics during the interviews could be attributed to a lack of knowledge.

Another explanation could be that rather than indicating a lack of knowledge, teachers’ hesitancy to talk about standardized testing policies and practices was related to their beliefs about the testing process. That is, the teachers may know and understand the testing process very well, but they may not fully endorse the tests because of their negative views of the policies or tests in general. Their perceptions of external pressure related to the tests may also contribute to their views on the testing process. As they did in observed lessons, teachers talked about testing in a way that made it clear that the process was externally imposed on them and their students. Teachers frequently used the pronoun “they” to refer to the test makers and/or graders and were articulate about the pressure they felt due to the tests. Therefore, in analyzing the relation to teaching practice, it is difficult to separate the impact of knowledge, beliefs, and perceptions of pressure, as is often the case with research on teachers’ knowledge and beliefs (Woolfolk Hoy, Davis, & Pape, 2006). Consequently, I will consider these constructs together in the remainder of the discussion.

Whether indicative of knowledge or beliefs, the difference in the amount and quality of discussion about classroom assessments as opposed to standardized testing in the interviews was reflected in the observed classroom lessons. As discussed above, teachers communicated to students the importance of showing what they know and demonstrating growth during classroom assessment review lessons but did not provide
any justification for the importance of doing well on standardized tests in the observed lessons. Teachers did demonstrate knowledge of test-taking strategies and scoring policies by having students practice writing in the boxes and answering all parts of complex questions, for example, but they did not divulge any information about the purpose or meaning of the tests or policies.

Teachers acknowledged the differences in the ways that they communicated to students about the different types of tests. That is, in interviews, teachers reported that they realized that students were given different messages about the tests, conveying that the standardized tests really were important and that students had to take them seriously. They focused on the test preparation more intensely with less “fun” than they did for regular classroom assessment reviews. This was consistent with their beliefs about the tests being important and students needing to focus and do their best, though the source of importance was more about external pressure and less about personal importance. In this way, the teachers were using the term “importance” to refer to what expectancy-value theory considers utility value rather than attainment/importance value (Eccles and Wigfield, 2002). Teachers talked openly about their feelings of pressure and stress related to the way the test scores were used to evaluate them as teachers. While they did not explain these feelings to their students during any observed lessons, their emphasis on preparing the students for the tests could be understood as a reflection of their perception of this pressure.

Finally, teachers were vocal, both in their interviews and in observed lessons, about their disagreement with aspects of the testing policy and/or process. Teachers made negative comments about the tests during lessons, such as calling the rule about writing
only in the box “dumb” or stating that they would not choose to evaluate students in the same way as the test does. These comments were honest reflections of the teachers’ beliefs about the standardized tests. All teachers expressed frustration about the test being too long, the grading policies being too rigid, the lack of individualization, and the inappropriateness of the use of test scores as the singular means of evaluating the teachers and school.

These results are consistent with the limited literature on teachers’ views of standardized testing (i.e. Assaf, 2008; Cimbricz, 2002) and add to that literature by linking teachers’ beliefs with their practices in the classroom. In particular, this study demonstrates that teachers’ instructional practices are generally reflective of their knowledge, beliefs, and perceptions of external pressure related to testing.

**Students’ Motivational Perceptions Across Assessment Contexts**

The third and final research question for this study asked how students perceived the classroom motivational climates surrounding regular math assessments and standardized testing. It was hypothesized that there would be a difference in student perceptions between the two contexts that would mirror the differences in teacher instructional practices discovered in the first part of this study. Short, simple surveys were used to measure student perceptions related to classroom goal structure and expectancy-value theory. The surveys were administered immediately following each classroom observation and asked students to answer the questions while thinking about the test that they were going to be taking soon, either a classroom assessment (for the first two observations) or a standardized test (for the third observation).
In contrast to the hypothesis, results from the surveys indicated that students did not perceive significant differences between the classroom motivational climates for the two testing contexts, with two exceptions: They responded with higher agreement to the statement “Trying hard is really important” in the standardized testing context and to the statement “It’s ok to make mistakes as long as you are learning” in the classroom assessment context. Both of those items assessed perceptions of mastery goal structures, thus the results indicate that students had inconsistent perceptions of mastery goal structures between the two contexts, because responses were higher for one item in the standardized testing context and the other item in the classroom assessment context.

