




4-2013

Making Grades Matter: Connections Between Teacher Grading Practices and Attention to State Assessment

Gregory D. Warsen

Western Michigan University, gwarsen@kvilleps.org

Follow this and additional works at: <http://scholarworks.wmich.edu/dissertations>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Secondary Education and Teaching Commons](#)

Recommended Citation

Warsen, Gregory D., "Making Grades Matter: Connections Between Teacher Grading Practices and Attention to State Assessment" (2013). *Dissertations*. 156.

<http://scholarworks.wmich.edu/dissertations/156>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.



MAKING GRADES MATTER: CONNECTIONS BETWEEN TEACHER GRADING
PRACTICES AND ATTENTION TO STATE ASSESSMENT

by

Gregory D. Warsen

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Educational Leadership, Research and Technology
Western Michigan University
April 2013

Doctoral Committee:

Patricia Reeves, Ed.D., Chair
Walter Burt, Ph.D.
Michael Stearns, Ed.D.

MAKING GRADES MATTER: CONNECTIONS BETWEEN TEACHER GRADING PRACTICES AND ATTENTION TO STATE ASSESSMENT

Gregory D. Warsen, Ph.D.

Western Michigan University, 2013

Research suggests that traditional grading practices are fraught with subjective problems and that many factors go into grading that have little, if anything, to do with what a student knows or is able to do. More recent research, however, has made connections between teacher-assigned grades and subsequent performance on the American College Test using correlational studies. This study reinforces and extends that work by, first, testing the relationship between grade point averages (GPAs) and ACT scores for four graduating high school classes in two case study high schools. Then, this study qualitatively examines teacher thinking and decision making around planning instruction, assessment of student learning, and grading practice with math and English language arts (ELA) teachers from the case high schools. Finally, this study examines how teachers react and respond when presented with correlational analyses of student grades and ACT scores over four graduating classes and asked for their reflections and interpretations of those correlations.

The results of a Pearson product–moment analysis found that correlations between math and ELA grades for the four graduating classes tested to be moderately positive and significant. Qualitative findings from interviews with ELA and math teachers from the case study schools indicate a high degree of intentionality on the part of

teachers to connect their instruction, assessment, and grading decisions to state standards and to position students for successful performance on the ACT. When positive and significant correlations between grades and ACT scores were presented to teachers, they voiced an expectation that this would be the case. Some teacher disenfranchisement from less autonomous decision making in these areas was also noted.

Copyright by
Gregory D. Warsen
2013

ACKNOWLEDGMENTS

It was my false impression that the research and preparation of a doctoral dissertation was an individual sport, but I have learned otherwise through this process. I am truly indebted to many.

First and foremost, I am grateful to my wife and family for their encouragement and support through this process. Numerous classes, hours of homework, and generalized stress had to be tolerated, and my family did so with grace, class, and humor. My wife Susan is an invaluable, inspiring life partner, and our children Andy, Abby, and Grace make up my favorite team. I am also grateful for the gifts and opportunities given to me through a charmed childhood by my parents, Duard and Sandy Warsen. I was raised to chase dreams.

Also, of critical importance on this road is my advisor, Dr. Pat Reeves, of Western Michigan University. Her thoughtful insights and gracious criticism delivered with utmost tact guided me to the finish line in a way that I hope every doctoral student could experience. Rounding out my committee are Dr. Burt of Western Michigan University and Dr. Stearns of Grand Valley State University, both of whom have been supportive.

The Board of Education and staff with whom I work also deserve much gratitude. I have made my professional home with them for some time and have been honored to practice educational leadership in their company.

Acknowledgments—Continued

I would also extend thanks to the participants within this study: eight teachers who gave of their time and insights to form the important ideas and additions to the research. They, like most teachers, are my heroes.

Finally, I am grateful to God for the opportunity and intellect to accomplish this work. The doors opened, and all I had to do was walk through.

Gregory D. Warsen

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER	
I. INTRODUCTION	1
The Conceptual Framework.....	2
Focus of the Study	6
The Problem.....	6
Purpose.....	9
Rationale	10
Research Questions.....	12
Methodology Overview	14
Definition of Terms.....	15
Limitations and Delimitations.....	16
Summary.....	17
II. LITERATURE REVIEW	18
Broader Frame: Assessment	18
Summative Assessment for Accountability	19
Formative Assessment for Teaching and Learning	21
Grades: The Gray in the Middle	22

Table of Contents—Continued

CHAPTER		
	General Grading Practice	23
	Grading Practice and Educational Measurement	24
	Relationship Between Grades and Outside Measurement	27
	Secondary Grades and ACT as Predictors of College Readiness	29
	Study Rationale	30
	Summary	31
III.	METHODOLOGY	33
	Study Methodology and Rationale for Mixed Method	33
	Study Methods and Procedures	36
	Research Design, Phase I: Quantitative	36
	Research Design, Phase II: Qualitative	38
	Data Collection Methods, Procedures, and Instrumentation	38
	Sampling, Subjects, Access, and Setting	39
	Data Analysis Processes and Procedures	40
	The Researcher	45
	Limitations and Delimitations	46
	Activities and Timeline	47
IV.	FINDINGS	49
	Purpose and Research Questions	49
	Purpose	49

Table of Contents—Continued

CHAPTER		
	Research Questions	50
	Description of Participants.....	51
	Schools	51
	Students.....	51
	Participants.....	52
	Results.....	53
	Quantitative Results: Research Question 3.....	53
	Qualitative Results: Research Questions 1, 2, and 4	57
	Chapter Summary	88
V.	SUMMARY, CONCLUSIONS AND RECOMMENDATIONS.....	89
	Summary of Major Findings.....	89
	Research Question 3: Quantitative.....	89
	Research Question 1: Qualitative.....	91
	Research Question 2: Qualitative.....	93
	Research Question 4: Qualitative.....	94
	Conceptual Frame Revisited.....	96
	Conclusions.....	98
	Recommendations.....	102
	For Practice	102
	For Further Research.....	103

Table of Contents—Continued

CHAPTER

Reflections 104

REFERENCES 108

APPENDICES

A. Interview Protocol for School A, ELA 116

B. Interview Protocol for School B, ELA 119

C. Interview Protocol for School A, Math..... 122

D. Interview Protocol for School B, Math..... 125

E. Consent Form..... 128

F. Human Subjects Institutional Review Board Letter of Approval..... 130

LIST OF TABLES

1. Distribution of Student Participants Across Racial/Ethnic Categories by School	51
2. Comparison of Enrollment, ACT Scores, and Graduation Rates by School	52
3. Pertinent Participant Characteristics by School Group.....	52
4. Letter Grade to GPA Conversion Table.....	54
5. Correlations Between GPA and ACT Scores for ELA and Math by School.....	55
6. Aggregate Correlations Between ELA and Math ACT Scores and GPA by School	56
7. Numbers of Typological Themes and Subthemes Found in the Data by Research Question	59
8. Quotes Illustrating Expectations for Positive Correlations Between ACT and GPA.....	85
9. Summary of All Themes and Subthemes by Research Question and Typology	87

LIST OF FIGURES

1. Conceptual frame	5
2. Broader frame of assessment	19
3. Conceptual frame: Findings	97

CHAPTER I

INTRODUCTION

When I began my educational career as a 10th grade English teacher, I vividly recall a staff meeting in which the principal looked directly at me and indicated that the knowledge and skills students would need to do well on the state assessment would have to be developed in my classroom. Not having tenure and wanting to be successful, I took his admonitions to heart. I built classroom assessments and made additions to the final exam that mirrored the state test. To support those, I tried to design instructional and assessment practices that would enable my sophomores to do well on the Michigan Educational Assessment Program (MEAP) tests, and hopefully the grades I assigned to students captured their ability to do so. Back then (the early 1990s), however, the only thing at stake for the student was a missing endorsement, in the form of a sticker, on his or her diploma if state tests were not passed. At that time, most of my colleagues with tenure and experience realized that, while they wanted students to perform well, their job was secure regardless of outcome.

The MEAP as well as many other state assessments across the United States have, no doubt, gone through revision since that time. What has not disappeared, however, is the accountability movement. The advent and implementation of No Child Left Behind (NCLB, 2002) has served to sharpen the sanctions of low-performing schools, and public education must worry about much more than the loss of a diploma sticker. But does that increased accountability translate into changed instructional practice? Specifically, are

teachers making a stronger connection between the state accountability assessments (for Michigan high schools, the American College Test or ACT) and their instructional planning and assessment or grading practices now that student performance on the state assessments has been taken to a high-stakes level for the school?

The Conceptual Framework

In an era of NCLB accountability and under the pressure to reach adequate yearly progress (or AYP), states have had to impose annual assessments on districts in order to measure levels of proficiency within the student body at various points. Some states are currently using or strongly considering use of the ACT as part of the high school assessment along with other state developed assessments or by itself (V. J. Dean, personal communication, June 15, 2009).

State assessments, however constructed, seek to measure the degree of mastery of the content and performance expectations of the students they serve so progress or lack of it can be seen in each school. This tracking of progress is one of the key components of NCLB (2002) and probably is not going away anytime soon. President Obama and his current education secretary, Arne Duncan, have espoused the value of excellence at all levels of education, and it would be politically ill-advised for them to do otherwise (Duncan, 2009).

Inasmuch as states seek to measure content and performance expectations, teachers make daily decisions about instruction, the assessment of and for learning, and ultimately how to assign grades. These grades have historically formed the currency between teacher and student, especially in the high school setting. At this level, grade point average becomes a reported data point on college applications and can have impact

on college acceptance and scholarship dollars. As one researcher found, even at the 8th grade level, grades take on more meaning in the minds of students because of this level's proximity to high school (Wilson & Corbet, 2001). The research on grading practice, however, is not kind.

When seen through the lens of the research in educational measurement, reviews on grading practice are not seen favorably. Cicmanec (2001) has edited an entire book of essays on the disconnect between the education measurement community and K-12 education. "Hodgepodge Grading: Endorsed by Students and Teachers Alike" (Cross & Frary, 1999) reports the many non-academic factors that go into K-12 grading practice such as attitude, effort, and participation.

This point gets sharpened by Friedman (2000), who goes a bit further to suggest that teachers and researchers work together to develop solutions that are sound in terms of both measurement and practice. Friedman also points to concerns about teachers violating the Family Educational Rights and Privacy Act (FERPA) when they allow students to orally report out their own or others grades in class. Additionally, Gustafson (2005) gets at some of the causes of this disconnect between K-12 classroom practice and standards of educational measurement. For instance, surveying 253 teachers on their philosophical beliefs and values in grading practice, Gustafson found that differences among teachers could be associated with grade level and number of years of experience.

Researchers have also attempted to determine the impact of faulty grading practice. J. D. Allen (2005) takes specific aim at the issue of validity or the lack thereof in the grading schemas of educators and the subsequent communication of those grades to others. Allen looks at three specific areas. First, the miscommunication and confusing

purposes of grades that permeate the current systems is addressed. This miscommunication can really hurt students whose knowledge and skill level are under reported. Second, Allen points to lack of pre-service training as a major reason behind the lack of validity in teacher grading practice. According to his research, less than one third of the teacher training programs in the U.S. require an assessment course for pre-service teachers. Third, he notes that the communication of this problem from educational measurement experts to K-12 teachers, teacher preparation programs, and teacher development planners and trainers has been happening for almost a century with little or no impact on practice.

The research suggests that grades, as typically generated from the current status of K-12 classroom assessment practice, have questionable validity and reliability and that teachers receive little or no pre-service training on appropriate assessment and grading practice. While the literature offers important insights and guidance for improved assessment and grading practice, studies indicate that little of that guidance is permeating actual teacher practice. Moreover, the literature does not offer much guidance on how teachers actually reconcile their grading practice. For instance, teachers may not be aware of how their grades do or do not correlate with scores on state assessments. In most schools, the staff does not systematically examine trends in correlation between core content high school grade point averages and state assessment scores in those same areas. The current literature on teacher grading practices does little to reveal if teachers consciously reflect on state assessments when planning instruction, assessing learning or assigning grades to their students. As opposed to using a theoretical framework, this

study instead uses a conceptual framework. The graphic in Figure 1 provides a conceptual framework for examining these issues further through this study.

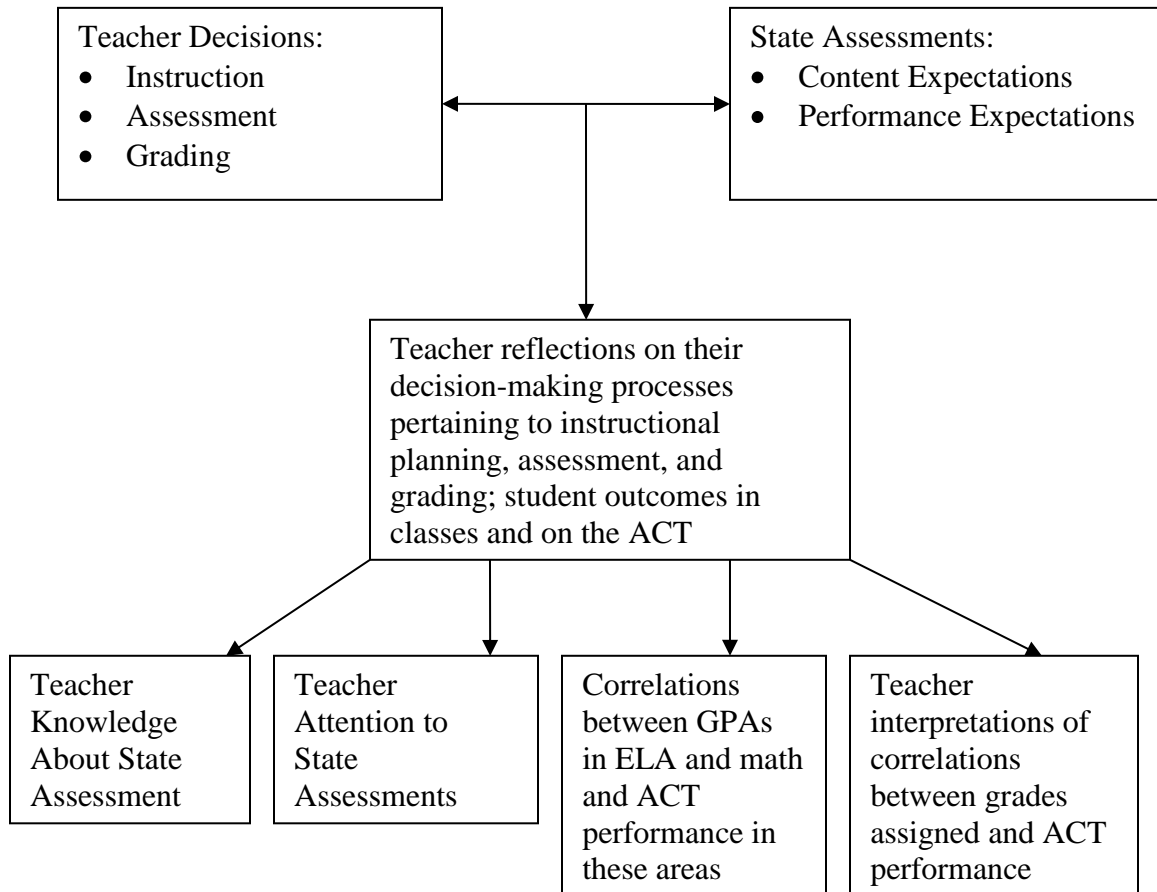


Figure 1. Conceptual frame.

Teachers making daily decisions about instruction, assessment, and grading is a given. State departments of education also make decisions about content and performance expectations that will be communicated to schools within accountability systems. The assessments within these systems that provide data to evaluate student achievement within and across schools and school districts become the arbiter of school success. These two decision-making processes take place independent of one another—the first taking

place in each teacher's classroom through the normal process of planning and delivering instruction and assessment. The second decision-making process for establishing state content standards and measures may include representatives from the teaching ranks, but the process takes place in a policy and political environment which is disconnected from the day-to-day reality of actual classrooms (Mazzeo, 2001). There are a number of studies that examine both of these decision making processes, but there is little insight from those studies on how teachers make the connection between the work they do in the classroom to assess learning and the measures and means prescribed by state law to judge school success.

Through a mixed-method approach, this study qualitatively examined teachers' thinking and reflections on how they make daily classroom decisions in light of state expectations. Quantitatively, this work examined degree of association between grades assigned by teachers and the corresponding student performance on state assessment, or more specifically, the ACT.

Focus of the Study

The Problem

An article in the *Washington Post* cites a number of students who received passing grades in their English and math classes but failed to pass their state's proficiency test (Shapira, 2006). The article is far from a controlled, well researched study, but it does point to the general issue under consideration in this proposed project. While it cannot be assumed that state assessments are the definitive measure of student learning, they are commonly accepted as one indicator. It is also commonly accepted that state assessments have a different purpose than classroom assessments. For instance, Reeves (2008)

discusses how state assessments provide important information to educators about the general state of learning against the state curriculum standards, but are insufficient to provide teachers real-time data on student progress in a fashion that actually informs instruction. Schmoker (2002) refers to this second type of assessment as assessment for (as opposed to about) learning. What is not clear is how teachers interpret the relationship between both types of assessment and how they use that interpretation to guide their instruction, assessment, and grading practices.

This has led a few researchers to ask what the relationship is between student performance on state assessments and grading practice. At the elementary level, L. F. Lambert (2002) found a “moderate, positive and linear” relationship ($r = .70$) between the grades given to 2,119 third graders and their performance on the Texas state assessment in reading. Another study in Florida (Dittmar, 2005) found a similar relationship on that state’s assessment, but this relationship weakened with teachers who factored homework into the reading grade. Bauer (1998) examined this relationship between the High School Proficiency Test (HSPT) in Michigan at the secondary level and found moderate positive correlations with grade point averages as well. These studies illustrate that moderately positive correlations between grades and state assessments have been identified, but that those correlations may be weaker or stronger based on what teachers factor into grades. These studies do not, however, provide any understanding of how teachers link their students’ performance on state assessments with either their instructional or grading decisions and practices.

We know that much of what goes into a grade can have little to do with knowledge and skill attainment, but how and to what extent do teachers consider how the

state will be assessing students as they create and assess student work and assign grades to their students? If the consideration is low, or even nonexistent, why is that? Little is known about how teachers think about the connection between state tests and their classroom instruction or grading practices. There is also evidence that teacher instruction is often weakly aligned to official district curriculum and, thus, the state curriculum (Mazzeo, 2001). Because this lack of alignment puts students and the school at risk for poor performance on high stakes state assessments, it is important to learn more about how teachers make the decisions and choices that shape their classroom instruction, assessment, and grading practices.

Numerous studies and conversations in the literature focus on the general disconnect between state and federal policy and actual teacher practice (Gustafson, 2005; Mazzeo, 2001). State and federal policy regarding state assessments is framed within a political context driven by accountability concerns. On the other hand, teacher practice is formed through the interaction of a teacher's experience and his or her strongly held assumptions and beliefs (Cicmanek, 1999; Cross & Frary, 1999). Additionally, studies have already isolated a disconnect between classroom instruction and state curriculum standards. The two very different contexts and frames for impacting teacher assessment and grading practice (state and federal policy vs. teacher experience) could make it less likely to find moderate or strong positive correlations between results on state assessments and grades; yet, at least a few studies have found just that. Theoretically, researchers have posited that classroom grades and achievement on state assessments should serve to confirm each other, at least to the extent that both measure attainment of a set of generally accepted curriculum expectations (Marzano, 2006). The problem is that

studies that examine the relationship between classroom grades and results on state assessments have yet to delve into how teachers understand, interpret, and act on the relationship between state assessments as measures of student learning and the grades they assign to interpret the achievement and learning attainment of their students.

Without a better understanding of the thinking and decision-making processes teachers use, it will be difficult to change the pattern of disconnect between state and federal policy and actual teacher practice pertaining to how teachers assess learning and translate that assessment into grades.

Purpose

This study was designed to examine how teachers make decisions about instruction, assessment, and grading with a focus on identifying how they connect or do not connect these decisions with the state assessments for their curriculum area. Since the state assessments are derived from the state curriculum standards, this study also provided some insight into how teachers consider the curriculum standards represented by the state assessments in their instructional planning and assessment of student learning. To carry out this examination of teacher thinking and decision making, this study asked core content teachers in math and English language arts (ELA) from two schools to reflect on their decision-making process with regard to instructional planning, assessment, and grading.

This study also captured the participant teachers' response to the results of an analysis of the relationship between student grades for these two curriculum areas and the results of the state assessments for those areas over a 4-year period. By exploring both the teachers' frame of reference for making instruction, assessment, and grading decisions

and their interpretation of the meaning of how grades relate to state achievement scores in their content area, this study shed some light on the nature of how the two contexts for measuring student learning (state and federal political context and local classroom teacher experience context) interact with one another. This study also looked for indicators of how correlations between grades and state assessments are occurring.

Rationale

The primary purpose of this work was to examine teachers' thinking as they reflect on how they make decisions about planning instruction, creating student work or classroom assessments, and grading. Additionally, this study looked at how teachers connect these decisions to their knowledge of and experiences with the state assessment for their area of the curriculum. Teachers spend inordinate amounts of time and energy to give students grades; these marks form the currency of a high school education. The parallel reality, at the state level, is the amount of resource that goes into developing, administering, and reporting the results of state assessments. Both the investment by teachers in their classroom grading practices and the investment in state accountability measures of learning attainment made the question of how teachers are or are not linking the two worth further examination.

Most educators understand that the accountability standards and sanctions embedded in NCLB are just the opening salvo of a national shift in education policy. Because state departments of education and, thus, local districts are held accountable by the U.S. Department of Education, the inquiry of this work has potential implications at the policy level as well. It has been well-established in the literature that policy mandates, alone, do not achieve fundamental changes at "street level" (Mazzeo, 2001). For policy

makers, it is important to know how policy initiatives like NCLB are impacting or not impacting what actually happens at the classroom level.

