

MATHEMATICS TEACHERS' AND PRINCIPALS' RESPONSES TO THE USE OF STUDENT
GROWTH DATA ON TEACHER EVALUATION INSTRUMENTS IN THE STATE OF MICHIGAN

By

Michael Henry Morissette

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Mathematics Education—Doctor of Philosophy

2020

ABSTRACT

MATHEMATICS TEACHERS' AND PRINCIPALS' RESPONSES TO THE USE OF STUDENT GROWTH DATA ON TEACHER EVALUATION INSTRUMENTS IN THE STATE OF MICHIGAN

By

Michael Henry Morissette

In 2011, the State of Michigan, along with other states, passed legislation that mandated standardized test data be used on teacher evaluations starting in the 2013-2014 academic year. Subsequent legislation was passed that altered specific weights for the data on the evaluations and when schools would have to implement changes to the evaluations. In some states, value-added models are used to determine growth on the evaluations, and it was recommended those models be used in Michigan as well. The state elected to not use them and left it up to individual schools to figure out how to measure growth.

In this study, I interviewed five mathematics teachers and four principals to ascertain what barriers existed as they attempted to implement the changes to teacher evaluations in Michigan. Specifically, I looked for evidence of the following barriers: disagreement, money, knowledge, other people, materials time, apathy, and stress. After analyzing my interview data, I found evidence of all of the barriers, with the bulk of the interview data coded as disagreement. Principals and teachers disagreed with how teachers were being evaluated. Both groups did not feel standardized test data should be used in a high-stakes matter. Principals felt the current system required too much of their time and that teachers focused too much on their scores. The teachers indicated they did not find the data from their evaluations to be useful to them. As a

result, they did not use their evaluations to inform their practice. Instead of the current system, principals advocated for a simpler system where they could just have conversations with teachers about their practice instead of talking about a rubric. Teachers just asked that, regardless of what system is in place, that they be evaluated fairly and the data from their evaluations are meaningful to them.

Copyright by
MICHAEL HENRY MORISSETTE
2020

This dissertation is dedicated
to my wife Nicole Pfeifer. Thank you for all the support during this process. I couldn't
have done it without you
I love you.

ACKNOWLEDGMENTS

Nobody can get through this PhD journey alone. If it was not for my peers in PRIME and the professors I had at Michigan State University, I would not have made it. The first person I would like to single out and thank is Kevin Lawrence. Kevin put up with me as a roommate for four years at The Pines and helped me grow as a person more than he would ever know. He was also there for me when I hit rock bottom mentally during my first year in the program and I owe him my life. Also, thanks for being my best man in my wedding. It meant a lot that you were there that day in Hell.

The second person on the list is my wife Nicole Pfeifer. She has been by my side since September 2014 and has been extremely supportive emotionally. She is also a good reminder that there are more important things in life than what I choose to do for a living. Before anything else, my number one priority is to be a good husband to her.

Person number three is Dan Clark. He helped me get involved in the GEU, was on my practicum committee, and helped code data for this dissertation. More importantly, he has been a good friend since I moved to the Lansing area. I even had the pleasure of standing in his wedding.

As for professors that I had at MSU, there are three that come to mind that deserve to be thanked. Beth Herbel-Eisenmann, Corey Drake, and Mike Steele have all played important parts in getting me through this program. Some of that support was academic as they are all great at giving feedback on my writing, but their emotional support was infinitely more important. I definitely would have walked away during my first year if it was not for them.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1: INTRODUCTION	1
What is mathematics education to me?	1
Why should mathematics educators care?	1
Background Information	3
School-level accountability	4
Unintended consequences of NCLB accountability legislation	6
Teacher accountability	7
Value-added models	8
Summary	9
Overview of dissertation	10
CHAPTER 2: TEACHER EVALAUTION REFORM	11
Standardized observation protocols	11
Types of test-based models in education	12
Concerns regarding the use of test-based models	13
Possible effects of the use of test-based models on teacher evaluations	17
Summary	24
CHAPTER 3: CONCEPTUAL FRAMEWORK	25
Defining barriers	25
Considering strategic responses	27
Relating various barriers to implementation	28
Examples of responses to barriers	30
Application of framework to other reforms	33
Indifference/Apathy	33
Human resource: Evidence of attention to knowledge	33
Material resource: Evidence of attention to money	36
Material resource: Evidence of insufficient time for implementation	36
Human resource: Evidence of other people	38
Barriers: Evidence of disagreement	38
Barriers: Evidence of stress	39
Summary	39
CHAPTER 4: METHOD	41
Research Design	42
Method	42

Description of the case	42
Data collection instruments	45
Data collection (surveys)	46
Data analysis (surveys)	47
Data collection (interviews)	49
Data analysis (interviews)	50
Summary	53
CHAPTER 5: FINDINGS	55
Disagreement	57
Money	60
Materials	63
Time	65
Other people	68
Knowledge	69
Apathy	72
Stress	74
Revisiting research questions	77
Summary	80
CHAPTER 6: DISCUSSION	82
Summary of dissertation	82
Revisiting literature	83
Addressing barriers	90
Potential options for teacher evaluations in Michigan	92
What can university-level mathematics educators do?	94
Positionality statement	95
Lessons learned	97
Limitations	99
Suggestions for future research	99
APPENDICES	
APPENDIX A: Principal Survey	102
APPENDIX B: Mathematics Teacher Survey	104
APPENDIX C: Principal Interview Protocol	107
APPENDIX D: Mathematics Teacher Interview Protocol	108
APPENDIX E: Survey Data	109
REFERENCES	112

LIST OF TABLES

Table 3.1: Definitions of barriers	26
Table 4.1: Mapping of research sub-questions to survey questions	48
Table 4.2: Example of coding interview text regarding barriers	52
Table 4.3: Example of coding interview text regarding responses to barriers	52
Table 5.1: Barrier codes (frequency by individual)	56
Table 5.2: Responses to barrier codes (frequency by individual)	56
Table E.1: Amount of assistance the principal provides with using standardized test scores to improve mathematics instruction	109
Table E.2: Does your school district use standardized tests (e.g. ACT, SAT, MEAP, MME, MSTEP) to help measure student growth on teacher evaluations?	109
Table E.3: Does your school use a status model or growth model when using student growth data on teacher evaluations?	109
Table E.4: Emphasis principals put on making sure standardized test scores are high.	110
Table E.5: To what extent do you encourage your teachers to use standardized test results to inform their practice?	110
Table E.6: To what extent do you believe standardized tests capture your students' knowledge of mathematics?	110
Table E.7: How good of an indicator do you believe standardized test results are of the quality of your teaching?	110
Table E.8: Since the beginning of your career, how often have you attended various professional development activities/sessions/conferences etc NOT provided by your school district?	111
Table E.9: Purpose of evaluation (Teacher responses)	111
Table E.10: Purpose of evaluation (Principal responses)	111

LIST OF FIGURES

Figure 3.1: Implementation Barriers

30

CHAPTER 1: INTRODUCTION

What is mathematics education to me?

Prior to beginning my doctoral studies, I really did not give the question much thought as I believed (and still believe) that anything that affects the teaching and/or learning of mathematics is fair game. It was not until I read about the debate of what should and should not be counted as mathematics education research in my proseminar classes at Michigan State University (see Lester & Lambdin's (2003) piece for a good overview of the development of the mathematics education research community) and when I had my own experience with someone pushing back against my idea of what mathematics education is that I realized there are various opinions regarding what mathematics education is. I was shocked when I found out there are some that believe that if the mathematics is not the primary focus of the research, then it does not count as mathematics education research. If you are in that camp, then this dissertation probably will not be to your liking because I focus on recent teacher evaluation reforms. If you have a broader interpretation of what mathematics education research is, like me, then feel free to proceed.

Why should mathematics educators care?

In recent years, the use of student growth data has become more prevalent on teacher evaluations, among other reforms. One reason to care about these reforms to teacher evaluations is that mathematics teachers, and other core teachers, have the potential for being evaluated differently and more harshly than teachers in non-core content areas as student growth data are used on most evaluations now (this will be discussed in more detail later) despite the fact that not all content areas are not tested with standardized tests. One could

argue that it is easier to show growth on teacher-generated pre- and post-tests than it is on a standardized test such as the SAT. As a result, it would be easier for non-core teachers to show growth and keep their jobs than core teachers. This could potentially have negative repercussions regarding the supply of mathematics teachers as some who really want to teach may choose to teach art, physical education, industrial arts, etc. to avoid the pressure of having to make gains on standardized tests. For those who insist on having a mathematics career, they may steer clear of teaching mathematics and choose a different route. Mathematics teacher educators need to think of ways to support and recruit prospective mathematics teachers to ensure that school staffing needs are met and that they are able to effectively respond to the added pressure that comes with teaching mathematics. This study helps in this area as it identifies specific issues or barriers that currently exist for mathematics teachers and knowing these issues is necessary in order to address them.

In addition to supporting our prospective mathematics teachers, we also need to be concerned with supporting our practicing mathematics teachers as they likely have different needs now than they had in the past regarding continuing education and professional development opportunities. With the added pressure on mathematics teachers to show student growth as measured by standardized tests, there is a greater need now for mathematics educators at the university level to foster relationships with the practicing K-12 teachers and figure out what aspects of instruction they find to be challenging. With this information, we may need to offer different graduate classes and make a larger effort to offer a variety of professional development activities in schools and at conferences to meet those needs than we currently might be doing. Although, we should not stop there. I argue that I,

and my future colleagues, must do more to influence educational policy at the state and national level than we currently are. Practicing teachers need allies in their corner fighting for them to improve the working conditions and overall climate in education so we can reverse the current trend of less people choosing education as a career. It is way too easy to throw your hands up into the air and think this is not worth the effort because policymakers will not listen anyway. In order to be better informed so that we can be good allies, in this dissertation, I will discuss some changes mathematics teachers and principals would like to see in the policy world to help them do their jobs better.

Background information

“We need, first of all, for there to be accountability, for there to be somebody who is responsible for enforcing standards and holding people’s feet to the fire” (Granholm, 2003).

-Jennifer Granholm, Former Governor of the State of Michigan

Most people have likely heard something along these lines from a variety of politicians on television, in print media, in social media, or on the radio any time public monies are spent. In these messages, the concern is that we get the “most bang for our buck” and that tax-payer money is spent wisely and efficiently. In public education, standardized tests have been used as a way to hold schools and teachers accountable as some claim that student performance on these tests is an indicator of school and teacher quality. Based on students’ performance on national (e.g. NAEP, SAT, ACT) and international standardized tests (e.g. PISA), policy makers have argued that schools and teachers have not been doing their jobs to the best of their abilities as student performance lags behind students in other countries around the world. In

response to this perception, they have made efforts to hold schools' and teachers' feet to the fire, as Governor Granholm stated, by having consequences for not showing improvement in student standardized test results (Anagnostopoulos, Rutledge, & Jacobsen, 2013; Dee & Jacob, 2011; Hanushek & Raymond, 2005; State of Michigan, 2011a; State of Michigan 2011b). What follows is a brief discussion of some work that has been done regarding school and teacher accountability. In the next chapter, I provide more detail regarding how standardized tests have been used to hold teachers accountable.

School-level accountability

The idea of holding schools and teachers accountable for what students learn is not a new one. Individual states (e.g. Connecticut and North Carolina) started implementing accountability systems more than twenty years ago (Hanushek & Raymond, 2005). In these early accountability systems, the states would typically create a set of standards for what students should learn and require standardized tests linked to these standards to assess how much learning took place. From the test results, the state would create a rating system for the schools within the state; typically the schools would get a letter grade (Hanushek & Raymond, 2005). Some states, for example, Connecticut, North Carolina, and Texas, attached consequences, both positive (e.g. teacher bonuses and vouchers for parents) and negative (e.g. closing schools), to these ratings. Other states like Mississippi, Indiana, and Kansas, however, did not.

To assess the impact of the accountability, with and without consequences, on achievement, Hanushek and Raymond (2005) conducted a study using National Assessment of Educational Progress (NAEP) mathematics data. To isolate the effect of accountability on

achievement, these authors looked at the growth in performance within states between fourth and eighth grade and also used growth models with state fixed effects to help control for the effects of other policies within each state. After performing their analysis of the NAEP data, they found a positive and statistically significant impact on student performance in the states that attached consequences to the ratings, but found no effect on student performance in states in which schools received a letter grade with no accompanying consequences. Other studies have reported similar findings to Hanushek and Raymond's (2005) findings. Jacob (2004), for example, found that mathematics and reading achievement significantly increased following the introduction of a consequential accountability policy in Chicago Public Schools. Additionally, Winters and Cowen (2012) studied New York City Public Schools and found that receiving a letter grade other than an F had no significant effect on student achievement. The results of these studies suggest that if policy makers want to effect change in schools via accountability, then any system put in place should have some consequences (e.g. teacher bonuses, vouchers for parents, closing schools) associated with it. This is all assuming, however, that the change we are looking for is better results on standardized tests. Do better results on standardized tests equate to better teaching and students learning the material?

One of the better-known, large-scale, consequential school accountability systems was put in place on the federal level with the reauthorization of the Elementary and Secondary Education Act in 2001, better known as the No Child Left Behind (NCLB) Act (U.S. Department of Education, 2014). Analyzing and discussing all that NCLB did is beyond the scope of this study, but one relevant aspect of the law is the emphasis on ensuring that all students were proficient in reading and mathematics by 2013-2014. To achieve this target of 100% proficiency, schools

had to reach a specific target proficiency levels each year known as adequate yearly progress (AYP). To illustrate, consider a school in the 2002-2003 school year in which 40% of its students were identified as proficient based on standardized tests the previous year. In order to reach 100% proficiency by 2013-2014, 45% of the students would need to be proficient in mathematics by the end of the 2002-2003 school year, 50% by the end of the 2003-2004 school year, and so on until they reached 100% in 2013-2014. If this school did not reach the indicated target score by the end of the year, then they failed AYP for that year. If the school failed AYP for two or more years, then there were consequences for that school. These consequences included allowing parents to transfer their children to non-failing schools, being forced to supply tutoring to students, replacing staff, implementing new curriculum, having outside experts come in to advise administrators at the school, extending the school day and/or year, and restructuring the internal organization of the schools (U.S. Department of Education, 2014). With this added pressure from NCLB, it may not be surprising to hear that schools tried a variety of strategies to avoid consequences.

Unintended consequences of NCLB accountability legislation

Even though NCLB had an admirable goal of having all students proficient in mathematics and reading by the 2013-2014 school year, the law had some unintended consequences. One of the unintended consequences was that, in order to make AYP each year, some schools would focus instructional time on those students who were just below the proficiency threshold, which some literature has referred to as the “bubble kids” (Dee & Jacob, 2011). The idea behind focusing on these students was that, with just some added attention, it would be relatively easy to get them above the proficiency threshold and, thus, make AYP for

that year. The additional time that was spent on these students had to come from somewhere. Usually it came from instructional time that was originally devoted to the high-performing students and the lowest-performing students (Dee & Jacob, 2011). Even though it could be looked at as a bad practice, allocating the focus away from high-performing students made sense as high-performing students would likely remain above the proficiency threshold anyway. In this system, it may also make sense that a school would not allocate time and resources to put a lot of effort into the lowest-performing students as they would not likely breach the proficiency threshold even with the added attention. In addition to focusing on this group of students who are on the border of testing at proficient, some schools shifted their stronger teachers to grades that were tested and weaker ones to grades that were not tested (Grissom, Kalogrides, & Loeb, 2013). This is a practice that I will discuss in more detail in my discussion of strategic response in Chapter 3. Other schools would literally “feed towards the test” where they would increase glucose in the school lunches during testing windows with the hope that it would improve student performance (Figlio & Winicki, 2005). Even with this gaming of the system, schools did not reach 100% proficiency. Given this, policy makers have turned their attention to teachers as it has been known for decades that teachers have the largest effect of any in-school factor on student standardized test scores (Coleman, 1966).

Teacher accountability

In the almost two decades since NCLB was implemented, the quantity of data that can be collected and analyzed has increased as a result of advances in technology. As data collection and analysis capabilities have improved, the focus of accountability in the United States has been transitioning from schools to individual teachers. Until the mid-1990s there

really was not an effective means of isolating a given teacher's effect on student achievement, but with the advent of value-added models, some argue that it now can be done (Sanders & Rivers, 1996). Several states have recently passed legislation that tie student achievement and growth data to teacher evaluations while others are already reversing course and stripping the mandate that standardized tests be used in the evaluation (Holloway-Libell & Collins, 2014; Maine Education Association, 2019). Some of the reasons for focusing on student growth are to limit the practice of focusing on students who are close to scoring proficient and not to penalize teachers for the varying levels of students that come into their classrooms. By attempting to address students' varying levels and backgrounds, one could argue that focusing on students' growth is a fairer practice than getting a certain percentage of students above a certain proficiency threshold as this focus rewards progress instead of reaching a particular target, but growth models are not without fault.

Value-added models

To address the issues with simple growth models, the studies that have been done recently (since about 1996) regarding tying student achievement data to individual teachers have been done using value-added models (NRC & NAE, 2010; Sanders & Rivers, 1996). Value-added models (VAMs) are statistical models that try to isolate a given teacher's contributions to student achievement and growth by controlling for variables over which the teacher has no control such as a student's socioeconomic status or prior achievement (NRC & NAE, 2010). The way these variables are typically controlled for is by using multiple years of test data for each student (though some VAMs only make use of one year of data). Using this past data, an expected score for a given standardized test is created and compared to the actual data from

that test for each student that a given teacher is responsible for. These comparisons are then aggregated in some way and then a value-added rating is assigned to a teacher.

Even though it can be argued that VAMS are fairer than simple growth models, it should be noted that several concerns about their use have been identified regarding the standardized tests, measurement error and validity, data analysis, equity, and generalizability among other things, all of which will be discussed in more detail in the next chapter (Cohen & Moffitt, 2009; Jackson, 2012; MET Project, 2013; NRC & NAE, 2010).

Summary

In sum, the federal government, along with state and local governments, have implemented various reforms in recent decades in response to a perceived crisis in American education at the K-12 level. One reform was to hold schools and teachers accountable for student learning. A method that was implemented to hold schools accountable for student learning was to assign letter grades to the schools, but researchers found that assigning letter grades did not have an effect on student performance on standardized tests if the grades were not accompanied by consequences. Regarding consequences, with the passage of NCLB at the federal level, schools and states were under more pressure to ensure that all students learn than at any other point in American history. Failure to ensure that all students learn, as measured by AYP, was accompanied by a variety of sanctions. In response to the added pressure, schools shifted resources to concentrate on what some literature has called “bubble kids” among other strategic responses in order to attain AYP and avoid sanctions. Recently the focus of accountability has transitioned to individual teachers. Several states have now passed laws that tie student achievement and growth data to teacher evaluations. Currently, most of

the research done regarding teacher accountability focuses on value-added models. Little is known regarding how people at the ground level have implemented these laws, especially in contexts that do not use value-added models.

Overview of dissertation

In this chapter I briefly discussed my beliefs regarding mathematics education research, established why we should care about teacher evaluation reform, and I provided a brief overview of recent efforts to use policy to improve schools in the United States. In the next chapter, I go more in-depth regarding teacher evaluation reform legislation, with an emphasis on test-based models for measuring student growth. In Chapter 3, I detail a conceptual framework that I created over the past few years. The purpose of the framework is to be able to apply it to a given policy, not just teacher evaluation policy, and identify barriers to implementation during the implementation stage. Based on the barriers that are present and the responses to those barriers, I can make recommendations to policymakers for the next iteration of a policy. In Chapter 4, I describe my methods for conducting my study. In Chapter 5 I share my results which are organized based on my conceptual framework. In my final chapter, I will discuss the results and provide recommendations for the next iteration of teacher evaluation policy.

CHAPTER 2: TEACHER EVALUATION REFORM

Prior to teacher evaluation reform legislation that was passed earlier this decade, it was mostly up to individual school districts to generate the instruments they would use for the purpose of evaluating their teachers. This was usually done in consultation with local teacher unions as part of the collective bargaining process. The new legislation that was passed in the State of Michigan, and elsewhere, attempted to standardize the instruments used. To do this more easily, legislation that made the evaluation instruments a prohibitive subject of bargaining was passed (State of Michigan, 2011(c)). This legislation would, in theory, allow a state to impose an evaluation format on the school districts. These imposed formats required schools to use a standardized observation protocol and several states also instituted a mandate that student growth data be tied to the evaluations as well. The reasoning for using student growth data is largely to introduce variation in the outcomes from evaluation as previously most teachers received positive evaluations even when principals knew some of their teachers had job performance issues (Weisberg et al, 2009).

