Fall 12-2016

Stress, Burnout, and Attrition: Implications of Student Performance Data on Math Teacher Effectiveness Evaluations

Becky L. Piscitella

STRESS, BURNOUT, AND ATTRITION: IMPLICATIONS OF STUDENT PERFORMANCE DATA ON MATH TEACHER EFFECTIVENESS EVALUATIONS

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Education

Becky L. Piscitella

Indiana University of Pennsylvania

December 2016

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of Professional Studies in Education

We hereby approve the dissertation of

Becky L. Piscitella

Candidate for the degree of Doctor of Education

10-13-16 Signature on file
Valeri R. Helterbran, Ed.D.
Professor of Professional Studies in Education,
Retired, Advisor

10-13-16 Signature on file
Edel Reilly, D.Ed.
Professor of Mathematics

10-13-16 Signature on file
Mark Twiest, Ph.D.
Professor of Professional Studies in Education

ACCEPTED

Signature on file
Randy L. Martin, Ph.D.
Dean
School of Graduate Studies and Research

Title: Stress, Burnout, and Attrition: Implications of Student Performance Data on Math Teacher Effectiveness Evaluations

Author: Becky L. Piscitella

Dissertation Chair: Dr. Valeri R. Helterbran

Dissertation Committee: Dr. Edel Reilly
Dr. Mark Twiest

This causal-comparative/quasi-experimental design study examines the impact that high-stakes testing stress related to the inclusion of student standardized testing data in teacher evaluations has on burnout and potential attrition rates among current Pennsylvania math teachers. Participants completed three surveys including: demographic questions, the Educators Test Stress Inventory (ETSI), and the Malasch Burnout Inventory-Educators Survey (MBI-ES). The analysis showed no real difference in levels of high-stakes testing, teacher burnout, or potential attrition between evaluated Teachers of Record (TOR) and not evaluated Non-Teachers of Record (NTOR). However, both groups reported high levels of performance pressure from administration, and burnout levels for both groups were "high" compared to the national standard. Positve correlations were found among all pairings of high-stakes teacher stress, teacher burnout, and potential attrition rates. Participant groups reporting statistically significant higher levels of high-stakes testing stress included: elementary level teachers, teachers with 11-20 years of experience, and teachers at low-income schools. Participant groups reporting statistically significant higher levels of burnout included: female teachers and those with 11-20 years of teaching experience. Those reporting high rates of potential attrition plans include the 11-20 years of experience participant group.

ACKNOWLEDGEMENTS

Thank you Lord for gifting me with a mind to reason, calculate, and create. Thank you for the patience and drive to complete a journey in faith and to fulfill a purpose for which I do not yet see the end. Thank you to my husband, Tom, and my children, Zane, Veta, and Mia, for your patience in my occassional absence on this journey. Thank you to my parents, Dave and Carol Cadwallader, for instilling the importance of education in my life. Thank you to my husband's parents, Tom, Jr. and Roxanne Piscitella, for helping out with the children in my absence.

I would like to give a special thanks to Dr. Valeri Helterbran for her guidance, attention to detail, timeliness, and frequent words of wisdom throughout this process. You are a giant among the giants, a professional woman who has paved the way for others. May we stand great upon your shoulders. I would also like to thank my committee members, Dr. Edel Reilly, Dr. Mark Twiest, and Dr. George Beiger, for their expert guidance. A special thanks also goes to Dr. Christoph Maier for quantitative research guidance as well as Dr. Lund and Dr. Lund for their online statistics research assistance.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Education reform has dominated the headlines for decades. The 1957 launch of the Soviet satellite, Sputnik, sparked a monumental shift in math and science education. Federal funding increased and scientists introduced lab experiments to the classroom (Bracey, 2007). After the release of the 1983 *Nation at Risk* report, concerns of academic underachievement persisted (National Commission on Excellence in Education). Despite fears that mass mediocrity threatened the ability of the United States to compete globally, workers remained competitive. Enacted to derail perceived mediocrity, President Regan's plan demanded more homework, more discipline, and more accountability (Elliot & Dweck, 2007). Most recently, reforming teacher accountability has become a focal point in strengthening student achievement. The *No Child Left Behind Act* (2002) began by calling for highly qualified teachers in classrooms. To that end, many states have initiated new teacher evaluations, or are in the process of creating them. Many include or are scheduled to include student standardized testing data in a teacher's overall performance evaluation for the teachers of certain tested subjects. In 2009, *Race to the Top* grants (Civic Impulse, 2014) required implementation of new teacher evaluation systems that included student performance data as a significant factor. Whereas some aspects of the teacher evaluation are under the influence of the teacher, many others are not. On the surface, it may appear that students' standardized test scores directly relate to the performance of the teacher; however, research indicates other factors beyond a teacher's control influence student achievement much more than any teacher (Darling-Hammond, 2013; Newton, Darling-Hammond, Haertel & Thomas, 2010). These factors include student backgrounds, rates of development, learning differences, language difficulties, family support, and learning disabilities

(Elliot & Dweck, 2007). Also, high stakes tests measure performance outcomes instead of target behaviors such as distinctive teaching approaches and strategies that a teacher can control (Elliot & Dweck, 2007). Because student test data can lower teachers' evaluations, consequences for the affected teachers may result in humiliation, decrease in pay, or eventually job loss. The outcome of this scenario, knowing that despite a teacher's best efforts, factors beyond the teacher's control may negatively influence student standardized test outcomes, may result in increased chronic stress. Whereby, chronic stress can lead to burnout, characterized by depression, disengagement, exhaustion, and eventually attrition, that is, leaving the field of education (Maslach, 1993; Maslach, Scaufeli & Leiter, 2001).

Many math teachers are among the specific educators whose students are tested and whose scores are included in their performance evaluation. Hence, math teachers may experience a disproportionate amount of chronic stress and, therefore, burnout and teacher attrition rates (Maslach, 1993; Maslach, Scaufeli & Leiter, 2001). Currently, 36 states have initiated new teacher evaluation systems and 30 of those states include standardized state test results as student growth measures in tested subjects and grades (Doherty & Jacobs, 2013). To that end, the practice of using standardized testing as a means of evaluating teacher effectiveness has potential to create increased teacher chronic work stress, increased teacher burnout, and increased attrition rates for affected teachers (Chang, 2009; Malasch, Schaufeli, & Leiter, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011).

Statement of the Problem

Newly reformed teacher evaluation systems create concerns for increased teacher workload and feelings of discouragement or emotional exhaustion, especially for those incorporating state standardized test scores through value-added measures to evaluate individual

teacher effectiveness (Malasch, Schaufeli, & Leiter, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011). Such evaluations generate prolonged or chronic teacher stress, leading to potential teacher burnout (Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011), and increased rates of teacher attrition (Chang, 2009). Currently, 30 states in the United States evaluate Teachers of Record (TOR), those who are accountable for standardized test scores of students, using the high-stakes test score as a piece of their teacher evaluation (Doherty & Jacobs, 2013). Evaluation differs for Non-Teachers of Record (NTOR), who are not accountable for standardized test scores of students in teacher evaluations. Recently, Pennsylvania became one of the 30 states to use high-stakes testing scores in conjunction with *PA Core Standards* to evaluate teacher effectiveness for some math teachers. Because these high-stakes tests measure performance outcomes and not teacher behaviors that are controllable, this is shown to create a large amount of chronic stress (Elliot & Dweck, 2007). TOR evaluation scores can be scored lower when student test results are added to the evaluation, potentially resulting in dismissal from teaching and subsequent loss of income (Doherty & Jacobs, 2013). This threat to only the TOR can increase teacher stress and cause a teacher to perceive themselves as being in an inequitable situation compared to the NTOR (Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011; Tye & O'Brien, 2002). Uncertainty generates feelings of discouragement, and further increases chronic teacher stress (Malasch, Schaufeli, & Leiter, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011), while also increasing the potential for greater rates of burnout and teacher attrition (Chang, 2009).