The finding that students responded with higher agreement to the item about mistakes in the classroom assessment context than in the standardized testing context is consistent with what would be expected from the literature and from the observations and interviews in this study. Achievement goal theory suggests that treating mistakes as opportunities for learning is consistent with a mastery goal structure, and mastery goals are more likely to be supported in classroom assessment contexts than standardized testing contexts (i.e. Anderman & Anderman, 2010; Kowalski & Marietta-Brown, 2011).

Throughout this study, teachers praised students for mistakes in the classroom assessment lessons, but warned them not to make mistakes in the standardized testing preparation lessons. This distinction might send messages to students about the different purposes of the different types of assessments. For example, it is good to make mistakes in the classroom assessment context because those mistakes help students to learn and improve, whereas it is not acceptable to make mistakes on the standardized test because the purpose of that test is only to demonstrate competency. Teachers acknowledged these
differences and their potential impact on students’ motivation in their interviews. Thus, it is not surprising that students perceived this difference in the same way as the researcher.

The other finding was that students responded higher to the item “Trying hard is very important” in the standardized testing context than in the classroom assessment context. This item was intended to assess perceptions of mastery goal structures, as an emphasis on effort (i.e. “trying hard”) rather than only achievement is a feature of mastery goal supportive classrooms (Ames, 1992). Therefore, results indicate that students perceived more emphasis on mastery goal structure in the standardized testing context than the classroom assessment context, in contrast with what would be expected based on the literature and the other findings from this study (i.e. Anderman & Anderman, 2010; Anderman, Anderman, Yough, & Gimbert, 2010; Kowalski & Marietta Brown, 2011). However, the wording of this item does not specify that effort is independent of performance, so students could have understood the item to mean that it was important to try hard in order to do well on the test. The item also includes the words “very important,” which may have influenced students’ responses. Results from the observations and interviews revealed that teachers used the word “important” to talk about the standardized tests much more than classroom assessments. Thus, this item may have been problematic in terms of students’ interpretations and responses.

For the remainder of the items and measures on the survey, students did not report perceived differences in the motivational climates of the two assessment contexts. One explanation is that the students actually did not perceive a difference between the contexts. This explanation is unlikely due to the drastic differences noticed by the observers in both teachers’ practices and students’ engagement during the lessons.
Another potential explanation could be that the survey items were not written in a way that adequately assessed the perceptions being measured. For example, the item about the importance of trying hard on the test may not have been interpreted by students as reflecting a mastery goal structure, as explained above. Upon further review of the survey, items measuring task values and expectancy for success represented more of individual perceptions than perceptions of classroom motivational climate, as intended. For example, the item reading “Taking this math test will be useful for me in the future” assessed a personal belief about the task value of the test. Perhaps it could have been worded in a way to more accurately gage the classroom motivational climate related to task value, such as “In this class, we are told that taking this math test will be useful in the future.” In future studies, the wording of survey items should be more carefully considered and, especially with young participants, cognitive interviewing to assess the validity of survey items would be beneficial.

Another likely explanation for the lack of significant findings is that students were simply too young to differentiate cognitively between the levels of agreement on the survey. Although efforts were made to address this issue, such as going through practice questions and responses with the students, using a pictorial smiley-face scale rather than a numbered Likert-type scale, and reading the survey aloud slowly while students completed it, it is possible that the responses were not true indications of students’ perceptions. While survey research has been conducted with students in third grade and even younger (e.g., Wigfield, et al., 1997), there is evidence that this type of inquiry may tend toward low reliability and validity due to the considerable difficulty that young participants have in comprehending intended meanings of survey questions (Hartigan,
Furthermore, while the current survey was intentionally kept short to be cognitively manageable for the young students, it may have had higher reliability and been able to detect more differences if it had been longer with more items for each measure. These limitations could be considered as contributing to the lack of significant findings from the survey phase of this study.