Standardized observation protocols

Regarding standardized observation protocols most states rely on protocols such as Charlotte Danielson's (2014) Framework for Teaching, Marzano's (2014) Teacher Evaluation Model, The Thoughtful Classroom (Silver, Strong, & Associates, 2014), or 5 Dimensions of Teaching and Learning (Center for Educational Leadership, 2014). The State of Michigan, for example, leaves it up to individual districts to choose one of these four protocols to use. These protocols are intended to measure aspects of quality teaching that are not captured by standardized tests. For example, the Marzano (2014) model evaluates teachers' performance

in four domains: “Classroom Strategies and Behaviors”, “Planning and Preparing”, “Reflecting on Teaching”, and “Collegiality and Professionalism”. Similarly, the Danielson Framework (2014) evaluates teachers in the following four domains: “Planning and Preparation”, “The Classroom Environment”, “Instruction”, and “Professional Responsibilities”. I do not plan to critique how well these protocols measure what they say they measure as the focus of this dissertation is on the student growth aspect of teacher evaluation reform and not the observation part. In the remainder of this chapter, I describe a range of test-based methods for tying student growth data to teacher evaluations.

Types of test-based models in education

In order to measure a particular teacher’s effect on a student, test-based evaluation models are typically used. These models fall within one of two primary categories, status models or growth models (Ladd & Lauen, 2010). Status models look at student performance at a given point in time and compare that performance to a target score. For example, they are used when a school or state assesses what percentage of their students score at a proficient level in a given year, such as what was done when determining AYP under NCLB. Growth models measure student achievement by tracking the improvement or decline in test scores of students within a given school year or from one year to the next. This can be done with the same group of students (e.g. comparing scores of a group of students at the beginning of a school year and at the end) or by looking at different group of students (e.g. comparing scores of this year’s 11th graders to last year’s 11th graders). Probably the most common example of growth models is when groups of students take pre- and post-tests to measure how much students learned over the course of a specified unit of time. A problem identified with basic

growth models is that they do not control factors that affect student achievement which schools and teachers have no ability to affect such as a student's family background or socioeconomic status (NRC & NAE, 2010). For example, a student could show a large gain from the pre-test to a post-test over a unit on quadratic functions. One might assume this was due to the instruction the student received in class, but it could very well be due to outside tutoring that this student had access to because the student's parents had the financial means to provide their child a tutor. To address this issue with basic growth models, value-added models, a more complex growth model, can be used. Value-added models are statistical models that attempt to isolate a school's or teacher's effect on student achievement by controlling for all the variables that affect student achievement, but that are out of the control of the school and teacher (NRC & NAE, 2010). This is normally done by taking at least two years of students' test scores and controlling for various student and school-level variables. A value-added estimate for a teacher is then created by comparing observed student data to expected student data. Thus, if the observed value is greater than the expected value, then one could make an argument that the teacher was an effective teacher. If they were close, then the teacher could be classified as an average teacher. If the observed was less than the expected value, then the teacher could be classified as an ineffective teacher.

Concerns regarding the use of test-based models

Several concerns about using test-based models to evaluate teacher performance can be found in current literature. Some concerns are regarding using test scores to evaluate teachers (NRC & NAE, 2010; Cohen & Moffitt, 2009). If teachers are to be evaluated based on student test scores and if the test covers only a small percentage of what a given teacher

teaches, then some question how the score can be an accurate measure of how well the teacher performed. Maybe the students would have performed significantly better if a different subset of the content were tested instead? Also, according to Cohen & Moffitt (2009), in most studies dealing with teacher quality there are significant differences between expert observers' opinions of what good teaching is and what test results state. For example, a teacher could have several cram sessions with their students emphasizing doing procedures right before the standardized test is given and these sessions could have the short-term effect of boosting test scores, but this teaching does not emphasize deep, conceptual understanding and likely will not translate into long-term learning. Another concern is related to the fact that standardized tests generally do not measure important non-content related aspects of teaching such as fostering intellectual curiosity, self-esteem, student motivation, persistence in tackling difficult tasks, or the ability to collaborate well with others (NRC & NAE, 2010; Jackson, 2012).

Other concerns relate to measurement error and validity. Given that test items are a subset of the entire set of possible relevant questions that could be asked and the test is given at one particular time, a student may perform slightly better or worse if they took the test on a different day due to a variety of reasons (NRC & NAE, 2010; Cohen & Moffitt, 2009). Also, if growth models are used that rely on longitudinal data, as is normally done with value-added models, the sample used to generate a score for a given teacher may be rather small because of missing data due to students moving in or out of a given district, student absences or imperfect record matching. Having a small sample size increases the error and decreases the precision of these estimates. Another problem that occurs when longitudinal data are used happens often at the high school level where students are not normally tested annually. No

Child Left Behind only requires students to be tested once during their high school career (Goldhaber, Goldschmidt, & Tseng, 2013). Given this, it would be hard to assign a score to an Algebra 2 teacher for example as there would most likely be a two- to three-year gap in time between standardized tests.

Regarding validity, Cohen & Moffitt (2009) argue that results obtained by using growth models are somewhat invalid because students are not randomly assigned to teachers. Some teachers typically get higher-achieving students (as defined by standardized test scores) and some typically get lower-achieving students. For example, more-senior mathematics teachers normally get to teach higher-level mathematics classes and less-senior teachers normally teach remedial-level mathematics. Thus, the less-senior teacher may appear to be worse than the more-senior teacher, not because of their ability to teach, but because of the students each teacher has. On the other hand, teachers that teach only advanced classes may appear to be worse than they are if their students do not show significant growth. One could argue that it is hard to show improvement if one consistently teaches students who are in the 95th percentile or higher. In addition to the lack of random assignment, Goldhaber et al (2013) found that issues can arise even between very similar growth models. For example, different types of value-added models exist. When comparing results from a “Student Fixed-Effects Model” and a “Student Fixed-Effects with Lagged Score Model” the authors found significant variations in teacher quality as reported by the models. Some teachers that were reported as being in the lowest quintile in one model were reported as being in the highest quintile using the other model and vice versa. These results are troubling if data from these models are used in high-

stakes decisions as good teachers could possibly lose their jobs after being mislabeled as poor ones and poor teachers could keep their jobs after being mislabeled as good teachers.

The concerns to this point deal primarily with the instrument used to acquire data, but other concerns deal with analyzing the data collected. Some researchers question how causal inferences can be made given teachers and students are not randomly assigned even after controlling for prior student achievement (NRC & NAE, 2010). Others question how other factors that affect student achievement in a particular class can be teased out (Jackson, 2012; NRC & NAE, 2010). For example, if a student is taking a mathematics course and a physics course at the same time, what the student learns in physics is going to have an effect on the student's knowledge of mathematics. When that student takes a standardized test that assesses their knowledge of mathematics, the student's mathematics teacher will get all of the credit for what that student knows. The student's mathematics teacher would also get all of the credit if the student has a tutor, receives help from another teacher, or attends private test preparation sessions. If these things cannot be teased out, then scores assigned to a particular teacher will not be accurate.

Equity issues are another concern regarding using test-based models. More specifically, some researchers have expressed concerns about using test-based models given the gap in resources that exists between low-poverty and high-poverty schools. Students and teachers in high-poverty schools typically have fewer educational resources, larger class sizes, weaker leadership, and more student and teacher mobility (Cohen & Moffitt, 2009). As Cohen & Moffitt (2009) argue, this inequality "could reduce the value that teachers in poorly resourced schools add to students' scores" and "teachers would be penalized for conditions beyond their

control, thus erroneously reducing the quality scores and eroding the scheme's legitimacy and political and legal standing" (Cohen & Moffitt, 2009, p. 203).

Another concern regarding the use of test-based models deals with the assumption that results of studies done at the elementary school level are generalizable to the middle and high school levels. C. Kirabo Jackson (2012) addressed this issue in his study entitled "Teacher Quality at the High School Level: The Importance of Accounting for Tracks." Jackson (2012) stated, "because elementary-school students are typically exposed to one teacher and all are on the same academic track, while secondary-school students are exposed to several teachers and placed into different tracks, methodologies designed for elementary-school teachers may be inappropriate for measuring teacher quality in other grades" (p. 2). If tracking effects are not controlled for, this study found that the importance of high school teachers may be overstated by approximately 50%. When controlling for tracking effects, he found that "a one standard deviation increase in algebra teacher quality is associated 0.08 σ higher test scores" (Jackson, 2012, p. 24). This finding suggests that teacher quality at the high school level has very little effect on student standardized test scores as a .08 standard deviation increase in test scores is virtually no change at all.

Possible effects of the use of test-based models on teacher evaluations

The practice of tying student achievement data to teacher evaluations can affect, both positively and negatively, teachers and teacher education in a variety of ways. With this increased emphasis on accountability, one could hypothesize that teachers' stress levels would increase significantly. Some may say that this is a good thing and that the increased stress would cause teachers to try harder; these people are assuming teachers are not giving 100%

effort now. This increase in effort, this line of thinking purports, would then lead to an increase in student achievement. I would argue, on the other hand, that this increased stress could be detrimental to teachers' performance in the classroom. I base this belief on my own experience as a classroom teacher and also on a study done by Drake and Sherin (2006).

Drake and Sherin (2006) investigated two teachers' implementation of a standards-based mathematics curriculum. As part of that study, the authors interviewed the two participants, Beth and Linda. When interviewing Beth, she said, "Weaknesses? That when I'm overstressed I go back to the way I always did, that's it. I lose sight of where I'm going and I just deal with the here and now" (p. 165). In other words, when Beth felt overstressed, she resorted to teaching in a more traditional way rather than teaching in a way that was more aligned with the ideals of the mathematics education reform movement. Like Beth, I also resorted to teaching in a more traditional way when I was stressed. Extra stress caused me to go into a survival mode where I would do just enough preparation to get through the next day. If other teachers react similarly to stress, then the added stress that comes from the increased emphasis on accountability could negatively affect what happens in the classroom.

In addition to affecting what happens in one's classroom, tying student achievement data to teacher evaluations can also affect a teacher's interactions with their colleagues and their colleagues' students. If teachers are to be evaluated using value-added models, as the Michigan Council of Educator Effectiveness (2013) recommends for the State of Michigan, then why would a teacher want to help their colleagues become better teachers or assist their colleagues' students in any way? This very thing came up as an issue in Guenther's dissertation (2019): "findings reveal, at the very least, this current evaluation system does not encourage

teachers to work together to improve their practice. At its most consequential, it appears to be encouraging isolationism and creating adversarial relationships among some teachers” (p. 6). Personally, this finding resonated with me a lot as when I taught, I was rather close with a lot of my students and I always told my students that I would always help them if they needed it, even if I was no longer their teacher. Several students took me up on the offer during my career. They would come in after school to get help on their mathematics homework because their teacher would not or could not stay after school on that given day (my schedule was more flexible given that I had no family obligations). If I was still teaching and was evaluated using test-based models, I am not sure if I would continue this practice in an environment where competition is fostered. It would be against my rational self-interest to do as my actions could actually cause me to lose my job.

In order to address this possible issue and to foster collaboration, the MCEE (2013), when making recommendations regarding how to implement the changes in teacher evaluation policy, stated that school-level value-added models (e.g., a mathematics department would get a value-added score for the group) may be used as part of a teacher’s evaluation. They also suggested, however, that this score can only comprise 10% of the teacher’s evaluation. Thus, taking the 2015-2016 school-year as an example, 40% of the teacher’s evaluation would still be determined by their individual score. One may question whether the 10% generated from school-level data is large enough to foster collaboration given 40% of the evaluation would still be generated from one’s own practice.

A possible effect on teacher education programs that needs to be addressed if test-based models are used on teacher evaluations is regarding the placement of student

teachers/interns in classrooms. If a teacher is going to be evaluated using value-added models, then would it not be against their rational self-interest to allow a novice to teach their students? That student teacher is likely going to affect the supervising teacher's score on their evaluation. Given this and the likelihood that it will be impossible to tease out the student teacher's effect on the students, more teachers may be unwilling to accept student teachers/interns in their classroom. That is, unless the state agrees to suspend the requirement that test-based models be used for any teacher who agrees to have a student teacher. This, however, would create its own set of problems as then every teacher would likely want to have a student teacher and there could be some hostility between colleagues in a school. Those who do not get a student teacher would think that they were being treated unfairly, especially if the same teachers always had student teachers placed in their classrooms.

To this point, I have only thus far addressed possible negative effects of using test-based models. There are, however, possible positive effects as well. The use of test-based models could require teachers to reflect on their practice and consider ways to improve it. One possible change that teachers might make is related to the textbooks they use since "mathematics has a long history of being driven by the textbook" (Remillard, 2005, p. 214). If teachers wish to keep their jobs, they may feel that it is necessary to find and use written curriculum materials that have been correlated with increases in student achievement. Standards-based curricula have been shown to do just that (Thompson & Senk, 2001).

In addition to changing the textbooks they use, teachers may decide to carefully analyze standardized test data to look for areas where students struggle and then make changes to their practice based on what they find. The use of this data is something Cavanna (2016)

investigated in her dissertation. She found that teachers did take part in data digs as a result of the changes in teacher evaluation policy, but the teachers did not find standardized test data to be useful in day-to-day lesson planning and teaching. Instead, the teachers reported that the data they had a hand in collecting themselves was more useful (e.g. formative assessments, videorecordings of their teaching, and student work) for these purposes than standardized tests. Cavanna also reported that the teachers did not get a lot of support from administration regarding what to do with the standardized test data during the data digs so one might conjecture that teachers may find that data more useful if they had more support in how that information might be useful to their practice. Alternatively, different kinds of data that are closer to teachers' everyday practice, like that collected when doing action research, may be more relevant and useful.

Another viable alternative could be to have teachers look to other successful schools and teachers for ideas on how to improve their practice. Kitchen, DePree, Celedon-Pattichis, and Brinkerhoff (2007) share the characteristics of nine successful schools that serve students who are poor and specific teachers' practices that led to high student achievement as measured by standardized tests. As for the specific characteristics and practices, the authors were able to identify three major themes including "high expectations and sustained support for academic excellence" (p. 33), "challenging mathematical content and high-level instruction" (p. 77), and "the importance of building relationships" (p. 115). Regarding the first theme, the authors found that all of the schools made teaching and learning the priorities over everything else. Administrators and teachers believed that all of their students were capable of learning. Failure

was not an option for them. Also, administrators were extremely supportive of their teachers.

Some specific examples of what was done in the schools included:

- Administrators handled most of the paperwork and problems with students and parents so their teachers would only have to worry about teaching.
- Students were provided with additional instructional time beyond their normal mathematics classes for remediation and also to challenge them.
- Teachers were given cell phones by the school so their students could call them at night for extra help.
- Schools provided a late bus so students could get extra help after school from their teachers (some of these teachers were paid extra to stay after school to tutor students).
- Teachers had access to a variety of different teaching resources. If the teachers wanted something for their classroom, all they needed to do was ask.
- Teachers were encouraged to and took part in sustained professional development activities. In many cases, the schools would pay for their teachers to attend conferences and workshops.

As for the second theme, Kitchen et al (2007) found that the schools in the study had a clear focus on developing students' problem-solving and critical thinking skills in addition to ensuring the students improved their proficiency with basic skills. There was also a focus on ensuring the students did well on standardized tests. Some practices that were common in the teachers' classrooms with these foci in mind included:

- Teachers had students work in groups in order to facilitate communication about mathematical ideas.

- Teachers did not let their textbook define what was going to be taught in their classrooms. The teachers viewed curriculum design and implementation as “an ongoing, dynamic process that should sustain high student expectations” (p. 83). Given this belief, several of the teachers would supplement their textbook and continuously plan, reflect, and alter what they did in their classrooms.
- Teachers incorporated tasks that were meaningful to their students.
- Teachers would analyze students’ standardized test scores and engage in “backward curriculum planning” (p 89). That is, the teachers would look at the scores, see where their students needed to improve, and then vertically align their middle school and high school mathematics curriculum to address the areas that were shown to be weak.

The final theme the authors identified addressed the importance of building relationships with fellow teachers as well as with students. When interviewed, the teachers stressed the value of being able to collaboratively plan lessons with their colleagues, creating a support system amongst the faculty, and holding each other accountable. Regarding the practice of collaboratively planning lessons, Kitchen et al (2007) stated, “for teachers searching for one magic bullet in this study that they could implement to impact their students’ learning and achievement, the collaborations that existed among participating faculties may be it” (p. 128). As for the relationships with the students, the teachers stated that it was important to show the students that they care about them and to show the students that they are human. If the students know that the teachers care about them, one could argue that the students might try harder in class so they do not disappoint the teacher.

Summary

In this chapter, I discussed changes in teacher evaluation policy. Specifically, several states now use standardized observation protocols and also use student growth data as a means of introducing variation in evaluation outcomes as there was not a lot of variation before. I discussed how test-based models can be used on teacher evaluations and also some issues identified with their use. This was followed by a discussion of some possible effects of using the test-based models on teacher evaluations, both positive and negative.

In the next chapter, I discuss a conceptual framework for analyzing the implementation of a given policy that I have been developing for the past few years. Its roots can be traced back to a concept called strategic response that came from literature regarding the implementation of No Child Left Behind. My reason for developing this framework is that I wanted to have something that I could use to analyze various educational policies with the hope of identifying specific implementation issues. Using this framework, implementation issues can be classified into a handful of categories that can be used to advise the next iteration of a given policy so that it is more successful, depending on what implementation issues are found.

CHAPTER 3: CONCEPTUAL FRAMEWORK

In this chapter I unpack a conceptual framework that helps us understand the barriers and related responses to policy implementation. Specifically, my framework describes various barriers that educational policy and mathematics education researchers have identified as commonly causing teachers not to implement a policy in ways intended by policymakers. When used as a lens to analyze the data I collected, this framework allowed me to identify specific areas where teacher evaluation reform in the State of Michigan broke down during the implementation phase. Having identified these barriers and teachers' responses to those barriers, I can now tailor recommendations to policymakers to make the next iteration of teacher evaluation reform policy more successful in improving the quality of instruction in our schools. What follows is a synthesis of the literature that informs my framework with a focus on the literature regarding policy implementation and a concept called strategic response. In this synthesis of literature, I discuss what others have found to be barriers to implementation and also how educators responded to various barriers that were present in other reforms. Following this discussion, I then apply the framework to literature of various reforms where the author(s)' focus was not implementation issues, but where evidence of implementation issues existed.

Defining barriers

After a policy is passed by legislators, some group or groups of people are charged with implementing it. Quite often, though, the policy is not implemented as the policymakers had intended due to a variety of factors. In this dissertation I call those factors *barriers* to intended implementation. Previous studies have identified several barriers to implementation that cause

a policy not to be implemented as policymakers had intended. These barriers include: implementers' indifference or apathy toward the policy, disagreement about how to achieve results, stress, and lack of resources (Drake & Sherin, 2006; Hope, 2002). For clarity, I provide my definitions for each of those barriers in Table 3.1 below. Although the table provides very short definitions for each of these terms, I next provide a longer description of "resources" as my definition includes terms that are not commonly understood.

Table 3.1:

Definitions of barriers

Term	Definition
<i>Indifference/Apathy</i>	The state of being completely uninterested in and unconcerned with a given policy.
<i>Disagreement</i>	Those charged with implementing a policy agree with the primary goal of the policy, but have different opinions than the policymakers regarding how to achieve the goal.
<i>Stress</i>	Long-term distress that causes internal psychological and/or physical tension which can lead to depression, anxiety, and other mental health issues. For the purposes of this study, I am ignoring eustress.
<i>Resources</i>	Any material, human, social and/or cultural asset that one uses to function.