Significance of the Problem

Chronic work stress, burnout syndrome, and subsequent teacher attrition have been problematic issues in education for four decades (Steinhardt, Smith-Jaggars, Faulk & Gloria,

2011; Tye & O'Brien, 2002). New teacher evaluation reform measures are not without controversy. Some have gone so far as to include public humiliation as a criticism through published lists of "ineffective" teachers. One such publication subsequently resulted in a teacher suicide in California (Pathe & Choe, 2013).

Chronic work stress and burnout occur when a teacher experiences the inability to cope with unbearable levels of stress and demands of work (Hamman & deMayo, 1982; Kyriacou & Sutcliffe, 1978). Both negatively affect mental health and physical health, whereby suffering teachers report experiencing exhaustion, poor performance, emotional exhaustion, emotional detachment, discouragement, reduced productivity and achievement (Ahola, et al., 2005; Malasch, Schaufeli, & Leiter, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011). Chronic teacher stress, an independent predictor of burnout (Chan, 2003), can result in depression (Bakker & Schaufeli, 2000), a primary cause of increased absenteeism and high attrition rates (Hammen & deMayo, 1982; Hastings & Bahm, 2003).

Attrition of new teachers originates in teacher education programs as only 50% advance to teach in classrooms and only 54% remain after the first five years (National Council on Teacher Quality, 2003). By this calculation, only 27% of new prospective teachers remain in the field beyond year 5. Additionally, 50% of active teachers are baby boomers approaching retirement age (Carroll & Foster, 2008). Another 13% leave for reasons other than retirement, as approximately 227,016 teachers move and 230,122 teachers leave the profession (Alliance for Excellent Education, 2014). Differences in school climate, including lack of peer collaboration, lack of administrative support, and absence of shared responsibility for success among colleagues contributes to attrition among schools with similar demographics (Alliance for Excellent Education, 2014).

According to Howard (2003), retention of highly qualified teachers is imperative to school improvement efforts and attrition negatively impacts student achievement (Ingersoll & Smith, 2003). Research regarding potential factors related to additional chronic work stress, teacher burnout, and teacher attrition, specifically among math teachers may contribute to the body of work dedicated to retaining quality teachers.

Purpose of the Study

The purpose of this quantitative study was to examine the impact that stress related to the inclusion of student standardized testing data in teacher evaluation has on burnout and potential attrition rates among current Pennsylvania math teachers who identify as either Teachers of Record (TOR) or Non-Teachers of Record (NTOR). In Pennsylvania, a TOR included any public school teacher responsible for teaching or co-teaching a mathematics class in grades 4 through 8 or Algebra 1. A NTOR included any public school teacher or co-teacher in grades Pre-K through 12 who was responsible for teaching a math class in any grade except 4 through 8 or Algebra 1. Assessment years and Teacher of Record parameters were set by the Pennsylvania Department of Education (2015). Teacher assignment in each school was determined by individual school administrators and not for the purpose of this study.

This study was important for (a) identifying teacher stress levels related to high-stakes testing, (b) comparing stress levels between TOR and NTOR, (c) identifying teacher burnout levels related to high-stakes testing, (d) comparing burnout levels between TOR and NTOR, (e) comparing potential teacher attrition levels between TOR and NTOR, (f) identifying any correlation among teacher stress, teacher burnout, and teacher attrition rates, and (g) identifying any correlation of effects of other factors (gender, school location, school socio-economic status, years of teaching experience, grade level of students) with stress, burnout, and attrition. A more

intricate understanding of the relationship of teacher stress, teacher burnout, and attrition rates related to high-stakes testing may influence government policy related to the mathematics or possibly STEM (Science, Technology, Engineering, and Math) fields regarding high-stakes testing of these content areas.

Methodology

The causal-comparative/quasi-experimental design was used to conduct this study. Causal-comparative/quasi-experimental research attempts to establish cause-effect relationships among the variables. These designs are analogous to actual experiments, but with some significant differences. The experimenter identifies an independent variable but does not manipulate it. The researcher measures effects of the independent variable on the dependent variable. Groups are naturally formed or pre-existing, not random assignments to a group by the researcher. Control groups are identified and exposed to the treatment variable. The researcher studies and compares exposed groups and groups who are not (Green, Camilli, & Elmore, 2006).

A quasi-experimental design was set up since participants were assigned to groups by legislators based on their teaching assignment. Participants are grouped by subject, math, and by the teaching assignment grades K through 3 or 4 through 8, and Algebra 1 or any other math course offered to grades 9 through 12. Teaching assignment determined whether one was a Teacher of Record (TOR), directly and indirectly evaluated with student performance data, or a Non-Teacher of Record (NTOR), only indirectly evaluated with student performance data, according to the Pennsylvania Department of Education specifications (Green, Camilli, & Elmore, 2006). TOR included those teaching grades 4 through 8 and Algebra 1. NTOR included those teaching grades K through 3 and any other math course offered to grades 9 through 12. All were current, full-time public school math teachers or elementary teachers

scheduled to teach mathematics as a part of their day. All TOR and NTOR completed demographic questions, the Educator Test Stress Inventory (ETSI) (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015), and The Maslach Burnout Inventory Educators Survey (MBI-ES) (Malasch & Jackson, 1986; Maslach, Jackson & Leiter, 1996; Malasch, Jackson & Schwab, 1986). All surveys were distributed to math teachers via the Pennsylvania State Education Association (PSEA) listserv of public school teachers with permission and administered via Qualtrics.

Definition of Terms

In this study, *chronic teacher stress* defines the complex cognitive process whereby a teacher is unable to cope with work demands for a prolonged period, compounded with frequent feelings of failure and disappointment, creating emotional exhaustion and apathy (Kyriacou & Sutcliffe, 1978). Teacher stressors include, but are not limited to role overload, inability to complete work load, lack of administrative and parent support, disruptive students, poor relationships with colleagues, and being evaluated with high stakes tests (Kyriacou, 2001; Montgomery & Rupp, 2005).

In this study, *teacher burnout*, comprised of three components: emotional exhaustion, depersonalization, and reduced personal accomplishment. It is defined as a psychological syndrome (Maslach, 1993; Maslach, Scaufeli & Leiter, 2001) generated by prolonged exposure to intolerable work situations, and emotional and interpersonal stressors, often without sufficient recovery time, leading to teacher disengagement in work (Cherniss, 1980; Hakanen, Bakker, & Schaufeli, 2006; Maslach & Leiter, 2008).