One notable finding related to students’ perceptions was the association between students’ prior achievement and their expectancy for success in the standardized testing context only. The results of the ANOVA indicated that students’ expectancy for success on standardized tests increased with higher performance on a different standardized test, the Terra Nova test, taken the previous year. These results are not surprising, as prior achievement is often found to be related to expectations for future success (Wigfield & Eccles, 2000). However, the association between prior achievement and expectancy for success was not present in the classroom assessment context. That is, students did not perceive their prior success on standardized testing to be predictive of their future success on classroom assessments. This supports the overall suggestion that students do perceive differences between the two contexts, as they do not see their performance on one type of test to be related to their performance on the other. The potential implications of this finding for student learning, motivation, and achievement merits further investigation in future research.

**Limitations**

In considering the current findings, it is important to note several limitations to this study. First of all, the small sample size of five classrooms and the lack of geographic diversity, in that they are all in the same school district, limit the generalizability of the
findings. The need for parental permission further limited the student sample size because several students failed to return signed permission forms to participate. Although this small sample size limits the generalizability of the quantitative analysis, this was an intentional decision in order to keep a manageable-sized study for the desired level of rich description from the qualitative data collected.

Another limitation was that the participants were not blind to the purpose of the study. In order to schedule observations during the necessary assessment activities, the teachers had to be informed of the focus of the study. Efforts were made to encourage them to teach normally and ignore the observer(s), such as verbally requesting that they not alter their lessons and visiting the classroom casually prior to observations in order to familiarize the students with our presence. There was no evidence to suggest that teachers changed their teaching as a result of the presence of the observer(s). Rather, both teachers and students seemed unaffected by visitors to the classroom. Nonetheless, it should be noted that the validity of the study might have been compromised if the teachers altered their practices due to the nature of the research.

Perhaps the most serious limitation to this study was posed by the young age of the students, discussed briefly above. The decisions to focus on third grade classrooms was important, because third grade is the first year of required state standardized testing and the implications of the introduction of that type of testing was of central interest in this study. However, third graders may have been too young to adequately complete the surveys. This limitation was addressed by keeping the survey short and cognitively appropriate for the age group and by administering the survey aloud as the students
completed it. Despite the efforts to circumvent this limitation, though, it may have contributed to the lack of significant findings from the survey portion of this study.

Finally, it should be noted that teachers are only one source of influence on students’ beliefs and motivation for assessments. This study is limited in that it does not factor in messages students receive from parents, school administrators, peers, or other influential sources about the meaning and purpose of different assessments. These other areas of influence should be examined in additional studies related to this topic.

**Direction for Further Research**

This study provided evidence of how teachers’ instructional practices differ between classroom assessment and standardized testing preparation contexts in third grade math classes, in a single school district. Further research should expand this investigation to examine differences between high and low performing schools and to look at additional grade levels. Prior research suggests that teacher’s instructional practices related to standardized testing vary by performance level of the school, with teachers in lower-performing schools feeling the pressure of sanctions and engaging in more “teaching to the test” and narrowing curriculum than teachers in higher-performing schools (e.g., Dooley & Assaf, 2009). Also, developmental differences in students’ perceptions of and motivation for testing have been documented (e.g., Paris, Lawton, Turner, & Roth, 1991) and it would be interesting to examine teachers’ practices across different grade levels to determine if and how teachers may contribute to students’ experiences. This study specifically focused on third grade because of the introduction of federally mandated standardized testing in that year, but exploring the differences in other grade levels would also be valuable, especially if older students might be more able
to distinguish and report differences in their perceptions of classroom motivational climate.

Students in this study did not report many significant differences in the perceived motivational climate between the two assessment contexts, but there were several limitations to the student survey component, as discussed above. Further research should investigate young students’ perceptions through different methodological approaches, such as interviews, focus groups, or video-simulated recall in order to better understand young students’ reactions to teachers’ instructional practices. Alternatively, the survey could be refined based on the results of this study to include questions that are more closely aligned with the types of differences observed in the two contexts and potential motivational impact of those differences. Subsequently, the survey could be used to study students’ perceptions of the motivational climate in a wider variety of grade levels to examine developmental differences as well as differences between the two contexts.