Most people, when asked what educational resources are, would probably state they are items such as calculators, books, computers, manipulatives, money, etc. Some researchers have classified these as material resources (Adler, 2000) or economic capital (Spillane, Hallet, & Diamond, 2003). A teacher's knowledge should also be considered a resource because it can affect how the teacher responds to a given policy; that is, one cannot do what one does not know how to do. Adler (2000) would categorize one's knowledge as a human resource. In

addition to material and human resources, time can be considered a resource as Adler (2000) classified it as a social and cultural resource. Other social resources could include trust and collaboration as well.

Considering strategic responses

The presence of barriers to implementation is only half of what we need to pay attention to when studying how a given policy is implemented. The other half is related to how those who are responsible for implementing a given policy respond to those barriers. One area of literature that informs my framework is regarding a concept called *strategic response*. *Strategic response*, as it has been used recently in the educational policy literature, has an almost nefarious connotation associated with it. For example, the term has been used to describe school and teacher responses to the mandates that came with No Child Left Behind (USDOE, 2014). Specifically, the phrase “*strategic response*” has been used in literature related to the following practices 1) focusing on students just below the proficiency threshold because of limited human and material resources (Dee & Jacob, 2011), 2) moving stronger teachers to tested grades at the elementary level to boost standardized test scores due to human resource barriers (Grissom, Kalogrides & Loeb, 2013), and 3) increasing calorie and glucose levels in school lunches around testing periods, possibly due to disagreement or resource issues, to boost students’ short-term cognitive ability (Figlio & Winicki, 2005). My definition of *strategic response* would include these responses under the umbrella of *strategic response*, but is broader as I would argue all responses by implementers should be considered as *strategic responses* as there is some thinking/strategizing involved in deciding if and how to deploy one’s resources to implement a policy.

Relating various barriers to implementation

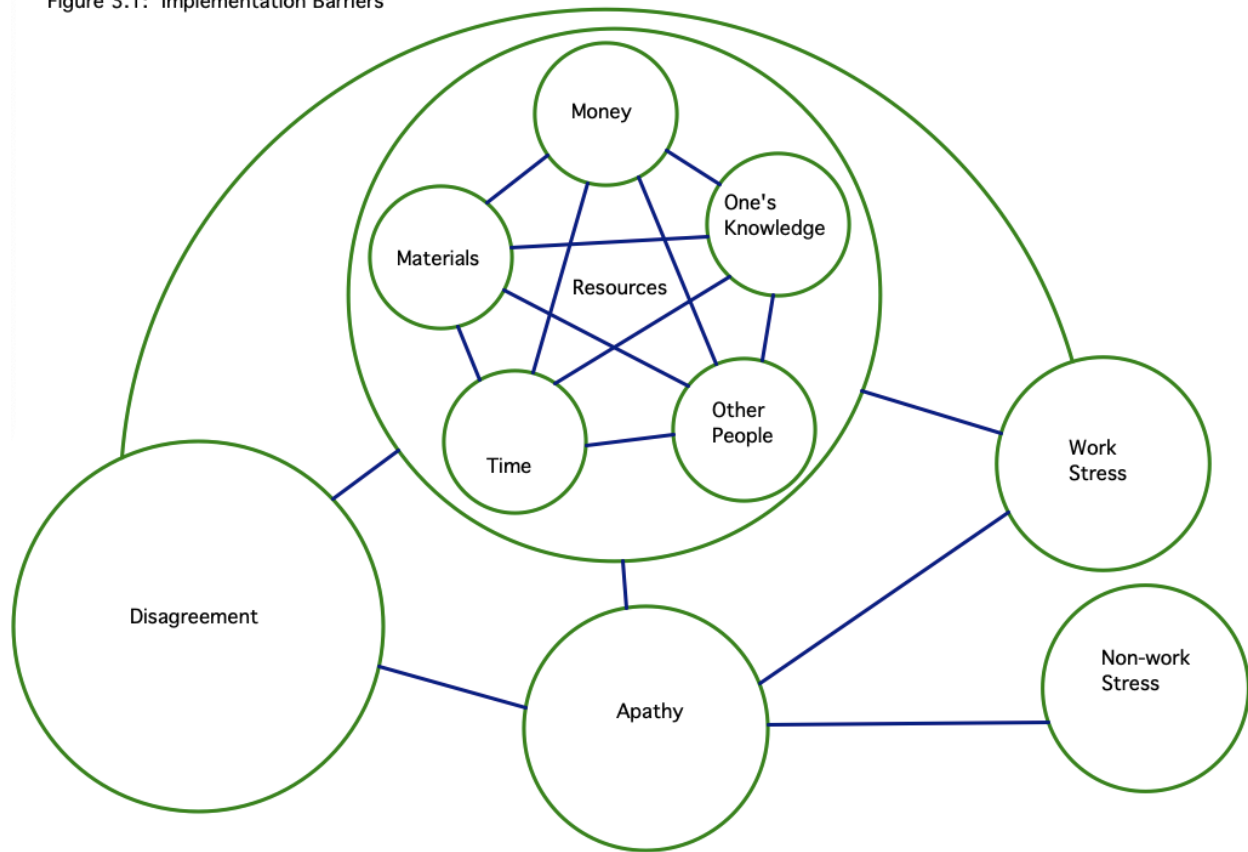
Figure 3.1 on page 30 is a representation of how I see the relationships among the various barriers to implementation. In the previous section, I explained the barriers so here I focus on explaining the relationships in the figure. Looking at the entire figure, all of the bubbles are connected to each other because there is not a neat, ordered linear relationship that exists among them. In previous iterations of this figure I tried to impose some order on this by having arrows pointing in various directions, but I elected to omit specific directional relationships on the figure as I came to realize, after a lot of thought, that some of these relationships between barriers can vary from person to person. That is, the relationship between two barriers could vary directly for Person A, but inversely for Person B. For example, let us consider the relationship between one's knowledge and time. If we were able to quantify knowledge regarding a specific topic, intervention, etc., then one might expect that the amount of time a person would need to learn something would increase if the amount of knowledge one has decreases. That may be true only to a certain point though. If one has no knowledge of how to do something, then one may devote no time to learn about it. This could also be due to some apathy/indifference which highlights the complexity that exists among all of these barriers.

In addition to the relationships between two barriers varying directly or inversely, a third barrier might act as a mediating or moderating variable. In the example in the previous paragraph, apathy/indifference could be considered as a mediating variable as the person's complete lack of knowledge caused them to feel apathetic which, in turn, caused the person to not devote any time at all to learning anything. An example of a barrier serving as a moderating

variable between two others would be if we were to consider non-work-related stress. By definition, it is not caused by factors at work, but it can have an effect on the relationship between two other variables. If someone has a lot of stress because of a chaotic situation at home, it may have an effect on the relationship between money and materials, for example. They may be so stressed that they cannot even think of how to spend available funds on available materials to implement a given policy.

Overall, what I would like readers to keep in mind from looking at this figure and considering barriers and relationships between them is that implementation at the ground level can get messy. It is important to think about how the various barriers (and relationships among those barriers) could be affecting the successful implementation of a policy as it is highly unlikely that only one barrier is the cause of all of the implementation issues. When offering recommendations for further iterations of a policy, it is important to devote the time and effort to consider the various ways multiple barriers could be affecting implementation and not just zero in on the one that comes to your attention first.

Figure 3.1: Implementation Barriers



Examples of responses to barriers

Identifying the barriers that exist is important, but responses to barriers are also important as knowing them can help in offering recommendations for future iterations of a policy. For example, lack of money could be identified as a barrier to implementation, but it may not be the primary issue causing implementation breakdown. There could be a toxic climate created by the people in an organization, the media, policymakers, etc. that is a bigger issue than the money. Throwing more money at the problem might help some, but it will not help as much as fixing the climate. Regarding potential responses for the barriers in Figure 3.1, it would be impossible to list all possible responses that could happen with a given barrier as policies are different and the people charged with implementing the policies are all different. What I will

do here is give a couple possible responses to each of the main barriers I discussed earlier.

Recommendations regarding how to address the barriers and responses would differ depending on what barriers exist and how one triages them.

If indifference or apathy toward the policy exists, educators might respond by ignoring a policy and not making any changes in their practice as a result. This might occur because the implementers believe the policy will be short-lived and the time and energy required to implement the policy would be a waste. An alternate response could be choosing to ignore the policy, but constantly complaining about how pointless it is to everyone they encounter. This complaining could wear down the others to the point where they decide not to bother trying to implement the given policy as well.

If lack of resources exists, educators might respond by trying to implement the policy to the best of their ability using what resources they do have, but fail to implement the policy as intended because the resources available are insufficient to get the job done. For example, a mathematics teacher could respond to a mandate that they teach different content in their Algebra 2 class the following year by finding a variety of materials on the internet. The teacher may not have the knowledge or time necessary to judge the quality of these activities or how to properly sequence the activities, however. Another potential response if lack of resources exists could be that an educator decides not to bother trying to implement a policy if they deem the lack of resources to be insurmountable. Thus, it may look like apathy at first glance, but the apathy is actually caused by the lack of resources.

If the disagreement about how to achieve results exists, educators could respond by changing their practice to achieve the goals of the given policy, but do so in a manner that is

inconsistent with what policymakers or others had in mind. For example, a middle school principal could mandate the use of a particular curriculum to boost standardized test scores, but a mathematics teacher might think that the given curriculum does not give their students enough practice and decides to have students do more practice problems from random worksheets they found online. Another potential response to disagreement could be that the educator bypasses their administrator and decides to speak with the school board president because the educator does not believe that what their administrator has planned will achieve the goal of a given policy.

If stress exists, either work or non-work related, educators may put in the minimum effort to just get by even though they may know of a better way to do something. For example, a mathematics teacher may know how to facilitate meaningful discussions in their classroom, but they choose to lecture instead because lecturing takes less planning and less planning might be all they can do given the stressors that exist. Another potential response to stress could be the educator decides to just resign and pursue a different career.

In sum, depending on the barrier that exists, educators can respond in a variety of ways. Some of those responses could be choosing to ignore a policy, rationing available resources and attempting to implement a policy to the best of their ability, responding to a policy by doing something that the implementer believes achieves the goal of a policy, but is inconsistent with the substance of a given policy itself, or by putting in minimal effort due to stress. What follows in the next section is my first attempt to use my framework with some qualitative data from journal articles.

Application of framework to other reforms

Ideally, I would have liked to code some interview transcripts to test and further clarify my framework, but I decided that attempting to code qualitative data from articles would be sufficient for my first attempt. Thus, I searched for studies of various reforms by inputting keywords into Google Scholar, accessed the articles using ProQuest through MSU's library, and looked for evidence of the barriers and responses to barriers listed above in the text of the articles. Specifically, I searched for and applied my framework to articles regarding inclusion of special education students in the general education classroom, Algebra for All, the Common Core State Standards, and teacher evaluation reform. I did not do an exhaustive search for articles regarding these forms, however. I just sampled a few so I could have some data to work with for a trial run. Results from this search and subsequent coding follow.

Indifference/Apathy

As I read the articles, I was not able to find evidence of every barrier in Figure 3.1. This did not surprise me as my focus on implementation barriers is not the focus the various authors had in mind when they wrote their respective articles. Additionally, I was not surprised that I was unable to find evidence that suggested teachers were apathetic or indifferent towards any of the reforms given the results of Spilane and Zeuli's (1999) study, for example, found most teachers try to implement policies even when they may disagree with them. This, however, was not the case for every person as I will report later in this dissertation regarding administrators' choices regarding tying student achievement data to teacher evaluations.

Human resource: Evidence of attention to knowledge

Regarding the other barriers, the bulk of the evidence I found in the articles related to lack of resources. Of the various resources I discussed earlier, the primary issue that resulted in implementation challenges in three of the reforms I looked at was gaps in knowledge, specifically pedagogical knowledge. Largely due to inadequate professional development opportunities and lack of attention in undergraduate and graduate classes, mathematics teachers struggled teaching more heterogeneous classes that resulted from inclusion and Algebra for All policies (Allensworth, Nomi, Montgomery, & Lee, 2009; Desimone & Parmar, 2006a; Desimone & Parmar, 2006b; Gamoran & Hannigan, 2000; Loveless, 2008). Regarding the implementation of the *Common Core State Standards* (CCSS, 2010), administrators reported concerns regarding teachers' content knowledge and "the teachers' ability to support the deeper learning that CCSS aims to encourage, especially in mathematics" (McLaughlin, Glaab, & Carrasco, 2014, p. 7). Teachers also had concerns regarding their content and pedagogical knowledge as evidenced in the following excerpt from Porter, Fusarelli and Fusarelli's (2015) study:

Mr. Harner acknowledged the complexities involved in making the shift. When asked about the potential challenges, he noted as follows: The biggest challenge is interpreting what that means in terms of teacher actions. You hear a lot of people say, "Oh, you're going to have to change the way you teach, you have to change the way you do things." Well, that's great to stand up and say, "Well, change the way you teach, change the way you do things," but we need to define exactly what that means in the classroom and we

actually have to help teachers understand exactly what kinds of things does that mean and how that impacts practice (p.123).

Administrators in the study also went into detail as to why teachers had concerns regarding their pedagogical knowledge, with one respondent explaining, “Teachers having to change how they always taught. To really teach these new standards well, you have to teach them differently and that is hard for teachers” (p. 127).

Given the existence of pedagogical knowledge gaps due, in part, to insufficient professional development and coursework in college, mathematics teachers assumed that special education students and low-achieving students without a specific learning disability could be taught in similar ways as there was practically no difference between the two groups of students (Desimone & Parmar, 2006a). Based on this assumption, mathematics teachers would often target their instruction towards the hypothetical middle student hoping that this would give the students at the bottom of the achievement distribution, as defined by standardized test results, a chance to succeed (Allensworth, Nomi, Montgomery, & Lee, 2009). Other mathematics teachers, those that devoted more attention to the students at the bottom of the achievement distribution, would slow the pace of instruction, skip difficult topics, focus on following procedures rather than focusing on problem solving, and put fewer questions on assessments assuming that these accommodations would help these students succeed (Stein, Kaufman, Sherman, & Hillen, 2012). As for how mathematics teachers responded to their pedagogical issues when implementing the *Common Core State Standards*, I was unable to find any descriptions of what teachers did in their classrooms. I did find the following quote that leads me to believe that teachers did not change their teaching habits at all in response to the

Common Core State Standards as they did not have the education or professional development to do so, “It’s just overwhelming too much at one time and not enough resources or training done in advance—not while you’re trying to implement!” (Porter, Fusarelli, & Fusarelli, 2015, p. 129). This quote also touches on another barrier, stress. One could argue that finding something overwhelming would cause one to also feel stressed.

Material resource: Evidence of attention to money

In addition to the knowledge gaps, I was able to find evidence of money issues, a material resource. Evidence of money issues could be found in discussions regarding large class sizes, lack of paraprofessionals to assist teachers in the classrooms, lack of up-to-date technology, and lack of quality classroom materials (Desimone & Parmar, 2006a; McLaughlin, Glaab, & Carrasco, 2014). As for how the teachers responded due to the lack of money, there was little direct evidence in the articles. Most of the discussion regarding money issues was at the district level rather than focusing on teachers and I do not have access to interview transcripts to look for teacher responses myself.

Material resource: Evidence of insufficient time for implementation

Time arose as a significant barrier in the research related to the inclusion of special education students and in the implementation of the *Common Core State Standards*, but not in the articles on Algebra for All or teacher evaluation reform. Regarding inclusion, time was an issue when it came to co-planning lessons with special education teachers (Desimone & Parmar, 2006a). Quite often, the mathematics teachers and special education teachers did not have a common planning period, which made collaboration with other people, another resource, on lessons difficult. In response to this, mathematics teachers generally planned their

lessons by themselves with no help from the special education teacher. This effectively cut off a valuable resource for the mathematics teachers as the special education teachers knew more about strategies that work with students with varying learning disabilities.

As for the *Common Core State Standards*, time appeared to be the largest barrier to effective implementation. Mathematics teachers felt unprepared and extremely stressed because of how fast they were expected to implement the CCSS. Two excerpts, for example, that touch on teachers' feelings as they were beginning to implement the CCSS follow:

Sayer principal Carlene Yeadon's comments pointed to how her teachers were feeling about starting the process: "As they begin the year. . . they are a little apprehensive. I think they're excited and nervous at the same time. They're building the plane while they're flying and they're trying to get the wheels on there so they won't crash" (Porter, Fusarelli, & Fusarelli, 2015, p. 122).

This quote seemed to indicate that teachers did not feel well-prepared to implement the CCSS as appropriate time was not devoted to professional development beforehand. The following quote also touched on this hurried implementation as well.

On the one hand, practitioners say that all aspects of CCSS implementation have been hampered by a lack of time. They have too little time to provide professional development, too little time to work on developing new curricula and instructional materials, and too little time to communicate with teachers, parents, and school board members. As one said: "Time, or lack thereof, appears to be the common enemy." (McLaughlin, Glaab, & Carrasco, 2014, p. 5).

In addition to feeling unprepared and stressed, teachers mentioned how difficult it was to evaluate the number of curriculum materials that claimed to be aligned with the CCSS, which is an issue that deals with both time and knowledge (McLaughlin, Glaab, & Carrasco, 2014). As for how the mathematics teachers responded to their issues with time, the articles did not discuss particular responses. It has been shown, however, that stress can negatively impact what happens in a classroom as teachers can go into survival mode and just do enough to get by, which is what happened to a teacher in a study by Drake and Sherin (2006).

Human resource: Evidence of other people

This was touched on briefly in the previous section, but it came up more when I looked at the articles regarding teacher evaluation reform. In evaluation systems that made use of value-added models, teachers became more isolated and did not want to help other teachers as helping other teachers was not in their rational self-interest (Darling-Hammond, 2015; Guenther, 2019; Johnson, 2015). It is not in their self-interest because helping others could potentially raise other teachers' value-added score in comparison to the person doing the helping. If that help increases the score of the teacher being assisted enough, a potential result could be that the teacher does the helping may lose his or her job.

Barriers: Evidence of disagreement

In the articles regarding inclusion, Algebra for All, and the Common Core, I was largely unable to find evidence of disagreement. The only exception was some teachers felt that special education students would be better served being taught in resource rooms (Desimone & Parmar, 2006a). In the articles regarding teacher evaluation reform, I was able to find more evidence regarding disagreement. Specifically, I found evidence that teachers believed their

evaluation focused too much on growth, that standardized test data did not adequately measure what students know, and there were several issues affecting students attending challenging schools (e.g. underfunded and high percentage of low-socioeconomic status students) that teachers cannot possibly help or control (Darling-Hammond, 2015; NRC & NAE, 2010). I was unable to find evidence of what teachers did as a result of these beliefs though.

Barriers: Evidence of stress

Stress was mentioned briefly a few times earlier as a barrier. Regarding teaching tested grades, one teacher said, “I’m scared I might lose my job if I teach in a transition grade level, because...my scores are going to drop” (Darling-Hammond, 2015, p. 134). Fearing for one’s job is clear evidence of work-related stress. As a response to the stress created by using VAMs on their evaluations, it was reported that teachers would try to avoid tested grades at the elementary level, avoid assignments (or even schools) where a large portion of the students have had low standardized test scores, and sometimes leave the profession entirely (Darling-Hammond, 2015; Johnson, 2015).

Summary

In this chapter I unpacked a framework to make sense of policy implementation issues. Specifically, I discussed various barriers to implementation and possible responses to those issues. On my first attempt to use the framework, I was able to find evidence of most of the barriers in journal articles pertaining to various reforms. Finding responses to these barriers in the articles was significantly harder, but this is likely because the authors did not focus on barriers and responses in their articles. I imagine I would have had an easier time finding responses to the barriers if I had access to interview transcripts. In the following chapter, I

discuss the design of my study in detail along with the specific research questions I wished to answer regarding teacher evaluation reform.

CHAPTER 4: METHOD

As the focus of accountability has transitioned from the school level to the individual teacher level and the use of student growth data has begun to be more common to assess teacher quality, it is of interest to know how the use of this data is affecting practice.