In this study, *Teacher of Record (TOR)* describes public school teachers who are individually accountable for state standardized test scores of students they teach by grade or by

subject as determined by the Pennsylvania Department of Education (2015) or any state department of education. These standardized test results are used as a part of the individual teacher evaluations, whereby each state determines the exact percentage. In Pennsylvania, the high-stakes testing results account for 15% of the individual evaluation and overall school results account for 15% of all teachers' evaluations. Therefore, high-stakes testing results account for a total of 30% of the evaluation for a teacher of record. In this study the TOR includes any PA teacher responsible for teaching or co-teaching a mathematics class in grades 4 through 8 or Algebra 1. Students taking Algebra 1 may take the course as late as grade 11 and still test on the state Keystone Algebra 1 exam.

In this study, *Non-Teacher of Record (NTOR)* describes public school teachers who are not individually accountable for state standardized test scores of students they teach by grade or by subject as determined by the Pennsylvania Department of Education (2015) or any state department of education. These standardized test results are not used as a part of the individual teacher evaluations. In PA the NTOR would only have 15% of comprehensive school high-stakes testing results factored into their evaluation. In this study the NTOR refers to PA teachers or co-teachers in grades K through 12 who are responsible for teaching a math class in any grade except 4 through 8 or Algebra 1. Although these teachers do not get an individual score, they are indirectly responsible for the school score, or 15% that all teachers in the district receive on their evaluation. NTOR are responsible for teaching math courses before grade 4, and after Algebra 1 to students who have not passed the Algebra 1 test. Teachers after Algebra 1 may also teach remediation courses to failing students who must pass the Algebra 1 exam by the end of grade 11.

Given that the word potential means existing in possibility, this study defines *potential teacher attrition* as the possibility that a teacher may leave the field of education. An attempt is made to measure the level of potential or possibility that a teacher may leave the field of education.

Assumptions

This study included the following assumptions. The first assumption was that teacher stress and burnout will continue to be issues as the teaching profession continues to be more demanding. The continuance is partially resulting from increased external performance pressure and controls due to high-stakes tests and new teacher evaluations (Doherty & Jacobs, 2013; Malasch, Schaufeli, & Leiter, 2001; Ryan & Deci, 2000; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011). The second assumption was that survey questions were answered correctly by the participants. Accuracy of data is assumed, however, participants may have incorrectly responded to questions due to misunderstanding the questions or intentionally misrepresenting the experience. A third assumption is that the findings in this study are more relevant to teachers in public schools. Public school teachers participate in state teacher evaluations that may include standardized test scores as a part of the evaluation.

Delimitations

Delimitations are decisions made by the researcher about which items will not be researched (Leedy & Ormrod, 2010). The study was restricted to public school teachers in the state of Pennsylvania, whereby private school teachers and public school teachers in other states were excluded. The study was further confined to K-12 teachers. An additional delimitation included the use of closed-ended survey items, thereby limiting responses of participants.

Limitations

Limitations are weaknesses or potential problems that exist within a study (Creswell, 2008) that may, however, assist in further replications of studies as the additional need for research will exist. This study included the following four limitations. First, the public school teachers in the sample were voluntary participants and could withdraw at any time during the study, which could affect the sample size and statistical validity. Another limitation is that findings may not be generalized to teachers in other countries, particularly those under different governing laws and regulations. However, the participants were a representative sample and findings may be generalized to the public school teaching population of the United States. An additional limitation was the method of data collection, as quantitative data, although objective, does not contribute in-depth explanations of emotions or responses. A final limitation was that the study relied on self-reported data.

Theoretical Framework

Motivation from the perspective of Self-Determination Theory (SDT) establishes the theoretical framework for this study. This contemporary theory, conceptualized by Deci and Ryan (1985), addresses issues of motivation for both students and teachers. In the case of high-stakes testing, a teacher's emotional well-being is affected by external controls that influence both student and teacher motivation.

SDT suggests that motivation can vary between controlled motivations (external) and autonomous motivations (intrinsic). Controlled motivation, in the form of external rewards or punishments, associates with negative emotions, decreased persistence, and decreased quality of learning. Autonomous motivation, defined as well-internalized values or intrinsic motivation,

engenders increased persistence, increased quality of learning, and a positive emotional experience (Grolnick & Ryan, 1987; Ryan & La Guardia, 1999; Ryan, Stiller, & Lynch, 1994).

Self-determination theory suggests highly negative outcomes for students following attempts to implement external controls to improve their achievement (Deci & Ryan 1985; Ryan & Deci, 2000). Currently, high-stakes test results determine teacher effectiveness. In turn, teachers are inadvertently focusing student attention on this external control. Concerns of unintended consequences include decrease in persistence by the students, decrease in quality learning, and decreased ability to challenge students of different achievement levels (Deci & Ryan, 2002). Students with disabilities create individual concern as all students are expected to reach benchmarks simultaneously despite learning differences, backgrounds, and developmental rates (Deci & Ryan 1985; Ryan & Deci, 2000).

A study conducted by Deci, Spiegel, Ryan, Koestner, and Kauffman (1982) pressured teachers to produce high achieving students. Teachers became more controlling and less tolerant of student autonomy. Teachers delivered more lecture based lessons, more criticism, and a misuse of praise. Each of these practices links to an adverse impact on learning. A follow-up study by Flink, Boggiano, and Barrett (1990) introduced new elementary curriculum to a variety of schools. As predicted, in line with SDT, the more the teachers were pressured to perform toward higher standards, the more controlling their instructional behaviors became. In return, their students scored more poorly on objective based tests than teachers who were not pressured.

SDT also considers self-esteem related threats to external motivation. Because high-stakes tests require all students to reach the same benchmark at the same time, it again diminishes internal motivation (Elliot & Dweck, 2007). People are optimally challenged, or most intrinsically motivated, when tasks are within reach (Elliot & Dweck, 2007). If a student is

not confident or unsure of success, effort may be withdrawn to protect self-esteem. It may be easier for a student to say "I did not try" instead of "I could not pass" in the presence of peers. Failure does not motivate people, teachers included. To that end, other negative practices have ensued. Curricula are narrowed to tested items, more controlling classroom techniques are utilized, and an increasingly negative experience occurs daily for teachers and students (Elliot & Dweck, 2007). In the perspective of SDT, teachers, under pressure to increase student achievement, may be unknowingly controlling students, thereby negatively impacting their learning and ultimately their achievement performance. Under the circumstances, teachers may unintentionally create a situation likely to produce more high-stakes testing stress for themselves.

SDT also suggests that controlled or external motivations can thwart teacher motivation. For teachers, social and cultural work-related factors either facilitate or undermine initiative (Deci & Ryan, 2002). When conditions support autonomy, competence, and relatedness, teachers experience the highest quality motivational environment and sense of emotional well-being (Deci & Ryan, 2002). Teaching performance, persistence, and creativity enhance. However, if any of the three conditions are unsupported for a teacher, a detrimental impact to well-being ensues in that environment (Deci & Ryan, 2002).

Autonomy reflects the need to feel a sense of self and responsibility for one's behavior. Competence reflects the need to interact and succeed at optimally challenging tasks. Relatedness concerns the degree to which an individual socially integrates and feels accepted (Deci & Ryan, 1985). Feeling external pressure toward an action or outcome, such as pressure to produce student achievement on high-stakes tests, tends to undermine autonomy (Deci & Ryan, 1985). A lack of autonomy or control has been a consistent predictor of work related stress and burnout (Skaalvik & Skaalvik, 2009). A recent study identified the perceived thwarting of autonomy,

competence, and relatedness as separate and significant indicators of teacher burnout (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014). In this high-stakes testing environment, teachers felt pressure to teach a certain way, thereby decreasing autonomy. Repeatedly confronted with evaluations based on student performance data, teachers reported feeling inadequate and incompetent (diminished competence). Some also experienced feelings of rejection by peers and colleagues. Another study found encroachment on professional autonomy further dismantled professional competence, and ultimately led to teacher attrition (Bouwma-Gearhart, 2010).