It is difficult to imagine the amount of time, energy, and financial resource at the federal, state, and local level that goes into the development and implementation of a federal policy such as NCLB. I would contend that this level of investment demands some investigation of impact, and the questions such as, “In what ways are teachers adapting their thinking, planning, and practice to more closely align with the student achievement outcome expectations explicit in the state assessments and the state/federal accountability standards?” shed some light on an important issue. Furthermore, policy writers at both the state and federal levels might be interested to see which aspects of NCLB have had the most, least, or any impact on teachers’ practice. Also, the MEAP tests as well as other state assessments have become high stakes for both schools and students. The list of sanctions for schools is clear and progressively inconvenient if proficiency scores fall below expectations. Students stand to gain or lose scholarship money for failing to meet certain levels of achievement.

It is certainly in students’ interests to see if teachers are aligning instruction, assessment, and grading practices to mirror the skill and knowledge level needed for success on the American College Test (ACT). As noted above, the connection between the content knowledge and proficiencies embedded in the ACT and other state assessments and the instruction, assessments, and grading practices students experience in their high school courses has received little or no attention in the published literature. Because of the stakes for student success based on state assessments, this issue was of significant interest.

To take it one step further, this issue and the question of alignment between teacher decisions regarding instruction, assessment and grading and performance expectations for success on state assessments fall under the broader umbrella of how educators use data in general. With the amount of time that most secondary teachers put into the calculation of grades, what is their experience in making these decisions? How do they come to the final grade and what criteria do they employ to get there? Moreover, how are they shaping classroom instruction and student work or classroom assessments to create the data against which grades are assessed? Finally, what is the status of either the alignment or disconnect between the classroom generated data on student learning and the state assessment generated data on student achievement?

Beyond policy questions mentioned above, this study filled a gap in the research as well. The American College Test (ACT) has done extensive research on the relationship between the number of core classes taken and performance on their assessment (ACT, 2006). What has not been researched extensively is the connection between the assessment of student learning, i.e., grades that students earn in the core areas of English language arts and math, and student performance on state assessments, i.e., the ACT. Studying this adds to the research literature and has far-reaching practical relevance as well, especially as it pertains to the influence of the performance expectations embedded in state assessments and the decisions teachers make about instruction, assessment, and grading.

Research Questions

In a No Child Left Behind (NCLB, 2002) era in which the reporting of tests results is demanded with possible sanctions to follow poor performance, this project

helped to discover if teachers feel any increased urgency to make the connections between their teaching and grading practices and the expectations of state assessments more solid. Did teachers see this as just one more federal law or initiative that is likely to fall away, or were they beginning to alter their thinking and practice as it relates to instructional planning, assessing student learning, and providing feedback to students through grades?

This study was guided by a set of research questions that engaged math and English language arts high school teachers in an examination of their thinking and decision-making processes relative to planning instruction, assessing learning, and assigning grades. In this examination, the teachers were also asked to interpret an analysis between 4 years of grades from their school in math and ELA and the state assessment results for those same years in those content areas. The specific research questions were the following:

1. How do teachers describe their decision-making process with regard to planning instruction, assessing student learning, and grading practice?
2. How do teachers describe and interpret the relationship between state assessments for their content area and their instruction, classroom assessments, and grading practice?
3. What is the relationship between grades given in English language arts and math over a 4-year period (2006-2009) in the participants' high schools and performance in those same core areas on the ACT?

4. How do teachers interpret the relationship between the grades given in core subjects and student performance in the same core curriculum areas on the ACT assessment for their school?

Posing a set of questions to educators around the above queries shed some light and revealed some meaningful patterns within and among the responses.

Methodology Overview

For this work, I employed a mixed-method approach, using qualitative as well as quantitative methods. Qualitatively, this study was conducted through the phenomenological approach within two case-specific schools, utilizing a semi-structured interview protocol. Creswell (2007) defines a phenomenological study as describing, “The meaning for several individuals of their lived experiences of a concept or a phenomenon” (p. 57). As I delved into the experience of educators with regard to how they make decisions about instruction, classroom assessments, and grading and the way they connect those decisions with state assessments, this approach fit the purpose well. The common phenomenon to teachers at the high school level is the experience of planning instruction, assessing student learning, and assigning grades to their students and how they arrived at the decisions related to those processes. Beyond that, I sought to discover what bearing, if any, the requirements of state assessment are now having on teachers as they carry out those functions. The examination of a set of experiences that is common to all high school teachers against the backdrop of a major policy shift (i.e., the measurement of and accountability for universal proficiency as measured by state assessments) fit well with a phenomenological approach, which can ferret out themes that

have potential transferability. These themes also offered direction to further study and/or policy level work.

Quantitatively, I examined the presence and direction of a correlation between the grades teachers assigned in core areas and the students' corresponding performance on the ACT in similar core areas. Put simply, did the "B student" in math class score higher on the math portion of the ACT than the "D student" and to what degree does the pattern hold up? Because Pearson's correlation has been widely used in social sciences to show relationships of this sort, it was a good fit here as well (Glass & Hopkins, 1996). The findings from this analysis were incorporated into the phenomenological inquiry with teacher participants who interpreted the meaning of those correlations for their own teaching and assessments practices.

In terms of scope, the state of Michigan offers a unique opportunity to examine the impact of giving the ACT to all students. In the spring of 2007, the Michigan Department of Education started giving the ACT as a major portion of its state assessment to all high school 11th graders. Prior to this, only students who elected to take the ACT (presumably those planning on a college career) were exposed to it. Comparing the correlations for the graduates of 2006 and 2007 to those in 2008 and 2009 showed us the degree and direction of change in those correlations. Put another way, I was able to see if increasing the number of students taking the ACT in two high schools caused the correlation between ACT scores and grade to change.

Definitions of Terms

The following terms deserved definition given the focus of this project.

State Assessment: This refers to mandated assessments given by a state for the express purpose of monitoring and accountability within the parameters of No Child Left Behind (NCLB, 2002). Specific to this study is the Michigan Merit Exam given to all high school juniors, with retests given to seniors as needed. It is comprised of both the American College Test (ACT) and content-based questions from the Michigan Educational Assessment Program (MEAP).

Grading Practice: This refers to the set of decisions, policies, and procedures employed by teachers to assign a letter grade to their students (Marzano, 2000). Specific to this study, my queries will look at both the practices of teachers and how those practices developed. Put another way, how have educators come to their decisions about how to grade students?

Limitations and Delimitations

This study is limited by the organizational relationship between the researcher and the participants (I am employed in the central office of one of the two schools involved in the study). Participants may have felt compelled to give answers that they thought I wanted to hear. To reduce the impact of this, I held the interviews in the work area of the participant or in a neutral location (e.g., a conference room) so they were in an environment in which they felt comfortable. Furthermore, I stressed to the participants my gratitude for their involvement and a request for honest responses. The sample size of eight instructors was clearly not large, so insights from the data need to be seen in that light. Closely related, this project's scope of two schools limits its transferability. Finally, transferability is also limited to educators in schools with similar size and demographics.

Quantitatively, correlations do not indicate a causal relationship, so if connections between grades assigned and ACT performance are strong, weak, or non-existent, this will not give us the factors behind the relationship. That might be the work of another project.

Summary

The focus of this study was on teachers' decision making on instructional planning, assessment, and grading. The purpose was to reveal the degree of connection between these teacher decisions and the policy decisions made at the state level on accountability systems within a NCLB environment. A mixed method was employed to examine teacher decisions qualitatively and the association between assigned grades and ACT performance quantitatively. The next chapter of this proposal further examines the research base that supports this work. Chapter III articulates the precise methods by which the study was executed.

CHAPTER II

LITERATURE REVIEW

Nearly everyone who has had contact with the K-12 education system in America has experienced the phenomenon of receiving a grade. Grading is so much a part of the educational process that few teachers and students give the process of grading or the resulting marks much thought. Interestingly, however, some researchers have queried the bases behind grading in terms of what they may or may not mean and if they are or are not valid and reliable.

Published research, pertinent to this specific area of inquiry, centers on three main areas: general grading practice, the relationship between grading practice and educational measurement, and the relationship between grades and state assessment. After framing this research within the broader context of assessment, each of these main areas are examined through the literature, followed by a rationale for the present study.

Broader Frame: Assessment

Before delving into the specifics of this study and the research that most closely supports or mirrors it, I will set the stage of this inquiry within the broader framework of the practice of assessment. Consider Figure 2. What this figure illustrates is that assessment has a broad set of purposes and functions that could be seen on a continuum from accountability measures that are almost purely summative to the variety of informal classroom assessments used in the process of teaching and learning that are almost purely formative. Stiggins (2002) draws out the two poles of this continuum well by pointing out

that we have plenty of systems in place for summative evaluation (assessment of learning) but few that really serve the goals of formative assessment (assessment for learning). He further advocates that formative assessment must involve students in such a way that they see their own progress and are invited to continue learning. In another work, Stiggins and Duke (2008) describe assessment used at either the classroom or institutional level, the former used to help students and teachers make decisions about the next step in learning and the latter used to help district leaders and policy makers determine the percentage of students meeting standards.

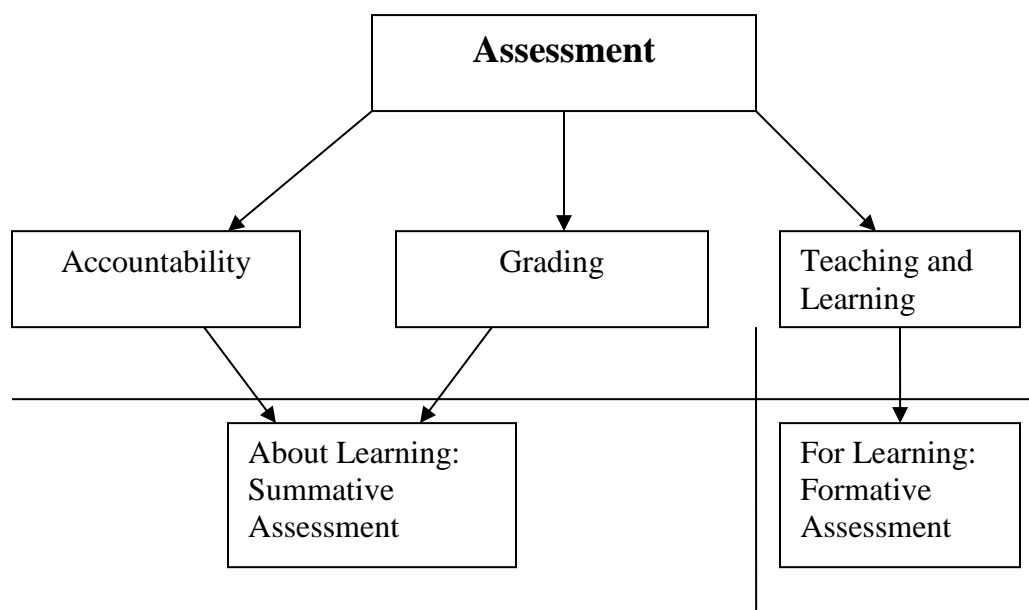


Figure 2. Broader frame of assessment.

Summative Assessment for Accountability

The accountability purposes of assessment of educational standards have been well developed in the political and policy history of education. The standards movement arose from *A Nation at Risk* (National Commission on Excellence in Education [NCEE],

1983). Commissioned by the Reagan administration and published in 1983, it outlines the dire state of American education as perceived by its authors. Chief among its concerns is the rapidly deteriorating quality of the end product of our educational system: the high school graduate. The rhetoric used in the report borders on inflammatory with claims that the current mediocrity in the educational system could be compared to “an act of war” (NCEE, 1983) had it been imposed upon the United States by an unfriendly country. Prominent among its 38 recommendations is a call to standards and expectations.

On the heels of *A Nation at Risk*, *Goals 2000* captured the education agenda of the senior President George Bush. Standards received more attention and specificity as it called for student mastery of specific standards in grades 4, 8, and 12 (Marzano & Kendall, 2007). Years later, the younger President Bush signed the *No Child Left Behind Act* (2002), which brought the standards and accountability movement to what it is today with mandatory testing and benchmarks on the road to 100% proficiency. Our preoccupation with the political desires for increased standardized test scores may very well short-circuit improved student learning, however, if that is where the focus begins and ends. Accountability testing has its limitations (Stiggins, 2002).

For example, Marzano (2003) points out that state-level tests cannot possibly capture all of the student achievement that may have taken place within a given classroom. These tests are by definition samples of the standards at any given grade level and schools would be unwise to use them as the only indicator of learning. Also, accountability pressures can inhibit teaching for meaning, setting up a tension between teaching for learning and teaching for accountability purposes. The standards movement, therefore, can lead to superficial coverage as opposed to deep learning (McTigh, Seif, &

Wiggins, 2004). This point gets underscored by Schmoker (2009) when he argues that an unintended consequence of the data culture, especially the data of state accountability tests, is that it can preoccupy educators away from true, deep learning. Test preparation can be the cause behind short-term gains without any real reform. Moreover, for some high-performing schools, high test scores actually stop the improvement because of a lack of motivation on the part of the school or the community to continue to grow.

Formative Assessment for Teaching and Learning

On the left side of the continuum in Figure 2 are the formative purposes of assessment designed for teaching and learning. Marzano (2006) further clarifies formative assessment from its summative counterpart by noting that formative assessment takes place while the knowledge or skill is being learned, while summative assessment takes place at the end of learning. As such, formative assessment has much deeper use as part of the learning process.

For example, Reeves (2006) makes the case the frequent formative assessments can provide valuable feedback for students and teachers alike. He also argues that if we define tests as summative, then schools generally test too frequently and don't assess (formatively) enough. Schmoker (2006) extends this by suggesting that formative assessments should be held in common by teams for learning that can lead to needed instructional change at the classroom level. Teachers can then quickly see if their collective efforts at improvement are paying off. More recently, Reeves (2009) has made the point not only that formative assessment can provide information both to students and teachers about learning, but it can also become rewarding to both, sustaining the learning for the student and perhaps a school initiative for the teacher.

Grades: The Gray in the Middle

So what do we make of grades? Educators, both teachers and principals, have little or no training in assessment literacy. There is no requirement at the state or federal level to make this a priority, and therefore learning may be inappropriately measured on a regular basis (Stiggins, 2002). Might then teachers easily confuse the purposes of assessment and fail to distinguish between formative and summative? Might they also give letter grades, clearly a summative measure if given after supposed learning, and consider this formative feedback?

Educators getting beyond grades and asking why a given percentage of students are succeeding at a given task is an important step in improving the teaching and learning process (Schmoker, 2003). Reeves (2003) furthers this point by showing that a characteristic of high performing schools is the frequent use of teacher made formative assessments. Poor performance on these assessments is not merely placed in a grade book while teachers move to the next unit. Rather, students are given multiple opportunities to master the material and get re-assessed (Reeves, 2003). He goes on to argue that if feedback is all summative (i.e., put in the grade book), some students would then be “happy with a ‘D’ and unmotivated by an ‘F’” (p. 4). With healthy assessment practices, however, poor performance is met with more coaching and assessment instead of a final low grade from which there is no return. The value of formative assessment on learning is also recognized in higher education (Turner, VanderHeide, & Fynwever, 2011).

Stiggins (2002) echoes a similar theme when he argues that some students approach a more difficult set of standards and testing with the belief that they will succeed because they have in the past. Others will not, based on previous difficulty.

Grades almost become a self-fulfilling prophecy. Much of the literature to which we now turn reinforces that general grading practice is flawed at best.

General Grading Practice

In the broadest strokes, the research is not kind to the practice of grading in K-12 schools because several factors other than actual achievement come into play. In one study (Goldwater & Nutt, 1999), researchers concerned about the subjectivity of grading found that if the family of origin for both the teacher and that teacher's student was similar, students tended to get higher grades. In another, Cicemanic (1999) discovered that factors such as school size, classroom size, and the number of at-risk students had measurable impacts on grading practice.

This issue of non-academic factors influencing grades is common to both elementary and secondary teachers. B. T. Smith (1999) conducted a qualitative study with elementary teachers asking questions such as

Why do teachers grade a student's work? What purpose does grading serve? To what extent does a teacher grade homework, class participation, daily assignments, extra credit, projects, quizzes, and tests? Are grades obtained differently for different students? What is the basis of the formula teachers use to arrive at a student's report card grade? What is the school's grading policy? What is the division's grading policy? Do teachers always follow the policies? Are non-academic factors considered when assigning grades? (p. 7)

Findings indicated that many non-academic factors are considered in a final grade. In a similar vein, McMillan, Myran, and Workman (2002) examined 900 teachers of third, fourth, and fifth grades, attempting to find a relationship between assessment and grade

level. No such relationship was found. Randall and Engelhard (2010) also established that elementary teachers will use non-academic factors in grading for borderline students, even if it means deviating from district guidelines.

Secondary teachers don't fare much better when trying to draw a straight line between grades and actual knowledge and skill. High school teachers in one study of three high schools were highly inconsistent when their grading practices were compared to each other (Reed, 1996). Some of this inconsistency may stem from the fact that many teachers base their grading practices not on research but on the practices of colleagues next door (Podgett-Harrison, 2000) or how competitive the environment happens to be (Bonesronning, 2004).

Another researcher looking at secondary grading practices conducted a study in which she attempted, with some success, to change the grading practice in a middle school setting. In pre-and post- interviews following exposure to some of the research in educational measurement, teachers did report altering their grading practice as a result of exposure to research (Lott, 1997). Noteworthy as a limitation, however, is that the researcher in this case was also the principal of the 17 teachers interviewed.

Grading Practice and Educational Measurement

As noted in Chapter I, Cicmanec (2001) provides a review of a book of essays dedicated to the connection of grading practices and standards-based education, or more accurately, the lack thereof. Four of the essays get into the problem of grading and the disconnect between what the educational measurement community would designate valid practice and the actual practice of teachers in K-12 education. Two of the essays give case studies of schools that are trying to make substantive changes to their grading and

reporting systems to make them more standards-based. Cross and Frary (1999) address a similar issue in their peer-reviewed research report presented at the Annual Meeting of the National Council on Measurement in Education. These researchers conducted a survey with 310 teachers and 7,367 students in the same school system to assess the “hodgepodge” effect of attitude, effort, and participation as it relates to the grades students receive. Results of the survey confirmed prior research in that both students and teachers confirmed the use of many non-academic factors in the assigning of grades. The authors speculate a bit on how those in the educational measurement community might better communicate some of their findings to teachers and school administrators (Cross & Frary, 1999).

Friedman (2000) reinforces the conclusion that teachers are using many non-academic factors in their grading practices. He suggests that teachers and researchers work together to develop solutions that are sound in terms of both measurement and practice. He also cites FERPA violations when teachers have students report audibly on their grades following a peer-graded assessment. Furthermore, as noted above, Gustafson (2005) as well found philosophical gaps among teachers and their grading practice depending on which grades they taught and how long they had been teaching.

Research has also delved into some of the causes of faulty grading practice. As noted in Chapter I, Allen (2005) calls into question the validity of grading, as have many others. He goes further, however, to observe that the reason many educators engage in grading practice of questionable validity is that they are merely doing what was done to them. Absent pre-service training in this area, new teachers tend to rest on grading practices that were used on them as students or that their colleagues across the hall

employ (Allen, 2005). It is not difficult to see how this continues from one generation of educators to another.

So what's the answer? The educational measurement community does not stop by simply pointing to the gap between practice and research; they have put forth many recommendations for reform. In his article "Beyond Assessment to Best Grading Practice: Practical Guidelines," Carlson (2003) examined five areas of criterion-referenced grading practice that provide options to educators in their attempt to find better ways to measure students' knowledge and skill. First, "*learning objectives* must be directly related to content, written in clear language that students can understand, and clearly measurable" (p. 508). Second, grading that means something to the student will have the most significant impact on learning. Third, "the number and variety of assessment measures employed affect the validity of a student's grade" (p. 509). Fourth, the use of portfolio assessment can work, provided it is well designed and communicated to students and parents (see also Ediger, 2000). Fifth, care must be taken when giving group grades for cooperative learning projects. Research cited above would imply that many of these recommendations are not currently in practice.

Another prominent researcher and theorist making recommendations in this area is Robert J. Marzano, whose ideas are put forth in *Classroom Assessment and Grading That Work* (2006). In this well researched work, he advocates a balance between formative and summative assessment, the judicious use of rubrics, and the unpacking of state standards into measurement topics or similar areas of learning that can be assessed. He also posits the idea of learning that takes place over time and honoring the student's individual rate of learning by capturing it upon mastery, as opposed to averaging scores

from the beginning, middle, and end of the learning as many educators practice. Marzano and Hefelbower (2011) extend these ideas further by advocating that students be given the chance to retake assessments to capture improved performance, a practice that doesn't always happen. The irony here is that education is one of the few industries that does not recognize a learning curve. Doug Reeves (2009) strikes a similar chord when he argues that improved grading practice can not only reduce failure rates but also improve attendance, school climate, and even teacher morale.

Relationship Between Grades and Outside Measurement

Despite the tense relationship between K-12 education in terms of grading practice and the education measurement community in terms of recommended grading practice, some work has been done in trying to connect the two. For instance, the background on the development of assessment policy in the United States shows an interesting progression. In a similar vein to Marzano and Kendall (2007) noted above, Mazzeo (2001) tracks this progression when she notes that three policy frameworks have been developed since 1965: "examination, guidance, and accountability" (p. 367). She begins by pointing out that accountability testing is a fairly recent occurrence, not really seen prior to 1965. In defining her terms, she notes that states can and have used tests either to get information about the schools or to bring about change in the schools. On some level, state testing policies have a symbolic value in that they show the public that government is doing its job by monitoring schools' performance.

Marzano (2006) has also focused efforts on ways to enhance the validity and reliability of teacher-assigned grades by promoting the use of assessment rubrics by teachers as opposed to using a point-based system, which can be fraught with

inconsistencies. He goes so far as to argue that common rubrics used among teachers will increase the interrater reliability from .294 to .719, an increase that teachers would be wise to give notice.