Specifically, this study aims to answer the following research questions:

1. What evidence of implementation barriers exists as mathematics teachers and principals in the State of Michigan attempt to tie student growth data to teacher evaluations?
 - a. What do mathematics teachers and principals believe the purpose(s) of evaluating teachers is(are)?
 - b. What is mathematics teachers' understanding of how standardized test data and other student growth data are used on their evaluations?
 - c. What do mathematics teachers and principals identify as pros and cons of how student growth is measured on their evaluation?
2. How have mathematics teachers and principals responded to the implementation barriers that do exist?
 - a. How do mathematics teachers analyze standardized test and other student growth data?
 - b. What do mathematics teachers do to address students' weaknesses as identified by standardized test and other student growth data?
 - c. What steps do teachers take to improve their teaching as a result of the standardized test and other student growth data?

Research Design

Method

To answer the previously mentioned research questions, I used case study methods. Depending on the study, a case can be an individual, group of people, institution, neighborhood, program, culture, region, nation-state or even a stage in a person's life (Glesne, 2011; Patton, 2002). In this study, the case studied was one state in the United States of America, the State of Michigan, that has mandated student growth data be used to evaluate teachers. Typically, in a case study, data are acquired via a variety of procedures such as observations, interviews, and the collection of documents (Creswell, 2014; Glesne, 2011; Patton, 2002). A variety of data from various methods are collected, compared, and contrasted, known as triangulation, in order to improve the trustworthiness of a given study (Glesne, 2011). For this study, I began with surveys because with a large enough sample size, they could allow me to say meaningful things about the population and make claims with a degree of confidence. I then followed this with interviews because they allowed me to get very detailed data to be able to better understand the relationships among variables. I decided against collecting documents because I had done that in an earlier study in which I analyzed teacher evaluation instruments in the State of Michigan (Morissette, 2014).

Description of the case

On July 19, 2011, Governor Rick Snyder of Michigan signed a package of four bills (Public Act 100, Public Act 101, Public Act 102, and Public Act 103) into law that were designed to reform education in the State of Michigan. These new laws made significant changes to teacher tenure (State of Michigan, 2011a; State of Michigan, 2011b), teacher evaluations (State of

Michigan, 2011d), and collective bargaining (State of Michigan, 2011c). Addressing all of the changes these laws made in one study would be a daunting task; although they all should be addressed at some point given the potential issues for teachers and students that may arise as a result of these changes. For this study, I focused on the practice of tying student achievement and growth data to teacher evaluations which was part of PA 102. PA 102, Section 1249 3(b) states,

For the annual year-end evaluation for the 2013-2014 school year, at least 25% of the annual year-end evaluation shall be based on student growth and assessment data. For the annual year-end evaluation for the 2014-2015 school year, at least 40% of the annual year-end evaluation shall be based on student growth and assessment data. Beginning with the annual year-end evaluation for the 2015-2016 school year, at least 50% of the annual year-end evaluation shall be based on student growth and assessment data. The student growth and assessment data to be used for the school administrator annual year-end evaluation are the aggregate student growth and assessment data that are used in teacher annual year-end evaluations in each school in which the school administrator works as an administrator or, for a central-office level school administrator, for the entire school district or intermediate school district.

(“Enrolled House Bill No. 4627”, 2011, p. 4)

After this piece of legislation was passed, the Michigan Council for Educator Effectiveness (MCEE) (2013) published a document entitled *Building an Improvement-Focused System of Educator Evaluation in Michigan: Final Recommendations*. This document detailed how incorporating student growth data into teacher evaluations should occur. According to the

MCEE, data regarding teachers' practices should be collected using an observation instrument adopted by the state, either Charlotte Danielson's (2014) Framework for Teaching, Marzano's (2014) Teacher Evaluation Model, The Thoughtful Classroom (Silver, Strong & Associates, 2014), or 5 Dimensions of Teaching and Learning (Center for Educational Leadership, 2014). The MCEE suggested that data regarding student growth should be a combination of value-added scores for teachers (for the core content areas) and changes in students' achievement, with value-added scores consisting of at least half of a given teacher's student growth component on their evaluation.

When implementing PA 102, the State of Michigan adopted some of the MCEE's recommendations and not others. One recommendation the state did not adopt was the use of value-added models to measure student growth on standardized test. Instead, the state left it up to the individual school districts to figure out how to measure student growth. To get an idea of how school districts measured growth, I conducted a study in which I collected and analyzed evaluation instruments from around the state (Morissette, 2014). I found large variation in evidence used to measure growth. Some districts used standardized tests using basic growth models. Others used student attendance rate, homework completion rate, semester exams, course grades, pass/fail rate, and graduation rates among other things to measure student growth. No school district took it upon themselves to use value-added models, which is not surprising given the difficulty in generating value-added scores for an individual teacher.

Data collection instruments

This study made use of both survey and interview data. The surveys were created in Fall 2017 with assistance from the Office of Survey Research (OSR) at Michigan State University. Specifically, I started by looking at publicly-available surveys from the Mathematics and Institutional Setting of Teaching study at Vanderbilt University to get ideas regarding survey design (MIST, 2010). I selected a few of their questions verbatim (Questions 15 and 16 in Appendix A and questions 9-14, 16 in Appendix B) that I thought would be helpful for my study and augmented those with questions I created. The questions I selected from the MIST study were related to the type and frequency of assistance teachers received from the principal and fellow teachers. I had hypothesized that teachers would not use standardized test data to inform their practice if they did not receive any assistance from others and I was hoping I could say something about that from the survey data I collected. The questions I created on the survey were more directly related to my research questions.

Once I had rough drafts of my mathematics teacher and principal surveys, I sent them to OSR for feedback. A person at OSR sent me my documents back about a week later with comments regarding recommended wording changes in questions and in my Likert scales. Once I received feedback from them, I made the changes they recommended and then sent those documents to my co-advisors for more feedback. After making edits based on their feedback, I then created the surveys in Qualtrics and these were the surveys that I sent to mathematics teachers and principals throughout the State of Michigan (see Appendices A and B).

After the surveys were completed by the participants, I created semi-structured interview protocols based primarily on my research questions and partially on some of the data generated by my surveys (see Appendices C and D). For example, many (12 out of 17) of the principals surveyed did not have a background in mathematics and I was curious how that affected the evaluation process in the eyes of both the teachers and the principals. Once my protocols were completed, I then sent them to my co-advisors for feedback. Prior to conducting my interviews, I piloted the mathematics teacher protocol with a fellow graduate student at MSU that had recent secondary mathematics teaching experience. Based on their feedback, I generated some probing questions that I could use in my interviews. These questions were specific to each of the contexts where my subjects taught as I tried to do some research about the schools and/or the individuals before the interviews.

Data collection (surveys)

Beginning in the second week of September 2018, I employed a stratified random sampling strategy to select schools to include in the study. Specifically, I used the Michigan High School Athletic Association's 2018-2019 enrollment list where high schools throughout the state are split into four different classes based on their size (MHSAA, 2018). I did this to ensure representation from different sized schools as schools of different sizes could have different challenges while implementing the changes to teacher evaluations. For example, administrators from smaller schools could have issues evaluating all of their teachers because they typically have other duties to perform that principals at larger schools do not have. On the other hand, administrators from large schools could have issues evaluating all of their teachers because of the large number of teachers they likely have on staff. From the list of schools, I

randomly sampled seven Class A, seven Class B, seven Class C, and seven Class D schools. If a school with “Catholic” or “Christian” in the name was sampled, I threw it out and sampled again as these schools do not have to follow the same evaluation rules as the traditional public schools. Once I had my sample, I looked up contact data on the schools’ websites and sent the principals an email to ask if they would be willing to participate in my study. In the email I specified that participation would entail taking a survey that would take approximately 15 minutes to complete, included links to both the principal and mathematics teacher survey on Qualtrics, and offered an option to participate in a follow-up interview if they wished to do so. I also wrote that all participants that did the survey and the interview had the choice of a \$25 Amazon or Meijer gift card as compensation for their time. In addition to asking for participation on their part, I asked that the principal forward the email to one of the mathematics teachers in the building to see if there was interest in participating in my study. Once I sent the email, I waited three to four days and sent a follow-up thank you/reminder email. This process was then repeated five more times so that I ended up contacting a total of 168 high schools (the entire population of MHSAA-affiliated schools consists of 748 schools which includes the religious schools I threw out if they were sampled). Despite these multiple attempts to contact principals and teachers, these efforts resulted in responses from 15 teachers and 19 principals.

Data analysis (surveys)

After the survey data were collected, the original plan was to see if there were any significant differences in responses between principals and teachers for similar questions on the surveys by comparing means with an independent samples t-test and also to see if there

were any significant differences among principals and teachers based on school size. It was hypothesized that schools could have different challenges implementing the changes to teacher evaluation policy based on their size. One issue was raised on the previous page, but other hypotheses I had were that smaller schools would have fewer professional development opportunities than larger schools as they have less money since funding is tied to student enrollment and smaller schools would have fewer human and material resources than larger schools. If true, this would mean smaller schools would have a harder time implementing the changes in teacher evaluations and would likely need more support in order to implement the changes effectively. Given the low number of people that completed the surveys (15 teachers and 19 principals), however, I was unable to determine if there were any significant differences among the schools based on their size. I did not want to completely omit the work I did with the surveys, so I decided to calculate descriptive statistics for what I had regarding my research questions. These results are not reported in the body of this dissertation as I elected to focus only on my interview data, but they can be found in Appendix E. To provide an idea of what survey questions I wanted to use to partially answer my research questions, a brief breakdown is included in Table 4.1 below.

Table 4.1:
Mapping of research sub-questions to survey questions

Research Sub-question	Survey Question
1(a) (Math teachers' and principals' beliefs regarding the purpose of evaluation)	T15, P14 (about purpose of evaluation)
1(b) (Math teachers' understanding of how ST test data and other growth data are used on their evaluations)	T7, T8, P7, P8 (regarding use of standardized tests on the evaluations and asking if they use a status or growth model)
1(c) (Pros and cons of how student growth is measured at their school)	T26, P9 (about advantages and disadvantages of how they measure growth)
2(a) (How do math teachers analyze ST and other growth data)	T14, P15 (both regarding assistance principals give teachers in using standardized test results to inform instruction)

Table 4.1 (cont'd)

2(b) (What math teachers do about students' weaknesses as identified by the data).	T19 (beliefs regarding how well ST measure what students know)
2(c) (What math teachers do to improve their practice as a result of ST and other growth data)	T20, T22 (beliefs regarding how well ST reflect teacher ability and if they use ST data to inform PD)
*Note: In the second column, P# refers to a specific question on the principal survey and T# refers to a specific question on the mathematics teacher survey.	

Several of the questions on the surveys were not used in answering the research questions as I found a lot of the survey data, especially a lot of the demographic data, to be not very helpful due to the small sample size. I had wanted to be able to make claims using school size, education of administrators, length of time worked in the school, etc. as ways to sort and compare the survey data, but the numbers just were not there to do any of it. In the future I think I need to offer a larger incentive to get those numbers and also be more purposeful and realistic about the questions I ask in the survey.

Data collection (interviews)

Even though I was not able to do any t-tests with my survey data, the survey did serve well as an interview recruitment tool, although most of my volunteers were from Class A schools. It was a challenge to get people from smaller schools to talk to me, which I found to be a bit discouraging as I thought this would limit my ability to compare responses from small and large schools. From my list of volunteers, I set up and conducted semi-structured interviews in November and December 2018 with 2-Class A principals, 1-Class B principal, and 1-Class C principal. No Class D principals agreed to be interviewed. Two of the principals I spoke to have a PhD (Principal A2 and Principal B) and one was a former mathematics teacher (Principal B). Regarding mathematics teachers, I conducted interviews with 2-Class A, 1-Class B, 1-Class C,

and 1-Class D teachers. Both Class A teachers and the Class C teacher were veteran teachers with over 10 years of experience each. Teacher A2 also happened to be the mathematics department head at her school. Teachers B and D had less teaching experience, but both were tenured. Also, a pair of the Class A principals and teachers, the class B pair, and the class C pair worked at the same school. Given I was not able to really use my survey and who ended up being my interview participants were, I do not feel I am able to make statistically significant claims regarding how school size affects the implementation of this teacher evaluation policy as I am unable to triangulate the data. I do, however, feel I can discuss how some schools responded to the policy and the barriers that they encountered while doing so. This still has value as it highlights things policymakers should be paying attention to in order to make this reform more successful in improving education in the State of Michigan.

Data analysis (interviews)

Once the interviews were completed, I immediately uploaded all of the audio files to an auto-transcription website (Temi.com) in order to cut down on the time required to transcribe them. Once I received the transcripts from the website, I listened to each interview and corrected any transcription errors I noticed as the website had some issues with how I pronounced particular words. Each transcript took approximately two hours to edit for precision to try to capture, as closely as possible, what participants said. Once the editing was done, I then coded each transcript based on my conceptual framework that I discussed in the previous chapter. Specifically, I used the following codes for excerpts of text that exhibited evidence of the various barriers discussed previously: money, materials, time, other people, knowledge, apathy, disagreement and stress. If I noticed an excerpt exhibited evidence of a

response to one of the barriers, I used the same codes, but put a “-R” after the code. Excerpts coded varied in length from a single sentence to an entire speaking turn depending on if one or multiple ideas were discussed in a turn.

Based on the recommendation of my committee, I had another colleague code the transcripts with the same coding scheme for inter-rater reliability. After my colleague coded each transcript, he sent the transcript back to me for feedback. We generally agreed on most of the codes (only disagreed on 16 of the 403 codes) and the agreement increased as my colleague coded more transcripts. Most of the disagreement at the beginning came from having slightly different ideas regarding what the code “other people” meant, but this was cleared up through discussion. We came to a consensus that we would code any excerpt that indicated a desire to work with their colleagues (other teachers or administrators) or university professors as other people. My colleague thought my other codes were rather self-explanatory. Any time money, time, or stress were specifically mentioned as a barrier, that excerpt got the corresponding code. If traditional classroom materials were mentioned, the excerpt got a materials code. Although, after discussing it with my colleague, we broadened this to include the standardized tests themselves and the standardized test data. For the knowledge code, we looked for instances where someone mentioned they “don’t know” or “don’t understand” something. As for disagreement, we used that code any time someone indicated they “don’t think” or “don’t believe” something. We also had to infer disagreement from what was said as sometimes the interviewees exhibited disagreement without using those exact phrases. For example, when referring to measuring good teaching, one of the participants stated, “so pretend that’s a thing”. By saying this, it seems the person does not believe you can measure

what good teaching is. As for apathy, we looked for phrases or words such as “I don’t care”, “I’m not going to”, “waste”, “joke”, basically anything that essentially indicated the person was not going to expend a lot of effort in implementing the policy. Some specific examples of what we coded for each barrier and responses to barriers are below in Tables 4.2 and 4.3 respectively.

Table 4.2:
Example of coding of interview text regarding barriers

Code	Example Text
Money	“I mean there’s a lot of money that the state puts into those ISDs and I, you know, it doesn’t funnel out to the schools very much”-Teacher A1
Materials	“I found that it becomes cumbersome in a way that at least the MSTEP and the SAT data that we get, isn’t as specific as you would like it to be.” -Teacher D
Time	“it is almost like a hoop to jump through that, you know, waste 20 minutes of a pretest day”-Teacher A1
Other People	“One thing that I believe in is that collaboration is the key component to making quality teachers and quality classrooms. And so anything that makes you compete to be better with other teachers is counterproductive.”-Teacher A1
Knowledge	“You can check all the boxes you want, but in order to be effective you can’t just check a box without explaining to me why because I’m not going to grow from it. I don’t know what any of those little things mean.” -Teacher C
Apathy	“It just has to be just the right number so that it doesn’t get us to highly effective status in some of the numbers. Where they’re change? You’re like, oh that’s just bullcrap. You know. So right now, the effective system is nothing but a huge joke.” -Teacher C
Disagreement	“If we could measure what good teaching is, right? Like it. So pretend that that’s a thing. If we could measure that. If a good teacher is a good teacher.” -Teacher A1
Stress	“I mean, their workload is insane. Our workload is insane.”-Teacher A2

Table 4.3:
Example of coding of interview text regarding responses to barriers

Code	Example Text
Money-R	“Uh, I mean I write our Title II (A) proposals, so anything that I know that’s upcoming, that is a need for me or one of my teachers, I’m going to write it in there so that there’s funding for it.” -Teacher A2
Materials-R	“We do like a half day every month where we do a data dig collaborating with our coworkers to try and see if we can create some things (tasks and materials)” -Teacher D. This was also coded as time-R and other people-R

Table 4.3 (cont'd)

Time-R	"Originally, what I did was I pre- and post-tested the crap out of everything. I found I couldn't do it on every chapter because I was wasting too much time so I would do basic ideas instead on every unit or maybe two to three chapters." - Teacher C
Other People-R	"That's another thing is that was one of the marks I've found with really good teachers. They are harder on themselves than I would ever be. When I get someone like that, I try and provide as many resources as I can and just stay out of their way."-Principal C
Knowledge-R	"I'm also offered free professional development through the College Board for pretty much everything I want. Um, and I take advantage of that regularly."- Teacher A2
Apathy-R	"I don't know. I take my PTO days off then every time". -Teacher C
Disagreement-R	"We talk a lot about that at our local level of how do we take over the story about our schools, how do we take over the story in the press, on social media, or whatever to continue to push out all the amazing things that go on every day." - Principal A2
Stress-R	"We started doing 10-minute walk throughs." -Principal C in response to losing all of his assistant principals

After the interview data were coded, I then analyzed all of the excerpts and looked for common themes. Specifically, I looked to see if multiple people mentioned the same barriers or responses to barriers. I felt this was an appropriate next step as my overall goal is to provide some recommendations to policymakers to make the next iteration of this policy better. Finding some common barriers or responses to barriers among my participants would provide me with data to back up these recommendations.

Summary

In this chapter I discussed my method I used to answer the following research questions:

1. What evidence of implementation barriers exists as mathematics teachers and principals in the State of Michigan attempt to tie student growth data to teacher evaluations?
 - a) What do mathematics teachers and principals believe the purpose(s) of evaluating teachers is(are)?

- b) What is mathematics teachers' understanding of how standardized test data and other student growth data are used on their evaluations?
 - c) What do mathematics teachers and principals identify as pros and cons of how student growth is measured on their evaluation?
2. How have mathematics teachers and principals responded to the implementation barriers that do exist?
- a). How do mathematics teachers analyze standardized test and other student growth data?
 - b). What do mathematics teachers do to address students' weaknesses as identified by standardized test and other student growth data?
 - c). What steps do teachers take to improve their teaching as a result of the standardized test and other student growth data?

I did two rounds of data collection and analysis where I surveyed and interviewed principals and teachers throughout the State of Michigan. Given my low sample size, I elected to not do independent samples t-tests, so my surveys were not as informative as I hoped they would be. Thus, I elected to not include most of the data I collected from the surveys. I did, however, get a lot of rich interview data and I will report my findings from these interviews in the next chapter using my codes in Tables 4.2 and 4.3 as a framework to organize the chapter.

CHAPTER 5: FINDINGS

As I stated in the previous chapter, since I did not get enough people to respond to my surveys to say anything meaningful from that data, I elected to not include the survey results in the body of this dissertation. Some of the survey data, however, are presented in Appendix E. In this chapter, though, I focus on my interview data. First, I discuss the frequency of codes. This section is followed by my findings pertaining to each of the barriers to implementation that I discussed in Chapters 3 and 4 and then a discussion of my research questions.

Tables 5.1 and 5.2 present the frequencies of each barrier and responses to barriers were coded for each of my interview participants. For me, it is not surprising that the quantity of codes is greater for the teachers than the principals as my interviews with the teachers were longer than the interviews with the principals. Therefore, I focus instead on the relative frequencies of the codes. For example, disagreement was the most common barrier coded for both principals and teachers. I share some specific quotes later in this chapter, but barriers related to disagreement that the principals and teacher reported was mostly about using standardized test data on teacher evaluations and having an evaluation format imposed on the schools. As for the responses to barriers listed in Table 5.2, Knowledge-R was coded more than the others, especially for principals. This finding might seem odd given knowledge was not coded that often as a barrier for principals. One explanation for Knowledge-R being coded more than Knowledge is that the principals did not discuss what they personally struggled with and I did not push them for this information during the interviews. Instead, principals focused on what they were doing to address their teachers' knowledge issues via professional

development and working with others in the building. So, these actions on the principals' part were responses to teachers' knowledge issues, but not their own knowledge issues.