This study adds to the growing body of work relating teacher stress, burnout and attrition related to student performance data on teacher effectiveness evaluations through the lens of SDT. Some studies reveal increased external pressure on teachers to improve student performance diminishes autonomy for students. In turn, these students perform more poorly, potentially generating more teacher stress (Elliot & Dweck, 2007). Several studies link the demands of high-stakes testing to loss of autonomy, reduced competence, and diminished relatedness in light of repeated performance evaluations (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Fernet, Guay, Senecal & Austin, 2012; Skaalvik & Skaalvik, 2009). This thwarting of autonomy, competence, and relatedness further correlates with increased teacher stress, burnout, and attrition (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Bouwma-Gearhart, 2010; Skaalvik & Skaalvik, 2009). This study contributes to the research by (a) identifying differences in teacher stress levels between TOR (more external controls) and NTOR (less external controls) regarding high-stakes testing results and the new teacher evaluation system, (b) identifying differences in levels of teacher burnout between TOR (more external controls) and NTOR (less external controls) regarding high-stakes testing and the new teacher evaluation

system, (c) identifying differences in levels of potential teacher attrition between TOR (more external controls) and NTOR (less external controls) regarding high-stakes testing and the new teacher evaluation system, and (d) identifying correlations among high-stakes testing teacher stress, teacher burnout, and potential teacher attrition rates.

Questions to be Researched

Three research questions were developed to guide the research. Chapter three contains seven hypotheses to be tested.

Research Question 1

Does high-stakes testing as a component of teacher evaluation lead to higher levels of stress, leading to burnout, and eventually attrition.

Research Question 2

Does high-stakes testing as a component of teacher evaluation lead to a correlation among stress, burnout, and attrition?

Research Question 3

Are other factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) effecting stress, burnout, and teacher attrition?

Summary

Education reform transformed over the last half century into the growing trend of accountability in the form of teacher effectiveness calculated through high-stakes testing results. However, these tests measure performance outcomes, not behavior. Students' achievement is shown to be influenced by many factors independent of the teacher and more than the teacher (Darling-Hammond, 2013; Newton, Darling-Hammond, Haertel & Thomas, 2010). 30 states, including Pennsylvania, include high-stakes testing data for a teacher's assigned students as a

piece of the individual's evaluation if they are a teacher of record (Doherty & Jacobs, 2013). This high-stakes testing tied to evaluation has potential to create large amounts of chronic stress (Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011; Tye & O'Brien, 2002). In Pennsylvania, math teachers are particularly vulnerable as almost all in grades 4 through 9 and some in grades 10 and 11 are teachers of record (Pennsylvania Department of Education, 2015). Chronic teacher stress can lead to teacher burnout and eventually teacher attrition (Maslach, 1993; Maslach, Scaufeli & Leiter, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011; Tye & O'Brien, 2002), creating a potential math teacher crisis in PA. Supplementary detail will follow in chapter two in the review of the literature. Chapter three details procedures and chapter four describes data and analysis. Chapter five closes with conclusions and recommendations.

CHAPTER 2

REVIEW OF RELATED LITERATURE

In this review of the literature, education reform was explored from three perspectives. National education reform policies and practices were reviewed, as well as education reform in the individual states. Education reform policies for the state of Pennsylvania, where the research was conducted, included specific detail about teacher effectiveness evaluations. The review also included a pilot teacher evaluation, and information about chronic teacher stress, teacher burnout, teacher attrition, and Self-Determination Theory. The chapter concludes with a summary.

Education Reform in the Nation

The history of educational accountability in the United States reveals over a century of external evaluation. School accreditation in the early 20th century consisted of voluntary compliance by school leaders to enact standard resources and processes (King & Rohmer-Hirt, 2011; Nevo, 2001). National goals were first set in 1918 in the *Cardinal Principles of Secondary Education*, followed by *What High Schools Ought to Teach* in 1940 (King & Rohmer-Hirt, 2011; Nevo, 2001). Neither involved commitment or mandated measuring to determine effectiveness. As standardized tests became available, some schools used them at key pre-collegiate points to self-assess district effectiveness (King & Rohmer-Hirt, 2011).

Program evaluation initiated in the 1960s with the federal antipoverty and education programs such as *Head Start* and *Follow Through*. Policy mandated extensive program evaluation accompany federal funding for these and similar federal programs (Patton 2008). In 1965, the *Elementary and Secondary Education Act* (ESEA) included evaluation requirements, a significant factor leading to program evaluation as a field (King & Rohmer-Hirt, 2011;

Fitzpatrick, Sanders, & Worthen, 2011). Almost simultaneously, the National Assessment of Educational Progress (NAEP) launched in 1964 and 1969 began annual assessments of American students in various subjects (NAEP, 2013). These assessments and reports continue annually.

After the release of A *Nation at Risk* in 1983 by the National Commission on Excellence in Education, fear of mediocrity was followed by dramatic growth in in educational accountability in the United States (McNamara & O'Hara, 2008). *The Goals 2000: Educate America Act* of 1994 added clear consequences for schools and states (King & Rohmer-Hirt, 2011). In 2002, the ESEA, renamed *No Child Left Behind* (NCLB), called greater public attention to high-stakes testing results in schools. NCLB also required testing of all students in grades 3 through 8 and selected high school grades, required inclusion of special needs students and English language learners, and required assessment of subgroups within the population for Adequate Yearly Progress (AYP). For the first time, school districts, accountable for failing results, would progress through a chain of severe consequences to eventually be taken over and reconfigured (King & Rohmer-Hirt, 2011).

In 2009, federal money through *Race to the Top* grants (Civic Impulse, 2014) now required implementation of new teacher evaluation systems that included student performance data as a significant factor. In 2010, the Obama administration released *A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act*. States must have "college-ready and career-ready" ("CCR") standards and assess students for growth measures to continue receiving Title I funding (Kress, Zechmann & Scmitten, 2011). Additional grants are proposed to help states develop higher order thinking questions on assessments instead of multiple choice, create tests for other subjects, and assess whether existing tests in reading and math are at the college-ready or career-ready level (Kress, Zechmann & Scmitten, 2011). For the

lowest 5% performing schools in any state, three options involve closing the school. Of these, either a new principal will be hired who can hire back up to half of the previous teachers, the school will reopen under new leadership with all new teachers, or students will be sent to a better school. A fourth option replaces the principal, maintains the teachers, implements professional development, extends learning time, and launches other recovery strategies (Kress, Zechmann & Scmitten, 2011). On December 10, 2015 President Obama signed *Every Student Succeeds Act* (ESSA) maintaining the plan for support, funding, and intervention of the lowest 5% performing schools (U.S. Department of Education, 2015). This bipartisan bill intended to fix NCLB also includes funding for quality pre-k and protections for high-needs students. The bill introduces a competitive program for innovation and evidence building, replication of high-quality charter schools, and encouraging support systems for vulnerable communities (U.S. Department of Education, 2015). The bill maintains inclusion of high academic standards for all, college and career ready standards, and continued annual statewide assessments with the promise of decreased disruption of class time for testing (U.S. Department of Education, 2015).