The specific relationship between how students perform on state assessments and what kind of grades they received in their K-12 environment has received some attention, but more work here is warranted. For example, L. F. Lambert (2002) examined the relationship between grades given to 2,119 third graders on end-of-course reading tests and their subsequent performance on the state reading assessment in Texas and found a correlation of $r = .62, p < .01$. Of additional interest, however, this study also found that most of the students who failed the state test had passed end-of-course assessments designed by the school. This reinforces the disconnect between the standards established at the state level and the grading decisions being made at the classroom level.

Kristine Dittmar (2005) conducted a similar study in Florida examining the reading report card grades of 1,064 third grade students and their performance on Florida's state assessment in reading. Her work found similar positive correlations between the two, but differences arose between schools. Among other factors, students for whom homework was factored as part of their grade had a weaker relationship between teacher assigned grades and performance on the state test. Dittmar also found that teachers who focused on reading instruction developed more specifically for each student had a higher correlation than those that did not. Put another way, differentiated instruction and individualized instruction proved effective as measured by the state's summative assessment. Her methodology was a correlation analysis of grades and test performance and examination of teacher practices and beliefs using a survey method.

Secondary Grades and ACT as Predictors of College Readiness

The focus of this research project was at the high school level, employing a similar correlation analysis, but it extended the work by examining both English language arts and math grades to ACT performance. Furthermore, teacher practice and beliefs were studied through qualitative means.

At the secondary level, much has been done on the relationship between performance on tests such as the American College Test and subsequent performance in college. For example, R. D. Peters (2001) looked ACT performance and high school grade point average as predictors of how students would fare in a freshman psychology class. While high school grade point average and ACT scores were not the relationship being examined, these were among the best predictors of success. It was also noteworthy to see ACT and grades in the same sentence within a controlled study. In a similar endeavor, Noble and Sawyer (2002) examined high school grade point average and ACT performance to predict different levels of achievement in college. Their work found that high school grade point averages compared well with those in college at the lower levels (2.0 to 3.0 on a 4-point scale), but not at higher levels. ACT scores, however, were an accurate predictor of all levels of college grade point average. Establishing a stronger relationship between high school grades and ACT scores might make grades an even better predictor of college performance.

The American College Test has also done extensive work on the number of core courses taken and performance on its assessments (ACT, 2006), but what they have not done is look at the relationship between high school grade point averages in core classes

and performance on related subtests of the ACT, presumably because they do not have access to this information.

Finally, perhaps the study that most closely mirrors the objectives of this work was conducted by A. J. Bowers (2007) in his dissertation from Michigan State University. In his work, he actually correlated teacher assigned grades and ACT scores for two high schools in the Midwest. Interestingly, he found strong correlations in one, but not the other. Although he extended his work quantitatively using a cluster analysis of correlations, he did not have a qualitative piece that might perhaps ferret out the differences in the correlations between the two high schools under his microscope. This project extended that work by replicating the quantitative side by finding correlations between ACT and teacher assigned grades, but then extending into the qualitative realm through a case specific phenomenology.

Study Rationale

As alluded to in Chapter I, the era of NCLB has brought an increased level of accountability in terms of student outcomes or, more specifically, student achievement as measured by state assessment. Because of this federal pressure, states have imposed high-stakes testing (NCLB, 2002) to comply with this legislation. Furthermore, states are using the ACT as a major part of their high school testing regimen (ACT, 2006).

State assessment seeks to measure content and performance expectations published by state departments of education. At the same time, teachers must make daily decisions about which instructional activities will best meet learning objectives, which assessments will give the fullest picture of the extent to which students have met those objectives, and, finally, what grades to assign. Determining through a mixed-method

approach what the relationship is between policies the state adopts and decisions the teacher makes helps us understand connections between the two.

Even though grades have been a hallmark of the experience of most students and teachers in the K-12 environment, the research cited above on grading practice does not put much confidence in either the reliability or validity of this staple of the educational process. That said, both Dittmar (2005) and Bowers (2007) found some decent correlations between grades and state assessment in some schools but not others. What this study sought to determine then, in broadest strokes, was the following:

1. What are the trends or correlations between ACT and teacher-assigned grades in two high schools?
2. How do teachers reflect on state assessment when planning, assessing, or assigning grades to their students?

Summary

Much is known about how K-12 teachers assign grades to their students, but almost too much is known about the number of non-academic factors that impair the validity of those grades. Moreover, tracking the relationship between ACT scores and grading practice gives K-12 educators one way to measure and/or enhance the validity of their overall assessment practice. And in cases or schools where that correlation is weak, it gives rise to some healthy questions and dialogue.

Finally, this endeavor may appear to rest on two assumptions that should be addressed. First, the assumption that a tight alignment between high school grades and ACT scores is desirable may be apparent, and many may not agree with that as it would narrow the curriculum and perhaps encourage teaching to the test. This study, however, is

more descriptive than prescriptive. It merely sought to find out what is, not what should be.

Another assumption that one might infer from this study is that high-stakes testing has an inherent value. While that case could and has been made by proponents of NCLB, the work proposed here rests not on the value of high-stakes testing or lack thereof. I am making the assumption, however, that the era of accountability, measured in large part by testing, is not disappearing anytime soon. Put another way, the accountability side of the continuum in Figure 2 is well developed and is not going away.

Chapter III delves into the specific methods by which this study was executed.

CHAPTER III

METHODOLOGY

My purpose in Chapter III is to provide a well-researched rationale of the method by which this study will be executed. Furthermore, this chapter gives a detailed description of the research methodology and analysis steps that were used to answer the research questions. Given the mixed-method design, two distinct phases will be described.

Study Methodology and Rationale for Mixed Method

The problem presented in Chapter I is to assess how teachers make a connection to state assessment standards as they make decisions about planning instruction, assessing student work, and assigning grades. A mixed-method approach was used to examine this issue. In the broadest terms, Stake (1995) differentiates qualitative and quantitative research methods by referring to the former as looking for understanding of complex relationships and the latter as seeking explanation and control. This study employed a mixed-methods design as it made use of both a quantitative method (i.e., a correlation analysis) and a qualitative method (i.e., a phenomenological method within case-specific schools). Creswell (2003) notes that mixed-methods approaches have gained utility in the research literature. As a method, it fit well for the research questions posed by this study:

1. How do teachers describe their decision-making process with regard to planning instruction, assessing student learning, and grading practice?

2. How do teachers describe and interpret the relationship between state assessments for their content area and their instruction, classroom assessments, and grading practice?
3. What is the relationship between grades given in English language arts and math over a 4-year period (2006-2009) in the participants' high schools and performance in those same core areas on the ACT?
4. How do teachers interpret the relationship between the grades given in core subjects and student performance in the same core curriculum areas on the ACT assessment for their school?

Questions 1, 2, and 4 lent themselves to qualitative inquiry, while question 3 was addressed quantitatively.

Creswell (2007) tells us that the philosophical assumptions that undergird qualitative approaches embrace what reality is (i.e., an ontological perspective) as well as the nature of knowledge (i.e., an epistemic perspective). Further assumptions include the values embraced by the researcher (i.e., the axiology), the language chosen by the researcher (i.e., rhetoric) as well as the methods used (methodology). Morrow (2005) further helps us understand the qualitative approach in general when she describes it as *idiographic* and *emic*, meaning that it looks at one individual or a small number of individuals and tries to draw out patterns and themes.

Within the qualitative tradition, many options are available for use by the researcher to achieve the desired knowledge ends including narrative, grounded theory, case study, ethnography, and phenomenology (Creswell, 2007). I used phenomenological methods for the qualitative portion of this study as it “describes the meaning for several

individuals of their lived experiences of a concept or phenomenon” (Creswell, 2007, p. 57). Moustakas (1994) further defines the phenomenological approach as appropriate for anything that presents itself to the consciousness, either real or imagined. The common and very real phenomenon that many of the teachers experience is the need to assign grades to their students. What they have perhaps not done is consciously consider why and how they arrive at those grades.

In another sense, however, this study combined two research methods within the qualitative tradition: phenomenology and case study—the former because the study examined how individual teachers interpret (or make sense) of their experiences with the grading and assessment process, and the latter because the study focused on the relationship between grades and state test scores in two schools (or two cases). Stake (1995) differentiates cases that are instrumental or intrinsic in his discussion of how to choose particular cases. Instrumental cases are those cases that, if examined, will lead to some understanding of a particular research question within a specific context.

The two schools selected for this study are instrumental in that they offer an accessible context in which to study the questions that guide this study and they are of size, demographics, and program configuration. The cases are not instrumental because they necessarily generalize to all schools, but because they are similar to each other and offered the researcher access to the elements under study, i.e., the relationship between grades and state assessment scores and the way in which teachers experience grading and understand the relationship between grades and state assessment scores. Thus, if differences presented between the two in either the qualitative or quantitative findings, further investigation would yield insights into why these differences are present in the

data. If, on the other hand, few differences presented in the findings gleaned from the two cases, future studies might be designed to replicate this study with a more random sample of districts with similar characteristics to investigate the transferability of the findings. As seen in the findings, the two schools were more similar than different.

Within a mixed-method design, Creswell (2003) further notes that the researcher must decide on the sequencing of the data collection. I began with the quantitative data collection so that the correlations between GPAs and ACT performance could be integrated into the qualitative interviews. This enabled participants to reflect on both their thought processes around planning instruction, assessing student learning, and assigning grades, as well as how those grades assigned actually correlated with a state assessment, the ACT.

Study Methods and Procedures

This work took place in two successive phases: quantitative followed by qualitative.

Research Design, Phase I: Quantitative

The first phase of this study examined the relationship between recorded semester/course grades in English language arts and mathematics and recorded ACT test scores for the same core subjects for 4 years of graduating students (2006-2009), in each of the two case study schools. Specifically, Phase I of this study sought to answer the question: What is the relationship between grades posted for graduates of the 2006-2009 classes from the case study high schools in English language arts and math over a 4-year period (2006-2009) and the posted ACT scores for the same students in those same core curricular areas on the ACT assessment? As noted in Chapter I, beginning with the spring

of 2007 (or for the 2008 graduates), *all* high school students were required to take the ACT as part of Michigan's assessment program. In prior years, only the college-bound students who elected to take the test had scores on their high school transcript. Both of the case study schools provided access to the transcripts and ACT scores for all students in the listed graduating classes; however, this study used only students for which ACT scores were available to match against the grades for courses in the two subject areas.

To obtain these data, high school transcripts were examined for the students in the above listed graduating classes for the following sets of data: GPA for math classes, GPA for English classes, ACT score for math, and ACT for English. Because these data were used only in the aggregate form for the study, this fell well within the guidelines stipulated by the Family Educational Rights and Privacy Act (FERPA), which lists research studies as a legitimate reason for release of data. Furthermore, specific transcript data connected to an individual student were not used or needed. The data obtained for the correlations between ACT subset scores and GPAs in math and English were needed in aggregate form for the study. The relationship between the math or English language arts GPA the ACT scores for math and English was tested using the Pearson product-moment correlation analysis (Glass & Hopkins, 1996). This yielded four correlations for the four graduating classes mentioned above:

1. Math GPAs and ACT scores in math for School A
2. English GPAs and ACT scores for English for School A
3. Math GPAs and ACT math scores for School B
4. English GPAs and ACT for English scores for School B

These values were integrated into the semi-structured interview protocol along with a brief explanation of what it meant, specifically that the correlations were positive and significant. I could not assume that every participant was familiar with research that employs correlations, so the interview protocol included a common explanation of what a correlation is and what it is not (see Appendices A–D).

Research Design, Phase II: Qualitative

For this work, I also examined the issues above qualitatively, specifically employing a phenomenological method within two case-specific schools. As noted, the common phenomenon under consideration was how teachers make decisions about instruction, classroom assessments, and grading, and the way they do or do not connect those decisions with state assessments. Clark Moustakas (1994) also provides support for this methodological decision when he notes that phenomenological inquiry focuses on the wholeness of an experience as opposed to merely its discrete pieces or parts. Allowing teachers a forum in which to explain their thought process behind planning, assessment, and grading gave a full, rich description of that phenomenon. Participant teachers also had the benefit of reflecting on how they have made these decisions.

Data Collection Methods, Procedures, and Instrumentation

For this study, I used a semi-structured interview lasting 30 to 45 minutes (see Appendices A–D). This approach was employed so that participants were invited to share their experience of planning instruction and assessment, the assigning of grades, and coming to know (or not know) the requirements of state assessment. As the second research question above suggests, I was particularly interested in what impact, if any, knowledge of the state assessment had on a teacher's decision-making process.

Participants received the questions electronically in advance and were invited to take notes on possible responses. During the interview, responses were audio-recorded and later transcribed (Marshall & Rossman, 2006). Also during the interview, correlations between the grades students were assigned and their performance on the corresponding portion of the ACT were shared and explained. Participants were then given an opportunity to reflect on this correlation. This served as the point of merger between the qualitative and quantitative methods and yielded a rich description from participants on their experience in making instructional decisions as well as one potential outcome of those decisions (i.e., the correlation between GPAs and ACT performance).

Sampling, Subjects, Access, and Setting

Participants for the qualitative portion of this study were eight teachers selected from two mid-size high schools in the Midwest with student enrollments between 400 and 700 students. The schools are highly diverse ethnically with socioeconomic levels in the middle- to lower-income levels. Free and reduced lunch rates for School A is 67% and for School B is 63% (MI School Data, 2012). Experience levels of the teacher participants ranged from first-year instructors to veteran staff of 20 or more years of experience. At least four teachers were invited to participate from the math and English departments of both schools, using criterion sampling, which avoided bias in the participant recruitment, given my role in the one of the districts (I serve as a central office administrator in the district of which this high school is a part). The key criterion for recruiting and selecting the eight teachers was the willingness of the participant and his or her involvement with the students who comprise the sample set generating the correlations. Specifically, prospective participants were included in this study only if they

had taught the students whose transcripts were used to obtain the correlations. Teachers in School B were accessed for the invitation to participate and the criteria for selection by the researcher via phone or email after gaining permission from the superintendent; for School A, a building secretary extended the initial invitation, so as to avoid any coercion by me as both the researcher and a district administrator. By having a building secretary handle all the recruitment information, I increased the likelihood that any participant's decision to be involved in the study was truly voluntary. The building secretary described the invitation as coming from a university student researcher, and my identity as the researcher was provided to them if they responded positively to the invitation through the secretary. If still interested, the prospective participants were provided the consent information and could make contact with me as the researcher to confirm their interest.

The recruitment process yielded eight study participants: five from school A (three in math, two in ELA) and three from school B (two in math, one in ELA). Interviews were conducted in teachers' classrooms or a location of their choice before or after school hours, depending on their availability. This also yielded a rich set of observations and classroom artifacts.

Data Analysis Processes and Procedures

Phase I. To identify the relationship between the students' grade point averages (GPA) in math and English language arts and their ACT scores in those same core content areas, I pulled the data from student transcripts in both schools for the graduating classes of 2006-2009 into an Excel spreadsheet and used Excel to run a statistical analysis using the Pearson product-moment test. The test was conducted to examine the

relationship between the calculated GPAs for each student from their math and English language arts courses and the subset ACT scores in math and English language arts.

Math and ELA GPAs were determined by entering grades assigned in those classes into an Excel spreadsheet and computing a simple GPA on a 4-point scale. Corresponding ACT subset scores for math and ELA were also entered into the spreadsheet. After loading the data into the Excel spreadsheet, the Pearson product-moment analysis determined correlations based on matched pairs of GPA and ACT scores for each of the two subject areas. The findings from this analysis were then used in the qualitative interviews (see Appendices A–D). To determine that correlations were positive and significant ($p < .01$), the correlation values were compared to a correlation chart in Glass and Hopkins (1996, p. 641, Table J).

Phase II. The qualitative analysis was more complex. Patton (2002) points out that guidelines exist to help the qualitative researcher, but the data can be such that the exact pattern of analysis can change. For this reason, I used two forms of analysis to ensure that I could see the critical features (or salient points) embedded in the data in more than one way. The analysis of qualitative data is a recursive process and calls for the researcher to move with patience through multiple steps, noting insights gleaned from the data obtained from each step (Richards, 2009). In qualitative data analysis, the process begins with transcription and unfolds in a spiraled fashion until the researcher is satisfied that the data are sufficiently reduced and organized in such a way that what is available in the data to address the study purpose and questions is distilled and converted to themes or constructs that allow the researcher to explain and/or describe how the data

respond to the study questions. In the next sections, I detail the steps I followed in the analysis to establish credibility and conformability (Mertens, 1998) for the findings.

Step 1 – Immersion in the data. One initial step of immersing myself in the data was to transcribe the data from the audio recordings. Further immersion involved reading over the transcribed data sets several times as a way of making myself increasingly familiar with each interview (Marshall & Rossman, 2006). This process is critical for forming and recording overall impressions from the data in researcher memos and using those impressions to check against the yield from the two coding approaches used in this study.

Step 2 – The coding process. Analysis also involved examination of the data both deductively and inductively for salient points that served as building blocks for the verification of data categories and, from those categories, themes and subthemes that address the purpose of the study. I used the research questions and questions from the interview protocol to develop the frame or lens for the deductive phase of the analysis. This yielded deductive typologies around issues of planning instruction, assessment of student learning, grading, and the impact of state assessment.

I then engaged in coding the data. This is a more formal step of organization in which labels were placed on the various units of data, called salient points or in vivo codes. This process was aided by using color coding for the in vivo codes based on typologies of planning instruction (yellow), assessment of student learning (green), and grading practice (blue), so that in the process, new insights (red) could arise (Marshall & Rossman, 2006).

After deriving in vivo codes (salient points) from the data sets based on the typological lens, I placed the yield aside and began phase 2 of the data analysis—the inductive phase. For this phase, I set the deductive typologies aside and re-examined the data sets for additional salient points in the form of in vivo codes that did not emerge from the deductive phase of the analysis. These codes were added to those derived from the deductive phase, and I moved into phase 3, which is the process of further reducing the data by forming categories and, from those categories, themes and subthemes.

The examination of the data inductively helps avoid constricting deeper insights within the data by the prefigured typologies used in the deductive phase of the analysis. The deductive phase, on the other hand, ensures that the data are analyzed in a way that conforms or tightly with the purpose of the study and the research questions. By combining both the deductive and inductive forms of analysis, I was able to avoid dislocation from the study focus (Saldana, 2009), while maintaining sufficient opening for either deeper or tangential understandings to emerge from the data where they might have been masked by the tighter deductive analysis frame.

Step 3 – Distilling themes and subthemes. Since the coding process was both deductive and inductive with an emphasis on the latter, the deductive analysis shaped preset categories of codes, while the inductive analysis offered categories that could either fit under one of the preset coding categories or stand alone as a new coding category. Finding categories and then examining how those categories suggest themes or subthemes is the interpretive step in which the researcher, through further engagement with the data, organizes and reorganizes codes into categories until the categories begin to solidify and themes begin to emerge (Patton, 2002). In some ways, it was the most

difficult step, but it was also the most effective in attaining deeper understanding and insight. Through the inductive analysis process, I became further involved in the data to discover additional common elements and nuanced understandings of the deductive findings that may have been missed in a purely deductive analysis (Marshall & Rossman, 2006).

By way of example, although the interview questions were focused on the major topics of planning instruction, assessment of learning, grading practice, and state assessment, salient points from the responses yielded insights into how teachers make decisions in these important areas. These insights could also be compared to participant reflections on the correlation between GPAs given in their content area and student performance on that section of the ACT. The deductive approach provided some initial structure to the data analysis process while the inductive approach allowed for the freedom to group and regroup salient points or *in vivo* codes until the grouping process yielded useful categories for addressing the study purpose and questions, and the categories became useful devices for discovering the themes latent in the data (Saldana, 2009).

Following the coding process described above, I grouped categories of responses into major and minor themes. Next, I wrote analytic notes to assist me in conceptualizing both the main and subthemes from participant responses (Creswell, 2007; Moustakas, 1994). Once these themes were captured, I was then able to compare them to the major issues brought forth in the research questions, *i.e.*, the planning of instruction, the assessment of student learning, the assignment of grades, and the impact of state assessment. Finally, I offered interpretations. In this step, as the researcher I was able to

share my interpretations of what was happening in the data and what it might mean (Marshall & Rossman, 2006). It is here in the mixed-methods approach that rich insights began to emerge as themes related to grading and state assessment were compared to actual correlations between GPA and ACT performance.

The Researcher

As a former teacher, principal, and current central office administrator, I have had the experience with the focus of this study in each of these roles. I have also experienced changes in my professional opinions with regard to the planning of instruction, assessment of student learning, and especially grading practice. As a classroom teacher, I viewed grading as the summative total of work completed and an evaluation of skill and knowledge gained. In my current administrative role, I would like to see grades as having a stronger connection to a measurement of knowledge and skill gained, with perhaps a separate report to stakeholders on work ethic (e.g., timeliness, work completion, etc.). As noted above in the introduction, I also made conscious attempts to connect instruction and assessment to state standards and assessments.

My role in the district has advantages as well as drawbacks. The latter are very apparent. Participants may have felt compelled to answer questions in ways they perceived that I wished to hear. To lessen the impact of these potential issues, I left the logistics of the interview completely up to them in terms of location, time, etc. These issues could have been larger, however. In my work in curriculum, much of the focus has been on the kindergarten through eighth grade levels, so my contact with high school teachers has not been as pronounced as it has with elementary and middle school staff. Furthermore, my role in the district had the advantage of access, giving me the needed

availability of transcript information to get the necessary data to calculate the correlations of GPA and ACT. Finally, since I was conducting the study in two schools, one in the district in which I am employed and one in a district in which I am not, comparisons of responses were readily available to check for bias based on my role, and in fact this did not seem apparent as common themes applied across both schools.