Table 5.1:
Barrier codes (frequency by individual)

Code	Princ A1	Princ A2	Princ B	Princ C	Teach A1	Teach A2	Teach B	Teach C	Teach D
Money	0	0	0	2	9	2	0	4	9
Materials	0	2	0	0	6	0	0	4	5
Time	3	1	2	4	4	9	2	8	7
Other	2	1	1	1	7	5	2	10	8
People									
Knowledge	1	1	1	2	4	1	1	3	3
Apathy	0	2	0	1	8	2	4	6	4
Disagree	7	6	5	6	27	16	8	21	12
Stress	5	1	2	2	1	12	0	1	2

Table 5.2:
Responses to barriers codes (frequency by individual)

Code	Princ A1	Princ A2	Princ B	Princ C	Teach A1	Teach A2	Teach B	Teach C	Teach D
Money-R	1	0	0	2	0	1	1	0	0
Materials-R	2	0	1	3	0	0	2	1	3
Time-R	0	1	1	0	1	3	0	2	3
Other	2	5	2	3	1	2	1	2	1
People-R									
Knowledge	11	4	6	6	1	4	2	1	8
-R									

Table 5.2 (cont'd)

Apathy-R	0	0	0	0	0	0	0	2	1
Disagree-R	0	1	0	0	5	2	1	1	0
Stress-R	1	0	0	3	0	0	0	0	0

What follows now is my interview data pertaining to each of the barriers, starting with disagreement. These are barriers that exist that make implementing teacher evaluation reform difficult, not necessarily barriers caused by the reform. In each of these sections, I incorporate responses to the given barrier if evidence of a response existed in the data. Some of the barriers listed did not have responses as I did not follow up with my participants as I should have in the moment and would be something I would focus on if I were to do another round of interviews.

Disagreement

As was mentioned in the previous section of this chapter, the majority of the interview data that was coded related to disagreement. My participants found the idea of using standardized tests in a high-stakes manner on teacher evaluations troublesome. They also had issues with the observation rubrics that were used, how student growth was measured, and the idea that good teaching can be measured. Regarding the use of standardized tests in a high-stakes manner, every principal and teacher I interviewed was against this idea because of a variety of factors that affect students' standardized test scores that are beyond any teacher's control. For example, Teacher D stated, "It's hard to have them take one test, one snapshot, on one day. Who knows what they had breakfast? Who knows where they slept last night?" In

addition to mentioning factors that are outside of teachers control, others had issues with using the tests because they did not believe standardized tests actually measured what a student did and did not know. Principal C, for example, said, "It's a sound bite and it doesn't really measure what we're asking it to measure" and Teacher A1 said, "So I don't believe that it (standardized tests) tells us how much understanding or knowledge a student has about math. I have a really hard time believing that students who score high or low on that (standardized tests) match with students who really get things, who really understand big ideas and things like that." If principals and teachers do not believe that standardized tests should be used on teacher evaluations, it would be hard to argue that teachers will actually use the data generated from those tests to inform their practice and it is also doubtful that principals will push teachers to use data from standardized tests.

As for the observation rubrics, I go in more detail about these in the section regarding knowledge when I mention a quote from Teacher C said she did not understand what all the little check boxes meant and in the section regarding time when principals states that the rubrics were time-consuming to fill out. Specific to disagreement, the principals all thought the current evaluation process was too complicated and should be simplified. For example, Principal A1 stated,

Part of me wishes that the labels didn't exist, that it would be satisfactory or unsatisfactory as opposed to what we have now because teachers are not that different than students. They have a mindset where they want the gold star, they want to be the A, so I find that instead of having a conversation with a teacher, if they're getting

effective versus highly effective, the conversation becomes why didn't I get highly effective as opposed to how can I improve my teaching?

Principal A2, also advocating for a simpler system, said,

I think if we were given a little more latitude so that if all I had to do was have conversations with the teachers about how they establish purpose for learning in their classroom and it was just one ranking instead of five subcategories with rankings thing around how they do student engagement instead of five subcategories for student engagement, I feel like you can be a little more authentic with, with teachers.

Essentially, these principals seem to be advocating for a system more like what we had in Michigan prior to our current evaluation system.

Regarding how growth is currently measured, Principal A2 said his district tries to “insulate our teachers by and large from that measurement as we feel there is more value in the observation piece of eval process” and the others reported they incorporated a variety of sources of data (e.g. SAT, NWEA, course grades, pre- and post-tests) to measure growth so that their teachers were not affected a lot by any one measure of growth. As for the teachers, they all had issues with how their districts measured growth. Most of them reported their districts were comparing different groups of students to measure growth by using last year’s SAT data and comparing it to this year’s data. Teacher A1 was the only one who did not have an issue regarding how his district used the SAT to help measure growth, but he said his district did not use the data from that test at all and not using that data was consistent with his beliefs regarding standardized testing. In addition to using the SAT, the teachers all stated their schools relied on them to create pre- and post-tests to measure growth, but most found this to

a meaningless exercise. For example, Teacher A2 said, “So they (her students) did really bad the first time. Then the second time they did really well. And that means I’m a great teacher. Um, no” and Teacher C stated, “But I mean, in all honesty, if you're looking at growth, no crap, they're going to grow. They didn't know anything about it before and now I taught them. So they're going to know it. So how meaningful that data is, you know, it's, it's kind of a farce.”

As for being able to measure what good teaching was, two of my participants, Principal A1 and Teacher A1, questioned whether this could even be done. Principal A1 said,

They tried to make it data based, which is very difficult to do in something that is almost more of an art than a science like teaching. So instead of being able to just go in and rely on an administrator’s best practice understanding of education and then having a good conversation. It is script, code, notes, gathered data in those areas.

Teacher A1 also explicitly stated, “I also don't believe that you could actually clinically measure what an effective teacher is.” When there was evidence that both principals and teachers questioned whether good teaching can be measured, it is questionable how useable the data generated from the evaluations were as both parties might be just going through the process thinking of it as performing a compliance task for the state. If this is the case, it is doubtful that teachers use evaluation data to inform their practice As far as my participants go, that seemed to be the case as all of the teachers said, in one way or another, that the purpose of evaluation was just accountability and not to make teachers better.

Money

After analyzing my interview data, money arose as a barrier at the individual, school, local, and state levels. At the individual level, all of the teachers mentioned how poorly paid

they are and some potential ramifications this has for their development as teachers. Teacher A2, for example, said, “I know my tax returns went backwards from 2008 until 2016...my tax guy was like, what the hell are you doing?” Teacher D, reflecting on his pay and how that affects his development as a teacher said, “I sit back and I look at the state of things and look at the benefits that I gain from getting a Master's degree and the only thing I see is that it's going to take me 16 years to pay off the debt...if I have a Master's degree, that makes me less desirable somehow than someone who's brand new because they don't have pay me as much.” Thus, Teacher D did not think it was worth getting a Master's degree because he could not afford the cost of it on his current salary and, if he were to lose his current job, it would be hard for him to find employment elsewhere as his experience and education would make him a less-desirable candidate given the way teacher salary schedules are structured. By not getting his Master's degree, one could argue this would be at odds with the goal of teacher evaluation reform, which is to improve teacher quality.

At the school level, Teacher A1 specifically mentioned how funding issues resulted in increased class sizes. Larger classes can be more challenging to teach than smaller classes, which could have a negative effect on student standardized test scores and teachers' evaluations. Teachers C and D mentioned how their schools have no money available for their teachers in order to attend external professional development opportunities. These professional development activities would be important to attend if we expect teachers to improve their practice as a result of their evaluations. Teacher C said, “They cut off all of our funding for anything like that. So, if we go we have to pay on our own or write our own grants or whatever.” Thus, she elects to not attend any external professional development. Teacher

D, on the other hand, said he finds going to conferences important, so he ends up just paying for it out of his pocket. Teachers A1, A2, and B said their schools did make funding available for external professional development so it could be a school size issue; larger schools sometimes have larger budgets since school funding is tied to the number of pupils in attendance. Another potential school-level funding issue was raised by Principal A1 as she said her district has been funneling more resources to the elementary schools recently to try to address low reading standardized test results at the 3rd grade level. Since the district has a finite budget, that money has to come from somewhere, so the middle schools and high schools have to make do with less.

At the local level, Teacher D said, “We're trying and trying to get the community to help to pay for some improvements and there was no money in the community, so the community won't vote for it.” Since the community would not support a sinking fund millage for repairs and maintenance at the school, the money needed for these repairs and maintenance has to come from the school's general fund. As a result, there was less money to pay for teachers and materials for the classroom.

As for the state level, Teacher A1 and Principal C both has issues with the Intermediate School District model in the state. Principal C stated,

A lot of our money flows through ISDs in Michigan. Is that the best use of our money to have that be the conduit to come to us? Um, I don't know. I think our ISD in particular has the second highest budget of any educational institution in our area, yet our ISD building has very few students at it. I'm not poking at them. I'm just wondering if that

model hasn't passed, isn't passed its time and more of that money should come directly to the districts.

In addition to questioning the ISD model, Principal C said they had issues with the amount of money that flows out of the state for standardized tests. He said,

The amount of money that our state is sending down south to the Alabama area for standardized testing. I think some of the numbers I heard is over the last eight years it's like \$38 million that leaves Michigan and goes to another state because we can't figure out testing. So, we're gonna buy testing from you guys. I think that is one of the most ridiculous things I've ever heard.

Principal C argued that if we could figure out standardized testing in this state, we could save money that could be spent elsewhere in education instead of contracting that out to another state. For example, that money could be used for teacher professional development, hiring more teachers in order to decrease class sizes, or purchasing new classroom materials. All of these things could help teachers become more effective at teaching mathematics.

Materials

The evidence I found regarding materials was related to standardized tests and curriculum materials. Regarding standardized tests, several of the principals and teachers mentioned that the lack of an annual standardized test at the high school level makes showing growth using standardized test data difficult. Principal A2, for example said,

Technically there really is no year-to-year standardized tests they would take like there are in the lower levels in the elementary school. We have a MSTEP test at the 11th grade, well that's great, but we can't compare it to where you were at the end of tenth

grade or at the end of ninth grade, so for us to use standardized test data, which is something we're struggling with right now.

This lack of year-to-year standardized test data makes it difficult for high school principals to measure growth on teachers' evaluations using standardized test data as there is no baseline data to compare results to. The only practical response, if schools are mandated to use standardized test data, is to compare scores on the MSTEP with different groups of students (e.g. last year's juniors vs this year's juniors). Another response, which was mentioned by all of my participants, to this lack of consistent standardized test data was to use some type of pre-test/post-test structure where scores were compared before and after units on teacher-created pre- and post-tests. In addition to these tests, my participants from my smaller schools (Teachers C and D and Principal C) reported they also used a computer-based standardized test called the Northwest Evaluation Association (NWEA) test to help measure growth due to the lack of year-to-year data from the MSTEP. These participants had positive things to say about the NWEA test. For example, Teacher C said, "I like it because it breaks it down by, I can see students' strengths and weakness and that's way more helpful to me as a whole. They struggle at this. I can work on that more." By being able to see where individual students happen to struggle, teachers can theoretically target interventions to address individual student needs, and thus, be able to increase student standardized test score results.

As for curriculum issues, Teacher A1 and Teacher D both said they did not have any mathematics textbooks. Everything they did, they pieced together from a variety of print and online sources. Doing this could potentially have negative effects on student achievement and teachers' evaluations as key topics may not be taught or taught in time before students are

assessed on standardized tests. Teacher A1 did not see this as a big issue, though, as his school had a long history of working with mathematics education faculty at the local university and he feels his students are prepared. Teacher D, on the other hand, was not as confident and would welcome some textbooks as he is responsible for several different mathematics and science classes at his school.

Time

The evidence I found related to time relate to lack of time to devote to evaluations due to multiple responsibilities, the amount of time required to do each evaluation, the amount of class time used to administer standardized tests, the timing regarding when standardized test data are available, and lack of time to collaborate with other teachers and/or administrators. Regarding lack of time to devote to evaluations, all of the principals mentioned how time-consuming evaluation was. If the principal did not have adequate help, it got even more difficult to do all that was needed to be done regarding evaluations. Principal A1 stated,

I had been in a situation where our assistant principal took a different position halfway through the year and I had an interim who was not allowed to do evaluations because he hadn't been trained. Suddenly I had a greater increase in the observations and, as much as I tried to give the same level of feedback that I had before, I just didn't have the time to do so.

Principal A2, who had adequate help to do all his evaluations, said, "I would say that on average I spend four to five hours a week working through something eval related, whether it's in the classroom watching teachers or it's meeting with teachers, you know, those types of things." So, with adequate help he still devoted 1/8 of his work week to evaluations. Principal C said he

started doing shorter 10-minute walkthroughs instead of observing a whole class because he had no help and this allowed him get to all of the teachers. Teacher C, a teacher at Principal C's school, indicated this may not be the truth though. She said, "I haven't had a classroom visit except for once when he came into a math lab for like 10 minutes, which the math lab, it's not, you know, you're not going to see how I teach. Um, other than that, I haven't seen him in my class in several years." If principals are still not doing the required observations, as Teacher C claims her principal is not, then one would have to question how accurate the data on the evaluations are and the usefulness of that data for teachers.

Regarding the time required to do the evaluations, the principals I spoke with all indicated how much more time consuming the new evaluation requirements were. They had to do more evaluations of teachers than they did before and the instruments they had to fill out were longer than in the past. For example, Principal A1 stated, "it's very time consuming to be in a classroom and have to write down everything that they say and then go through a complicated coding system which may have 25 to 30 different areas of coding." All of the principals advocated for a simpler evaluation, in part, to save time given all of the other aspects of their jobs that needed attention.

As for the standardized tests themselves, the principals and teachers indicated that a significant amount of class time was lost because of the administration of standardized tests and preparing for the tests. Several days of instructional time were lost because of the MSTEP, SAT, ACT Workkeys, and also teacher-generated pre- and post-tests. For those schools that also administered the NWEA, even more time was required. Teacher D stated,

It's (the NWEA math test) 52 questions. A lot of times they'll get to question 40, which means that I've got to pick it up on the next day, which means that I've got a class and a half each. Each time that is time wasted and they get so burnt out on it that it's very hard to find a way to motivate them to take it in a way that truly is representative of their math abilities because they're just not feeling it that day.

One could argue the time that was devoted to standardized testing might be better used actually teaching students as teachers reported the data from the MSTEP and SAT did not even get back in time to do anything meaningful with it and the data from the pre- and post-tests is “largely a joke” (Teacher C).

The final time-related barrier I found evidence of related to collaboration time, which is something that would be important to have if we want teachers to work on their weak areas as identified in their evaluations. For example, teachers could observe how another colleague teaches a particular lesson that they may have had difficulties teaching in the past and then debrief the lesson afterwards. As for the evidence I found, Teachers C and D indicated they did not have any time in the school day to collaborate with their peers. Teacher A2, on the other hand, indicated she did have time to collaborate with her peers, but that was mostly because she is the department head and was given time out of the school day to work with other teachers. She also reported that her school had something called “late-start Wednesdays” where teachers have built-in meeting and professional development time each Wednesday in the morning and the students came in later on in the day.

Other people

The barriers I found regarding other people largely related to collaboration issues for teachers and not enough help for principals. Of all the barriers, this seemed to be the one where school size appeared to have an effect. Teachers at the large schools (A1, A2, and B) said they had opportunities to observe their colleagues teach whereas the teachers at the smaller schools did not (C and D). The reasons given for not being able to collaborate had to do with a lack of common planning time, which was mentioned in the previous section, and also the lack of other mathematics teachers to collaborate with as smaller schools often only have one or two mathematics teachers in the building. In addition to not being able to collaborate with other math teachers, Teachers C and D said they did not have a mathematics coach to collaborate with either. Teachers A1, A2, and B also did not have a mathematics coach, but they at least had other teachers to observe and work with. Teacher A1 also said he his school has had a history of collaborating with mathematics education faculty at a nearby university whereas the others have never worked with university mathematics education faculty.

Regarding collaboration and the new evaluation system, Teacher A1 said,

One thing that I believe in is that collaboration is the key component to making quality teachers and quality classrooms. And so anything that makes you compete to be better with other teachers, I feel is counterproductive. Right? Like if I feel like I have to be better than these five other classrooms that I'm not so sure I want to share my sweet project with them and things like that. Right. I want my scores to look better.

Essentially Teacher A1 was saying the new evaluation system disincentivizes collaboration, which is at odds with what he believes makes quality teachers and the goal of changing evaluation policy, assuming that goal is to improve the quality of instruction.

As for help for principals, Principal C stated,

We've made a conscious decision in this district that the money we were going to spend is going to be on our teachers. We were going to try and keep our teacher to student ratio the best we possibly could. So we had to make some sacrifices. When I started this job, I had two counselors and an assistant principal. Now I'm down to one counselor and no assistant principal. That can make this job really difficult.

This lack of help did not allow this principal to devote the time he would like to teacher evaluations, given all of the other aspects of his job that needed attention as well. It is possible that this would also be the case at even smaller schools as well, given many of the principals at class D schools in Michigan also happen to be the superintendent of the district. As I reported Chapter 4, however, I was unable to interview a principal at a class D school so I cannot confirm this.

Knowledge

The evidence I found regarding knowledge as a barrier related to principals' knowledge of mathematics, teachers' knowledge of the evaluation rubrics, teachers' pedagogical knowledge, teachers' understanding of standardized test data, and legislators' knowledge regarding education. As for principals' knowledge of mathematics, the principals all felt they did not need a background in mathematics to be able to evaluate whether the mathematics

teachers were effective. For example, Principal B stated, “Good teaching is good teaching whether it's calculus or world history or gym”. Principal A2 stated that,

In some respects, I feel like I can gauge the quality of the effectiveness of a lesson in an area where maybe I am not as comfortable with the content easier than an area where I know the content. So I think the reason for that is if it's not content that I'm entirely comfortable with and I can go into a classroom and I can observe what's going on and I can see learning targets and look at performance tasks and the success criteria together and I can actually walk out feeling like I understand it better than I did when I went in.

That tells me the teacher probably was doing a pretty darn good job.

Other principals, Principal C and Principal A1, stated something along the same lines where they downplayed the importance of knowing the content in order to be able to evaluate their teachers effectively. The teachers, on the other hand, all felt having a content background was important. For example, Teacher B stated, “My administration, as much as I like them, they were not math teachers so their instructional feedback can only really be behavior management type stuff, which I don't usually have a problem with.” Even though principals and teachers seem to have some disagreement regarding whether or not a background in mathematics was important to being able to evaluate mathematics teachers, the principals at the larger schools did respond to their lack of content knowledge and tried to divide the administrative staff based on their backgrounds. For example, Principal A2 stated,

We go by department by and large. Part of that is our admin staff has a pretty good variety of teaching background and while I don't think you have to have content

knowledge or experience teaching a specific content to be an evaluator of that type of teaching, I think it does bring some credibility to the conversation with teachers.

This response indicates that principals feel, at least to some degree, that having a background in the content is important if teachers are going to take the results of their evaluations seriously.

Regarding teachers' knowledge of the evaluation rubrics, Teacher C stated, "You can check all the boxes you want, but in order to be effective you can't just check a box without explaining to me why because I'm not going to grow from it. I don't know what any of those little things mean." This quotation indicated that, without some professional development time being devoted to the evaluation rubrics, teachers may not use the data from their evaluations to inform their practice because they do not understand the rubric.

As for teachers' pedagogical knowledge, Principal B stated that "Some of the veteran teachers think that the newer teachers get better scores than them sometimes because you know, they're trained in it." To address this perception, Principal B said he encourages his teachers to attend external professional development opportunities, such as going to the MCTM and NCTM conferences. The school also helped pay for these conferences. Although, as was discussed in the section regarding money, the teachers at the smaller schools did not get funding assistance to attend these conferences.

Regarding teachers' knowledge of standardized test data, Teachers A1, A2, C, and D all said their administrators responded to other teachers' lack of knowledge, and possibly their own, of quantitative data by having the mathematics teachers help make sense of standardized test for the other teachers on staff. The mathematics teachers said they had to work with

other teachers when their districts devoted internal professional development time to data digs. Teacher C indicated helping the other teachers make sense of the data was difficult though. She said,

We tried. We had a meeting with those of us at the high school that received this (standardized test) information. I was the only math teacher in there and somebody else was in there too that, that at least understood the basics and so. But the two of us trying to help all of them (other teachers) was a little bit tricky. Some of them are old dogs and not wanting to learn any new tricks.