Education Reform in the States

Different states use different approaches to teacher evaluations. Teachers, whose students are not subject to standardized testing, can be evaluated for student achievement by formative assessments, Student Learning Objectives (SLOs), national assessments like the ACT and SAT, student surveys, classroom observations, teacher self-assessments, lesson plan reviews, and school-wide value added measures (Hull, 2013). Evaluations for teachers whose students are subject to standardized tests can include test scores as well (Hull, 2013). The Measures of Effective Teaching study recommends using a weighted and balanced formula of student achievement (SA), classroom observations (CO), and student surveys (SS) (MET, 2013). MET

researchers argue either a SA/CO/SS ratio of 50%:25%:25% or 33%:33%:33% will prevent teachers from over emphasizing any one area and still represent an accurate measure of effective teaching. Baker, Oluwole, and Green (2013) caution that weighted values and numeric importance should be field tested first, and be assigned by quality of fair account of performance.

Teacher evaluation systems throughout the states vary in structure. The District of Columbia and 11 states, including Pennsylvania, operate under a single teacher evaluation system (Doherty & Jacobs, 2013). A presumptive state model exists for ten states, whereby the state provides a model that the districts may decline by being approved to use a comparable evaluation system (Doherty & Jacobs, 2013). The other 27 states with evaluation systems provide guidelines or criteria that districts can use to design their systems and still adhere to state policies. Additionally, 11 of these states also provide a state model for districts to use as an option (Doherty & Jacobs, 2013). Only two states lack a statewide policy for teacher evaluation (Doherty & Jacobs, 2013).

Statistical methods for correlating student performance data to teacher performance vary. These methods can be described as either a value-added model or a student growth percentile (Hull, 2013). Value-added models (VAM) attempt to isolate the impact a teacher has on student achievement, independent of outside factors (Hull, 2013). Student growth percentiles (SGP) compare student academic progress to other students (Hull, 2013). The VAM recognizes that student growth varies from year to year and from student to student (Hull, 2013). This model attempts to isolate individual teacher effectiveness on student growth (Baker, Oluwole & Green, 2013) and more accurately does so than SGP (Ehlert, Koedel, Parsons & Podgursky, 2013). However, VAMs require multiple years of data and complex statistical tools to calculate, and advanced knowledge of statistics to interpret (Hull, 2013). Research also shows VAMs have

limitations and fail to provide reliable and accurate measures of teacher effectiveness based on student performance data, opposite of what researchers had hoped (Darling-Hammond, 2013). Teachers who clearly enable student learning or whose students are demonstrating notable success are receiving low or varying VAM ratings (Sentell, 2012; Watanabe, 2012). Standards based exams only measure grade-level standards and are less valid for special education students, English language learners, and new immigrants. Additionally, VAMs also assume that students are assigned randomly to teachers and tests measure a complete range of achievement, not containing a low ceiling and high floor (Darling-Hammond, 2013; Sentell, 2012; Watanabe, 2012). Statistically speaking, the same teacher will appear more effective when teaching gifted students than special education students and English language learners (Amrein-Beardsley & Collins, 2012; Newton, Darling-Hammond, Haertel & Thomas, 2010). One study found teachers who ranked in the top quartile based on student performance data from state tests, ranked in the lowest 50% based on higher-order thinking standardized tests, and vice versa (Bill & Melinda Gates Foundation, 2010; Rothstein, 2011). Another study found no strong correlation between teacher's effects on high-stakes tests and low-stakes tests, and teacher's effects on high-stakes exams diminish more quickly (Corcoran, Jennings & Beveridge, 2011). Research also shows highly inaccurate ratings, specifically when changing exams, changing content of exams, or changing quality of exam questions (Bill & Melinda Gates Foundation, 2010; Corcoran, Jennings & Beveridge, 2011; Darling-Hammond, 2013; Rothstein, 2011; Sentell, 2012; Watanabe, 2012). In turn, VAMs can produce largely inaccurate rating scores for teachers (Darling-Hammond, 2013; Sentell, 2012; Watanabe, 2012). Cases of highly respected teachers in New York and Houston have been dismissed after switching from teaching high-achieving students one year to

newly mainstreamed students the following year (Newton, Darling-Hammond, Haertel & Thomas, 2010).

SGPs, easier to calculate and understand, measure the growth of an individual student from one year to the next (Hull, 2013). SGPs tend to be more popular because they are cost effective and more accurate at evaluating teachers than student test scores (Hull, 2013). Conversely, SGPs fail to be as accurate at isolating individual teacher effectiveness on student growth (Baker, Oluwole & Green, 2013) and fail to distinguish growth differences among students (Hull, 2013). Currently, 12 states, including Pennsylvania, use value-added models and 17 states use student growth percentiles (Hull, 2013; MET, 2013).

Student Learning Objectives (SLO), written by teachers for their individual courses and students, are gaining popularity for evaluating teachers whose students are not subject to standardized testing (Hull, 2013). Evaluation by SLO includes a teacher's professional input, reflection on individual teaching practices, and evaluation of student progress (Hull, 2013). SLO evaluation can be applied to TOR and NTOR. The process of evaluation by SLO is not standardized and their effect on evaluations in the long term is unknown (Baker, Oluwole & Green, 2013; Hull, 2013; MET, 2013). SLOs also create concerns of teachers setting low goals for some students and create comparison teacher problems throughout districts and states (Baker, Oluwole & Green, 2013; Hull, 2013). Twenty states currently require or allow SLOs to be a part of the teacher evaluation (Doherty & Jacobs, 2013).

Classroom observation evaluations provide a rich measure of instructional practice and teaching feedback (Hull, 2013). The principal maintains responsibility for classroom observations in most states, despite lack of time and proper training. Research shows classroom evaluations are most effective when conducted by multiple trained individuals, multiple times a

year, using new research-based protocols (Baker, Oluwole & Green, 2013; Hull, 2013; MET, 2013). Multiple classroom observations are required in 25 states, and only 15 states specify that some classroom observations must be unannounced (Doherty & Jacobs, 2013). Feedback to teachers concerning evaluations is required in 22 states (Doherty & Jacobs, 2013). Doherty and Jacobs report 44 states and the District of Columbia include classroom observations as a part of their teacher effectiveness evaluation (2013) and only 8 of these states and the District of Columbia use the evaluation to grant teacher tenure. Pennsylvania requires classroom observations in the teacher evaluation as well as mandatory feedback (Pennsylvania Department of Education, 2015). Researchers found 41% of teachers across the nation thought their classroom evaluation was "just a formality," 32% thought it was "well-intentioned but not particularly helpful," and only 26% thought it was "useful and effective" (Darling-Hammond, 2013; Duffet, Farkas, Rothertham & Silva, 2008). As for the granting of tenure, 69% of teachers believe it is a formality, not based on whether one displays good teaching skills (Darling-Hammond, 2013; Duffet, Farkas, Rothertham & Silva, 2008).