Summary

In summary, this study provided evidence of how teachers’ instructional practices related to assessment differed between classroom assessment and standardized testing preparation contexts. Teachers engaged in more mastery goal supportive teaching strategies in the classroom assessment context and more performance goal supportive teaching strategies in the standardized testing context. They also spoke differently about the value of the two types of tests and about their expectations for students’ success. Thus, the introduction of standardized testing in third grade classrooms represents an area of observable changes in teachers’ practices that have important theoretical implications for students’ motivation. The differences were reflective of teachers’ knowledge, beliefs,
and perceptions of external pressure related to the different types of assessments. Student surveys provided little evidence to support the hypothesis that they would perceive the classroom motivational climate as being different in the two contexts, however, severe limitations to the survey design may have contributed to the lack of significant findings. Further research should continue to explore the important motivational implications of teachers’ instructional practices related to classroom assessments and standardized testing.
References


Appendix A: IRB Approval
October 5, 2011

Protocol Number: 2011B0349
Protocol Title: TEACHERS' INFLUENCE ON STUDENT MOTIVATION IN DIFFERENT ASSESSMENT CONTEXTS, Lynley Anderman, Monica Kowalski, Policy and Leadership
Type of Review: Initial Review—Expedited
IRB Staff Contact: Jacob R. Stoddard
Phone: 614-292-0526
Email: stoddard.13@osu.edu

Dear Dr. Anderman,

The Behavioral and Social Sciences IRB APPROVED BY EXPEDITED REVIEW the above referenced research. The Board was able to provide expedited approval under 45 CFR 46.110(b)(1) because the research meets the applicability criteria and one or more categories of research eligible for expedited review, as indicated below.

Date of IRB Approval: October 4, 2011
Date of IRB Approval Expiration: September 19, 2012
Expedited Review Category: 7

In addition; the protocol has been approved for the inclusion of children (permission of one parent sufficient).

If applicable, informed consent (and HIPAA research authorization) must be obtained from subjects or their legally authorized representatives and documented prior to research involvement. The IRB-approved consent form and process must be used.

Changes in the research (e.g., recruitment procedures, advertisements, enrollment numbers, etc.) or informed consent process must be approved by the IRB before they are implemented (except where necessary to eliminate apparent immediate hazards to subjects).

This approval is valid for one year from the date of IRB review when approval is granted or modifications are required. The approval will no longer be in effect on the date listed above as the IRB expiration date. A Continuing Review application must be approved within this interval to avoid expiration of IRB approval and cessation of all research activities. A final report must be provided to the IRB and all records relating to the research (including signed consent forms) must be retained and available for audit for at least 3 years after the research has ended.

It is the responsibility of all investigators and research staff to promptly report to the IRB any serious, unexpected and related adverse events and potential unanticipated problems involving risks to subjects or others.

This approval is issued under The Ohio State University’s OHRP Federalwide Assurance #00006378. All forms and procedures can be found on the ORRP website – www.orrp.osu.edu. Please feel free to contact the IRB staff contact listed above with any questions or concerns.
Shari R. Speer, PhD, Chair
Behavioral and Social Sciences Institutional Review Board

Behavioral and Social Sciences Institutional Review Board

Office of Responsible Research Practices
300 Research Administration Building
1960 Kenny Road
Columbus, OH 43210-1063
Phone (614) 688-8457
Fax (614) 688-0366

www.orrp.osu.edu
Appendix B: Email to Recruit Participants
Dear _____________,

My name is Monica Kowalski and I am a PhD candidate in Educational Psychology at The Ohio State University. I am writing to request your participation in my dissertation research. My topic is the influence of teachers’ instructional practices on student motivation in different assessment contexts. I am conducting my research in third grade classrooms during the 2011-2012 school year in a variety of local schools and I have identified your school as a desirable site for this research. I hope you will consider allowing me to include your class in my study.