Limitations and Delimitations

Creswell (2003) describes delimitations as those aspects of the study design that the researcher is aware of at the outset. With the phenomenological inquiry, one delimitation can be the interview method itself. There is no way to know if participants are truthful or cooperative, even if they provide signed consent (Marshall & Rossman, 2006). This delimiting factor can be mitigated, however, by field testing the interview protocol and process, which is a step included in this study and by member checking—a procedure for allowing participants to review the transcript of their interview and make any additions or corrections. This step did not yield significant changes to the transcripts of this study, which may suggest that participants responded authentically in the initial interview. Another mitigating factor was that the interviews were very low risk for the participants and there was no sign of participant discomfort portrayed by the participants during their participation in the study.

Another delimitation is that qualitative research does not generalize to broader populations like its quantitative counterpart. Given the scope of this study, we learned some things that are happening within these two case-specific schools regarding some teachers' experience of assigning grades, but we would be hard pressed to say this same thing would happen in another setting. Finally, as noted, my role within the district at one

of the schools may have limited or biased responses, but again there was no sign of reticence on the part of study participants during the interview process.

Creswell (2003) tells us that limitations are those items that take place during the study and sometimes these are hard to predict. I estimated in the proposal that the some participants might have answers to interview questions that are rich in detail and provide interesting insights, but others may not, and those interviews may be relatively short and data poor. This was in fact the case. While most interviews were data rich, one in particular was brief by comparison. I also would have hoped for a larger participant pool in the study. Only half of the 16 teachers invited chose to become participants. Of those who chose to participate, however, all followed through.

Regardless of these design delimitations and limitations, execution of this study did yield rich insights into teachers' experience of connecting state assessment results to their own assessment and grading practices.

Activities and Timeline

Upon Human Subjects Institutional Review Board (HSIRB) approval, I sought permission from the two districts to do the study. This involved meetings with the superintendents from each school district to describe the study, answer questions, and gain permission. The first task was to gain access to the transcripts for the four graduating classes needed (2006, 2007, 2008, and 2009). Grades and corresponding ACT performance data were obtained from these documents to get the needed data sets to run the four statistical analyses described above in the Data Analysis section. These analyses, using the Pearson product–moment test yielded correlational values, which were used to

shape question 5 of the interview protocol (see Appendices A–D). I completed this step in the summer and fall of 2012.

The next step was to contact participants via email to request their participation in this study. I stressed that participation was completely voluntary and willingness to do so or choosing not to do so had no impact on them professionally. As noted, in School A in which I hold a central office role, I worked through a building secretary on the initial invitation. Once consent was given via email, I sent participants the questions in advance electronically as well as the consent form, encouraged them to take a brief look at these, and perhaps jot down some notes on their initial reactions. In the same email, I asked for times when they could be available to interview. I suggested locations for the interview that would put them at ease (i.e., their own classroom or a neutral location such as a restaurant), but the final decision on location was left to the participant.

Upon arriving to the interview, I showed them the consent form (see Appendix B) and reviewed it to make sure they again realized that their participation was completely voluntary and they had a chance to get answers to any questions before deciding whether to consent to participating in the study. I then reminded participants who completed the consent process that the interview was to be recorded, but that their responses were to be kept in strictest confidence and that no personally identifying information would be included in the written results of the study. I completed these interviews in the fall of 2012.

Once the interviews had taken place, I began the process of analyzing the data described above. Data analysis and final report preparation took place in the winter of 2012-13 with the defense in the winter of 2013.

CHAPTER IV

FINDINGS

In this chapter, I will review the purpose and research questions of the study, describe the participants, and articulate the results of both the quantitative and qualitative phases of the study. The chapter will close with a summary of the major findings and draw connections between the quantitative and qualitative results.

Purpose and Research Questions

Purpose

As explained in Chapter I, this study was designed to examine how teachers make decisions about instruction, assessment, and grading with a focus on identifying how they connect or do not connect these decisions with the state assessments for their curriculum area. Since the state assessments are derived from the state curriculum standards, this study sought to provide some insight into how teachers consider the curriculum standards represented by the state assessments in their instructional planning and assessment of student learning. To carry out this examination of teacher thinking and decision making, this study asked core content teachers in math and English language arts from two schools to reflect on their decision-making process with regard to instructional planning, assessment, and grading. This study also captured the participant teachers' response to the results of an analysis of the relationship between student grades for these two curriculum areas and the results of the state assessments for those areas over a 4-year period. By exploring both the teachers' frame of reference for making instruction,

assessment, and grading decisions and their interpretation of the meaning of how grades relate to state achievement scores in their content area, this study shed some light on the nature of how the two contexts for measuring student learning (state and federal political context and local classroom teacher experience context) interact with one another. This study also helped us look for indicators of how correlations between grades and state assessments are occurring.

Research Questions

More specifically, this study attempted to answer the following research questions:

1. How do teachers describe their decision-making process with regard to planning instruction, assessing student learning, and grading practice?
2. How do teachers describe and interpret the relationship between state assessments for their content area and their instruction, classroom assessments, and grading practice?
3. What is the relationship between grades given in English language arts and math over a 4-year period (2006-2009) in the participants' high schools and performance in those same core areas on the ACT?
4. How do teachers interpret the relationship between the grades given in core subjects and student performance in the same core curriculum areas on the ACT assessment for their school?

Research question 3 is addressed below in the quantitative results, while questions 1, 2, and 4 are addressed in the qualitative results.

Description of Participants

Schools

Study participants were math and ELA teachers from two neighboring Midwestern high schools, School A and School B, both of which share similar characteristics. Both are situated close to an urban center with demographic characteristics that mirror that urban center: free and reduced lunch rates in the 60-70% range, high ethnic diversity, and a socioeconomic status predominantly in the lower- to middle-class range.

Students

Table 1 shows both the similarities and differences between the student populations of the two schools in the study in terms of ethnicity.

Table 1

Distribution of Student Participants Across Racial/Ethnic Categories by School

	African American	American Indian	Asian	Hispanic	Hawaiian	2 or more races	White
School A	21.4%	< 5%	< 5%	27.72%	< 5%	< 5%	41.23%
School B	23.6%	< 5%	6.84%	36.60%	< 5%	< 5%	31.15%

Comparisons of enrollment, ACT composite scores and graduation rates are as follows in Table 2.

Table 2

Comparison of Enrollment, ACT Scores, and Graduation Rates by School

	Enrollment	ACT Composite	Graduation Rate
School A	570	18.9	87.1%
School B	581	17.2%	86.0%

(MI School Data, 2012)

Participants

In order to preserve the anonymity of the participants, specific descriptions of each one will not be provided. Below, however, are some characteristics of the group as a whole. In total, the invitation to participate was given to 16 teachers, eight from each school. Participants were either ELA or math teachers at School A or School B. Table 3 shows a breakdown of pertinent participant characteristics.

Table 3

Pertinent Participant Characteristics by School Group

	Participant Total	Gender	Content Area
School A	5	3 male; 2 female	2 ELA; 3 math
School B	3	1 male; 2 female	1 ELA; 2 math

Experience levels ranged from teachers entering the field within their first 5 years, of which there were two, to those with 15-20 years or more of experience, of which there were six.

As each interview was conducted, each participant was given a specific code and corresponding abbreviation to classify them within the school and content area while at the same time protecting their identity. Codes, with corresponding abbreviations, were the following:

School A, ELA, participant one	SA.ELA.P1
School A, ELA, participant two	SA.ELA.P2
School A, Math, participant one	SA.M.P1
School A, Math, participant two	SA.M.P2
School A, Math, participant three	SA.M.P3
School B, ELA, participant one	SB.ELA.P1
School B, Math, participant one	SB.M.P1
School B, Math, participant two	SB.M.P2

In the results that follow, specific responses will be attributed to the above abbreviations, and masculine pronouns will be used in every case to ensure anonymity.

Results

The results of the above study will be presented in two major sections: quantitative and qualitative. The qualitative section will be supported by the research questions and the themes and subthemes that address them, respectively.

Quantitative Results: Research Question 3

A Pearson product–moment analysis was completed on the recorded grades in the content areas of math and ELA converted to a content-specific grade point average (GPA) for the graduating classes of 2006, 2007, 2008 and 2009 to address research question 3: What is the relationship between grades given in English language arts and

math over a 4-year period (2006-2009) in the participants' high schools and performance in those same areas on the ACT?

Data collection. To obtain the needed data, I pulled and reviewed individual transcripts. For each student in the respective graduating classes, grades were recorded into an Excel spreadsheet. Letter grades were given standard numerical values based on a 4.0 grade point scale. If an honors or advanced placement course was listed on the transcript, a 5.0 grade point scale was used, which was consistent with the method both schools used to calculate a student's grade point average (GPA). Table 4 shows the GPA equivalents used for letter grades pulled for ELA and math courses from student transcripts for each school for the graduating classes (2006-2009).

Table 4

Letter Grade to GPA Conversion Table

Letter Grade	Standard GPA Equivalent	Honors/AP GPA Equivalent
A	4.0	5.0
A-	3.7	4.7
B+	3.3	4.3
B	3.0	4.0
B-	2.7	3.7
C+	2.3	3.3
C	2.0	3.0
C-	1.7	2.7
D+	1.3	2.3
D	1.0	2.0
D-	0.7	1.7
E	0.0	0.0

GPA equivalents for math and ELA were entered for each student with an N size of 368 for School A and 331 for School B across four graduating classes from 2006-2009. Students had to have an ACT score on file in order to be included in the data set. Corresponding ACT subset scores for math and ELA were recorded and entered into the Excel spreadsheet as well for each student.

Data analysis. Content-specific grade point averages were then calculated in math and ELA for each student. These math and ELA GPAs were then matched with the corresponding ACT subset score in math and ELA for each graduating class, as well as a total for all four graduating classes. The data were then analyzed for correlational values using the Pearson product–moment test. Table 5 presents the yield from the analysis for each school by graduating class.

Table 5

Correlations Between GPA and ACT Scores for ELA and Math by School

	06 Math	06 ELA	07 Math	07 ELA	08 Math	08 ELA	09 Math	09 ELA
Sch. A	0.731***	0.668***	0.743***	0.601***	0.452***	0.320**	0.723***	0.623***
N	43	43	38	38	87	87	95	95
Sch. B	0.587***	0.710***	0.457***	0.600***	0.550***	0.523***	0.550***	0.610***
N	56	56	61	61	104	104	103	103

** $p < .01$. *** $p < .001$.

(Glass & Hopkins, 1996, p. 641, critical values for r in Table J)

Most correlations are significantly at the .001 confidence level with the exception of 2008 ELA for School A, which was significant at the .01 level. Most would be characterized as “moderate positive” according to Hinkle, Wiersma and Jurs (1998)

because they fall within the range of .50 to .70 (p. 118) except for those in the .70 to .90 range (high positive) and those in the .30 to .50 range (low positive).

Finally, the Pearson product–moment test was run on the combined ACT and GPA averages across all four graduating classes (2006-2009) for each school yielding the correlations listed in Table 6 below.

Table 6

Aggregate Correlations Between ELA and Math ACT Scores and GPA by School

	2006-09 Math	2006-09 ELA
School A	0.693***	0.584***
<i>N</i>	368	368
School B	0.543***	0.600***
<i>N</i>	331	331

** $p < .01$. *** $p < .001$.

(Glass & Hopkins, 1996, p. 641, critical values for r in Table J)

All aggregate correlations are significant at the .001 confidence level using critical values for r listed in Glass and Hopkins (1996, p. 641, Table J) and can be characterized as “moderate positive” according to Hinkle et al. (1998) because they fall within the range of .50 to .70 (p. 118). The correlations in Table 6 above were used in Phase II of this study to assist in conducting the semi-structured interviews for the qualitative portion of the study.

Discussion and connection to research question 3. The moderately positive correlations found in the analysis of the relationship between ELA and math GPA and ACT scores derived from student transcripts for the four graduating classes (2006-2009)

of the two study high schools reinforce and replicate findings from studies cited in the literature review (Bower, 2007; Lambert, 2002), which looked at similar relationships between grades and standardized tests. This study's findings also point to a moderately positive and significant correlation between the GPA data set and the corresponding ACT subset scores used in this study. This relationship holds true for each class measured in each school as well as the four graduating classes taken as a whole.

Of some interest is a decrease in correlation strength in math and ELA for the classes of 2007 and 2008 in School A, and in math in School B for the same years. For School A, in math, the decrease in correlation is from 0.743 to 0.452, and in ELA the drop is from 0.601 to 0.320. The *N* size increases for both schools from 2007 to 2008 due to a legislative change in Michigan, which required all high school juniors to take the ACT as opposed to previous years when the ACT was optional for students and, generally, completed by those students considering college after high school. A drop in correlation value is not, however, seen in School B for ELA, and the drop in math for School B is less pronounced (0.600 to 0.523).

This analysis addresses the third research question in that there is a moderate positive correlation (at the .001 confidence level for most) between assigned grades and corresponding ACT subset scores in math and ELA for these two schools.

Qualitative Results: Research Questions 1, 2, and 4

Once the correlations had been obtained for the aggregate relationship between ACT and GPAs for each school, the correlation values were infused into the semi-structured interview protocol (see Appendix A). Interview participants were given the correlation that corresponded with both their content area and school.

Data collection. Potential participants for interview were selected on the basis of membership to the faculty of School A or B and having taught classes in the content areas of math and ELA. Invitations were emailed to participants using a standard script. In School A, in which I serve as a central office administrator, the building secretary sent the email and identified the request as coming from a researcher from Western Michigan University. The purpose of this step was to preserve the voluntary nature of participation and remove any hint of coercion given my role in the district. This was apparently successful, as three potential participants from School A declined participation.

Participants were sent the interview questions in advance. All participants were met at a location of their choosing, often their classroom, but some requested to meet in a conference room. Interviews were recorded with the written consent of the participant. Their anonymity was assured, and they were encouraged to be as honest as possible in their responses. At the end of the interview, participants were again reminded of the confidentiality of their responses, paid a \$25 honorarium, and thanked for their participation. Interestingly, two of the participants wanted their honorarium donated to the school, and this request was honored.

Data analysis. Data analysis began with the transcription of audio recordings to Microsoft Word documents. Following this, deductive typologies were developed based on the research questions and supported by the interview questions. Then, participant responses were read and reread for in vivo codes that fit the typologies that were identified from the research questions and/or that emerged from the inductive phase of the analysis. In vivo quotes within each typology and other categories formed by the inductive analysis were further reduced to themes and subthemes, a process that produced

15 themes and four subthemes for research question 1, four themes for research question 2, and four themes for research question 4. Table 7 summarizes the yield from the analysis in terms of numbers of typological and subtheme elements found and associated with each research question. The themes and subthemes will be identified and discussed in the subsections of this discussion devoted to each research question.

Table 7

Numbers of Typological Themes and Subthemes Found in the Data by Research Question

	Typologies	Themes (T) + Subthemes (ST)
Research question 1	4	15 T + 4 ST
Research question 2	4	4 T
Research question 4	2	4 T

Results for research question 1. How do teachers describe their decision-making process with regard to planning instruction, assessing student learning, and grading practice? This research question is addressed from responses to interview questions 1, 2, and 3 as follows:

1. Describe your experience and/or decision-making process with regard to planning instruction. Or, how did you go about deciding how to plan instruction?
2. Describe your experience and/or decision-making process with regard to assessment. Or, how did you go about making decisions about how to assess student learning?

3. What factors went into a student's grade in your class? How did you come to decide on those factors?

The deductive typologies for research question 1 were planning instruction, assessment of student learning, and grading practice. An “other” category was also available for responses that fell outside of these typologies but had emergent themes worth noting.

Planning instruction. Within the typology of planning instruction, the following themes and subthemes emerged:

Theme 1-A: Student Awareness

Theme 1-B: Student Progress

Theme 1-C: Curriculum Alignment

Subtheme 1-C.1: State Standards

Subtheme 1-C.2: Pacing

Subtheme 1-C.3: Instructional Methods

Subtheme 1-C.4: Relevance

Observations around the theme of student awareness indicated an understanding of the diverse demographic and instructional realities served by both School A and School B. For example, SB.M.P2 shows an awareness for one piece of the ethnic diversity served by the school when he notes, “Because we have a huge Hispanic [population] . . . we say things like ‘reciprocal.’ I’m sure that they use that term at some point in grade schools, but you can’t say that if you know we’ve got kids from everywhere.” Another example of a challenging awareness of the range of student skill, and therefore the complexity of instruction, is illustrated by SA.M.P3:

One of the biggest obstacles I think in teaching, especially in teaching math, is in a room full of, say, 28 kids you will have maybe 7 that get it, like this [snapped finger], “OK, let’s move on.” And you might have 10, maybe 7 to 10, that they are completely confused. And then everyone else is somewhere in between there where they’re OK, could use a little more.

SA.M.P2 indicates the adjustments his department made as a result of being aware that students come to a math course at very different levels: “The whole math department came up with the curriculum that would best fit . . . each student.” SA.ELA.P1 hits a similar chord when he indicates, “I’ve adapted that [diverse student skill levels] to my classroom and our clientele here at [School A].”

Closely related to the themes of student awareness is the theme of student progress, which took the form of knowing how the students were progressing and adjusting instruction or affirming that growth. SA.M.P3 put it well: “Because no matter what you plan, you have to read the students and the assessment, whether they’re written or observed or whatever, and decide, are they ready to proceed?” SB.M.P2 struck a similar theme in terms of adjusting instruction to fit student progress, but then also stressed a need to affirm that progress, no matter the form it took: “So if we can get them to understand that it doesn’t matter where you start, we’re going to teach you something, you’re going to improve.”

The theme of curriculum alignment within the typology of planning instruction took on at least four forms: alignment to state standards, alignment to pacing, alignment of instructional methods, and alignment to relevance. I’ll illustrate them in that order.

Curriculum alignment to state standards was definitely of interest. Although this particular question stem did not mention state standards (“Describe your experience and/or decision-making process with regard to planning instruction.” Or, “How did you go about deciding how to plan instruction?”), it was nonetheless a prevalent theme. Participants did have all interview questions in advance, so they had time to think about the questions and bring forth whatever that question prompted by way of a response from them, even if that response went beyond the strict parameters of the question. Clearly, aligning instruction to state curriculum standards is an integral part of the way they think about and make decisions in the instructional planning process.

By way of background, the state of Michigan is transitioning to the Common Core Curriculum, along with most of the United States so references to “Common Core” are, in fact, references to state standards. SA.ELA.P1 allocated some of his summer to the effort of aligning his instructional plans to the Common Core Curriculum as illustrated by this statement: “So I’m in today [summer] because [of] the Common Core. I need to start combing over that to make sure my lessons prepare the kids for what the common core wants.”

Keeping the Common Core front and center was also important to SA.M.P1, so much so that “when I’m preparing for a specific unit, I look at those Common Core standards.” In School B, the Common Core had an impact on textbook use for math: “This year with the Common Core, I’ve revamped that and found the holes or the gaps and tried to match it to the current textbook” (SB.M.P1). SB.ELA.P1 used an interesting metaphor to describe the role of state standards: “And we’re given the whole state curriculum and this is my Bible [a three-inch, three-ring binder with state curriculum

guidelines supplemented with items found by the teacher] and this is what I work from.”

This same participant later described additional steps he takes to keep state standards in focus when planning: “And then I always keep my little cheat sheet up here by my computer for the four main strands (from the Common Core) that we need to be looking at.”

Another subtheme that emerged relating to theme of curriculum alignment was that of pacing, or the time allotted to get students to master standards in the curriculum. SA.M.P1 characterized it this way: “So then we kind of divide the total number of days out by how many units I have, to figure out how many days roughly I should take per unit.” SB.M.P2 noted the same idea and connected it to both the subthemes of Common Core alignment and pacing as follows:

Because this [textbook] is new, my question will be speed. And I think this first year that we’re doing this is going to be, you know . . . we’re using out best judgment . . . trying to follow the Common Core standards, so we have go get from, you know, A to B.

SB.ELA.P1 pointed to pacing in the construction of curriculum maps in the following:

“How long it’s going to take.” He stressed that this is a key consideration. Finally, SB.M.P1 captured the tension between pacing and the theme of student progress (or lack thereof) when he stated, “I still have kids and I’ve at this point I’m thinking, ‘Oh dear, are we ever going to get these kids past chapter one?’ And here we are testing our third unit.”

The theme of curriculum alignment also played out in comments referring to the subtheme of instructional methods. For example, SB.M.P2 commented on the specific attempt to connect method to the ACT: “So what I try to do, and it’s looking toward the

ACT, one of the possible handout sheet is what they [textbook publisher] call ‘the plan.’ And it’s for every section . . . it’s the thinking process.” In another example of instructional method, SA.ELA.P1 noted that, “For each unit . . . writing is the biggest thing for me,” which is not surprising coming from an English language arts teacher. Another teacher in this discipline, SA.ELA.P2, noted the use of instructional method from more of a process perspective, “Basic pedagogy, how am I going to teach where the kids are going to learn . . . I change that” based on how the students are responding to the instruction.

One math teacher, SA.M.P3, noted the use of technology integration within instruction: “I take them to the computer lab where I spend way too much time designing things that they can do online because I think it’s more comfortable for them than the constant paper and pencil and lecture” and also cited the relative failure of lecture only as an instructional technique: “I can talk all day, and they can copy it down, and they feel really good about it. And then they go to work on their own, and it falls to pieces” (SA.M.P3). Both schools also had built in intervention strategies into their instructional methods, and SA.M.P3 noted its value: “It [after school intervention] is working, just very time-consuming,” as well as it not being a cure-all: “and the kids, there’s few kids in there that after three weeks of trying to pass my very first chapter test.”

The final noteworthy subtheme of the theme of curriculum alignment was that of relevance, or a sense that curriculum had to, in some way, relate to life in general, a high-stakes test, or both. For example, SA.ELA.P2 noted the impact of this in curriculum choices: “I also want to use relevancy to modern society and . . . teaching the *Tale of Two Cities* [would] be pretty tough to make relevant with students.” Ostensibly, according this

this participant, if students cannot connect to the literature choice, its instructional impact is diminished.