Some states are now using performance assessments or portfolios to predict teacher effectiveness (Darling-Hammond, 2013). Researchers found the BEST portfolio used for 2nd year teachers in Connecticut was the only qualification among several to predict teacher effectiveness (Darling-Hammond, 2013; Pecheone & Chung, 2006). This portfolio required teachers to analyze their work and their students' work, and enabled teachers to make adjustments as necessary. The self-assessment and reflection aspect of the process helped teachers develop effectiveness (Darling-Hammond, 2013; Pecheone & Chung, 2006; Pecheone & Stansbury, 1996). Universities also find this practice useful for helping teaching candidates prepare and develop (Darling-Hammond, 2013; Hanby, 2011).

Some of the most ambitious teacher evaluation plans, present in 19 states and the District of Columbia, include professional development based on evaluation outcomes (Doherty & Jacobs, 2013). Professional development offered less than 14 hours a year on specific content had no effect on student performance. However, student achievement increased by 21 percentile points after teachers participated in about 50 hours of quality professional development over a span of 6 to 12 months (Wei, Darling-Hammond, Andree, Richardson & Orphanos, 2009). Research shows that almost all teachers participate in professional development, yet teachers rarely concentrate on any topic for more than two days (Darling-Hammond & Richardson, 2009). Additionally, when schools create time and productive working relationships among teachers, research shows an increase in willingness to share strategies and try new things, greater uniformity in instruction, and a more successful practice of problem solving (Darling-Hammond, 2013). The best systems in Europe and Asia include common time for teachers to collaborate, while most schools in the United States provide little or no time for collaboration (Darling-Hammond, 2013).

Other areas addressed by teacher evaluations include ineffectiveness, layoffs, teacher compensation, and teacher preparation. In 25 states and the District of Columbia, ineffective teachers are required to participate in an improvement plan (Doherty & Jacobs, 2013). Of those, 22 states and the District of Columbia may dismiss teachers after repeated ineffective evaluations. Florida, Hawaii, Indiana, Louisiana, Utah, and the District of Columbia have policies that link teacher compensation to teacher evaluations and student performance data (Doherty & Jacobs, 2013, O'Shaughnessy, 2013). As of 2013, 14 states and the District of Columbia include layoff contingences in teacher evaluation policies instead of tenure or years of service (Doherty & Jacobs, 2013). A small group of 8 states connect student performance data to

teacher evaluations and the institutions where those teachers were trained, placing an evaluation rating on the college or university linked to each evaluated teacher (Doherty & Jacobs, 2013). Pennsylvania is one of 8 states that report teacher effectiveness data by school (Doherty & Jacobs, 2013).

Tenure and licensure are also included in some teacher effectiveness policies. Evaluation ratings factor into tenure decisions for 18 states and the District of Columbia (Doherty & Jacobs, 2013). Evaluations determine teaching licensure advancement in 8 states (Doherty & Jacobs, 2013). Delaware is the only state with a policy for license reciprocity that is, transferring a teaching license from state to state (Doherty & Jacobs, 2013). Delaware requires three years of "successful" teaching in another state, including two satisfactory evaluations similar to the system set up for teachers in Delaware.

By the 2014-2015 school year, 48 states and the District of Columbia approved College and Career Ready (CCR) standards, and 46 states and the District of Columbia administered CCR assessments in line with federal regulations (Education Commission of the States, 2014). Nebraska adopted partial CCR standards and Oklahoma has not yet adopted CCR standards. Massachusetts has administered partial CCR assessments and Iowa, Nebraska, and Oklahoma have not begun CCR assessments (Education Commission of the States, 2014).

In May of 2015, U.S. News & World Reports released a list of the best ranked schools in the United States. Schools were given gold or silver medal status based on courses offered and not CCR standards. Maryland ranked first with 28.9% of its schools earning a gold or silver medal (Morse, 2015). California ranked second with 27.2% earning medals, followed by Connecticut with 25.4%. Pennsylvania ranked 23rd with 11.5% of schools earning gold or silver

medals (Morse, 2015). Massachusetts ranked 4th, however, only partial CCR assessments were administered in the state (Education Commission of the States, 2014; Morse, 2015).

Education Reform in Pennsylvania

A year after the release of A Nation at Risk (1983), Pennsylvania implemented the Testing for Essential Learning and Literacy Skills (TELLS) test, the state's first mandated student competency exam in the 1984-85 school year (Pennsylvania State Education Association, 2015). In 1992 the Pennsylvania System of Schools Assessment (PSSA) officially began, initiating the first version of the current standardized test used (Pennsylvania State Education Association, 2015). The PA Chapter 4 Academic Standards and Assessments became effective in January 1999. Almost a decade later, the National Governors Association and the Council of Chief State School Officers initiated development of a core set of math and English academic standards. The following year, in June of 2010, the Common Core standards were finalized for math and English language arts (Pennsylvania State Education Association, 2015). From the inception of the PSSAs in 1992, the exams have adapted to reflect changes at the federal level. In 2013, Keystone Exams replaced the grade 11 PSSA with subject specific tests in Algebra 1, Biology, and Literature to continue to comply with federal regulations (Pennsylvania State Education Association, 2015).

According to the Pennsylvania State Education Association (PDE), Act 48 of 1999 required all licensed teachers in PA to obtain 180 continuing education credits or six credits of collegiate study every five years beginning July 1, 2000 to maintain licensure (2015). The goal was to establish expectations, increase the quality of professional development, and improve accountability. Following the NCLB in 2002, public attention increased to published high-stakes testing results in schools. Teacher accountability increased for the first time as school districts,

accountable for failing results, would progress through a chain of serious consequences to eventually be taken over and reconfigured (NCLB, 2002). In 2009, new teacher evaluation systems, that included student performance data as a significant factor, became linked to federal money through Race to the Top grants (Civic Impulse, 2014). In response, Pennsylvania passed Act 82 on June 30, 2012 enacting the Educator Effectiveness System in PA (PDE, 2015). After the 2010 release of *A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act*, PDE aligned the PA Core Standards and high school level Keystone Exams to meet "college-ready and career-ready" ("CCR") standards. Also, all students are now assessed for growth measures (Kress, Zechmann & Scmitten, 2011; PDE, 2015) to continue receiving Title I funding. All Keystone Exams and PSSAs in the school year 2014-2015 were designed to be more rigorous, require higher level thinking skills, and included higher cut scores for passing (PDE, 2015).

The Pennsylvania Teacher Effectiveness System of Act 82 (PDE, 2015) identifies four categories for Teachers of Record (TOR) and three categories for Non-teachers of Record (NTOR). The three categories common to both include Teacher Observation/ Practice, Building Level Data, and Elective Data/ Student Learning Objective (SLO). TORs also have a section of Teacher Specific Data, also known as the student performance data category. Teacher Observation/ Practice consists of 50% and Building Level Data/ Schools Performance Profile consists of 15% of the overall evaluation for TORs and NTORs. However, 15% of a TOR's evaluation comes from Teacher Specific Data and 20% comes from Elective Data/ SLO, while 35% of the NTOR's evaluation comes from the Elective Data/ SLO only. A teacher will receive an overall evaluation rating of distinguished, proficient, needs improvement, or failing (Doherty & Jacobs, 2013). Teacher evaluations in Pennsylvania are linked to licensure advancement,

public reporting of aggregate teacher data, improvement plans, and teacher dismissal (Doherty & Jacobs, 2013).