The study will involve observing and taking field notes during math lessons two times during normal classroom weeks and once during the month prior to OAT testing in the spring. I will be investigating typical classroom events and interactions, so I will not be evaluating or judging your teaching in any way. I will work with you to schedule appropriate observation times. Immediately following each observation, I will administer a short survey to your students to assess their levels of motivation. After all observations have taken place, I would like to interview you individually to learn more about your beliefs about assessments and standardized testing. I will not interfere with any classroom activities and I will use pseudonyms for the teachers and the schools in my reports, protecting the confidentiality of all participants. Your participation, and that of your students, will be completely voluntary and you will be able to withdraw at any time without penalty or repercussion. In appreciation for your participation, I will provide a small monetary contribution for your personal use to benefit you and your class (up to $100).

If you are willing to participate in this project, please contact me at the email or phone number listed below. I would be happy to answer any questions you may have.

Thank you for your consideration and I look forward to hearing from you,

Monica Kowalski
kowalski.96@osu.edu
614-570-9282
Appendix C: Consent Forms
# The Ohio State University Consent to Participate in Research

<table>
<thead>
<tr>
<th>Study Title:</th>
<th>Teachers’ Influence on Student Motivation in Different Assessment Contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher:</td>
<td>Lynley H. Anderman, Monica J. Kowalski</td>
</tr>
<tr>
<td>Sponsor:</td>
<td>N/A</td>
</tr>
</tbody>
</table>

This is a consent form for research participation. It contains important information about this study and what to expect if you decide to participate.

**Your participation is voluntary.**

Please consider the information carefully. Feel free to ask questions before making your decision whether or not to participate. If you decide to participate, you will be asked to sign this form and will receive a copy of the form.

**Purpose:**
This is a study to understand how teachers influence students’ motivation about math tests (classroom and standardized).

**Procedures/Tasks:**
A researcher will be observing your teaching during three math lessons this year. After each lesson, a short survey will be given to students asking them about their motivation for an upcoming math test. The survey will take 5-10 minutes. Later, the researchers would like to access students’ math grades and scores on math standardized tests to investigate how their motivation matches with their achievement. You will also be asked to participate in a short (approximately 30 minutes) one-on-one interview with a researcher about your beliefs about testing.

**Duration:**
All of the research will take place during the 2011-2012 school year. You may leave the study at any time. If you decide to stop participating in the study, there will be no penalty to you, and you will not lose any benefits to which you are otherwise entitled. Your decision will not affect your future relationship with The Ohio State University.

**Risks and Benefits:**
Participating in this study is an opportunity to help us understand how teachers can motivate students for testing. The risks of participating in this study are minimal. Your privacy and confidentiality are going to be protected.

**Confidentiality:**

Efforts will be made to keep your study-related information confidential. However, there may be circumstances where this information must be released. For example, personal information regarding your participation in this study may be disclosed if required by state law. Also, your records may be reviewed by the following groups (as applicable to the research):

- Office for Human Research Protections or other federal, state, or international regulatory agencies;
- The Ohio State University Institutional Review Board or Office of Responsible Research Practices;
- The sponsor, if any, or agency (including the Food and Drug Administration for FDA-regulated research) supporting the study.

**Incentives:**

In appreciation for your time, a small monetary incentive will be given at the end of the study intended for your personal use for your classroom. If you choose to leave the study early, a partial amount will be given.

**Participant Rights:**

You may refuse to participate in this study without penalty or loss of benefits to which you are otherwise entitled. If you are a student or employee at Ohio State, your decision will not affect your grades or employment status.

If you choose to participate in the study, you may discontinue participation at any time without penalty or loss of benefits. By signing this form, you do not give up any personal legal rights you may have as a participant in this study.

An Institutional Review Board responsible for human subjects research at The Ohio State University reviewed this research project and found it to be acceptable, according to applicable state and federal regulations and University policies designed to protect the rights and welfare of participants in research.

**Contacts and Questions:**

For questions, concerns, or complaints about the study you may contact Monica Kowalski 614-570-9282.

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you
may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.

If you are harmed as a result of participating in this study or for questions about a study-related harm, you may contact Monica Kowalski 614-570-9282.


**Signing the consent form**

I have read (or someone has read to me) this form and I am aware that I am being asked to participate in a research study. I have had the opportunity to ask questions and have had them answered to my satisfaction. I voluntarily agree to participate in this study.

I am not giving up any legal rights by signing this form. I will be given a copy of this form.