Other examples of the subtheme of relevance aimed at both life in general as well as making a connection to high-stakes testing such as SA.M.P2: “Then it’s basically using what I can do to relate to the real world while still making it relevant to testing.” Relevant connection to life after high school was also important to SB.ELA.P1 when he stated, “And just realizing that not everyone goes on to college, but 90% of us go on to some kind of schooling later on. What are you going to need to know how to do?”

Assessment of learning. In addition to the typology of planning instruction, assessment of student learning played a key role in the responses given by participants, especially in response to the second interview question (Describe your experience and/or decision-making process with regard to assessment. Or, how did you go about making decisions about how to assess student learning?). Within this typology, the following themes emerged:

Theme 1-D: Pretesting

Theme 1-E: Formative Assessment

Theme 1-F: Student Progress

Theme 1-G: ACT/MME/ Preparation and Alignment

Theme 1-H: Alignment to Instruction

Pretesting was a theme for participants as they commented on efforts to determine what students knew at the outset of a course. For example, SB.M.P2 noted his fondness for pretesting resources: “Then there’s a lot of practice problems in a pretest; I love this packet.” Furthermore, SA.ELA.P2 noted use of this assessment method as well: “Each

class is going to be different based on what your students are, how you teach them, and . . . you do your pretest and see where they're at." SB.M.P1 also made use of pretesting as a way to see what students needed instructionally, but also as a way for students to calibrate themselves to what the learning would entail: "I give them a pretest to see where they stand, and then it's good for the kids—they know what they still need to learn."

Formative assessment was also a theme within the typology of assessment of student learning. Participants either used the term in a similar vein as Stiggins (2002), i.e., assessment *for* learning, or described a practice that fit the concept well. For example, SA.ELA.P1 described a practice in which he would "meet with kids every other week and document their progress" by asking them questions about what they were reading, which would give information on ongoing progress. SA.M.P1 employed formative assessment in a similar way: "The quizzes aren't graded for me. They're graded by the students so they can see where they are." SB.M.P2 used assessment formatively as well to get a read on what math students thought certain processes should look like: "The lined paper is more when they work on problems on their own; that's their thoughts. They're doing what they think are the steps." He contrasted this activity with those that had more summative value.

SA.M.P3 would use formative assessment to make instructional decisions to prevent him from getting too far ahead of the class: "No matter what you plan, you have to read the students and the [formative] assessment whether they've written or observed or whatever and decide are they ready to proceed." SA.M.P2 used a very similar method:

The [check for understanding] is in front of me so it's all by themselves; no one else can help . . . And then when they turn it in I get a chance to see are there any spots that we need to go over or are students getting it down?

Not surprisingly, another theme within the assessment typology was connected to student progress or lack thereof. SA.M.P3 described the general understanding of how assessment should capture student progress when he said, "Of course the assessment should reflect their effort and their understanding." Closely related, but stressing the growth component, SB.M.P2 noted the importance of capturing and recognizing growth in all students, no matter where they start:

You know they may get 4 or 5 out of 20 right the first time, and 10 or 11 right [on the summative test]. And I think that's a significant difference, you know, it's not based on guessing. So we know they're learning something.

Also related to the theme of student progress (growth), SA.M.P3 commented on how he would have the students estimate how they thought they would perform on a given assessment prior to taking it: "It's something I've learned through just reading about successful teaching methods is to have the students gauge their own success as we go." When growth did not take place, SB.M.P1 described how the intervention system works in School B if sufficient progress is not made: "So the kids who do not, do not make 75% or higher on every assessment, they have to go into intervention and relearn the material and retest until they hit 75% or higher on every assessment."

Finally, SA.ELA.P1 captured the satisfaction of both teacher and student that growth can bring: "I think that by the end of the year they realize, 'Oh, my gosh, I can do so much more than what I previously believed.'"

A fourth theme within the assessment typology was an intention to align the assessment to the instruction that had taken place. For example, one of the ELA teachers for whom writing was a key instructional goal noted, “So, a lot of their assessment for my class is writing” (SA.ELA.P1). A math teacher, SA.M.P1, gave a more detailed example of how he makes assessment decisions connected to instruction:

If the “I can” statement is one where it’s more of a memorization type statement, then I can use a question like a multiple-choice question or just a quick fill-in-the-blank question. If it’s one that I think takes more of a higher level thinking, then that’s one where I want to have either a multi-step question or more of a free response question for them.

SB.ELA.P1 noted how he aligns assessment with the instructional goals: “I always developed my curriculum from the top down: where do I want them to be, what do I want the students to be able to do, and what are the goals?” He further explains with an example: “So I have a rubric for their speaking skills, which is not on the ACT test, but I think it’s important for students to be able to speak well and to present something.” SB.M.P1 noted a similar practice: “It’s by topic and of course the lessons that I teach. I already know what the test is going to be because of the goals I want.” Finally, SB.M.P2 collaborates with colleagues to a similar end:

[A colleague] and I sit down at the beginning and either one of us will go through the exam view program [software] from this and make sure we get one or two questions out of each of the sections [referring to instructional units].

A final theme within the assessment typology was the connections participants made to the ACT or MME in terms of preparation or alignment. Recall that the interview question prompting these responses made no mention of the state test or curriculum.

SA.ELA.P1 illustrated with the following quote that the state test, the ACT, is clearly on the minds of both teacher and student:

We do a lot of timed readings after I feel like they're ready for it. Usually by the end of September, because of the pace, the kids will tell you flat out that they're not, that they don't feel that they read at the pace required for ACT.

That same participant got even more explicit when he expressed, "How am I going to get you ready for the March [when the test is given], and then college and then life?"

SA.M.P2 framed his instruction in terms of a target on a timeline: "Especially, you know, the junior year their ACT and MMEs."

As participants discussed assessment in general, the ACT continued to emerge. SB.M.P2 framed his assessment practices as something "that to me would correlate with better scores on the ACT because what it makes the kids do is really show all of their work." SB.ELA.P1 even framed the focus on ACT almost in terms of loss: "Lately, it's more test-driven than anything. What are you going to be expected to do on the ACT? It's a quandary."

Following the typologies of planning instruction and assessment of student learning, grading practices formed the third typology addressed under research question

1. Within this typology, four themes emerged:

Theme 1-I: Connection to Assessment

Theme 1-J: Weighting of the Summative

Theme 1-K: Student Impact

Theme 1-L: Connection to State Standards

The first theme, connection of grading practice to assessment, emerged in a more general sense with statements such as this one from SA.ELA.P1, who directed his students in writing assignments to “compare that [reading assignment] to what you’ve read in your independent reading book. And they have to write them up. That would be a grade, and it’s really easy to tell if they’re reading or not.” Another ELA teacher commented on the direct connection between assessment and grades with, “I’ve always graded kids based on assessment and what I mean by that is whatever assessment I give . . . I don’t give grades based on participation, good behavior or anything like that” (SA.ELA.P2).

SA.M.P1 commented on how homework fit into assessment and the student’s grade:

Now that homework category [of the grading scheme]. I know it says homework, but it’s not just homework. It’s all of our daily activities, our supplement activities from our formative quizzes, our practice in class, the actual homework, the homework quizzes we take at the end of each lesson after our formative quizzes; that’s all kind of linked in there.

SB.M.P2 discussed how retesting of an assessment tool fit into the grade as well, and how some students may take advantage of that: “And they get to keep the score of the retest. There’s not an average in there. So that’s a heck of an advantage. Some kids of course use that.”

The second theme within the typology of grading practice was the weighting of summative assessments, such as tests, and how that fit into a student's overall grade.

SA.ELA.P1, realizing the currency metaphor of grades, arranged it as follows:

If I didn't weight them, kids were opting out of writing assessment pieces, which is not good . . . So I had to start weighting my grades . . . at 45% for writing. "If you," and I make this very clear to my students, "if you choose not to do a written piece, you will sabotage your grade."

SA.ELA.P2 also tried to match the summative weighting to the amount of work involved for the student: "Some assessments might be worth more than others based on how much time you put into it in the classroom."

Math teachers in School A had come to consensus as a department on the weighting and placed more emphasis on summative measure. SA.M.P1 captured this thinking directly:

The only thing that was in the test section was our final test. I'm a firm believer that the final, the final product of what they do over the unit is the most important one . . . Basically those test scores, the 70% [of the grade] was just one test. And then probably maybe 40 other grades went into that 30% over the course of the unit.

SA.M.P3 reinforced this: "Within the math department, we set up the percentages all the same . . . 70% is tests and the other 30% is divided [into homework and quizzes]."

In School B, it was clear that the administration had made the decision on the weighting between summative measures and formative assessments such as homework. SB.M.P1 stated it this way: "This year we were told by our administrators that we have to

have 80% of our grade be tests and 10% homework and 10% quizzes. Everyone has to do it.” SB.M.P2 echoed this administrative decision as well: “The administration did [made the decision on grading]. It’s 10% for homework, 10% for quizzes, and 80% for unit tests. So there really is no factor in for . . . um, like how you act in class, nothing for that.”

A third theme under the typology of grading practice was the student impact or how parts of the grading practice affected students. For example, SA.ELA.P1 pointed to a concern regarding potential harmful effects of grading when she noted that grading had an impact on a student’s “self-esteem, the confidence. They feel like they’re going to be ripped apart, so why even bother doing it?” An ELA participant from School B hit on a similar theme: “So I try to make sure variety’s there, and there’s a sense of it’s OK to practice. And it’s OK to learn. And there’s not a penalty for doing that, and sometimes there’s a reward” (SB.ELA.P1).

Concern was also expressed regarding the impact of grading decisions on students. For example, SB.M.P1 noted the school decision to go to 80% of final grades being made up of summative tests: “Some of our kids are very poor test takers. And it’s very difficult for many of them to have 80% of their grade be tests. So, that’s challenging.” SB.M.P2 noted as well that even in light of that 80/20 split between summative measures and other formative measures, students “don’t grasp the importance of a quiz versus homework.”

The last theme of note in the grading practice typology was a connection to state standards. As with the connections participants made between assessment of student learning to ACT and MME preparation, it should be noted that the interview question

made no mention of state standards. Participants, however, did choose to bring it up. SB.ELA.P1 noted it in his choice of activities to grade: “I do writing every couple of weeks. We do some kind of practiced, timed writing, and they get credit if they do it.” Timed writing is a key component of the ACT. This participant further clarified, “I try to make sure that every assignment I do touches on one of those four strands (from the state) in some way.” SA.M.P1 hit this theme as well when advocating for a high percentage of summative assessment use in the final grade in terms of ACT and MME preparation:

I know some kind of debate as to which one should count more, but I think that the test is what should count the most because I think that prepares them then for ACT and MME by having a high-stakes test at the end. I mean they really [have to] show what they have learned.

Emergent themes outside of the typologies of planning, assessment, and grading. Some responses fell outside of the typologies of planning instruction, assessment of student learning, and grading practice, but nonetheless revealed pertinent themes in teacher thinking. Three themes emerged:

Theme 1-M: Student-Related Reflections

Theme 1-N: Teacher-Related Reflections

Theme 1-O: Disenfranchisement

I’ll first describe the theme of student-related reflections. Within student reflections, comments about students’ affective or emotional lives arose among participants. For example, SA.ELA.P1 said that he noticed most of his students did not like to read, and thus, “because they don’t like to read, I also feel like their competence in writing isn’t there.” He further noted that the experience of most students coming into an

ELA class is not positive: “The whole concept of English for most of my students, and it’s always been this way for 12, 13 years, is they hate it.”

Affective issues came up in math as well, according to SA.M.P3: “Sometimes it can be a motivational issue like . . . some kids feel that may be this [math] doesn’t pertain to their lives and they’re going to put forth minimal effort.” This participant had developed, however, ways to address this with students:

As the years go by, I’ve gotten really good at noticing, I can tell if you are a math person or not, right away, as far as with students. Some of the kids that say “math is so difficult for me,” and I can say, “Well, actually when I compare you to other students, math is not difficult for you” (SA.M.P3).

SB.ELA.P1, also desiring to make his content area engaging for students, noted, “I also know that it’s good for kids to do other things creatively, and to be able to express themselves in other ways.” In context, this comment was in reaction to a test focus in the curriculum.

SA.M.P3 would pair students in an effort to engage them, noting students’ desire to be social: “Kids do like to work in partnerships for the most part and they don’t, these kids don’t ever know they’re in the lower 10 and I’ve put you with the upper 7.”

SB.M.P2 also realized students’ desire for peer affiliation, even if it was in an after school intervention program:

They don’t care if they’re in intervention because all of their friends are there.

Then they’ll come in, study some more with their friends, and then take a test and get an 80. But you know what? They still got an 80.

Within the theme of student-related reflections was a sense for the diversity of the student population these two schools serve in terms of skill level and language. For example, SA.ELA.P1 noted, “Most of my students, the majority of them don’t read at level” and that “maybe they can’t spell.” SA.ELA.P2 noted the native language diversity in his experience: “I had a class last year where I had eight students who were ELL students and the needs of those students are . . . much different than the needs of a higher level class.” This language diversity was also seen in School B by SB.M.P2: “Some of these kids, you know, I have kids that are a week out of another country [and] don’t speak any English . . . We’re looking at a whole variety.” SA.M.P3 noted the more general diversity in skill level and the challenge that can present in grouping students for cooperative learning: “The hard part is the kids that really excel quickly don’t tend to be very good at teaching it or explaining it because it comes so easy to them. The middle kids are better at that.”

A variation on the theme of teacher-related reflections was the reference to reflections on themselves and their colleagues. Some teacher anxiety seemed to emerge in participant comments. SB.M.P1 rather honestly shared that in the transition to the Common Core, “I’m having to make changes. It’s hard, truthfully.” SB.M.P2 noted some challenges with regard to not having a textbook: “We don’t have the book yet. That’s the other thing . . . so I can’t say let’s work on problem, problem, problem.” Also, SA.M.P3 noted the challenge of serving a diverse skill level: “It’s . . . that’s probably the hardest for me is figuring out how to keep two ends of the spectrum engaged, learning and not to lose them.” This was even a concern if students were being served in an intervention program, and if not, how would teachers get students to the instructional finish line, as

noted by SB.M.P1: “It’s [staying after school in the intervention program] very discouraging too for some of those kids. I just wonder how am I going to get them through a school year, truthfully.”

Participants also noted a bit of teacher disenfranchisement (the final theme associated with research question 1), as certain decisions that used to be theirs were now made by someone else. For example, SB.ELA.P1 reflected that in conversations with colleagues, “we talked about all of the things that we used to do that were fun, that hooked kids into reading and writing and learning and fun, and we feel like we’ve lost a whole lot over the years because of the attention on the test.” SB.M.P2 noted a similar pattern with regard to grading decisions made by administration: “The only thing I can do is decide whether I want to call something, and I do this sometimes, instead of calling it ‘homework’ I’ll give them a take-home quiz.”

Discussion and connection to research question 1. Research question 1 sought to determine how teachers think about planning instruction, assessing student learning, and grading practice. It’s clear from the varied and multiple responses that teachers do put quite a bit of thought into these issues.

In the planning of instruction, participants discussed being aware of where students are from both a skill level and language diversity standpoint. They also kept a close eye on student progress as a means to inform and adjust their instruction. Attention was given to state standards and the transition to the Common Core, along with attention to pacing to make sure students and instruction stayed on track, and varied instructional methods such as technology integration to provide opportunities for all students to be successful.

Assessing student learning is no less complex. Most participants used some form of pretesting and/or formative assessment to track student progress from an assessment perspective. Participants connected assessment to instruction as well as the ACT and/or MME, even though those were not part of the interview question.

Participants connected grading practices to summative assessment, although it varied between schools as to who made the decision to weight summative assessments more heavily in computing the student grade. That said, participants had concerns on what impact this was having on students in terms of their self-esteem and perceptions of how successful they could be in a math or ELA setting.

Finally, student- and teacher-related reflections emerged as themes along with a sense of disenfranchisement as teachers deal with the loss of autonomy. Student-related reflections centered on the affective or emotional impact of various decisions made by the teacher, school, or state, as well as an awareness of student diversity. Teacher-related reflections centered on teacher anxiety related to getting students with challenges to be successful and their sense of disenfranchisement due to a loss of decision-making authority.

Analysis of the yield from the interview questions supporting research question 1 revealed a highly diverse set of themes and subthemes (19 in all) that were all related, revealing the complex nature of teaching and learning. Also of interest were the frequent connections participants made to state standards and assessment without any prompting from the question prompt. The interview questions for the next section, research question 2, are more direct by contrast.

Results for research question 2. Research question 2 (How do teachers describe and interpret the relationship between state assessments for their content area and their instruction, classroom assessments, and grading practice?) was addressed by interview question 4: “How did you use your knowledge of state assessment in the area in which you taught? What impact, if any, did this have on how you made decisions around the planning of instruction, the assessment of student learning, and the assigning of grades?” The deductive typologies for research question 2 are also planning instruction, assessment of student learning, and grading practice.

Typology of instructional planning. The two most prevalent themes under planning instruction were connections participants made between instructional planning and the ACT and connections participants made regarding aligning instruction to state standards.

Theme 2-A: Connections to the ACT

Theme 2-B: Connections to State Standards

Connection between instructional planning and the ACT was characterized by an overt effort on the part of teachers to align their instruction in such a way as to prepare students for success on the ACT. For example, SA.ELA.P2 described the connection this way:

Well, and we’ve all heard this before, the test drives your curriculum and we have to teach to the test or what the state requires us to do. The days of a doing a unit on lighthouses because I happen to like lighthouses, that doesn’t exist anymore.

SA.M.P1 described using the state standards for a similar ACT-targeting purpose, “So I kind of use those standards and make sure that my lessons are tailored to those standards to prepare them for their tests.”

Even SA.M.P2, who notes he teaches at the freshman level, integrates ACT awareness into his pedagogy:

How I did that, was even though they’re at the freshman level my first . . . anticipatory set for a class room every Mondays and Wednesdays were . . . I called it ACT prep. It’s one or two questions and I would pull them directly off of the ACT plan. So I would just go through and see what we were working on and then I would implement what it would look like in an ACT format.

SA.M.P3 was also keenly aware of the pacing needed to make sure students were on track to do well: “So I know by, say, October 17th that all of this has to be covered . . . because the unit tests reflect what’s needed on the state assessments.” Finally,

SB.ELA.P1 used the ACT as a way to answer an age-old question that teachers get from students: “And a lot of times the kids will say, ‘Why are we doing this?’ And I can say this is practice for the test.”

Complementing connections to the ACT was the theme of connections to state standards or an overt attempt on the part of teachers to align their instruction to what the state had published. Again, it is important to note that Michigan is in the midst of transition to the Common Core.

SA.M.P1 framed the connection to state standards and the ACT: “What I’m doing is getting them the knowledge they need, and if the states holding their end of the bargain by saying ‘this is all we’re going to test you on,’ then I’m preparing them for the test.”

SB.ELA.P1 was even more methodical in his approach to making sure instruction met state standards:

So when we went to the new state curriculum, we had to come up with our own lesson plans . . . But we had to really pick apart our textbook, our assignments that we do. Now that drives me, and like I say I keep a little cheat sheet cause I'm always asking, "Well this was fun to do in the past, but why did I have students do that?"

This participant even went so far as to point out, "So yeah, it [state curriculum] drives me with everything that I do because I feel I've been given a directive by the state of Michigan and that's pretty important" (SB.ELA.P1).

SB.M.P1 hit this theme as well: "Definitely in the planning of instruction no doubt because I have my whole curriculum designed around that Common Core" and SB.M.P2 felt that, as a team, the math department had made necessary connections: "So I think we have faith in math that we have picked a good process to go through that actually does cover all of the Common Core stuff."

Typology of assessment. The typology of assessment for research question 2 centered on the theme of ACT and MME preparation, not surprisingly given the interview question.

Theme 2-C: ACT/MME Preparation

Comments on assessment practices such as the following were common: "The kids do a timed reading from these little literature booklets that we have. I think they're about 400 words long. And then we map and graph. We do that once a week" (SA.ELA.P1). These timed readings were complemented by timed ACT writings.

SA.ELA.P2 noted the tight alignment that he perceived to be important:

“Everything has to be relevant to what is being assessed and what the state requires so you teach to the test, you teach skills that allow kids to do well on the test or the ACT, the MME or whatever the test might be, so that’s how I do it.” SA.M.P2 noted a similar focus in math, even at the lower levels, in terms of exposing students to assessment items similar to those found on forms of the ACT: “So then my freshman would start to see ‘ah, that’s what it’s going look like’ and then when they take the PLAN [a form of the ACT] then they say ‘hey that’s exactly . . .’ and then so on and so forth.”

SA.M.P3 commented on the alignment of the math department’s common assessments to state assessments: “And [names math department chair] is good about . . . you know we with our common assessments as far as the unit tests, they’re all the same. And all, everything matches up to the state assessments.” This effort is further supported with daily assessments: “So, yes, it’s [common assessments] already set up with the state assessment, all the points that are supposed to be hit are already in the unit assessment. And from this, we take this and design all of the daily assessments” (SA.M.P3).

SB.M.P2 was hopeful that intervention efforts in School B would prove fruitful on state assessments: “We won’t know that yet because we’ve just done it [intervention program] with the freshman last year, so now it’s freshman and sophomores. So it will be interesting to see if that actually is true when they start taking MMEs as juniors.”

Typology of grading practice. In addition to the typologies of instruction and assessment, grading practice also yielded some connections to the ACT and/or MME.

Theme 2-C: ACT/MME Preparation (repeated theme)

For example, SA.ELA.P1 noted a direct connection between some grading practices and the ACT: “And we’re actually giving a grade for those assessments because they are mostly ACT released at the junior level.” Further, SA.ELA.P1 pointed out that those assessments get weighted at progressively higher levels of impact on the overall grade: “Sometimes we weight the assessment. For the first quarter it might be 20%, and then second semester it’s 25%, and at 3rd quarter it’s 35%.” School B was also hopeful that increases in the threshold of what constitutes a passing grade would have positive effects on state assessment performance: “By requiring 70% instead of 60% to pass, that hopefully will correlate into better grade on the MME.” He further stated, “And hopefully that [decision of a 70% pass threshold] will correlate into good grades on the MME . . . There will be nobody, I can speak for math, that just gets passed on.”