The Teacher Observation/ Practice category (50%) follows the Danielson Framework teaching model, although the Marzano Teacher Evaluation Model is an acceptable substitute (PDE, 2015). A teacher submits a plan to the evaluator in advance. The evaluator observes the lesson, provides feedback at a future meeting, and the teacher has an opportunity to respond to the evaluator's remarks. The teacher is scored on Planning & Preparation (20%), Classroom Environment (30%), Instruction (30%), and Professional Responsibilities (20%). Evaluators in PA are offered training, however, they are not required to be effective teachers or obtain an evaluators certificate (Doherty & Jacobs, 2013). Multiple evaluators or observers are also not required, although research shows classroom evaluations are most effective when conducted by multiple trained individuals, multiple times a year, using new research-based protocols (Baker, Oluwole & Green, 2013; Doherty & Jacobs, 2013; Hull, 2013; MET, 2013). Classroom observations are required once a year, announced, and feedback must be given to complete the PDE online formal process (PDE, 2015).

Elective Data/Student Learning Objectives (SLO) represent 35% of the effective teacher evaluation for a NTOR and 20% for a TOR. The SLO process measures teacher effectiveness based on student achievement of content standards (PDE, 2015). Each SLO specifies individual teacher standards-based objectives that clearly define measureable goals to define student achievement. SLOs are approved by the district administration, placed into action, measured or assessed as specified, and reported electronically through PDE. The administrator's comments are attached and the teacher has an opportunity to respond before final electronic submission to PDE. The SLO process is not funded and offers little oversight (PDE, 2015). For teachers who

have failed to meet SLO objectives with real student data, this self-reporting process creates opportunity to falsify student data to satisfy the SLO objective. For example, a teacher could manipulate data by grading post-tests easy, changing scores, destroying original student work, or manipulating a few scores without anyone else knowing, thereby fraudulently maintaining the status of effective.

Teacher Specific Data represents 15% of the evaluation for TORs only, however, it directly affects Building Level Data/ Schools Performance Profile that represents 15% of the evaluation for both. To understand the significance, one must understand the testing policies and procedures specific to Pennsylvania. The following information is online and available through the Pennsylvania Department of Education (2015). All Pennsylvania students in grades 3 through 8 are tested in English language arts and math, and all students in grades 4 and 8 are tested in science. Keystone Exams in Algebra 1, Biology, and Literature are given at the end of the respective course as an end of course (EOC) exam. If a student does not score advanced or proficient on any Keystone exam, the student must retest on that exam until a passing score is achieved or until the student becomes a senior. Beginning with the graduating class of 2017, every student must pass 2 of the 3 Keystone Exams with an advanced or proficient score or complete a project as outlined by the state under direct supervision of a teacher to graduate (PDE, 2015). A project will not count as a passing score for a Teacher of Record.

These high-stakes tests account for 15% of the teacher effectiveness evaluation of the individual Teacher of Record (TOR) and influence the overall Building Level Data/ School Performance Profile results that account for 15% of all teachers' evaluations in a district. Therefore, high-stakes testing accounts for a total of 30% of the evaluation for a TOR. Conversely, high-stakes testing only accounts for 15% of the Building Level Data/ School

Performance Profile of the evaluation for a NTOR. The TOR includes any PA teacher responsible for teaching or co-teaching a tested class. Classes include English language arts and math in grades 4 through 8, science in grades 4 and 8, Algebra 1, Biology, or Literature. Students taking Algebra 1, Biology, and Literature may take the course as late as grade 11 and still test on the Keystone state exam. Students may retest until the winter of grade 12 to attempt to pass any Keystone Exam or complete a subject project as outlined by PDE (2015). However, passing Keystone exam scores only apply to teacher effectiveness evaluations the first time the test is taken. Passing Keystone exam scores and passing project scores only apply to the School Performance Profile when obtained by the last testing opportunity in the spring of grade 11. Passing scores obtained in grade 12 only count for graduation (delayed until the graduating class of 2019) as determined by PDE (2015).

A TOR must verify their roster of students online with the Pennsylvania Department of Education (2015). This process allows the teacher to verify which students' high-stakes test scores will count toward their teacher effectiveness evaluation and what percent of each student's instruction they are accountable for on the evaluation (Pennsylvania Department of Education, 2015). For example, a pair of co-teachers may claim 50% responsibility for each student in their class. A teacher must verify at least 11 students on a roster to receive a report or an "actual count" greater than 11 (Pennsylvania Department of Education, 2015). A teacher also needs an "active count" greater than 6. The active count is adding more than one student to get a total percentage of 100%, whereby, each 100% counts as an active student (Pennsylvania Department of Education, 2015). For example, if student A is taught 30% of the time and student B is taught 70% of the time by the TOR, then students A and B count as 1 active student. A student's score will not count in a teacher's evaluation if the instructional responsibility is less

than 10%, the student did not take the test, the student is a foreign exchange student, the student took the PASA (alternate assessment), the student is a first year English Language Learner, or the student did not have a sufficient testing history (Pennsylvania Department of Education, 2015). However, a student, who moves to a district a few days before testing begins, will be assigned up to 100% to the new Teacher of Record despite instructing the student for only a few days.

In Pennsylvania, student scores are not taken explicitly as pass or fail for a Teacher of Record. Instead a value-added model, the Pennsylvania Value Added Assessment System (PVAAS), charts individual student achievement and growth (Pennsylvania Department of Education, 2015). Achievement, a measure of performance at a certain point in time, differs from growth, a measure of progress across time in years or grades. For this reason, grade 3 teachers in English language arts and math are Not Teachers of Record, as this is a baseline year to obtain a record of achievement. PVAAS contends that a child is either gaining, maintaining, or falling behind. Students in Pennsylvania are expected to maintain or make gains academically each year (Pennsylvania Department of Education, 2015). Growth Standard Methodology calculates projected annual growth for students taking assessments in English language arts and math for grades 4 through 8. Predictive Methodology calculates projected growth for students taking grade 4 and 8 science tests and end of year Keystone Exams (Pennsylvania Department of Education, 2015). The PVAAS summary reports a calculated average of student performance data to determine whether a teacher's students met or exceeded the standard for academic growth.

In April of 2015, the Pennsylvania Department of Education (PDE) released an announcement titled *PVAAS and the Transition of PA's Assessment System*. The document states

that despite test changes, such as becoming aligned to PA Core Standards and increased rigor and difficulty, the PVAAS system is flexible enough to be fair. PDE claims PVAAS "will continue to provide fair, accurate, and meaningful value-added measures for PA districts, schools, and teachers," although research shows otherwise. Disrepancies ocurr when changing exams, changing content of exams, and changing quality of exam questions (Bill & Melinda Gates Foundation, 2010; Corcoran, Jennings & Beveridge, 2011; Darling-Hammond, 2013; Rothstein, 2011; Sentell, 2012; Watanabe, 2012). A follow-up email announcement (July 14, 2015) by the Pennsylvania Department of Education notified all school districts and teachers that testing completed by students in the 2014 - 2015 school year reflected the more rigorous PA Common Core standards and new cut scores. As a result, scores were 9% lower in English language arts and 34% lower in mathematics across the state. The announcement mentioned a possible delay of the Keystone graduation requirement, however, nothing was noted about reversing the decision to use the Keystone scores on the School Performance Profile or the Teacher Effectiveness Evaluation.