The other mathematics teachers did not indicate that it was difficult working with their colleagues, but also did not say it was easy.

The last piece of evidence regarding knowledge I encountered was regarding legislators' knowledge of best practices in education. The only participant that did not talk about this much was Teacher B as she said she tended to ignore politics. The others all questioned whether legislators actually knew what was best for the students in the state and if they actually consulted with experts in education. Principal C was the most vocal about legislators as he stated they have completely stepped over the line by imposing an evaluation system upon all the schools in the state. He claimed, "They don't know what's important in our area" and, as a result, should leave it up to the individual districts to figure out how to best evaluate their teachers.

Apathy

The evidence I found regarding apathy dealt with various aspects of standardized testing and also the teacher evaluation progress. Regarding the standardized tests, the principals and

teachers had concerns about students taking the tests seriously. For example, the NWEA test had no effect on anything for students so there was no incentive for them to take it seriously. Likewise, not every student in high school was planning on going to college so using a test like the SAT was problematic because it may not matter to this subset of the population if they performed well or not on it. Principal A2 addressed this when he stated, “I know darn well that 200 of those kids could care less about how well they do on the SAT because it's not meaningful to them. They know it's not meaningful to them in the sense that that's not their focus in their mindset and their drive or their view of success for themselves.” If students did not take the tests seriously, then it becomes problematic to use data from the tests to evaluate teachers.

In addition to the students not taking the standardized tests seriously, with the exception of Teacher A2, the rest of my participants thought that the bigger standardized tests like the SAT were a “waste of time” (Principal C) and they had “little faith in their results” (Teacher A1) and largely “do not use it (SAT data) for anything” (Teacher A1). These points were at odds a bit with what the participants reported in other parts of the interviews in which they claimed they spent professional development time on doing data digs, that is, unless, nothing was really done with the results of the data digs and they were only being done to appease some directive from their superintendents and/or school boards.

As far as the evaluations went, all of the teachers I interviewed found them not to be helpful in becoming a better teacher. They felt more like a “hoop to jump through” (Teacher D) and just “documentation for them (principals) to use to find your (teachers) way out the door” (Teacher A2). There really was not any reason to use them to get better, especially when some teachers were told they would never be rated as highly effective. For example, Teacher C

stated, "All I know is I know I'm not going to be rated highly effective. I'm going to be rated as effective because that's what the superintendent said we can be. So you can be the best teacher in the building or the worst teacher and we are only able to be effective. Nobody's allowed to be highly effective in our district." One can only imagine being told this would be completely demoralizing and you would not have any incentive to try harder as the effort would not result in any changes to your evaluation.

Stress

All of the evidence I found regarding stress was all related to work stress. None of my participants mentioned anything regarding issues at home that were impacting how they did their jobs. I maybe should have been expected this finding, though, as it was doubtful the teachers and principals would open up about their home lives in an interview with a complete stranger.

As for what was reported as causes of stress at work, my participants mentioned the lack of assistant principals, standardized tests, the current evaluation format, staff shortages, and being the mathematics department head. Regarding the lack of assistant principals, Principal C said he often had to, because of the lack of help, take on additional responsibilities as the athletic director and counselor if those people were busy or out of the building. This reallocation of his responsibilities took valuable time away from his other duties as principal and was a source of stress for him. Teacher D also discussed the lack of assistants as he said his principal was also the superintendent and had no help. Given all the aspects of this person's job, Teacher D said his principal often neglected his evaluation duties and ended up not having

face-to-face meetings with his staff members regarding their final evaluations. Instead, he said he sent the teachers an electronic file they were expected to digitally sign and that was all.

Regarding standardized tests and the current evaluation format, Principal B stated that putting high stakes on the tests “puts unnecessary stress and pressure” on the students and teachers in his building as students’ futures and teachers’ evaluations were dependent on “one score, on one test, on one day”. To help alleviate this stress, Principal B said his school does a lot of standardized test preparation prior to the administration of the MSTEP and SAT and the also incorporate other measures of growth on their evaluations. Likewise, Principal A1 said,

We’ve broken down growth into, we have a district score which is based on standardized testing. We have a building score which is based on our standardized testing for us it is the SAT. And then we also have a teacher growth piece, which for the high school is um, exam, pre, post data. So, we separate that so that we have some classroom data, some district data, and the reason why we chose to do that is we didn't want to get into a situation where growth data which could potentially impact layoff connects directly having the teachers compete.

Even though the intent of what Principal A1 said they did is to help teachers in the event of layoffs, individual data would still be what creates variation in evaluations as the district data would be the same for all teachers. If the individual data is determined by pre- and post-tests, this could incentivize teachers to make easy, and non-informative, post-tests in order to maintain their jobs. Teacher A2 mentioned something along these lines when she said,

People choose goals that they know they are going to achieve because it is directly tied to your evaluation, which in a layoff situation, luckily we haven't been in one here, but I

know most schools have. It's the bottom of the eval pile that gets laid off. It's lowest eval. They're ranking people essentially. And if you're in the bottom of the ranks, I mean that's tied to your livelihood. So am I, you know, as a smart person, am I going to pick a goal I'm likely to achieve? Absolutely. It's one less stressor.

Thus, the pre- and post-test process seemed to be a meaningless exercise for most teachers.

Regarding staff shortages and being the department head, Teacher A2 discussed how both caused her stress and affected her work. She said her school had an issue keeping teachers in the district and, as a result, she ended up teaching an overage where she is compensated for teaching during her prep hour and also during the hour that is supposed to be set aside for her department head duties. Since she was teaching during these two hours, she ended up doing the work she would normally do during her preparation time before school, after school, and at lunch and she found she had no time to devote to help out the newer teachers and they ended up leaving because they did not getting the support they needed. Teacher A2 also mentioned that her duties as the math department head put her in awkward positions at times as she felt she is part teacher and part administrator. Even though she would be a good person to evaluate other teachers, when asked if she did it at all she said, "No, intentionally. Absolutely not. I am a union member the same as my colleagues. So I as a teacher leader that is about the last thing I want to touch with a 20 foot pole." Teacher A2 was also the one who did the scheduling for the mathematics teachers and it was expected that the more senior teachers would get the advanced classes and the new teachers would teach the remedial ones. When she gave the new teachers the remedial classes though, she knew the students in those classes would not perform as well as the students in the advanced classes. Thus, the

teachers of those classes could get a poor evaluation as a result since her school did not use methods to control for factors that affect standardized test scores that teachers have no control over. This has the potential to increase stress and exacerbate their staffing problems.

Revisiting research questions

Now that I have shared the key findings regarding each of the barriers to implementation, I compile the findings together to answer my research questions. As a reminder, the research questions this study aimed to answer are as follows:

1. What evidence of implementation barriers exists as mathematics teachers and principals in the State of Michigan attempt to tie student growth data to teacher evaluations?
 - a. What do mathematics teachers and principals believe the purpose(s) of evaluating teachers is(are)?
 - b. What is mathematics teachers' understanding of how standardized test data and other student growth data are used on their evaluations?
 - c. What do mathematics teachers and principals identify as pros and cons of how student growth is measured on their evaluation?
2. How have mathematics teachers and principals responded to the implementation barriers that do exist?
 - a. How do mathematics teachers analyze standardized test and other student growth data?
 - b. What do mathematics teachers do to address students' weaknesses as identified by standardized test and other student growth data?

- c. What steps do teachers take to improve their teaching as a result of the standardized test and other student growth data?

Regarding research question 1(a), there seemed to be a difference of opinion between principals and teachers regarding the purpose of evaluation. The principals in my study believed the purpose was to help teachers grow, but the teachers felt it was largely a means to identify the weaker teachers in order to fire them. Since the teachers felt the purpose was not to help them grow, they did not use the data generated from the evaluations to identify areas where they could grow. Even if they wanted to, the teachers did not find the current evaluation format useful to them and some did not understand what all the checkboxes on the observation rubric even meant.

Regarding research question 1(b), all of the mathematics teachers were very familiar with how their districts measured growth. The teachers said their districts used several different pieces of data to help measure growth. These data sources included national and state standardized tests such as the MSTEP and SAT, the NWEA, course grades, and teacher-generated pre- and post-tests. All of the teachers had concerns regarding using standardized test data to measure growth as several of the schools were comparing SAT data from different groups of students (e.g. last year's juniors and this year's juniors) and not doing anything to control for factors that the teachers have no control over. The teachers also found using pre- and post-tests to be an easy way to manipulate the growth part of their evaluations, but found it to be a waste of time and a meaningless exercise. They just did it because they were required to.

Regarding question 1(c), principals and teachers both indicated that not relying on any one source of data and being able to show growth easily by using pre- and post-tests to be pros. As for cons, the biggest one was mentioned in the previous paragraph as the teachers found using pre- and post-tests to be a waste of valuable class time and was essentially a meaningless exercise.

Overall, there were several barriers to implementation identified from the interview data as described in this chapter. A discussion of these barriers and what might be done to address these barriers can be found in the next chapter. As for the second research question, I did not collect as much data regarding responses as I did regarding the actual barriers. I wove some of the responses into the discussions of the various barriers throughout the chapter, but I will answer my sub-questions under the second research question here.

Regarding question 2(a), if the mathematics teachers did analyze standardized test data, it was usually done during an internal professional development day where the mathematics teachers were relied upon to help make sense of the data for the whole staff. The teachers usually did not receive any assistance from their administration or outside experts. As for the other tests such as the NWEA, the mathematics teachers received printouts for each of their students regarding how they performed on the fall, winter, and spring administrations of the test and it was up to the individual teacher to make sense of the data.

Regarding question 2(b), the data from tests such as the SAT did not get to the schools in a timely manner, so teachers generally did not make radical changes based on it. Teacher D did report that he would look at his homemade curriculum and see if he had to reorganize it at all the following year to better address student weaknesses. As for the NWEA, this was the test

that seems to have the most promise as Teachers C and D said they do make changes in what and how they teach because they received the data from the test in a timely manner and the data was broken down by each student so they know who needed help in what area.

Regarding question 2(c), teachers from the larger schools reported they sought out external professional development to help work on their craft, but their decision to do this was not driven by standardized test data. They did it because they want to be better teachers. Teacher D said the same thing regarding conferences, but it is harder for him as he has to pay the cost of attending them out of his pocket. Teacher C elected to not do anything as she said there really was not anything worthwhile to go to in her area (she teaches in the Upper Peninsula of Michigan) and if she did want to go to something, she would have to pay for it herself.

Summary

In this chapter I shared my analysis of interview data in relation to the eight barriers to implementation that I identified in Chapter 3. I found evidence of all of the barriers and the largest barrier, based on the number of times excerpts were coded, was disagreement. Principals and teachers largely disagreed with how teachers are currently being evaluated in the State of Michigan and would like to see a simpler system in place. The current system is too time consuming to execute for the principals and not informative or meaningful for teachers. Those teachers that do attempt to work on their craft because they just feel the need to do so on their own. Their decisions are not informed or affected by standardized test or growth data.

In the next chapter I discuss my findings in more detail and offer some suggestions on what might be done if we want improve teacher quality, assuming that is the goal of teacher evaluation reform. I also discuss some ideas for future research.

CHAPTER 6: DISCUSSION

Summary of dissertation

In this study, I interviewed five mathematics teachers and four principals to ascertain what barriers existed as they attempted to implement the changes to teacher evaluations in Michigan. Specifically, I looked for evidence of the following barriers: disagreement, money, knowledge, other people, materials time, apathy, and stress. After analyzing my interview data, I found evidence of all of the barriers, with the bulk of the interview data coded as disagreement. Principals and mathematics teachers disagreed with how teachers were being evaluated. Both groups did not feel standardized test data should be used in a high-stakes matter. Principals felt the current system required too much of their time and that teachers focused too much on their scores. The teachers indicated they did not find the data from their evaluations to be useful to them. As a result, they did not use their evaluations to inform their practice. Instead of the current system, principals advocated for a simpler system where they could just have conversations with teachers about their practice instead of talking about a rubric. Teachers just asked that, regardless of what system is in place, that they be evaluated fairly and the data from their evaluations are meaningful to them.

In this chapter, I first revisit the literature from my first two chapters and connect it to my findings. Then I discuss what could be done to address the barriers mentioned in Chapter 5. This is followed by a discussion of some possible options for teacher evaluation policy in Michigan, including my recommendation for what my findings suggest we should do, especially if it is later determined that my findings are common throughout the state. Then I briefly

discuss how my identities affected this study. This is followed by some lessons I learned as I was working on my dissertation, limitations of this study, and suggestions for future research.

Revisiting literature

In the first chapter of this dissertation, I mentioned how various reforms have been implemented in recent decades in response to a perceived crisis in American education at the K-12 level. My specific focus was regarding those reforms that involved schools and teachers being held accountable for student learning. One method that was implemented to hold schools accountable for student learning was to assign letter grades to the schools, but researchers found that assigning letter grades did not have an effect on student performance on standardized tests if the grades were not accompanied by consequences (Hanushek & Raymond, 2005; Winters & Cowen, 2012). Given this finding, I hypothesized that consequences were necessary if a reform was expected to have a positive effect on student performance, but I did not investigate this hypothesis in this study as I did not have access to student standardized test data.

In recent years, several states passed laws that tied student achievement and growth data to teacher evaluations. I assumed a goal of these laws was to increase student performance by putting more pressure on teachers to improve their practice. The consequence for not improving, which would be determined by teachers' scores on their evaluations, would be loss of employment. In this study, however, I found that schools were trying to insulate their teachers from students' standardized test scores and allowing teachers to use other sources of information to show student growth. Some schools did not use standardized test data at all, even though it was mandated by the state. Others supplemented this data with

scores from other standardized tests, such as the NWEA, and with teacher-generated pre- and post-tests for which teachers could set their own growth goals. Also, it should be noted that no teacher or principal in the study reported that any teacher they knew lost their job as a result of their evaluation since the changes to teacher evaluations in Michigan took effect. Thus, this new teacher evaluation policy did not appear to have any real consequences in practice for my participants. If consequences are necessary for student performance to increase, then my findings suggest that it is unlikely this policy is positively affecting student standardized test scores given how it is being implemented in Michigan. This is assuming my findings are common throughout the state.

In the second chapter of this dissertation, I discussed the two big changes to teacher evaluations that have occurred as a result of recent reforms. Specifically, several states now use standardized observation protocols and also use student growth data as a means of introducing variation in evaluation outcomes as there was not a lot of variation before (Weisberg et al, 2009). Regarding the standardized observation protocols, the principals in my study found them to be extremely time-consuming to fill out and the teachers did not use the data from them to inform their practice. If teachers are not using the data from the protocols, then the use of data either needs to be seriously questioned or teachers need some professional development time allotted to understanding the protocols and learning how to use them to inform their practice.

As for the student growth component on the evaluations, in the second chapter I discussed how test-based models can be used on teacher evaluations and also some issues identified with their use. The participants in my study mentioned a lot of the same issues that

were raised in the literature. For example, one of my participants mentioned that using standardized test data on the evaluations disincentivizes collaboration, which is something Guenther (2019) raised in her dissertation. Also, several of my participants mentioned how the way their schools measure growth does not control for factors that affect student achievement, such as a student's family background, whether the students' basic needs are being met, or socioeconomic status which are all factors mentioned in the NRC and NAE's report (2010). Schools and teachers, however, have no ability to affect these particular factors. Although, the State of Michigan could have attempted to control for these factors if they had followed the MCEE's (2013) recommendation of using value-added models to measure growth, but they elected to ignore this recommendation and leave it up to the individual schools to figure out how to measure growth. One could argue that the lack of direction from the state regarding how to measure growth played a large role in the policy not being implemented as policymakers may have intended at the schools in my study.

In the second chapter, I ended by discussing a study regarding the characteristics of successful schools that served poor students. I did this because, if a goal of the teacher evaluation policy was for teachers to improve their practice, then these teachers might look to successful schools to see what they are doing in order to boost their evaluation scores. According to Kitchen, DePree, Celedon-Pattichis, and Brinkerhoff (2007), the nine schools they studied all had some commonalities that they argue led to higher student achievement as measured by standardized tests. Those characteristics included:

- Administrators handled most of the paperwork and problems with students and parents so their teachers would only have to worry about teaching.

- Students were provided with additional instructional time beyond their normal mathematics classes for remediation and also to challenge them.
- Teachers were given cell phones by the school so their students could call them at night for extra help.
- Schools provided a late bus so students could get extra help after school from their teachers (some of these teachers were paid extra to stay after school to tutor students).
- Teachers had access to a variety of different teaching resources. If the teachers wanted something for their classroom, all they needed to do was ask.
- Teachers were encouraged to and took part in sustained professional development activities. In many cases, the schools would pay for their teachers to attend conferences and workshops.

In addition to these characteristics, there were some common practices among the teachers at these schools:

- Teachers had students work in groups in order to facilitate communication about mathematical ideas.
- Teachers did not let their textbook define what was going to be taught in their classrooms. Several of the teachers would supplement their textbook and continuously plan, reflect, and alter what they did in their classrooms.
- Teachers incorporated tasks that were meaningful to their students.
- Teachers would analyze students' standardized test scores and engage in "backward curriculum planning" (p 89). That is, the teachers would look at the scores, see where their students needed to improve, and then vertically align

their middle school and high school mathematics curriculum to address the areas that were shown to be weak.

- Teachers worked collaboratively with their colleagues planning lessons, creating a support system amongst the faculty, and holding each other accountable.

If schools in Michigan try to emulate what was reported in the Kitchen et al.'s (2007) study, my findings regarding teacher evaluation reform suggest they may have difficulties with some of the characteristics and practices. One issue is regarding collaboration, as I mentioned earlier. Teachers in my study, and Guenther's (2019), indicated that using standardized test data on their evaluations disincentivizes collaboration as helping another teacher could artificially inflate their evaluation scores. In the event of layoffs, this could result in someone unfairly keeping or losing their job. Also, even though Kitchen et al (2007) did not argue that collaboration was the most important practice that led to the schools being successful, it was something that Teacher A1 felt was extremely important if teachers are to grow as professionals, which is an assumed goal of the current evaluation policy. This also resonates with my own experience as a high school teacher. I attribute a large portion of my learning as a high school mathematics teacher to observing others teach, co-planning with others, and listening to my colleagues' advice. Without this collaboration, I likely would not have made it past my first year of teaching; any policy that would discourage helpful collaborations. I would argue, needs to be seriously rethought.

Another potential difficulty schools may have is regarding the use of standardized test data. Teachers in Kitchen et al.'s (2007) study seemed to use the data from the tests to inform their practice a lot more than the teachers in my study or Cavanna's (2016) study. This could be

because the teachers in our studies did not value the standardized test data because they did not have a hand in creating it, as Cavanna (2016) suggested. This could also be because standardized tests were not meant to inform classroom instruction, but rather inform district-level evaluation and policy making (Shepard, Penuel, & Pellegrino, 2018). Another possible hypothesis is that it could be due to differences at the administrative level at the schools as the principals I interviewed did not seem to value the data from standardized tests and, as a result, they did not encourage their teachers to use the data in any way. Likewise, Cavanna (2016) stated the administrators at one of the schools in her study did not encourage their teachers to use the data to inform their instruction at all. If administrators actively encouraged their teachers to use the data, it may have been done instead of just doing a data dig one day and then doing little or nothing with the data afterwards. Speaking from my experience this past year teaching at an online school, I know my administrators' perceived lack of caring about our NWEA results, a test that other teachers in my study found to be helpful, played a large role in my, and my colleagues', lack of use of the data to inform our instruction.

Another characteristic of successful schools that some of my participants had issues with was they did not have access to and nor were they encouraged to take part in sustained professional development activities. As was reported earlier, the teachers at the smaller schools in my study reported that they had to pay for any professional development they attended. The teacher who taught in a very rural part of the state reported that the quality of professional development activities was lacking. Thus, it seems being a teacher at a small and/or rural school in Michigan could be detrimental to a teacher's professional growth. For example, during the first nine years of my teaching career, I taught at a small rural school in the

Upper Peninsula of Michigan and most of my professional development activities consisted of “teacher time” where I essentially sat in my classroom and graded papers as the administration had nothing prepared for us. The only time I had the opportunity to take part in meaningful professional development was when the local university received grant money and they recruited teachers to take part in a two-year lesson study project. After this project ended and I finished my Master’s degree, I felt as if my professional growth had plateaued and this feeling was one of the reasons I decided to quit my job and pursue my PhD at Michigan State University.