Discussion and connection to research question 2. As one might expect, explicit questions regarding the use of knowledge of state assessment yielded explicit answers and connections. From participants who work with a “cheat sheet” on their computer, to unit and daily assessments that drill down from common assessments mirroring the ACT, these participants believe they are focused on getting their students positioned to perform well on state assessment. The strength of this focus on preparation for state assessments suggests that there may be a narrowing effect on the curriculum of participants who now use state assessment and standards as primary criteria in making decisions about instruction, assessment, and grading practice.

Results for research question 4. Research question 4 (How do teachers interpret the relationship between the grades given in core subjects and student performance in the same core curriculum areas on the ACT assessment for their school?) was addressed by

interview question 5. Prior to the interview question, the concept of correlation was explained (see interview protocols, Appendices A–D) with visual examples and participants were given an opportunity to ask any clarifying questions. Participants were also given the specific correlation for their school and content area (see Table 6) and it was explained that the correlations obtained from student transcripts from the classes of 2006-2009 were positive and significant at a .01 level of significance (in other words, not likely to have occurred by chance). The interview question was then posed: What is your opinion of this correlation? Is this what you might have expected? As you reflect on it, what factors do you think may have contributed to it? The deductive typologies for research question 4 were assessment of student learning and correlation interpretations. A few participants made comments that fell within a planning instruction or grading practice typology, but these were not nearly as prevalent.

Typology of assessment of student learning. The major theme under the typology of assessment of student learning was the connections teachers made specifically to the ACT.

Theme 4-A: Connections to ACT

For example, SA.ELA.P1 made this statement connecting his work to state assessment and beyond: “I am preparing them for that state assessment, but I’m also reminding them: life, jobs, good people, an active citizen, and the right way. So, that’s what’s important too.” SA.ELA.P2 echoed a similar theme of preparation for the ACT: “I think students that have a familiarity with, um, the style of the ACT test and will do better, which is why we try to have tests that are similar to that in the class.”

SB.ELA.P1 noted that teachers are much more aware of what's on the ACT: "I also think that we as educators are much more attuned to what's on the ACT test and I think more people are teaching with that in mind, expecting students to do ACT-like work and practice in their classes." SB.M.P2 reinforced this point: "We've done a lot to actually prepare them specifically for an MME, for whatever it takes."

Typology of correlation interpretations. The other typology of note was correlation interpretations offered by participants. Themes here included response patterns that the positive result was expected, that reading ability was an important factor in student success, and that student motivation was also critical.

Theme 4-B: Expectation of a Positive Correlation

Theme 4-C: Reading Ability Impacts ACT Results

Theme 4-D: Motivation Impacts ACT Results

Illustrating the theme that positive correlation was expected included statements from all eight participants, as illustrated in Table 8 below.

In addition to interpreting the positive correlation as an expected relationship, some participants attributed factors that explained it such as reading ability or student motivation. Examples of reading ability included in vivo quotes such as this one from SA.ELA.P2: "The ACT is a reading test. Students who read well and can retain the information score much higher on the test." Another ELA teacher, SB.ELA.P1, made the same point: "The better reader you are, the higher your ACT score. So to me that makes sense. I'm not shocked or surprised that there is a correlation." Interestingly, a math teacher had a similar observation: "I think reading is the most important when it comes to these standardized tests. I mean, if you're not literate, then I couldn't imagine someone

being successful” (SA.M.P2). These three participants pointed to reading as an important factor in ACT success.

Table 8

Quotes Illustrating Expectations for Positive Correlations Between ACT and GPA

Participant	In Vivo Quote Illustrating an Expectation of Positive Correlation
SA.ELA.P1	I think it [positive correlation] makes sense to me because if kids are doing better reading wise and the writing’s improving and they’re analyzing.
SA.ELA.P2	I believe that there is a correlation between English scores and ACT scores the GPA in English class and I think that is a result of student’s ability to read.
SA.M.P1	It does not surprise me that there is a positive correlation just because so much of high school is the effort you put in.
SA.M.P2	So for an individual to be successful at math GPA wise or at least grade wise and then to see that kind of a similar relationship or at least a good correlation makes complete sense to me.
SA.M.P3	I would definitely assume there would be a positive correlation between actually both math and reading and a student’s assessment performance.
SB.ELA.P1	Well, I think it’s an honest correlation and it’s what I would expect because so much of the ACT is reading.
SB.M.P1	I think for the most part if the kids are doing well in high school, there’s a reason for it, and it’s going to be directly correlated to the ACT test.
SB.M.P2	You can be assured that there’s some kind of relationship between their scores that they get in high school and their ACT.

Another factor, according to some participants, however, is student motivation. SA.M.P1 pointed this out: “If you’re willing to put in the effort, and regardless of how intelligent you are, if you’re willing to put in the effort you’re going to succeed,” referring to both the ACT and GPA. SB.M.P1 pointed out that student motivation goes both ways in terms of success on the ACT:

So because of that [students caring about post high school learning] they take an interest in the ACT. And so they're going to try, to try to do well on the ACT because that's going to determine the types of schools they may or may not be able get into, possible scholarships. And so they care about it.

He also noted a lack of success on the ACT for students not motivated to do well:

For them [non-college bound student] to take the ACT test on top of it, to them it's just a waste of time. They feel like they're forced to take it. It used to be in the past, of course, they chose to take it. And now, all juniors are taking it.

That last insight points to the shift in Michigan's assessment practice of having all students take the ACT as part of state assessment.

Discussion and connection to research question 4. Given an opportunity to reflect on a positive correlation (at a .001 level of significance) between GPA and ACT results in ELA and math for four graduating classes at their high school, the most prevalent theme to emerge amongst the participants is that they expected this to be the case. Participants interpreted this relationship in terms of the efforts they and their colleagues had taken to connect their instruction and assessment decisions in such a way as to position students to do well, citing reading ability as one of the most important skills to impart. On a more realistic note, participants also acknowledged the role of student motivation, which, if connected to grades, would only strengthen the relationship between grades and ACT subset scores.

Table 9 below summarizes all of the themes and subthemes as well as their relationship to the typologies and research questions.

Table 9

Summary of All Themes and Subthemes by Research Question and Typology

Research Question	Typology	Theme
How do teachers describe their decision making process with regard to planning instruction, assessing student learning, and grading practice?	<i>Planning Instruction</i>	1-A: Student Awareness 1-B: Student Progress 1-C: Curriculum Alignment 1-C.1: State Standards 1-C.2: Pacing 1-C.3: Instructional Methods 1-C.4: Relevance
	<i>Assessment of Learning</i>	1-D: Pre-Testing 1-E: Formative Assessment 1-F: Student Progress 1-G: ACT/MME/ Preparation and Alignment 1-H: Alignment to Instruction
	<i>Grading Practices</i>	1-I: Connection to assessment 1-J: Weighting of the summative 1-K: Student impact 1-L: Connection to state standards
	<i>Emergent Themes</i>	1-M: Student related reflections 1-N: Teacher related reflections 1-O: Disenfranchisement
How do teachers describe and interpret the relationship between state assessments for their content area and their instruction, classroom assessments, and grading practice?	<i>Planning Instruction</i>	2-A: Connections to the ACT 2-B: Connections to State Standards
	<i>Assessment of Learning</i>	2-C: ACT/MME Preparation
	<i>Grading Practices</i>	2-C: ACT/MME Preparation
How do teachers interpret the relationship between the grades given in core subjects and student performance in the same core curriculum areas on the ACT assessment for their school?	<i>Assessment of Learning</i>	4-A: Connections to ACT
	<i>Teachers' Interpretations of Correlations</i>	4-B: Expectation of a Positive Correlation 4-C: Reading Ability Impacts ACT Results 4-D: Motivation Impacts ACT Results

Chapter Summary

This chapter reviewed the purpose and research questions of this study and described the participants, their students and schools, and the data collection techniques employed to answer the research questions. I then reviewed the quantitative (research question three) and qualitative (research questions one, two, and four) results.

Consistent with previous research, most correlations between ACT scores and GPA in ELA and math for the four graduating classes at the two study high schools were moderately positive at the .001 level of significance; thus, confirming that a relationship does exist between the grades given at Schools A and B and the subset scores on the ACT in math and ELA. This finding was complemented by the qualitative findings in which participants spoke of a high degree of intentionality in their decisions around planning instruction, assessing student learning, and grading practices, much of which participants described as focused on preparing students to be successful on the State assessment. Both schools had made decisions, albeit one at the faculty level and one at the administrative level, to increase the weight of the summative assessments in their grading schemes. When asked directly about use of their knowledge of state standards and assessments, participants revealed a highly focused approach to ACT preparation by gearing instruction, assessment, and grading to that state expectation. Not surprisingly, when presented with positive correlations between the grades their students had received and subsequent ACT performance, all participants had expected it.

But what does this all mean in the broader context? That will be the focus of the next chapter in which I will summarize the major findings, discuss implications, and offer recommendations for further research.

CHAPTER V

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

In this final chapter, I endeavor to provide a summary of the major findings, organized by research question, and identify the contributions of this work to the published research. I will also offer some interpretations of these findings and make some recommendations for educational practice and research.

Summary of Major Findings

This section is organized by the four research questions that drove the study, but I will begin with research question 3, the quantitative piece. Findings will then be summarized from research questions 1, 2, and 4 and then connected to the conceptual frame. This will be followed by implications of the findings and then recommendations for further research and educational practice.

Given that the purpose of this study was to examine how teachers make decisions about instruction, assessment, and grading with a focus on how they connect these decisions with state assessment for their curriculum area, the findings discussed below reach that purpose.

Research Question 3: Quantitative

Research question 3 poses the following: What is the relationship between grades given in English language arts and math over a 4-year period (2006-2009) in the participants' high schools and performance in those same core areas on the ACT?

In the most straightforward terms, this question can be answered in the affirmative: there is a moderately positive correlation between grades posted on student transcripts for math and ELA courses and the posted math and ELA subscores on the ACT test (the state MME assessment) for the four graduating classes of students from the two schools, and it was significant at the .001 level of confidence. This finding reinforces what was found in the literature review pertaining to correlation between grades and state assessment (Bowers, 2007; Dittmar, 2005; Lambert, 2002).

For example, both Lambert (2002) and Dittmar (2005) conducted correlational studies between the grades given to elementary students and their performance on standardized tests from Texas and Florida, respectively, and obtained positive results. Alex Bowers (2007) found positive correlations in high school grades and ACT scores as well. What none of these studies did, however, was extend their work into the qualitative realm to delve into the thinking and decision-making processes of teachers.

Also of some interest in this study, however, is the trend of these correlations over a 4-year time period with four graduating classes. As indicated in Chapter IV, correlations decreased for at least one of the two studied curriculum areas (math) at both schools between the graduating years of 2007 and 2008, and in school A, the correlation decreased in ELA as well (the ELA correlation for School B did not decrease). Contextually, the state of Michigan legislated that all high school juniors would take the ACT as part of their state assessment in the year 2008, almost doubling the number of students taking the test in these two schools from 99 students in 2007 to 191 in 2008.

Because states have only begun the use of the ACT for the high school component of the state assessment program in the last few years, studies reviewed for this

dissertation did not address the issue that expanding the population of students taking the ACT from primarily college-bound students to all students might have on a grades to ACT correlation. Thus, it is too soon for there to be much in the way of published studies that trace the trend over time in ACT scores and grades correlations as the assessment becomes more broadly used, rendering the results from this study without comparable findings on this issue. A possible explanation of this drop could be an implementation dip as high schools transitioned their instruction to prepare all students for the ACT, but future studies would need to replicate the correlations in other high schools to ascertain a trend.

Research Question 1: Qualitative

While some research exists on the quantitative aspects of the relationship between grades given to students and their performance on state assessments as seen in Lambert (2002), Dittmar (2005) and Bowers (2007), little qualitative research has been done in this area. The results of the present study, therefore, complement and add to the conversation by delving into the thought processes and decision making of teachers in the areas of planning instruction, assessment of student learning, and grading practice

Research question 1 poses the following: How do teachers describe their decision-making process with regard to planning instruction, assessing student learning, and grading practice?

When participants reflected on planning of instruction, being aware of the skill level of students was prominent in their thinking, as was the language diversity of the students they served. Also, paying attention to students' progress helped to inform teachers' instructional planning, which reinforced what is explained in the literature

review on formative assessment (Schmoker, 2009; Stiggins, 2002). Specifically, participants commented on the thoughtful use of formative assessment to inform their instructional decisions. Participants also commented on the use of pretesting as a means of knowing the knowledge and skill level of students entering their classroom.

While discussing instructional planning, assessment, and grading, the teachers in this study repeatedly referenced the state assessment (i.e., the ACT or MME), even though the state assessments were not directly referenced in the interview prompt. This runs counter to research done by Mazzeo (2001), who suggested a general disconnect between policy decisions and classroom teacher practice. Since state assessment systems and the ramifications of student performance on state assessments are determined by state (and through states, federal) policy, the teachers' intense focus on preparing students for the state tests and aligning their instruction and classroom assessments to the state tests suggests that, for these teachers, there is no disconnect. This finding does find support, however, in the findings from a more current research study done by Berryhill, Linney, and Fromewick (2009), who documented teacher burnout resulting from state accountability policy using a survey method with follow-up interviews for a subset of participants.

The teachers in this study indicated that their grading practices are increasingly influenced by district summative assessment, a recent shift in both schools. Participants also voiced concern regarding the negative impacts of a higher degree of focus on summative assessment and state assessment, such as teacher anxiety and self-esteem issues for lower-performing students. The increased weight of summative assessments in determining course grades, coupled with the teacher experience of anxiety and perception

of a negative impact on student self-esteem, adds to the current research on how teachers think about grading practices. By contrast, Cicmanec (2001) and others (Bonesronning, 2004; Goldwater & Nutt, 1999) discuss the many non-academic factors that have historically gone into grading practice.

The increased reliance on summative assessment for purposes of computing course grades reported by study participants may represent a shift from teachers' traditional theories, philosophies, and practices relative to assigning grades to student performance. Specifically, the findings from this study may suggest that teachers are beginning to either accept or even embrace the need for grades to be validated with explicit measures of curriculum mastery.

Research Question 2: Qualitative

Research question 2 posed the following: How do teachers describe and interpret the relationship between state assessments for their content area and their instruction, classroom assessments, and grading practice?

With interview questions asking directly about the role of state assessment, participant responses became more direct and the focus on state assessment is almost laser beam in some cases. This focus runs counter to the findings of Mazzeo (2001), who characterized the role of state testing at the onset of the accountability movement as having mere symbolic value. For participants in this study, that value had moved from the symbolic to the real.

Concerns about a narrowing curriculum came up as participants use state assessments and standards when making classroom decisions. This complements the current research as seen in the literature review as researchers cautioned educators

specifically that curriculum can narrow if too much emphasis is placed on summative state assessment (McTigh et al., 2004; Schmoker, 2009). Further examples of narrowed curriculum include Gulifoyle (2006), who notes that a key difference in NCLB from previous iterations of the federal education law is the feature of accountability for student performance on standardized tests with the possibility of sanctions for poor student performance levels and, more recently, for insufficient growth in student performance levels. Gulifoyle concluded that this policy shift to high-stakes accountability based on state assessments may have an impact on teacher decision-making regarding instruction, which limits the curriculum. Wantanabe (2007) echoes this point in his qualitative study, which found that high-stakes testing had the same narrowing effect on what is happening in classrooms.

Research Question 4: Qualitative

Research question 4 asked, “How do teachers interpret the relationship between the grades given in core subjects and student performance in the same core curriculum areas on the ACT assessment for their school?”

Participants were not surprised when presented with a moderately positive correlation between grades assigned in math and ELA and the ACT subset scores in these two areas for the same students. This finding adds to the conversation in the literature about how teachers view the connections between the grades they assign to students and students’ performance on state or national assessments. Specifically, although previous research has supported correlations between state assessments and grades to some extent, the addition of presenting teachers with this finding from their own school and hearing their subsequent reactions offers new insights on this issue. These teachers were not

surprised and even indicated that they expected there to be a strong positive correlation between the grades students earn in the classroom and the scores they achieve on state or other assessments.

This positive expectation by teachers is not difficult to understand when taken in the light of how prominently the data showed their intentionality of planning instruction, assessing student learning, and assigning grades that both prepare students for success on the state assessments and predict how they will do. This strong intentionality about alignment and positive expectation that alignment will lead to better student performance that teachers demonstrated in this study may suggest a fundamental shift in how teachers see the state assessment process as a more integral component of the educational process for their students.

Participants also pointed to the increased awareness on the part of teachers of what the ACT demands to achieve a good score and what types of instructional decisions they might make to support student performance on this state assessment. As noted above, this complements prior research findings that high stakes accountability policies and practices may be having a narrowing effect on the curriculum. On the other hand, this apparent level of accepting responsibility for how students perform on state assessments matched with the level of intentionality portrayed by the study teachers for planning instruction that will build students' capacity to perform well on the test may suggest a breakdown of the barriers early studies found in pushing reform initiatives down to the classroom level (e.g., Mazzeo, 2001).

Specific to student performance on the ACT, participants also pointed to student motivation and reading ability as important factors, both of which are supported by

previous research. Taylor (1982) supports that students who are motivated tend to perform better on standardized tests, and Popham (2006) and Stiggins, Schmeiser, and Ferguson (1978) connected reading ability to ACT performance. What this study offers, however, is a glimpse of how teachers are reacting to and responding to the premise that reading performance and motivation are key factors impacting student success on state assessments. The teachers interviewed in this study moved quickly from voicing their assumptions about the importance of reading skill and motivation to describing how they address these issues in their instructional planning.

Conceptual Frame Revisited

Given the findings summarized above, the conceptual frame from Chapter I could be reconsidered as seen in Figure 3 below.

Teachers' report of knowledge of what was on the ACT was strong as evidenced by phrases such as "highly attuned" in reference to an awareness of the content and skills needed for ACT performance. That knowledge also translates into attention and support with instruction, assessment, and grading practice decisions. Participants planned specific learning activities to foster success on the ACT in both content areas. Participants also reported constructing assessment examples at even the freshmen level that are similar to the ACT, with an almost laser beam focus as students near the testing window in the spring of their junior year. Moreover, teacher participants in both schools report using a higher percentage of summative assessments (that have been aligned to the ACT) in their grading schemes. These specific actions and decisions support teachers' report of knowledge of state assessment.

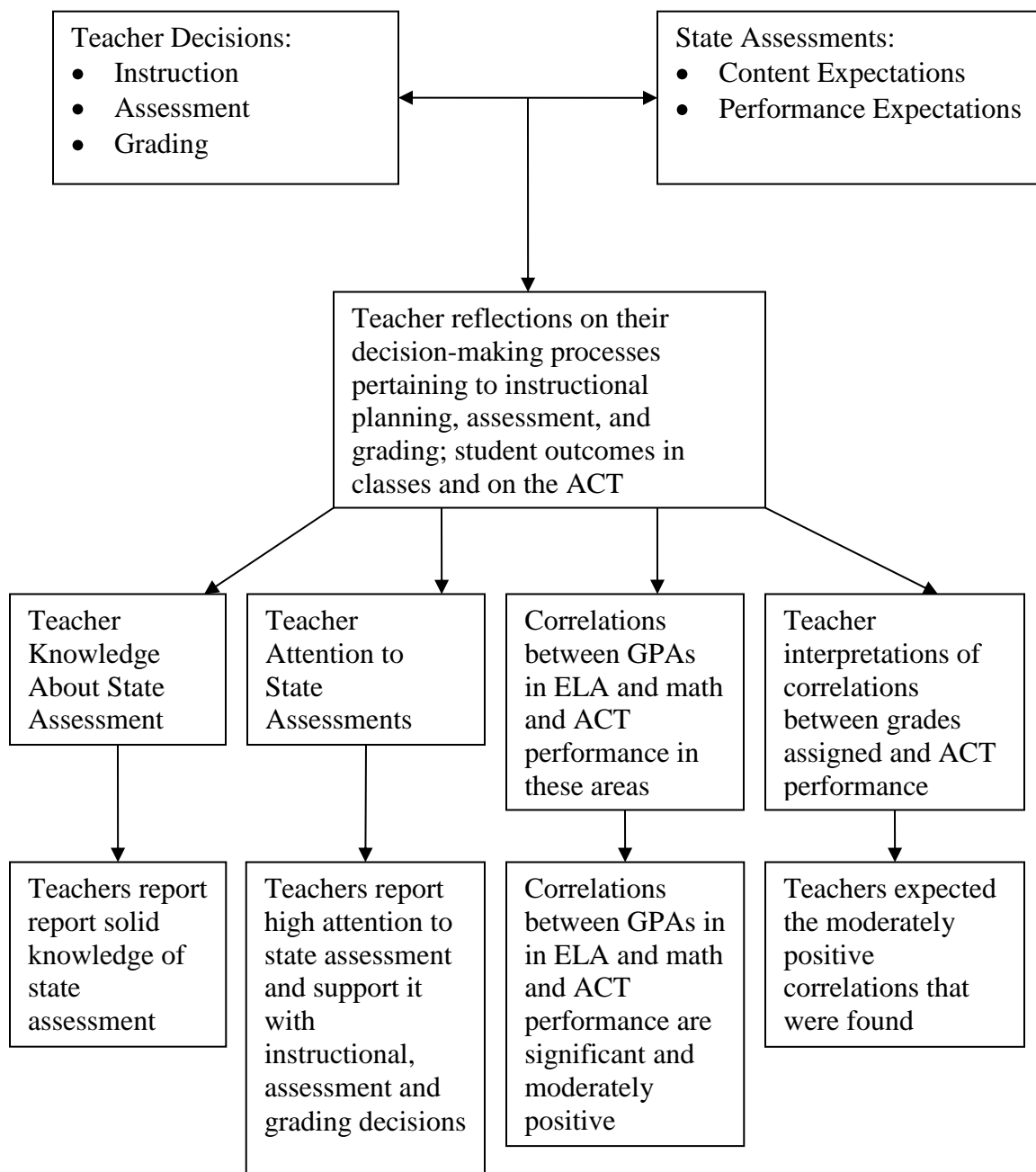


Figure 3. Conceptual frame: Findings.