<table>
<thead>
<tr>
<th>Printed name of subject</th>
<th>Signature of subject</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AM/PM</td>
</tr>
<tr>
<td></td>
<td>Date and time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Printed name of person authorized to consent for subject (when applicable)</th>
<th>Signature of person authorized to consent for subject (when applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AM/PM</td>
</tr>
<tr>
<td></td>
<td>Date and time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relationship to the subject</th>
<th>AM/PM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Date and time</td>
</tr>
</tbody>
</table>

**Investigator/Research Staff**

I have explained the research to the participant or his/her representative before requesting the signature(s) above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

<table>
<thead>
<tr>
<th>Printed name of person obtaining consent</th>
<th>Signature of person obtaining consent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AM/PM</td>
</tr>
<tr>
<td></td>
<td>Date and time</td>
</tr>
</tbody>
</table>
Dear Parent/Guardian,

My name is Monica Kowalski and I am a PhD student in Educational Psychology at The Ohio State University. Your child’s teacher has volunteered to participate in my dissertation study this year. I am studying how teachers influence students’ motivation for math assessments and standardized testing.

For this research, I will be observing your child’s teacher during three math lessons throughout the year. After each observation, I will ask the students to complete a simple, confidential survey. The survey asks about students’ motivation for upcoming math tests. I would also like to request access to student records to analyze standardized test scores as a variable in my research. I will keep all information confidential and will only use compiled data, not individual results. More information about this study is included in the attached form.

As a University student, I work under the review of the Institutional Review Board at OSU, which approves and oversees my research. It requires the attached official parental permission form to be signed for all participants under 18 years of age. If you agree to allow your child to participate in this study, please check “yes” on the attached form. Either way, please sign and return one copy of the form to school with your child.

Thank you for your consideration,

Monica Kowalski
Kowalski.96@osu.edu
614-570-9282
The Ohio State University Parental Permission
For Child’s Participation in Research

Study Title: Teachers’ Influence on Student Motivation in Different Assessment Contexts
Researcher: Monica Kowalski, Lynley Anderman, PhD
Sponsor: N/A

This is a parental permission form for research participation. It contains important information about this study and what to expect if you permit your child to participate.

Your child’s participation is voluntary.

Please consider the information carefully. Feel free to discuss the study with your friends and family and to ask questions before making your decision whether or not to permit your child to participate. If you permit your child to participate, you will be asked to sign this form and will receive a copy of the form.

Purpose:
This is a study to understand how teachers influence students’ motivation about math tests (classroom and standardized).

Procedures/Tasks and Duration:
A researcher will be observing your child’s teacher during three math lessons this year. After each lesson, a short survey will be given to students asking them about their motivation for an upcoming math test. The survey will take 5-10 minutes. Later, the researchers would like to access students’ math grades and scores on math standardized tests (2010-2011 Terra Novas and 2011-2012 OAT) to investigate how their motivation matches with their achievement.

Your child may leave the study at any time. If you or your child decides to stop participation in the study, there will be no penalty and neither you nor your child will lose any benefits to which you are otherwise entitled. Your decision will not affect your future relationship with The Ohio State University.

Risks and Benefits:
Participating in this study is an opportunity to help us understand how teachers can motivate students for testing. The risks of participating in this study are minimal. Your child’s privacy and confidentiality are going to be protected.
Confidentiality:
Efforts will be made to keep your child’s study-related information confidential. However, there may be circumstances where this information must be released. For example, personal information regarding your child’s participation in this study may be disclosed if required by state law. Also, your child’s records may be reviewed by the following groups (as applicable to the research):

- Office for Human Research Protections or other federal, state, or international regulatory agencies;
- The Ohio State University Institutional Review Board or Office of Responsible Research Practices;
- The sponsor, if any, or agency (including the Food and Drug Administration for FDA-regulated research) supporting the study.

Incentives:
There are no incentives offered for your child to participate in this research.

Participant Rights:
You or your child may refuse to participate in this study without penalty or loss of benefits to which you are otherwise entitled. If you or your child is a student or employee at Ohio State, your decision will not affect your grades or employment status.

If you and your child choose to participate in the study, you may discontinue participation at any time without penalty or loss of benefits. By signing this form, you do not give up any personal legal rights your child may have as a participant in this study.