The final characteristics of successful schools I would like to discuss both relate to time. In the successful schools, administrators handled a lot of paperwork and the problems with students and parents, allowing teachers to focus on teaching. With the increased demand on principals’ time due to the new evaluation system in Michigan, it is unlikely principals will be able to handle all of this, especially at smaller schools where principals may not have any assistants or where they also wear the hat of superintendent. Teachers will likely have to shoulder more of the administrative burden, which will make devoting additional time to instruction difficult.

Overall, the new evaluation system seems to present issues that make replicating what is done at successful schools, as identified by Kitchen et al (2007), difficult. Assuming that their findings are transferrable to other schools and contexts, then it seems that something needs to change with the current teacher evaluation policy as it seems to be fostering practices that are counterproductive to the goals of the policy.

Addressing barriers

In the previous chapter, I highlighted evidence of the various barriers from my interview data. As was mentioned earlier, I found evidence of all of the following barriers: disagreement, money, knowledge, other people, materials time, apathy, and stress. Knowing what barriers exist is helpful, but the work does not end there. It is also necessary to figure out what can be done to address the barriers. Thus, I would like to highlight a few of the barriers that could be the biggest roadblocks to implementing the new teacher evaluation policy and also give a possible way to address each barrier.

The first barrier that needs to be addressed is the overall disagreement that principals and teachers had regarding the use of standardized test data on teacher evaluations. Neither group believed that standardized test data should be used on the evaluations. As a result, standardized test data were either completely ignored or other data were used in addition to the standardized test data to measure growth in order to lessen its impact. If this disagreement is common throughout the state, then policymakers need to clearly articulate how and why educators need to use standardized test data. Without clear guidance, it is unlikely the data will be used as policymakers had intended and this policy will not have the desired effect on student achievement.

Another issue that was raised in the previous chapter that needs to be addressed is actually a combination of multiple barriers. Principals struggled to get all of their evaluations done, in addition to all of their other duties, because they lacked the time and personnel to do so. All of the principals mentioned how time-consuming it was to conduct all of the observations, fill out the observation rubrics, and have follow-up meetings with the teachers.

Some principals mentioned how their budgets have been shrinking and how they have lost assistant principals. Thus, they now have more time they need to devote to evaluations with fewer people to do it. An obvious solution to this would be to adequately fund our schools so principals can get more help and to not be as stressed out about all of the things they need to accomplish in a day. In addition to not being stressed, having fewer things to accomplish in a day could allow time for thoughtful practice instead of checking things off a list which could improve their performance as principals and also improve the quality of feedback they give their teachers. If funding cannot be increased, then maybe reduce the frequency of observations for tenured teachers to be more in line with previous teacher evaluation policies in Michigan where they were evaluated only once every three years.

A third issue that needs to be addressed relates to knowledge. There was evidence that teachers did not understand the observation rubrics. As a result, they did not use the data on the rubrics to inform their practice. If teachers are to use this data, then they need to understand “what all the little boxes mean” (Teacher C). This could be addressed by having each school devote professional development time to discussing, understanding and using the observation rubric.

Other barriers from Chapter 5, such as student apathy, certainly affect standardized test scores and teacher evaluation scores, but are harder to address. That does not mean they should be ignored though. It just means significant time, thought, and probably money need to be devoted to addressing them. If addressing the barriers to implementing this policy are not palatable to policymakers, then another option would be to change the policy itself. Possible options for what could be done follow in the next section.

Potential options for teacher evaluations in Michigan

One option for teacher evaluations in Michigan is to stay with the status quo. That is, continue using state-approved observation rubrics and leave it up to the schools to figure out how to measure growth. I would advise against this if what the teachers reported in my study is common throughout the state: they did not find the data from their evaluations useful, so they did not use the data to inform their practice at all. This would go against one of the assumed goals of teacher evaluation reform. The teachers and principals also questioned the use of standardized tests in a high-stakes manner given teachers cannot control for many of the factors that affect students' scores on these tests which I described earlier in this chapter and in Chapter 2. If these factors are not controlled for, teachers could be unfairly retained or terminated in the event layoffs are necessary.

Another option is to leave the current policy relatively intact, but follow the MCEE's (2013) recommendation of using value-added models statewide to measure growth instead of leaving it up to the individual schools to figure out how to measure growth. If policymakers insist that standardized test data be used, this is likely the fairest way to do it as value-added models attempt to control for the factors that teachers have no control over and allegedly are able to isolate a given teacher's effect. Many articles have been published in recent years regarding value-added models, though, that raise issues with their ability to do what they claim. For example, Goldhaber et al (2013) found that when comparing results from a "Student Fixed-Effects Model" and a "Student Fixed-Effects with Lagged Score Model" (two different value-added models) there were significant variations in teacher quality as reported by the models. Some teachers that were reported as being in the lowest quintile in one model were reported

as being in the highest quintile using the other model and vice versa. Given this, it seems there would still be some work to do with value-added models if they were to be used in a high-stakes manner. It could be something to pursue, however.

A third option could be to just throw out the growth component and keep the state-approved observation rubrics. This would help eliminate the temptation to teach towards the tests and allow principals and teachers to focus on what they value in education. An issue with this, though, is the teachers in my study did not find the data from the rubrics to be useful to them and some (e.g. Teacher C) did not understand “what all the check boxes mean” on the rubrics. This indicated that, if this is the option that would be chosen, time would need to be devoted to helping teachers understand the rubrics prior to their use. In addition to the teachers not finding the data useful, principals in my study indicated that filling out the rubrics was very time-consuming. Thus, it is likely schools would need more money to hire assistants for principals in order to carry out their evaluation duties.

A fourth option, even though it would likely be unpopular, would be to throw out the observation piece and just use standardized test data to evaluate teachers. The reason for evaluating teachers this way would be to address claims that there was little to no variation in evaluations in previous iterations of teacher evaluation policy (Weisberg et al., 2009). This would introduce variation into the evaluations and allow schools to rank their teachers from most to least effective, but there is evidence that different value-added models can generate different rankings with the same data so this may not be the best option. This option also emphasizes an aspect of school that my participants do not value at all and completely ignores the parts they do value.

My recommended option, assuming my findings are common throughout the state, would be to do something along the lines of what my participants wanted. The principals all advocated for a simpler system where they could just focus on conversations with teachers and not have to fill out unruly forms. The teachers wanted something meaningful to them. For example, Teacher A1 advocated for a system where teachers have to prove they are working on their craft in some way. Thus, I would suggest allowing individual schools more freedom in what observation rubrics they use. If they really want to use the ones that were approved by the state, then they could choose to do so. If they feel the need to create a simpler form, then that they could choose to do that. In addition to using these forms, the state may want to consider having teachers do some type of action research project every few years in order to demonstrate they are working on and learning from their craft. There could be other ways to prove they are working on their craft, but action research was one that I, and others, have found to be valuable as a high school teacher (Bennett, 2004; Mitchell, Reilly, & Logue; 2009). It is also something that some schools were doing already to show growth when I looked at evaluation instruments in a previous study (Morissette, 2014).

What can university-level mathematics educators do?

One thing that mathematics educators can do, assuming something along the lines of my recommended option happens, is to help practicing mathematics teachers with their professional growth beyond teaching graduate classes. At a minimum, I believe mathematics educators should be reaching out to local teachers to see what they are struggling with, what questions they have, and possibly be a guide as teachers work on their own action research projects. We can suggest different articles teachers may want to read or even act as a sounding

board as they think of potential solutions to whatever issue they may be trying to work through in their classrooms. Several mathematics educators already have partnerships with teachers and do work in schools, but in my ideal world, I would like to see everyone do this as I believe more frequent interactions between academics and K-12 teachers would be beneficial to all in the education field and it would help eliminate any perceived chasm between academia and the “real world”.

In addition to interacting with practicing teachers more, mathematics educators may consider getting involved in the political arena and try to influence policymakers, or actually become one. Several of my participants were fed up with the lack of respect the teaching profession gets and also how it seemed that policymakers would pass legislation regarding education without getting or listening to the opinions of people with a background in education. We need to get involved in politics and be persistent if we are going to be able to effect change. Regarding this, I need to listen to my own advice and get involved more as I have found I have wanted to give up on the education field several times in the past few years. I need to take what I have learned in this study and be persistently annoying until I am able to get the attention of policymakers. Once I have their attention, I want to help get a fairer and more meaningful teacher evaluation policy instituted in the state.

Positionality statement

There are a couple aspects of my identity that are worth mentioning in this dissertation as they had an effect on this study. The first aspect I feel that needs to be mentioned is that I have lived in the State of Michigan my entire life and I care very deeply about what happens in this state. My love for this state is the reason I decided to study teacher evaluation reform

here as opposed to any other state. Also, for the first 32 years of my life, I lived in a very rural area in the middle of the Upper Peninsula (UP). Even though I have lived in the greater Lansing area for 8 years now, I still identify as a Yooper. Yoopers generally feel they are often ignored by politicians in Lansing, basically because they can be. Only 3% of the state's population is in the UP and the perception that Yoopers have is their opinions and votes largely do not matter on anything that affects the whole state. In the MCEE's (2013) report, for example, there was no representation from the UP and I found this to be extremely bothersome. Since I am bothered every time the UP is ignored, I ensured I would have some representation from the UP when I selected participants to interview for my study. If I continue to do research regarding education-related policies in Michigan, I plan on continuing this practice of ensuring I have at least one participant from the UP as I strongly believe their opinions and concerns matter.

Another aspect of my identity that needs to be mentioned is that I was a secondary mathematics teacher in Michigan for nine years prior to going to graduate school and I am currently teaching at the secondary level in Michigan again as I finish this dissertation. The people in this study are more than just subjects to me. They are my colleagues and I want to see them treated fairly. When introducing myself to my potential interviewees, I discussed my teaching background and my overall goal of seeing a fair and meaningful evaluation system put in place in Michigan. I feel that doing this and letting them know I was "one of them" helped me get more honest and thorough answers during my interviews. Honestly, I do not believe that a teacher would flat out tell me their school does not use standardized tests on their evaluations if I was a policymaker, reporter, or another graduate student that had no experience teaching.

As I was working on this dissertation, where I am from and who I am were constantly on my mind. I wanted to make sure the concerns and opinions of people in a part of the state that gets ignored were told and I feel I was able to do that in addition to making sure people from other parts of the state were heard as well. Regarding my identity as a teacher, I tried to consciously withhold my beliefs and wishes regarding what changes I would like to see happen during the interviews as I did not want to influence the teachers' and principals' thinking. There were a few that asked about my beliefs, but I tabled my answer until after the interviews were over and these conversations were not recorded. I would like to say I felt these off-the-record conversations to be rather cathartic for me and my participants. Even if this dissertation does not affect any meaningful change at the state level, I was at least able to make my participants feel heard and that someone cares about what they think.

Lessons learned

During this whole dissertation process, I came to realize two big things. The first is that I feel I was underprepared regarding research methods. I had assumed that if I took the research courses listed under our program requirements, then that, along with what I learned in my research assistantship my first two years at MSU, would have been enough to prepare me to conduct my research. The problem with the courses in the program requirement, though, is they are mostly a survey of a variety of different methods and you do not really come out of the courses really knowing how to do any of those methods well. Looking back, I really should have asked to be on a different research assistantship after my first year and also taken a course or two regarding survey methods as I feel I failed in that area in this dissertation. In fact, I did not get enough people to respond to my surveys and ended up throwing out most of the data. I

only ended up reporting some basic descriptive statistics in an appendix. By doing this, I was unable to make any claims or use quantitative findings to further support my interview results. Thus, I could not make any firm recommendations for what to do with this policy as I really only have data from nine people. With data from only nine individuals, I doubt any policymaker would make any changes based on my results. My advice for people starting out in a doctoral program would be to try to figure out what type of study you would like to do for your dissertation as early as possible and then sit down with your guidance committee to figure out what course(s) to take in order to give you the best chance to succeed or possibly take additional coursework even when it is not required. Learning more things is never a bad thing.

The other thing I came to realize as I was working on this dissertation is that everyone seems to have a different idea regarding what needs to be in a dissertation, even within the same program. I had assumed that using a dissertation of one recent graduate of the program as a guide would be a good way to get me through this process. I was wrong and it resulted in extending my time of completion by a year. If I were to do this all over again, I would have had conversations with my committee regarding what they believed should be in a dissertation and why and also looked for dissertations of people that had all, or most, of the same committee members on it. Having these conversations early on and looking at multiple dissertations would be my advice to others as well as doing extensive revisions is a stressful and time-consuming process.

Regarding the lessons I learned, there is nothing I can really do to address the second one as that would require a time machine. The first, however, is a lesson I can do something about in the future. I plan on finding and reading books and articles regarding survey research

in order to patch any holes in my understanding. Then, if I am lucky enough to get a job at a university, I plan on seeking out individuals that have more experience with surveys to see if they would be willing to work together on a study and/or paper. Hopefully, with time and effort, I can get better and be able to effect change in the policy world.

Limitations

The biggest limitation I see in this study is that I had a rather small sample size of principals and teachers that participated. As a result, I was unable to perform my independent samples t-tests as I had planned and did not use my survey data at all. Thus, triangulation was difficult, but I did have another person code my data to try to improve the trustworthiness of my study. The other limitation I can think of is the teachers and principals I interviewed seemed like they wanted to talk to me because they did not like the current system and wanted to see changes. It is unclear how prevalent their views regarding the current evaluation system are among educators in the state. More work needs to be done to address the weaknesses in this study, but I feel I have a better idea regarding what to focus on in a follow-up study.

Suggestions for future research

If I were to follow up on this study, I would likely draft new surveys using my conceptual framework and findings as guides to see how common the barriers to implementation of this teacher evaluation policy are throughout the state. Specifically, I would be interested to know how many principals feel you do not need a background in a subject in order to evaluate it well as all of my principals said something along the lines of “good teaching is good teaching” and that they did not need to know the subject in order to judge if the teacher was performing well. I would also like to know how many mathematics teachers feel that this background is

important as all of my teachers in this study thought it was necessary to have this background to be able to evaluate them. This disagreement between the two groups of people seems like it could affect teachers' use of the evaluation data to inform their practice.

In addition to investigating the importance of content background, I would also like to know if other teachers throughout the state find the standardized observation protocols not to be helpful. If they are not using the data on those forms to inform their practice, are the protocols really worth using? Is there anything that can be done to encourage teachers to use this observational data to inform their practice?

Another thing I would like to investigate further is principals' and mathematics teachers' beliefs regarding the purpose of teacher evaluations. The principals in my study all believed that the purpose of evaluation was to help teachers grow as professionals, but the teachers in my study all thought the primary purpose was accountability-related. That is, they thought they were being evaluated in order to identify who the weaker teachers were in order to fire them. It would be reasonable to believe that if teachers think evaluation is just a sorting mechanism, then they may not look at the data from their evaluations to inform their practice.

APPENDICES

APPENDIX A: Principal Survey

1. Including your current job and all previous jobs where you have worked, how many TOTAL years will you have been a principal at the end of this school year?
2. By the end of this school year, how many total years will you have been the principal at your current school?
3. How many school districts have you been employed at?
4. Are you certified to teach mathematics in the State of Michigan? If so, was mathematics your major or your minor?
5. Do you have a Master's degree? If so, what was your major field of study for the advanced degree?
6. Do you have a Doctorate? If so, what was your major field of study for the advanced degree?
7. Does your school district use standardized tests (e.g. ACT, SAT, MEAP, MME, MSTEP) to help measure student growth on teacher evaluations (Y/N/Unsure)?
8. Which of the following models comes closest to the way your school uses standardized tests for the purpose of conducting teacher evaluations? (Note: include a brief description of a status and growth model on Qualtrics)?
9. What do you see as the advantages and disadvantages of using the model you use for measuring student growth with standardized tests for the purpose of conducting teacher evaluations (only fill in the box on the survey)?
10. How familiar are you with value-added growth models (not at all familiar, slightly familiar, moderately familiar, very familiar, extremely familiar)
11. How comfortable are you with evaluating the quality of instruction for your mathematics teachers (not at all, slightly, moderately, very, extremely)?
12. What parties (e.g. teacher union, superintendent, principal, individual teachers, ISD staff) were involved in the decision regarding how to measure student growth on evaluations?
13. How well do you think standardized tests measure student growth (not at all well, somewhat well, moderately well, very well, extremely well)
14. To what extent do you agree or disagree with the following statements (strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree)?
 - a) The purpose of teacher evaluation is to inform personnel decisions (e.g. hiring, termination, raises, and/or promotions).
 - b) The purpose of teacher evaluation is to inform professional development activities as evaluation helps identify areas of weakness for a given teacher.
 - c) The purpose of teacher evaluation reform is to identify and reward strong teaching (e.g. merit pay).
15. In general, how much assistance would you say you personally give your school's mathematics teachers with each of the following tasks (none, a little, some, quite a bit, a great deal)
Planning for instruction
 - a) Acquiring materials related to mathematics instruction.
 - b) Establishing classroom routines and procedures (e.g. collecting homework).
 - c) Matching the curriculum to standards.
 - d) Using standardized test scores to improve instruction.

- e) Identifying individuals who can share their expertise in mathematics (and/or mathematics teaching).
 - f) Understanding the central mathematical ideas in the curriculum.
16. To what extent do you expect your mathematics teachers to do the following things (not at all, a little, some, quite a bit, a great deal)?
- a) Adhere to a prescribed pacing in their instruction.
 - b) Make sure that their students' test scores are high.
 - c) Address the state/district standards and objectives.
 - d) Have whole classroom discussion in which students explain how they solved tasks.
 - e) Have small-group discussion in which students explain how they solved tasks.
 - f) Use the adopted curriculum as a basis for their instruction.
 - g) Keep their students quiet and disciplined during classroom instruction.
 - h) Use challenging, problem solving tasks with their students.
 - i) Use students' current mathematical thinking to inform their instruction.
 - j) Collaborate with other mathematics teachers.
 - k) Observe other mathematics teachers' instruction.
 - l) Use yourself as a resource when instructional problems arise.
 - m) Make their lesson plans available for review.
 - n) Assist other mathematics teachers in improving their instruction.
17. To what extent do you encourage your teachers to use standardized test results to inform their practice (not at all, a little, to a moderate extent, quite a bit, a great deal)
18. To what extent do you agree or disagree with the following statement: I have enough time to evaluate all of my school's teachers to the best of my ability (strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree strongly disagree).
19. Have you terminated any teachers as a result of their score on their evaluation? (Y/N)
20. If you are interested in being contacted in the future for a follow-up interview, please provide your contact information below. The information will ONLY be used for my research purposes and will not be shared with anyone. (First Name, Last Name, Phone Number, E-mail Address).

APPENDIX B: Mathematics Teacher Survey

1. Including your current job and all previous jobs where you have worked, how many TOTAL years will you have been a mathematics teacher at the end of this school year?
2. By the end of this school year, how many total years will you have been a teacher at your current school?
3. How many school districts have you been employed at?
4. Are you certified to teach mathematics in the State of Michigan? If so, was mathematics your major or your minor?
5. Do you have a Master's degree? If so, what was your major field of study for the advanced degree?
6. Do you have a Doctorate? If so, what was your major field of study for the advanced degree?
7. Does your school district use standardized tests (e.g. ACT, SAT, MEAP, MME, MSTEP) to help measure student growth on teacher evaluations (Y/N/Unsure)?
8. Which of the following models comes closer to the way your school uses standardized tests for the purpose of conducting teacher evaluations? (Note: include a brief description of each in survey)?
9. How often do teachers in your department do the following?
 - a) work together to develop curriculum and instructional materials (never, annually, monthly, weekly, daily)?
 - b) observe each other teach (never, annually, monthly, weekly, daily)?
 - c) offer advice or help to each other (never, annually, monthly, weekly, daily)?
 - d) share ideas on teaching (never, annually, monthly, weekly, daily)?
 - e) promote innovative teaching practices (never, annually, monthly, weekly, daily)?
10. This past school year, how many times have you received meaningful feedback on your performance from an administrator? (never, 1-2 times, 3-5 times, 6-10 times, more than 10 times)
11. This past school year, how many times have you received meaningful feedback on your performance from a fellow teacher? (never, 1-2 times, 3-5 times, 6-10 times, more than 10 times)
12. Does your school have a school-based mathematics coach (Y/N)?