Given the study's correlations, participants' report of their knowledge and attention to state assessment gains even more credibility. Correlations were moderately

positive and all aggregate correlations were significant at the .001 confidence level. Had the opposite been true (low or non-existent correlations), one might have reason to question participant assertions, but such was not the case.

Finally, perhaps one of the most encouraging findings of the study was every participant reporting that they expected to see this correlation. It came with a sense of pride and affirmed the intentional steps they took to align instruction, assessment, and grading decisions to position students well for ACT performance.

Conclusions

This study points to at least four implications, all of which are related. First, as explained in Chapter II, the education measurement community has a rather dismal view of grading as an overly subjective measure that does not truly capture the skill or knowledge base of students. If that were completely true, then one would think that correlations between grades given and subsequent ACT scores would be negative, low, and more attributable to chance, i.e., well over the .01 confidence level. Consistent with other studies, however, this study showed moderately positive correlations with a high level of confidence between ACT and grades.

Not only were the 4-year cumulative correlations positive and significant at the .001 level, all 16 subset correlations (i.e., four graduating years, two subject areas, and two schools) were also significant at that level (except one that was significant at the .01 level) and all but three were moderately positive. Moreover, both schools realized a combination of moderately positive correlations between ELA and math GPA and ELA scores for each graduating class studied over a 4-year period, with some drop in the r

values to “low positive” (Hinkle et al., 1998) for the 2 years following the shift to all students taking the test as their high school state assessment (MME).

Furthermore, the qualitative side of the study revealed that participants were intentional about aligning their curriculum with state standards and relying more heavily on summative assessments for grading compared to previous years, which may explain why these findings differ with the research in educational measurement. Put another way, veteran teachers in this study referred to having more freedoms in terms of their instructional choices earlier in their careers. Presumably, had correlations between grades and ACT scores been done extensively 15 years ago, perhaps the correlation values may have been lower if all students had been required to take the test.

Another reason that both greater emphasis on authentic grading practices and moderately positive correlations between test scores and GPA may co-exist in this study is that high school teachers may have an increased awareness of what is expected on the ACT since it became a state assessment and are responding to that clarity by investing significant effort to plan and deliver instruction that prepares more of their students to be successful in meeting the expectations inherent to doing well on the test. Participants in this study certainly voiced intention and cited specific classroom decisions that were designed to improve student performance on the ACT. Finally, it may be that students who engage in “good student” behaviors that have often been cited as part of the “hodgepodge” effect of grading (Cross & Frary, 1999), such as homework completion, extra credit, etc., may also have developed themselves intellectually along the way, thus doing well on both their grades and their ACT scores. One manifestation of this intellectual development could be good grades and another could easily be a higher ACT

score. Of course, then, the reverse would also be true: poor student behaviors having a relationship to poor grades and a low score on the ACT.

The problem with this argument explaining the moderate to strong relationship this study found between GPA and ACT scores is that teachers predominantly talked about teacher-controlled factors (like instructional planning and curriculum/assessment alignment) as opposed to factors inherent to the learner (like being “good” students). This suggests that teachers believe that they have the power to influence the congruence between grades and state test scores and the motivation to exercise that power through careful planning and delivery.

The second implication, closely related to the first, is that the quantitative and qualitative pieces of this work support each other. In many different ways, teachers voiced their desire and intention to align instruction, assessment, and grading practices to state standards and the ACT. Quantitatively, these statements were supported by moderately positive and non-random correlations. This internal consistency lends credibility to both findings.

Third, evidence of teacher anxiety regarding student performance on state assessment implies that this century’s two heavy-hitting federal education policies (No Child Left Behind [NCLB] and Race to the Top [RTTT]), with all their interrelated high-stakes features, carry weight with the study participants. To aid in our understanding of this, the current political context of education in Michigan is important to understand.

Several state laws have recently been considered and passed that affect education and could be serving to increase teacher anxiety. This has transpired recently, over a two-year time span (2011 to 2012). These laws have greatly reduced the protection of tenure

to the point that a termination of a tenured teacher can take place for any reason that is not arbitrary and capricious. The previous standard was just cause, a much more difficult level of proof. Furthermore, the seniority of teachers was eliminated as the primary criteria in layoff and recall decisions. Seniority can still be considered, but only if teachers are judged equally effective. Also, mandated annual evaluations for all teachers are now required, and by 2014, 50% of these evaluations must be related to student achievement measures. In addition, certain areas that had before been subjects of bargaining are now prohibited, such as how decisions are made on teacher discipline and discharge, how teacher evaluation is completed, how teacher compensation is connected to evaluations, and where teachers are placed in the school system. Finally, added to all of those changes, in December of 2012, the legislature passed and the governor signed a Right to Work bill, which will prohibit mandatory teacher union membership and the dues that accompany it. This will likely reduce the influence of the two teachers' unions in the state.

In two short years, therefore, teachers have seen their employment security decimated and a key piece of their evaluation tied to measures of student achievement. Comments by participants regarding the loss of instructional decision making and increased attention to student test scores reveal a certain anxiety and stress over student outcomes. What we have, then, in Michigan is an interesting interaction between federal policy (NCLB/RTTT) operationalized by a state decision to utilize ACT as a major piece of state assessment. All of this is taking place within a legislative and political context where there has been a significant loss of teacher job security and professional autonomy.

Given these realities, the next section addresses recommendations for practice and further research.

Recommendations

For Practice

In light of the degree to which teachers in this study were focused on student outcomes as measured by state assessment and how well that finding was supported by current research, educational leaders would do well to keep their interactions with teachers positive and supportive, especially in the state of Michigan. The broader societal context could easily be seen as less than supportive of educators in general, and leaders would be wise to factor in this reality when addressing staff issues. The negative effects of anxiety on performance is far from novel and has been documented on activities from swimming (Burton, 1988) to teachers as graduate students taking exams (Griffore, 1977). Closer to this study, however, was a research project done recently on the anxiety induced by evaluation. A group of researchers (Coy, O'Brien, Tabaczynski, Northern, & Carels, 2011) established that participants who received evaluation-induced anxiety showed decreased performance of working memory and an increase in negative self-dialogue. Connecting that insight to the new evaluation-heavy environment in which teachers find themselves should give educational leaders pause in how this is approached.

From a policy perspective, it is clear that the 2002 iteration of NCLB and its more recent partner, RTTT, have much more influence than previous counterparts, such as *A Nation at Risk* (NCEE, 1983) or *Goals 2000* (Marzano & Kendall, 2007). The policy's impact on teacher thinking and instructional decision making is clear, and we would be wise to heed the warnings of Marzano (2003), McTigh et al. (2004), and Schmoker

(2009) when they caution against a narrowing of the curriculum or a preoccupation with test scores at the expense of deep learning.

For Further Research

Like many studies, this one raised more questions than it answered and follow-up studies could be many.

For example, this study queried the content areas of math and ELA at the high school level, but that is but one slice of the K-12 instructional progression. Other content areas and grade levels could be examined in much the same way to see if findings hold true.

Another fruitful area of inquiry would be qualitative work on the impact of an avalanche of state legislation that decreases the autonomy of teachers individually and collectively. What effects on teacher thinking does reduced decision-making power have on professional self-worth? Would teachers today recommend the profession to college students and what would be the rationale for such a recommendation? What do veteran teachers of 15 or more years see as the most significant changes in the field?

Also, this study revealed two high schools that had both increased the weight of summative assessment into their grading practices. How pervasive is this kind of finding? Does it have an impact on the correlation between grades and state assessment if seen over a larger number of sample schools?

Finally, the quantitative results revealed a decrease in correlation between the years of 2007 and 2008 for one school when all high school juniors began taking the test as opposed to just those planning on a college career. It may be interesting to see if this is a pattern or an exception by examining this relationship with several high schools.

Reflections

In the early 1990s, when I began as an educator and wondered if I was perhaps one of the few educators that was fairly concerned about how students performed on state assessments, the educational setting in Michigan and the United States was vastly different. NCLB and RTTT did not exist, and teachers had a tremendous amount of individual professional autonomy as well as collective influence through their unions or associations.

Teachers today see a different world. Much of their official evaluation is tied to how their students perform on state measures. Furthermore, they might also hear palpable evidence of teacher anxiety amongst even veteran staff who do not know exactly how to get every student to mastery, but who may feel as if their professional careers depend on it. And if budgets get strained, as they often do, the degree of effectiveness as captured in a teacher's annual evaluation becomes paramount.

To see a high degree of curriculum alignment as evidenced by both quantitative results (correlation) and qualitative results (teacher interview) is, in a sense, encouraging. It tells us that teachers are giving heed to state assessments and working together with intentionality to position students for success on state assessment. The state assessment, in this case the ACT, has tremendous impact for the individual student in terms of college acceptance and access to scholarship dollars that can make the difference for many students as to whether they choose to engage in furthering their education. Given the work done by ACT (2006) and Peters (2001) on the predictive value of ACT and high school GPA on college performance, the knowledge and skills gained by students who

attain higher GPAs and ACT scores should translate into success in higher education. In that sense, this tight alignment to state assessment serves the student.

At the same time, to see teacher anxiety bordering on fear and a curriculum that has, by most reports, narrowed is a bit disheartening. Are teachers still seizing the teachable moment that often presents itself within the teacher and student interaction, or are they passing it by in servitude to a state assessment? Do teachers still see themselves as authors of a course of study for their students, or do they view their work as merely responding to a checklist of state directives that form “the Bible” of what should be taught? If answers to these questions fall away from treating teaching as a professional craft and more toward treating teaching as a response to a set of state directives, then something has been lost in education. Teachers need to be seen as the professionals that they are, and education needs to be recognized for the highly complex endeavor that it is. As seen in this study, teachers serve a highly diverse set of students in terms of ethnicity and skill diversity, and getting all of them to mastery is incredibly challenging.

Hopefully education will strive to find a balance on behalf of all of our students. Accountability certainly has a place and lends credibility to any profession. Doctors have board certification and licensure, and attorneys have a bar association. For those who work with children on a daily basis, accountability is an important ingredient. Taken too far, however, accountability can lead to a narrowed curriculum and, perhaps worse, communicate a subtext of mistrust to educators doing the work on a daily basis. To put it another way, the message from the state can be interpreted as, “We’re not sure you’re teaching what students need to learn, so we will test the kids and put a very high stake on it to make sure teachers are on the right track.”

Context is critical. State departments of education and, in many cases, legislatures should be partnering with educators as much as possible to create accountability systems that are fair, empower improvement efforts, and recognize those schools and districts that obtain academic growth as well as proficiency. Getting an immigrant student who has been in the United States for less than a week to the level of mastery needed for the ACT is a multi-year task, for example. It is also a task that both schools in this study confront, and though academic movement takes place, proficiency may not. A proficiency test will not recognize that academic growth.

Trust is also important for any professional. One of the education reform laws recently passed (Public Act 103 of 2011) in the state of Michigan took teacher evaluation out of the teacher contract by making it a prohibited subject of bargaining. This means that districts can unilaterally mandate any teacher evaluation instrument or process they deem worthy. Many districts decide, however, to bring teachers into the conversation anyway. For instance, my own district and Board of Education supports this approach because they recognize teachers are professionals and want to treat them as such. Our premise is that implementing state and federal mandates without exercising local decision-making prerogatives can demoralize the system and the professionals working in that system. Our choice is to include teachers in shaping evaluation and other educational processes in order to capture their best thinking.

As an educational leader, this study reminds me of the complexity of both teaching and learning. Further, it affirms my belief in the dedication and professional desire on the part of teachers to give their students a better future. If the participants in

this study represent the current reality and future of education in any way, I leave this study with sense of optimism for the future of the children they serve.

REFERENCES

- ACT, Inc. (2006). *College readiness: Benefits of a core curriculum*. Retrieved February 27, 2007, from <http://www.act.org/path/policy/index.html>
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(5), 218-223.
- Bauer, E. (1998, February). *HSPT results valid*. Paper presented at the Michigan School Testing Conference, Ann Arbor, MI.
- Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The effects of education accountability on teachers: Are policies too stress-provoking for their own good? *International Journal of Education Policy & Leadership*, 4(5).
- Bonesronning, H. (2004). Can effective teacher behavior be identified? *Economics of Education Review*, 23(3), 237-247.
- Bowers, A. (2007). *Grades and data driven decision making: issues of variance and student patterns*. Doctoral dissertation, Michigan State University, East Lansing.
- Burton, D. (1988). Do anxious swimmers swim slower? Reexamining the elusive anxiety-performance relationship. *Journal of Sports & Exercise Psychology*, 10(1), 45-61.
- Carlson, L. A. (2003). *Beyond assessment to best grading practice: Practical guidelines*. In Wall, J. E. & Waltz, G. R. (Eds.), *Measuring up: Assessment issues for teachers*

- (pp. 507-515). Greensboro, NC: ERIC Counseling and Student Services Clearinghouse.
- Cicmanec, K. M. (2001). Standards-based scoring and traditional grading practices. *Journal of Educational Measurement, 38*(2), 188-190.
- Cicmanec, K. B. M. (1999). High school mathematics teachers: Grading practice and pupil control ideology (Doctoral dissertation, Ohio University, 1999). Available from *Bell and Howell Information and Learning Company*. (UMI No. 9956769)
- Coy, B., O'Brien, W. H., Tabaczynski, T., Northern, J., & Carels, R. (2011). Associations between evaluation anxiety, cognitive interference and performance on working memory tasks. *Applied Cognitive Psychology, 25*(5), 823-832.
- Creswell, J. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. (2nd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education, 12*(1), 53-72.
- Dittmar, K. L. (2005). Factors affecting the alignment of grades and reading scores for third-grade students on the Florida Comprehensive Assessment Test (Doctoral dissertation, Florida International University, 2005). Available from *ProQuest Information and Learning Company*. (UMI No. 3206021)
- Duncan, A. (2009, February). Speech to the 91st Annual Meeting of the American Council on Education. Washington, DC.

- Ediger, M. (2000). *Assessment of student achievement and the curriculum*. Retrieved from <http://search.proquest.com/docview/62318816?accountid=15099>)
- Family Educational Rights and Privacy Act. 20 U.S.C. § 1232g; 34 CFR Part 99. U.S. Department of Education. Retrieved December 19, 2011.
- Friedman, S. J. (2000). Are teachers making the grade? *Educational Forum*, 64(4), 317-322.
- Glass, V. G., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Needham Heights, MA: Allyn and Bacon.
- Goldwater, O. D., & Nutt, R. L. (1999). Teachers' and students' work-culture variables associated with positive school outcome. *Adolescence*, 34(136), 653-664.
- Griffore, R. J. (1977). Correlation between performance on final examination and debilitating anxiety among professional educators. *Psychological Reports*, 40(3), 954.
- Gulifoyle, C. (2006). Is there life beyond NCLB? *Educational Leadership*, 64(3), 8-13.
- Gustafson, K. H. (2005). Beliefs and values of K-12 teachers underlying the grading process in North Dakota small rural schools (Doctoral dissertation, The University of North Dakota). Available from *Proquest Information and Learning Company*. (UMI No. 3220581)
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences*. Boston, MA: Houghton Mifflin.
- Lambert, L. F. (2002). The validity of end-of-course reading grades as indicators of TAAS performance for third-grade students in the Klein Independent School

- District (Doctoral Dissertation, Texas A&M University). Available from *Proquest Information and Learning Company*. (UMI No. 3049968)
- Lott, C. R. (1997). A case study analysis of teacher grading practices: Modification of grading practices at the classroom level based upon recommendations by measurement experts (Doctoral dissertation, Duquesne University). Available from *Bell and Howell Information Company*. (UMI No. 9721226)
- Marshall, C., & Rossman, G.B. (2006). *Designing qualitative research*. Thousand Oaks, CA: Sage.
- Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum.
- Marzano, R. J. (2003). Using data: Two wrongs and a right. *Educational Leadership*, 60(5), 56-60.
- Marzano, R. J. (2006). *Classroom assessment and grading that work*. Alexandria, VA: Association for Supervision and Curriculum.
- Marzano, R. J., & Hefelbower, T. (2011). Grades that show what students know. *Educational Leadership*, 69(3), 34-39.
- Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives* (2nd ed.). Thousand Oaks, CA: Sage.
- Mazzeo, C. (2001). Frameworks of state: Assessment policy in historical perspective. *Teachers College Record*, 103(3), 367-397.
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *Journal of Educational Research*, 95(4), 203-213.

- McTighe, J., Seif, E., & Wiggins, G. (2004). You can teach for meaning. *Educational Leadership*, 62(1), 26-31.
- Mertens, D. M. (1998). *Research methods in education and psychology*. Thousand Oaks, CA: Sage.
- MI School Data. (2012). Retrieved December 27, 2012, from <https://www.mischooldata.org>
- Morrow, S. (2005). Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counseling Psychology*, 52(2), 250-260.
- Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage.
- National Commission on Excellence in Education (NCEE). (1983). *A nation at risk: The imperative for educational reform*. Retrieved April 13, 2009, at <http://www.ed.gov/pubs/NatAtRisk/index.html>
- No Child Left Behind Act. (2002). U.S. Department of Education. Retrieved October 13, 2006, from <http://www.ed.gov/nclb/landing.jhtml>
- Noble, J., & Sawyer, R. (2002, June). *Accuracy of high school grades and college admissions test scores for predicting different levels of academic achievement in college*. Paper presented at the annual meeting of the Association for Institutional Research, Toronto, ON.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.
- Peters, R. D. (2001). Predicting the end of term status of community college general psychology students. *Journal of College Student Retention*, 2(2), 109-114.

- Podgett-Harrison, S. K. (2000). Influences on secondary teacher grading practices (Doctoral dissertation, University of Georgia). Available from *Bell & Howell Information and Learning*. (UMI No. 9994119)
- Popham, J. W. (2006). Branded by a test. *Educational Leadership*, 63(7), 86-87.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education: An International Journal of Research and Studies*, 26(7), 1372-1380.
- Reed, D. E. (1996). High school teachers' grading practices: A description of methods for collection and aggregation of grade information in three schools (Doctoral dissertation, University of California, Riverside). *Bell & Howell Information and Learning*. (UMI No. 9713921)
- Reeves, D. B. (2003). High performance in high poverty schools: 90/90/90 and beyond. *Center for Performance Assessment*, 1-20. Retrieved January 10, 2010, from http://gowmu.wmich.edu/cp/render.UserLayoutRootNode.uP?uP_tparam=utf&utf=http%3A%2F%2Fwww.wmich.edu%2Flibrary%2Farticles%2F
- Reeves, D. B. (2006). *The learning leader: How to focus school improvement for better results*. Alexandria, VA: Association for Supervision and Curriculum.
- Reeves, D. B. (2008). Looking deeper into data. *Educational Leadership*, 66(4), 89-90.
- Reeves, D. B. (2009). *Leading change in your school: How to conquer myths, build commitment, and get results*. Alexandria, VA: Association for Supervision and Curriculum.
- Richards, L. (2009). *Handling qualitative data: A practical guide* (2nd ed.). Thousand Oaks, CA.

- Saldana, J. (2009). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.
- Schmoker, M. (2002). Up and away. *Journal of Staff Development*, 23(2), 10-13.
- Schmoker, M. (2003). First things first: Demystifying data analysis. *Educational Leadership*, 60(5), 22-24.
- Schmoker, M. (2006). *Results now: How we can achieve unprecedented improvements in teaching and learning*. Alexandria, VA: Association for Supervision and Curriculum.
- Schmoker, M. (2009). Measuring what matters. *Educational Leadership*, 66(4), 70-74.
- Shapira, I. (2006, November 21). Those who pass classes but fail tests cry foul. *The Washington Post*. Retrieved November 21, 2006, from <http://washingtonpost.com>
- Smith, B. T. (1999). *Grading practices of elementary teachers* (Doctoral Dissertation, University of Virginia). Available from *Bell & Howell Information and Learning*. (UMI No. 9935036)
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Stiggins, R. J., & Duke, D. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan*, 90(4), 285-291.
- Stiggins, R. J., Schmeiser, C. B., & Ferguson, R. L. (1978). Validity of the ACT assessment as an indicator of reading ability. *Applied Psychological Measurement*, 2(3), 339-346.

- Taylor, C., & White K. R. (1982). The effect of reinforcement and training on group standardized test behavior. *Journal of Educational Measurement, 19*(3), 199-209.
- Turner, M., VanderHeide, K., & Fynwever, H. (2011). Motivations for and barriers to the implementation of diagnostic assessment practices: A case study. *Chemistry Education Research and Practice, 12*, 142-157. doi:10.1039/C1RP90019F
- Wantanabe, M. (2007). Displaced teacher and state priorities in a high-stakes accountability context. *Educational Policy, 21*(2), 311-368.
- Wilson, B. L., & Corbett, H. D. (2001). *Listening to urban youth*. Thousand Oaks, CA: Sage.