An Institutional Review Board responsible for human subjects research at The Ohio State University reviewed this research project and found it to be acceptable, according to applicable state and federal regulations and University policies designed to protect the rights and welfare of participants in research.

Contacts and Questions:
For questions, concerns, complaints, or if you feel your child has been harmed by participation in the study you may contact Monica Kowalski at 614-570-9282. For questions about your child’s rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.
Signing the parental permission form

I have read (or someone has read to me) this form and I am aware that I am being asked to provide permission for my child to participate in a research study. I have had the opportunity to ask questions and have had them answered to my satisfaction. I am not giving up any legal rights by signing this form. I will be given a copy of this form.

Please check one option below:

______ YES, I voluntarily agree to permit my child to participate in this study.

______ NO, I do not agree to permit my child to participate in this study

Printed name of subject

Printed name of person authorized to provide permission for subject

Signature of person authorized to provide permission for subject

Relationship to the subject

Date and time

Investigator/Research Staff

I have explained the research to the participant or his/her representative before requesting the signature(s) above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

Printed name of person obtaining consent

Signature of person obtaining consent

Date and time
Verbal Script for Assent
Teachers’ Influence on Student Motivation in Different Assessment Contexts

My name is ----------------- and I am a graduate student at The Ohio State University. I am here to conduct a study that will look at how students feel about taking math tests.

I will tell you a little bit about this study. Please stop me at any time if you have any questions. After I’ve told you a bit more about my project, you can decide whether or not you would like to participate.

I am doing this research as part of my studies in Educational Psychology at OSU. I will read you some questions about your feelings, called a survey, and ask you to circle your answers. It will take about ten minutes. It is really important that you think about each question and answer honestly so that I can really understand how you feel. Your survey does not have your name on it and I will not tell anyone what your answers are. You have an ID number on the survey, but that is just so that I can compare your answers if you take this again. I plan to use the answers to the questions on this survey to better understand how students feel and to help teachers learn better ways to teach.

Nothing special, good or bad, should happen to you for answering the questions on this survey. You do not get anything for participating, it is just up to you whether or not you want to do it. You can also stop at any time if you decide you don’t want to answer any more questions.

If you have questions, you are free to ask them now. If you have questions later, or if you feel that you have been harmed by participating in this study, you may contact me at Kowalski.96@osu.edu.

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.

I will now give you a survey if your parent or guardian has agreed to let you participate. If you want to participate in this study, please follow along as I read the directions and the questions for you to complete this survey. If you do not want to participate, or if you do not get a survey, you may work on something else (like reading a book or drawing a picture) at this time.
Appendix D: Student Survey
1. For this test, trying hard is very important.

2. For this test, really understanding the material is the main goal.

3. For this test, it’s ok to make mistakes as long as you are learning

4. For this test, getting the right answers is very important

5. For this test, it’s important to get a high score.
6. For this test, it’s important not to do worse than other students.

7. For this test, it’s important not to look dumb.

8. For me, doing well on this math test is important.

9. Taking this math test will be useful for me in the future.

10. How well do you expect to do on this math test?

11. How much do you like taking this math test?
Appendix E: Teacher Interview Protocol
Teacher Interview Protocol

How long have you been teaching? Where did you do your teacher prep program? What degree(s) do you hold?

Think back to your teacher education program. Can you describe what you remember about the type of instruction you received about assessment? About standardized testing?

What other sources have helped you develop your knowledge and beliefs about classroom and standardized testing?

What do you think is the main purpose of classroom assessment?

How well do you expect your students to do on most of their in class assessments? How well do you think they expect to do?

How do you try to talk to your students about assessment? What kind of messages do you think they hear?

What do you think is the main purpose of standardized testing?

How well do you expect your students to do on their OAA? How well do you think they expect to do?

How do you try to talk to your students about testing? What kind of messages do you think they hear?

What challenges do you face when it comes to assessment?

Thinking about the lessons that we observed (recap lesson activities), talk to me about your decision making in planning for the activities in the lessons. Compare/contrast the two contexts.

Additional questions related to specific instances in observations.