13. In the past year, how often have the following events occurred (never, 1-2 times, 3-5 times, 6-10 times, more than 10 times)?
- a) I discussed my teaching with a school principal or assistant principal.
 - b) A school principal or assistant principal observed my teaching (for at least 10 minutes).
 - c) A school principal or assistant principal provided me with feedback to improve my instruction after observing my teaching.
 - d) A school principal or assistant principal reviewed my students' work with me.
14. In general, in the past year, how much assistance would you say your principal gave you with each of the following tasks (none, a little, some, quite a bit, a great deal)
- g) Planning for instruction
 - h) Acquiring materials (e.g. manipulatives, textbooks, technology) related to mathematics instruction.
 - i) Establishing classroom routines and procedures (e.g. collecting homework).
 - j) Matching the curriculum to standards.
 - k) Using standardized test scores to improve instruction.
 - l) Identifying individuals who can share their expertise in mathematics (and/or mathematics teaching).
 - m) Understanding the central mathematical ideas in the curriculum.
15. To what extent do you agree or disagree with each of the following statements (strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree)?
- a) The purpose of my school principal (or assistant principal) visiting my classroom is to directly assist me in improving my teaching.
 - b) The purpose of my school principal (or assistant principal) visiting my classroom is to evaluate my teaching in terms of job performance.
 - c) My principal (or assistant principal) possesses a thorough knowledge of the curriculum and related instructional materials.
16. To what extent does your principal (or assistant principal) and fellow mathematics teachers expect you to do the following things (not at all, a little, some, quite a bit, a great deal)?
- o) Adhere to a prescribed pacing in my instruction.
 - p) Make sure that my students' test scores are high.
 - q) Address the state/district standards and objectives.
 - r) Have whole classroom discussion in which students explain how they solved tasks.
 - s) Have small-group discussion in which students explain how they solved tasks.
 - t) Use the adopted curriculum as a basis for my instruction.
 - u) Keep my students quiet and disciplined during classroom instruction.
 - v) Use challenging, problem solving tasks with my students.
 - w) Use students' current mathematical thinking to inform my instruction.
 - x) Collaborate with other mathematics teachers.
 - y) Observe other mathematics teachers' instruction.
 - z) Use him/her/them as a resource when instructional problems arise.

aa) Make my lesson plans available for review.

bb) Assist other mathematics teachers in improving their instruction.

17. To what extent have you made efforts to change your teaching based on your experience in professional development sessions (not at all, slightly, moderately, considerably, a great deal)?
18. To what extent have you made efforts to change your teaching based on your experience in courses you have taken since the beginning of your teaching career (not at all, slightly, moderately, considerably, a great deal)?
19. To what extent do you believe standardized tests capture your students' knowledge of mathematics (not at all, small, moderate, great)?
20. How good of an indicator do you believe standardized test results are of the quality of your teaching (not at all, small, moderate, great)?
21. How much of a teacher's evaluation do you think should be based on standardized test results (0%, 1-10%, 11-25%, 26-50%, >50%)?
22. How often do you attend state or national mathematics education conferences (e.g. MCTM or NCTM) (never, 1-2 times, 3-5 times, 6-10 times, more than 10 times)?
23. How often do you attend various professional development activities/sessions/conferences etc. NOT provided by your school district (never, 1-2 times, 3-5 times, 6-10 times, more than 10 times)?
24. How familiar are you with how teachers are evaluated in your district? (none, a little, some, quite a bit, a great deal)
25. Do you believe teacher evaluation reform has made the instruction in your classroom (a lot worse, a little worse, no change, a little better, a lot better)?
26. What do you see as the advantages and disadvantages of using the model you use for measuring student growth with standardized tests for the purpose of conducting teacher evaluations (only fill in the box on the survey)?
27. If you are interested in being contacted in the future for a follow-up interview, please provide your contact information below. The information will ONLY be used for my research purposes and will not be shared with anyone. (First Name, Last Name, Phone Number, E-mail Address).

APPENDIX C: Principal Interview Protocol

Beginning Script:

Before we get started, I'd like to thank you for not only filling out the survey but also for taking the time to talk with me further. I'm really interested in teacher evaluation and in the relationship between teacher evaluation and standardized test scores of students. Would it be okay if I recorded our conversation? This is mainly for my recollection and so that I can listen carefully now rather than have to write things down at this point in time. I will assign you a pseudonym so that this data cannot be traced back to you. And, I am the only person who will listen to and transcribe this.

Interview Questions (may ask others depending on how conversation goes):

1. What have your experiences with the teacher evaluation process been like? Have you noticed any changes in the process throughout your career? If so, what are those changes?
2. When you conduct formal observations of your mathematics teachers, what are some aspects of their teaching you focus on?
3. How do principals evaluate content-specific aspects of teaching if the subject was not their major or minor in college?
4. In your opinion, what is(are) the purpose(s) for evaluating teachers? (fire vs inform teaching)
5. What is your understanding of how student growth data are used on teacher evaluations? How do you show growth? Is it different for different subjects? What do you think the advantages and disadvantages are of measuring growth the way your district does?
6. What changes would you like to see happen regarding how teachers are evaluated? (student growth data, number of observations, who observes teachers)
7. What are your views regarding standardized testing? (frequency, how well it measures student knowledge, indicator of teacher effectiveness).
8. When your school gets their standardized test results from the state, what do you do with them? (nothing, instruct teachers to formulate a plan to address them, etc)
9. If standardized test data indicate a specific area is a weakness for your students, how do you respond?
10. How would you like your math teachers to use standardized test data?
11. How does standardized test data affect PD decisions (e.g. what you provide for staff)?
12. What questions do you have for me?

APPENDIX D: Mathematics Teacher Interview Protocol

Beginning Script:

Before we get started, I'd like to thank you for not only filling out the survey but also for taking the time to talk with me further. I'm really interested in teacher evaluation and in the relationship between teacher evaluation and standardized test scores of students. Would it be okay if I recorded our conversation? This is mainly for my recollection and so that I can listen carefully now rather than have to write things down at this point in time. I will assign you a pseudonym so that this data cannot be traced back to you. And, I am the only person who will listen to and transcribe this.

Interview Questions (may ask others depending on how conversation goes):

1. What have your experiences with the teacher evaluation process been like? Have you noticed any changes in the process throughout your career? If so, what are those changes?
2. In your opinion, what is(are) the purpose(s) for evaluating teachers? (fire vs inform teaching)
3. What is your understanding of how student growth data are used on teacher evaluations? How do you show growth? What do you think the advantages and disadvantages are of measuring growth the way your district does?
4. What changes would you like to see happen regarding how teachers are evaluated? (student growth data, number of observations, who observes you)
5. What are your views regarding standardized testing? (frequency, how well it measures student knowledge, indicator of teacher effectiveness).
6. When your school gets their standardized test results from the state, what do you do with them? (nothing, study with others, etc)
7. If standardized test data indicate a specific area is a weakness for your students, how do you respond?
8. How do you decide WHAT you are going to teach on a given day? (Do ST affect?)
9. How do you determine HOW you are going to teach on a given day? (Do ST affect?)
10. How does standardized test data affect PD decisions (what you seek out/what you provide for staff)? (e.g. geometry issue-seek out info regarding this)
11. Is there anything you would like to see your principal do to help you that he/she isn't already doing? (in evaluations, PD, dealing with parents or students)
12. What questions do you have for me?

APPENDIX E: Survey Data

Table E.1:

Amount of assistance the principal provides with using standardized test scores to improve mathematics instruction

Response	None	A little	Some	Quite a bit	A great deal	Total
Teachers	8	3	0	2	0	13
Principals	1	6	6	5	0	18

Table E.2:

Does your school district use standardized tests (e.g. ACT, SAT, MEAP, MME, MSTEP) to help measure student growth on teacher evaluations?

Teacher Response	Yes	No	Unsure	Total
	7	2	3	12

Table E.3:

Does your school use a status model or growth model when using student growth data on teacher evaluations?

Teacher Response	Status	Growth	Unsure	Total
	1	9	3	13

Table E.4:

Emphasis principals put on making sure standardized test scores are high.

Principal Response	None	A little	Some	Quite a bit	A great deal	Total
	2	4	8	1	1	16

Table E.5:

To what extent do you encourage your teachers to use standardized test results to inform their practice?

Principal Response	None	A little	Some	Quite a bit	A great deal	Total
	0	3	5	8	0	16

Table E.6:

To what extent do you believe standardized tests capture your students' knowledge of mathematics?

Teacher Response	Not at all	A little	Some	Quite a bit	A great deal	Total
	3	4	5	1	0	13

Table E.7:

How good of an indicator do you believe standardized test results are of the quality of your teaching?

Teacher Response	Extremely bad	Somewhat bad	Neither good nor bad	Somewhat good	Extremely good	Total
	4	1	5	3	0	13

Table E.8:

Since the beginning of your career, how often have you attended various professional development activities/sessions/conferences etc NOT provided by your school district?

Teacher Response	Never	1-2 times	3-5 times	6-10 times	More than 10 times	Total
	0	4	3	2	4	13

Table E.9:

Purpose of evaluation (Teacher responses)

Response	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree	Total
Improve Teaching	2	7	2	2	0	13
Accountability	8	4	0	1	0	13

Table E.10:

Purpose of evaluation (Principal responses)

Response	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree	Total
Improve Teaching	10	8	0	0	0	18
Accountability	0	7	3	4	4	18

REFERENCES

REFERENCES

- Adler, J. (2000). Conceptualising resources as a theme for teacher education. *Journal of Mathematics Teacher Education*, 3(3), 205-224.
- Allensworth, E., Nomi, T., Montgomery, N., & Lee, V. E. (2009). College preparatory curriculum for all: Academic consequences of requiring algebra and English I for ninth graders in Chicago. *Educational Evaluation and Policy Analysis*, 31(4), 367-391.
- Anagnostopoulos, D., Rutledge, S. A., & Jacobsen, R. (2013). *The infrastructure of accountability: Data use and the transformation of American education*. Cambridge: Harvard Education Press.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching what makes it special?. *Journal of Teacher Education*, 59(5), 389-407.
- Bennett, M. (2004). A review of the literature on the benefits and drawbacks of participatory action research. *First Peoples Child & Family Review: A Journal on Innovation and Best Practices in Aboriginal Child Welfare Administration, Research, Policy & Practice*, 1(1), 19-32.
- Bivona, K. N. (2002). Teacher Morale: The Impact of Teaching Experience, Workplace Conditions, and Workload.
- Cavanna, J. (2016). *Mathematics teachers' data use in practice: Considering accountability, action research, and agency*. (Doctoral Dissertation). Michigan State University, East Lansing, MI.
- Center for Educational Leadership. (2014). *5 dimensions of teaching and learning*. Retrieved from <http://www.k-12leadership.org/services/5-dimensions>.
- Chazan, D. (1996). Algebra for all students? *Journal of Mathematical Behavior*, 15, 455-477.
- Clark, P., Kirk, E., & Burriss, K. G. (2000). Review of research: All-day kindergarten. *Childhood Education*, 76(4), 228-231.
- Cohen, D. K. & Moffitt, S. L. (2009). *The ordeal of equality: Did federal regulation fix the schools?* Cambridge, MA: Harvard University Press
- Common Core State Standards Initiative. (2010). Common Core State Standards for Mathematics. Washington, DC: National Governors Association Center for Best Practices and

the Council of Chief State School Officers.

Coleman, J. S. (1966). Equality of educational opportunity.

Danielson, C. (2014). *The framework*. Retrieved from <http://danielsongroup.org/framework/>

Darling-Hammond, L. (2015). Can Value Added Add Value to Teacher Evaluation?. *Educational Researcher*, 44(2), 132-137.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.

DeSimone, J. R., & Parmar, R. S. (2006a). Issues and challenges for middle school mathematics teachers in inclusion classrooms. *School Science and Mathematics*, 106(8), 338-348.

DeSimone, J. R., & Parmar, R. S. (2006b). Middle school mathematics teachers' beliefs about inclusion of students with learning disabilities. *Learning Disabilities Research & Practice*, 21(2), 98-110.

Drake, C., & Sherin, M. G. (2006). Practicing change: Curriculum adaptation and teacher narrative in the context of mathematics education reform. *Curriculum Inquiry*, 36(2), 153-187.

Figlio, D. N., & Winicki, J. (2005). Food for thought: the effects of school accountability plans on school nutrition. *Journal of Public Economics*, 89(2), 381-394.

Fitz, J., Halpin, D., & Power, S. (1994). Implementation research and education policy: practice and prospects. *British Journal of Educational Studies*, 42(1), 53-69.

Fowler, W. J., & Walberg, H. J. (1991). School size, characteristics, and outcomes. *Educational evaluation and policy analysis*, 13(2), 189-202.

Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis*, 22(3), 241-254.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2013). Strategic Staffing: Examining the Class Assignments of Teachers and Students in Tested and Untested Grades and Subjects. In *American Education Finance and Policy Conference, New Orleans, LA*.

Granholm, J. (2003). CNN interview. Retrieved from <http://edition.cnn.com/TRANSCRIPTS/0308/18/ltn.02.html>

Guenther, A. (2019). *"How is this making my instruction better at all?: Centering teachers' voices and striving for humanization in an investigation of high-stakes evaluations"*. (Doctoral

- Dissertation). Michigan State University, East Lansing, MI.
- Hahs-Vaughn, D. L., & Lomax, R. G. (2013). *An introduction to statistical concepts*. Routledge.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance?. *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Herman, J., & Linn, R. (2013). On the Road to Assessing Deeper Learning: The Status of Smarter Balanced and PARCC Assessment Consortia. CRESST Report 823. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.
- Holloway-Libell, J., & Collins, C. (2014). VAM-Based teacher evaluation policies: Ideological foundations, policy mechanisms, and implications. *InterActions: UCLA Journal of Education and Information Studies*, 10(1).
- Honig, M. I. (2009). What works in defining “what works” in educational improvement: Lessons from educational policy implementation research. Directions for future research. *Handbook of educational policy research*. New York: Routledge, 333-347.
- Hope, W. C. (2002). Implementing educational policy: Some considerations for principals. *The Clearing House*, 76(1), 40-43.
- Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5), 761-796.
- Johnson, S. M. (2015). Will VAMS reinforce the walls of the egg-crate school?. *Educational Researcher*, 44(2), 117-126.
- Kirst, M. W. and Wirt, F. M. (2009). *The Political Dynamics of American Education*. (4th edition). New York: Teachers College Press.
- Kitchen, R., Depree, J., Celedon-Pattichis, S., & Brinkerhoff, J. (2007). *Mathematics education at highly effective schools that serve the poor: Strategies for change*. New York: Lawrence Erlbaum Associates
- Lester, F., & Lambdin, D. V. (2003). From amateur to professional: The emergence and maturation of the U.S. mathematics education research community. In G. M. A. Stanic & J. Kilpatrick (Eds.), *A history of school mathematics: Volume 2* (pp. 1629-1700). Reston, VA: NCTM.
- Loveless, T. (2008). *The misplaced math student: Lost in eighth-grade algebra*. Providence, RI: Brown Center for Education Policy.
- Maccini, P., & Gagnon, J. C. (2006). Mathematics instructional practices and assessment

- accommodations by secondary special and general educators. *Exceptional Children*, 72(2),
- Maine Education Association. (2019). *New law removes standardized test score requirement from teacher evaluation process*. Retrieved from <https://maineaa.org/news/new-law-removes-standardized-test-score-requirement-from-teacher-evaluation-process/217-234>.
- Marzano, R. (2014). *Marzano teacher evaluation*. Retrieved from <http://www.marzanoevaluation.com/>
- Matland, R. E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. *Journal of public administration research and theory*, 5(2), 145-174.
- McDonnell, L. M., & Elmore, R. F. (1987). Getting the job done: Alternative policy instruments. *Educational evaluation and policy analysis*, 9(2), 133-152.
- McLaughlin, M., Glaab, L., & Carrasco, I. H. (2014). *Implementing Common Core State Standards in California: A report from the field*. Stanford, CA: Policy Analysis for California Education.
- Michigan Council for Educator Effectiveness. (2013). *Building an improvement-focused system of educator evaluation in Michigan: Final recommendations*. Retrieved from <http://www.mcede.org/>
- Michigan Department of Education. (2006). *Michigan merit curriculum*. Retrieved from http://www.michigan.gov/documents/mde/New_MMC_one_pager_11.15.06_183755_7.pdf
- Mitchell, S. N., Reilly, R. C., & Logue, M. E. (2009). Benefits of collaborative action research for the beginning teacher. *Teaching and Teacher Education*, 25(2), 344-349.
- Morissette, M. (2014). *Using student achievement and growth data for teacher evaluations: An investigation of the implementation of Michigan PA 102 of 2011*. Unpublished manuscript, Program in Mathematics Education, Michigan State University, East Lansing, MI.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The future of children*, 113-127.
- National Research Council & National Academy of Education (2010). *Getting value out of value-added*. Braun, H., Chudowsky, N, & Koenig, J. (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Newman, J. W. (1998). *America's teachers: An introduction to education*. (3rd edition).

Addison-Wesley Longman, Inc.

Nomi, T. (2012). The unintended consequences of an algebra-for-all policy on high-skill students effects on instructional organization and students' academic outcomes. *Educational Evaluation and Policy Analysis*, 34(4), 489-505.

Porter, R. E., Fusarelli, L. D., & Fusarelli, B. C. (2015). Implementing the Common Core: How educators interpret curriculum reform. *Educational Policy*, 29(1), 111-139.

Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement.

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21-34.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 4-14.

Silver, Strong, et al (2014). *The thoughtful classroom*. Retrieved from <http://www.thoughtfulclassroom.com/>

Silver, E. A. (1997). "Algebra for All"--Increasing students' access to algebraic ideas, not just algebra courses. *Mathematics Teaching in the Middle School*, 2(4), 204-207.

Stein, M. K., Kaufman, J. H., Sherman, M., & Hillen, A. F. (2011). Algebra: A Challenge at the Crossroads of Policy and Practice. *Review of Educational Research*, 81(4), 453-492.

Spillane, J. P., Hallett, T., & Diamond, J. B. (2003). Forms of capital and the construction of leadership: Instructional leadership in urban elementary schools. *Sociology of Education*, 1-17.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387-431.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, 21(1), 1-27.

State of Michigan. (2011a). *Enrolled House bill no. 4625*. Retrieved from <http://www.legislature.mi.gov/documents/2011-2012/publicact/pdf/2011-PA-0101.pdf>

State of Michigan. (2011b). *Enrolled House bill no. 4626*. Retrieved from

<http://www.legislature.mi.gov/documents/2011-2012/publicact/pdf/2011-PA-0100.pdf>

State of Michigan. (2011c). Enrolled House bill no. 4628. Retrieved from
<http://www.legislature.mi.gov/documents/2011-2012/publicact/pdf/2011-PA-0103.pdf>

State of Michigan. (2011d). *Enrolled Senate bill no. 7*. Retrieved from
<http://www.legislature.mi.gov/documents/2011-2012/publicact/htm/2011-PA-0152.htm>

State of Michigan. (2012). *Enrolled Senate bill no. 1040*. Retrieved from
<http://www.legislature.mi.gov/documents/2011-2012/publicact/pdf/2012-PA-0300.pdf>

United States Department of Education. (2014). *No child left behind*. Retrieved from
<http://www2.ed.gov/nclb/landing.jhtml>

Vaughn, S., Bos, C. S., & Schumm, J. S. (2000). *Teaching exceptional, diverse, and at-risk students in the general education classroom* (2nd ed.). Boston: Allyn & Bacon.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *New Teacher Project*.

Winters, M. A., & Cowen, J. M. (2012). Grading New York Accountability and Student Proficiency in America's Largest School District. *Educational Evaluation and Policy Analysis*, 34(3), 313-327.