Appendix A

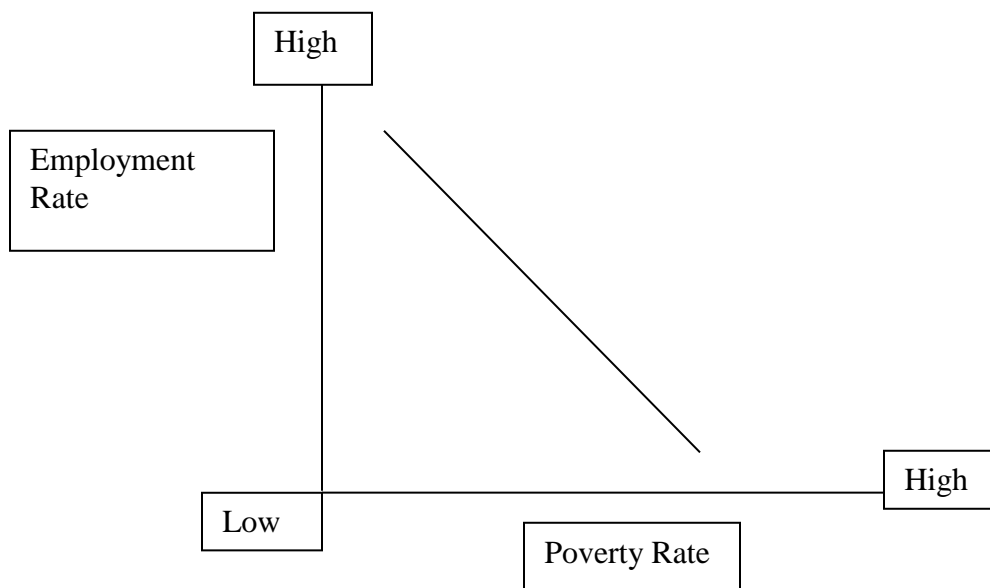
Interview Protocol for School A, ELA

Interview Protocol for School A, ELA

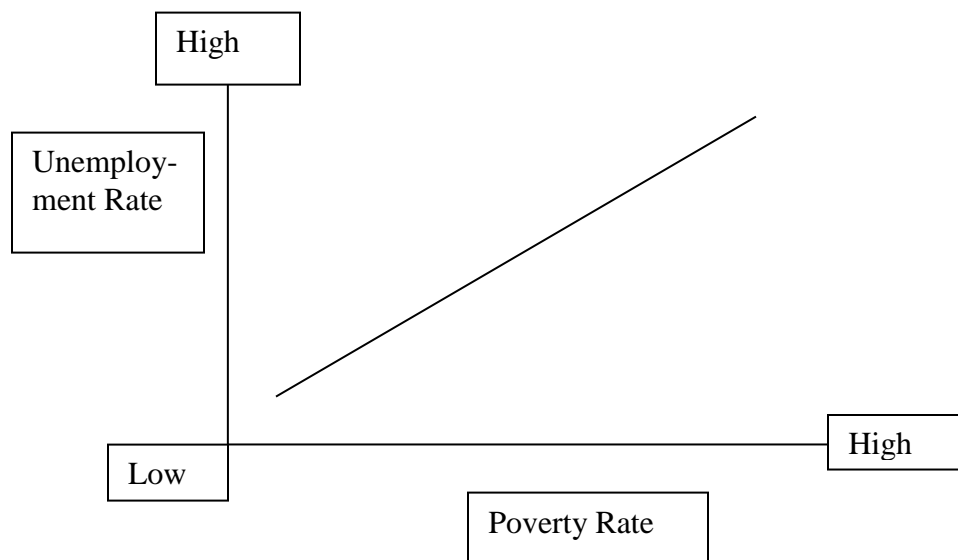
ELA Questions, School A

1. Describe your experience and/or decision making process with regard to planning instruction. Or how did you go about deciding how to plan instruction?
2. Describe your experience and/or decision making process with regard to assessment. Or, how did you go about making decisions about how to assess student learning?
3. What factors went into a student's grade in your class? How did you come to decide on these factors?
4. How did you use your knowledge of state assessment in the area in which you taught? What impact, if any, did this have on how you made decisions around the planning of instruction, the assessment of students learning, and the assigning of grades?
5. A correlation is a study of the relationship between two sets of data. The range for a correlation is -1.0 to $+1.0$, although these two extremes are rarely seen.

A negative relationship means that as one measure goes up, the other goes down (for example, as the overall employment rate rises, the poverty rate may well drop).



A positive correlation means that as one measure goes up, so does another (for example, as the unemployment rate rises, the poverty rate may also rise).



A correlation of zero shows that no relationship exists between two measures (for example, there may be no correlation or relationship between the poverty rate and the amount of rainfall in a given year).

Note: Correlations do not show cause, only relationship.

The correlation between the GPAs in English classes and subsequent performance on the ACT was **.58**. This level of correlation is significantly different from zero which means the correlation is positive. As GPA rises, so also does a student's performance on the ACT in English Language Arts.

What is your opinion of this correlation? Is it what you might have expected? As you reflect on it, what factors do you think may have contributed to it?

Appendix B

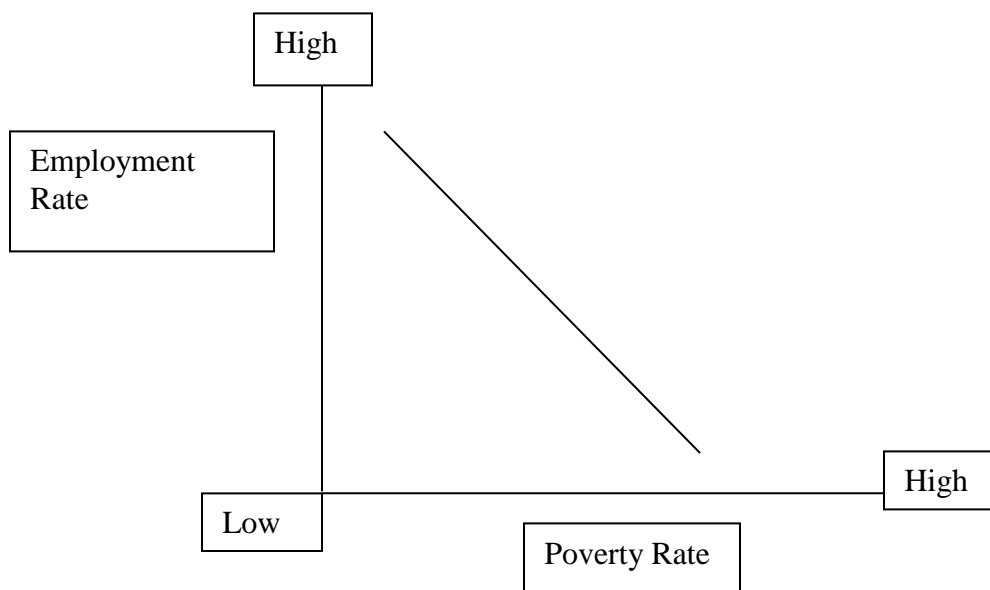
Interview Protocol for School B, ELA

Interview Protocol for School B, ELA

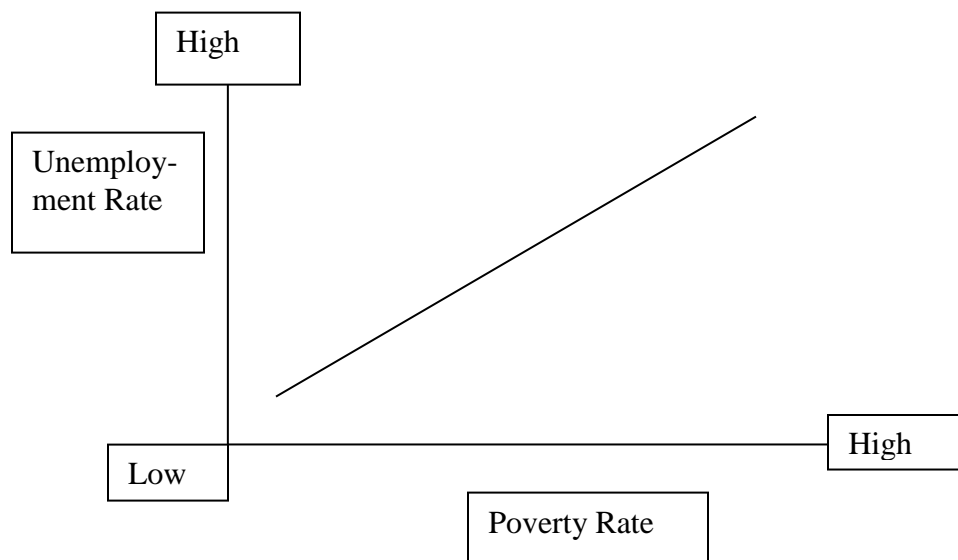
ELA Questions, School B

1. Describe your experience and/or decision making process with regard to planning instruction. Or how did you go about deciding how to plan instruction?
2. Describe your experience and/or decision making process with regard to assessment. Or, how did you go about making decisions about how to assess student learning?
3. What factors went into a student's grade in your class? How did you come to decide on these factors?
4. How did you use your knowledge of state assessment in the area in which you taught? What impact, if any, did this have on how you made decisions around the planning of instruction, the assessment of students learning, and the assigning of grades?
5. A correlation is a study of the relationship between two sets of data. The range for a correlation is -1.0 to $+1.0$, although these two extremes are rarely seen.

A negative relationship means that as one measure goes up, the other goes down (for example, as the overall employment rate rises, the poverty rate may well drop).



A positive correlation means that as one measure goes up, so does another (for example, as the unemployment rate rises, the poverty rate may also rise).



A correlation of zero shows that no relationship exists between two measures (for example, there may be no correlation or relationship between the poverty rate and the amount of rainfall in a given year).

Note: Correlations do not show cause, only relationship.

The correlation between the GPAs in English classes and subsequent performance on the ACT was **.60**. This level of correlation is significantly different from zero which means the correlation is positive. As GPA rises, so also does a student's performance on the ACT in English Language Arts.

What is your opinion of this correlation? Is it what you might have expected? As you reflect on it, what factors do you think may have contributed to it?

Appendix C

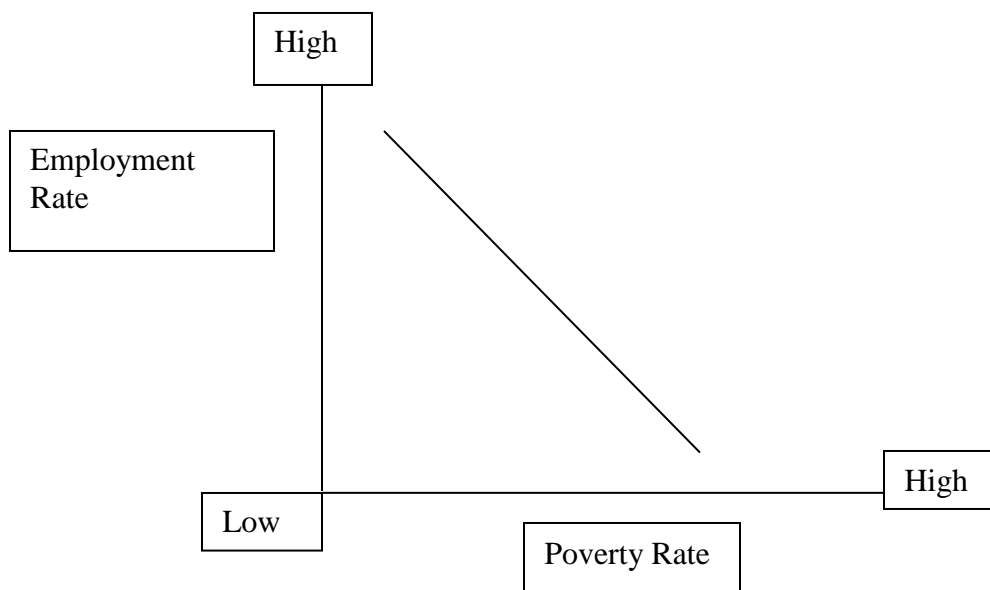
Interview Protocol for School A, Math

Interview Protocol for School A, Math

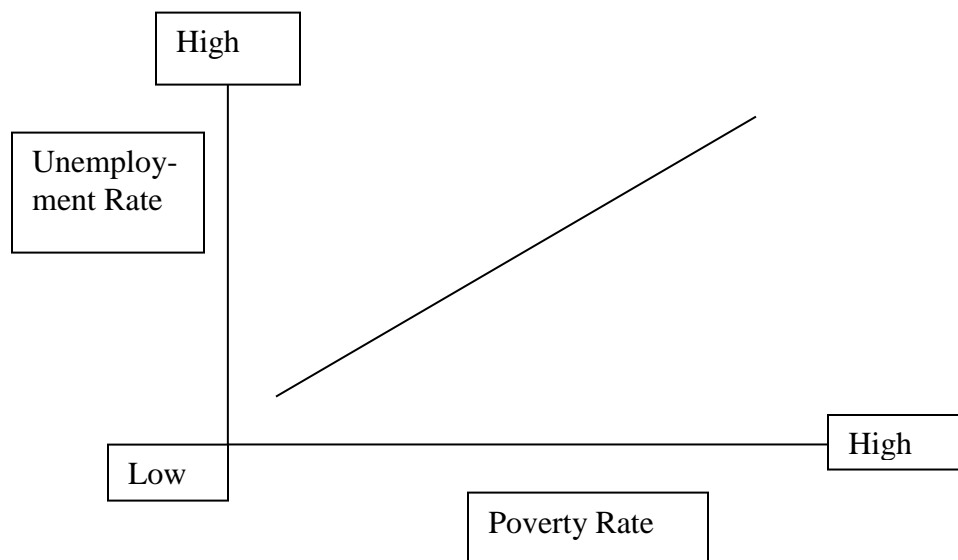
Math Questions, School A

1. Describe your experience and/or decision making process with regard to planning instruction. Or how did you go about deciding how to plan instruction?
2. Describe your experience and/or decision making process with regard to assessment. Or, how did you go about making decisions about how to assess student learning?
3. What factors went into a student's grade in your class? How did you come to decide on these factors?
4. How did you use your knowledge of state assessment in the area in which you taught? What impact, if any, did this have on how you made decisions around the planning of instruction, the assessment of students learning, and the assigning of grades?
5. A correlation is a study of the relationship between two sets of data. The range for a correlation is -1.0 to $+1.0$, although these two extremes are rarely seen.

A negative relationship means that as one measure goes up, the other goes down (for example, as the overall employment rate rises, the poverty rate may well drop).



A positive correlation means that as one measure goes up, so does another (for example, as the unemployment rate rises, the poverty rate may also rise).



A correlation of zero shows that no relationship exists between two measures (for example, there may be no correlation or relationship between the poverty rate and the amount of rainfall in a given year).

Note: Correlations do not show cause, only relationship.

The correlation between the GPAs in Math classes and subsequent performance on the ACT was **.69**. This level of correlation is significantly different from zero which means the correlation is positive. As GPA rises, so also does a student's performance on the ACT in Math.

What is your opinion of this correlation? Is it what you might have expected? As you reflect on it, what factors do you think may have contributed to it?

Appendix D

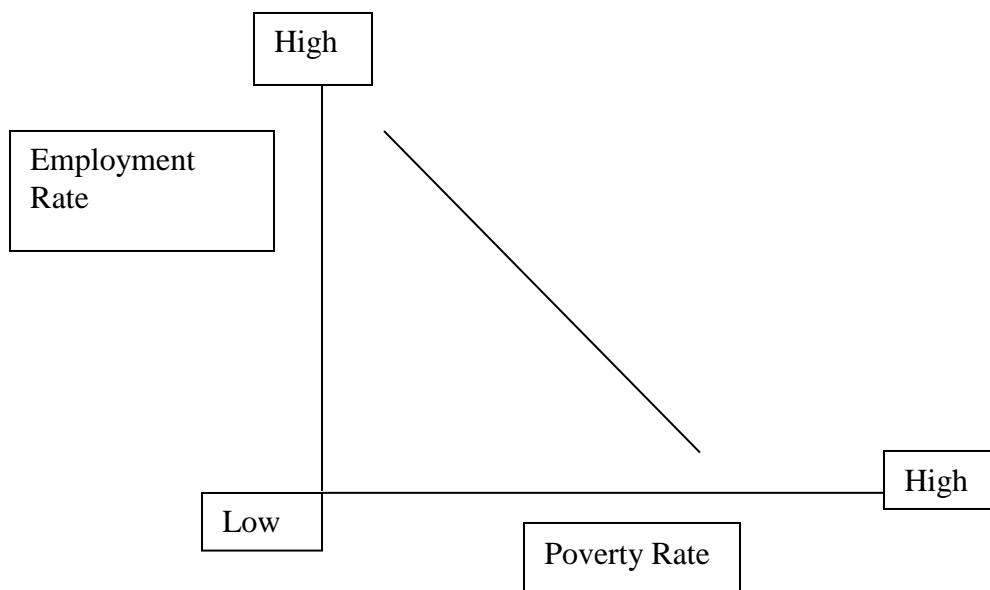
Interview Protocol for School B, Math

Interview Protocol for School B, Math

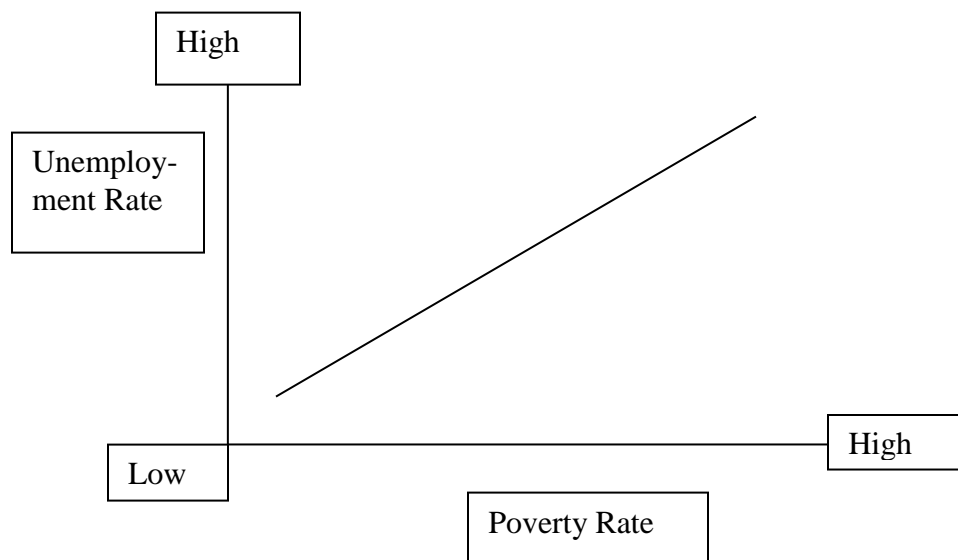
Math Questions, School B

1. Describe your experience and/or decision making process with regard to planning instruction. Or how did you go about deciding how to plan instruction?
2. Describe your experience and/or decision making process with regard to assessment. Or, how did you go about making decisions about how to assess student learning?
3. What factors went into a student's grade in your class? How did you come to decide on these factors?
4. How did you use your knowledge of state assessment in the area in which you taught? What impact, if any, did this have on how you made decisions around the planning of instruction, the assessment of students learning, and the assigning of grades?
5. A correlation is a study of the relationship between two sets of data. The range for a correlation is -1.0 to $+1.0$, although these two extremes are rarely seen.

A negative relationship means that as one measure goes up, the other goes down (for example, as the overall employment rate rises, the poverty rate may well drop).



A positive correlation means that as one measure goes up, so does another (for example, as the unemployment rate rises, the poverty rate may also rise).



A correlation of zero shows that no relationship exists between two measures (for example, there may be no correlation or relationship between the poverty rate and the amount of rainfall in a given year).

Note: Correlations do not show cause, only relationship.

The correlation between the GPAs in Math classes and subsequent performance on the ACT was **.54**. This level of correlation is significantly different from zero which means the correlation is positive. As GPA rises, so also does a student's performance on the ACT in Math.

What is your opinion of this correlation? Is it what you might have expected? As you reflect on it, what factors do you think may have contributed to it?

Appendix E
Consent Form

Consent Form for Making Grades Matter: Connections Between Teacher Grading Practices and Attention to State Assessment

Western Michigan University
Department of Teaching, Learning, and Leadership
College of Education

Principal Investigator: Patricia L. Reeves, Ed.D.

Student Investigator: Gregory Dale Warsen

I am invited to participate in a study entitled “Making Grades Matter: Connections Between Teacher Grading Practices and Attention to State Assessment.” The study is intended to examine teachers thinking as they make decisions about planning instruction, creating student work or classroom assessments, and grading. Additionally, this study will look at how teachers connect these decisions to their knowledge of state assessment and how student grades correlate to ACT performance.

I will be asked to participate in an individual interview lasting 30 to 45 minutes. During this time, the doctoral candidate researcher will ask predetermined questions that I have received in advance of the interview. As in all research, there may be unforeseen risks to the participant. If accidentally injury occurs, appropriate emergency measures will be taken; however, no compensation or treatment will be made available to me. I may benefit from this study by reflecting on my own and the teaching practice of others.

All information collected from me is confidential. This means that my name will not appear on papers on which this information is recorded. The interview will be tape recorded. All forms will be coded and the student researcher will keep a separate master list with the names of the participants and the corresponding code numbers. Once the data are collected and analyzed, the master list will be destroyed. All other forms will be retained for the duration of the project in a locked file in the student researcher’s home office. Following the completion of this study all data will be destroyed.

I may refuse to answer any questions or participate in the study. I may quit at any time during the study without prejudice or penalty, professional or otherwise. If I have questions or concerns about this study, I may contact the doctoral candidate researcher at 616/538-7460 or gwarzen@kvilleps.org or the principal investigator (Patricia L. Reeves) at patricia.reeves@wmich.edu.

This project has been registered as a class project with the Human Subjects Institutional Review Board at Western Michigan University.

My signature below indicates that I have read and/or had explained to me the purpose and requirements of the study and that I agree to participate.

Signature: _____

Date: _____

Appendix F

**Human Subject Institutional Review Board
Letter of Approval**

WESTERN MICHIGAN UNIVERSITY



Human Subjects Institutional Review Board

Date: March 13, 2012

To: Patricia Reeves, Principal Investigator
Greg Warsen, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair *Amy Naugle*

Re: HSIRB Project Number 11-12-11

This letter will serve as confirmation that your research project titled "Making Grades Matter: Connections between Teacher Grading Practices and Attention to State Assessment" has been **approved** under the **expedited** category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may **only** conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: March 13, 2013