On September 8 (PDE, 2015), Governor Wolf announced approval of a one year ESEA waiver by the U.S. Department of Education that permits omission of PSSA achievement data on teacher evaluations and School Performance Profiles from the 2015 year of key changes. One of the more drastic fluctuations occurred in Lancaster County schools that saw a massive decline in scores overall. Discrepancies in scores from 91% of students scoring proficient last year to 41% proficient this year at Conestoga Valley Middle School seem inexplicable to administrators in this consistently high performing school (Newhouse, 2015). However, the new standards push content to earlier years and present material at a higher reading level (PDE, 2015). Although

PSSA achievement scores are not a part of the evaluation of teachers this year, evaluations will still contain student growth measures, or PVAAS data (PDE, 2015).

In December of 2015, after the release of individual PVAAS data, reports of questionable discrepancies surfaced. According to two individual teacher PVAAS reports from PDE (2015) for 7th grade math teachers we will call 'A' and 'B' for the condition of anonymity, both received different scores despite sharing 50% responsibility for 143 students. The difference for teacher A was the inclusion of 100% responsibility for one additional student for a total of 144. Both teachers A and B report the additional student was scheduled to attend only one or two class periods daily, was placed in the course for "social reasons" despite lacking prerequisite skills required for the course, and was absent at least half of the school year (A & B, personal communication, December 8, 2015). As a result, teacher B with 143 students received a single year growth index of 0.2 (PDE, 2015). Teacher B with the additional student received a single growth index of -2.1 (PDE, 2015). A 4th grade math teacher in another district reports being questioned by administration after receiving a single year negative growth index. Upon review, these students scored so advanced in grade 3, the group lacked room for statistical growth on PVAAS calculations the following year (S. B., personal communication, December 20, 2015). This scenario resembles that of the New York teacher whose students performed so high in grade 7 that she was fired for failure to show growth in grade 8 grade (Newton, Darling-Hammond, Haertel & Thomas, 2010).

On June 16, 2015, Senate Bill 880 unanimously passed a vote in the PA Senate that delays the use of Keystone exams as a graduation requirement for two years, beginning with the graduating class of 2019 (Palochcko, 2015). On November 24, 2015 (Newhouse), the PA House of Representatives unanimously voted to delay testing with the provision the PA Department of

Education conducts research to find a possible alternative. The bill returned to the PA Senate on January 20, 2016 and passed with a unanimous vote (Niederberger & Crompton, 2016). Governor Tom Wolf agreed to pass the measure February 3, 2016 (Maisel) with the expectation that school districts and the Department of Education will work together to make recommendations for Keystone exams. Pedro Rivera, PA Secretary of Education, is careful to point out that this is not the end of standardized testing and fails to mention removal of high-stakes tests scores from teacher evaluations (Maisel, 2016). Although this relieves pressure from some students who may not have graduated, teachers, not students, maintain lone accountable for these scores.

To add to the miscalculation of PVAAS data, parents in Pennsylvania are opting their children out of standardized testing in increasing numbers. In 2011, 624 students opted out of PSSA tests. In 2015, at least 7,890 students opted out across the state (Hall, 2016). Some areas of PA host groups of parents who organize opt-out movements (Hall, 2016), whereby large numbers of students are not represented in school data. Opt-out groups claim this refusal represents a protest of damage done to public schools through overuse and misuse of standardized testing (Hall, 2016). Although some administrators defend the data obtained by testing, these parents intend to obtain attention from legislators to promote change by eliminating such tests (Hall, 2016).

Following the Vaule-Added Measure (VAM) evaluation decisions in Pennsylvania, a judge in New York ruled against current U.S. Secretary of Education, John King, and New York's similar VAM-based teacher evaluation system (Strauss, 2016). A teacher in New York filed a suit claiming that the evaluation she received was an unfair calculation. She charged that the 'value' added by the teacher to raise standardized test scores cannot be excluded from other

'values' added to the computer model prediction equation. Examples of these other 'values' or 'factors' include hunger, stress, abuse, and sickness (Strauss, 2016). Students are then measured against theoretical students in this computer model. As for the teacher, reports showed the students in her classes each year consistently scored higher than the state average in math and in English. The judge found the VAM system to be "Arbitrary and Capricious" (Strauss, 2016). He concluded the evaluation was biased against teachers with low-performing students or teachers with high-performing students (Strauss, 2016). Final remarks included strict adherence to a bell curve and drastic score ratings despite minimal differences in student scores from the previous year (Strauss, 2016). Unfortunately, despite repeated warnings of inaccuracy, data-driven individuals regard this as the perfect solution to evaluate teachers.

A Pilot Teacher Evaluation

In June 2009 the Bill and Melinda Gates Foundation began working with teacher effectiveness reform through the Intensive Partnership for Effective Teaching. The Foundation provided grant money to ensure the best teachers were available to the neediest students. Research sponsored by the Foundation was conducted by RAND Corporation and the American Institutes for Research. Four cities were studied, including Pittsburgh, PA and their 67 public schools (Stecher, Garet, Holtzman & Hamilton, 2012). Each of the four sites piloted new teacher reform evaluations and were chosen based on two characteristics that included a high percentage of free and reduced lunches and a high percentage of minority students (Stecher, Garet, Holtzman & Hamilton, 2012).

The foundation arranged for strategic planning partners to engage with schools and required submission of plans before receiving all of the grant money. Planning involved multiple stakeholders including administrators, teachers, teachers' unions, and community

members and lasted months before implementation (Stecher, Garet, Holtzman & Hamilton, 2012). The initial evaluation for the study did not include student performance data in the 2010-2011 school year, however, one site was adding it after the fact. In the future, the measures were intended to make key decisions including teacher retention, compensation, intervention, and teacher placement (Stecher, Garet, Holtzman & Hamilton, 2012).

Stecher, Garet, Holtzman, and Hamilton (2012) report that most teachers agreed the study maintained worthy goals, however, most also viewed the evaluation as punitive. Two-thirds of the observers reported receiving training and 70% of teachers felt the observer was qualified to observe. Over 80% of teachers felt the observation language was clear and they understood what the observer was expecting. Another 60% believe they had a clear idea of what the system considered 'good instruction' and some teachers felt it helped them improve specific aspects of their teaching. 60% to 70% believe teachers who already have students who are high performing fair better on evaluations. One-third of teachers believe they were being forced to teach students in ways that were not beneficial to students. Another 68% to 82% reported experiencing increased stress as a direct result of the evaluations, even without student performance data included.

Chronic Teacher Stress

Teaching is among the most stressful professions and 95% of school teachers encounter increasing work related stress throughout their careers (Chan, Chen & Chong, 2010; Gloria, Faulk & Steinhardt, 2013). Studies indicate 42% of teachers claim the profession is extremely stressful. 36% admit being stressed most of the time (Borg & Riding, 1991; Gloria, Faulk & Steinhardt, 2013; Kyriacou, 2001; Smith, Brice, Collins, Matthews & McNamara, 2000; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011). It is important to note that positive stress,

eustress, is necessary to motivate action. However, negative stress, or distress, typically defines teacher stress that associates with feelings of anger, tension, anxiety, and frustration (Kyriacou, 2001). Teachers experience stress when work demands appear threatening and teachers feel unable to cope with the loads (Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011). Chronic teacher stress defines the complex cognitive process whereby a teacher is unable to cope with work demands for a prolonged period, compounded with persistent feelings of failure and disappointment, creating emotional exhaustion and apathy (Kyriacou & Sutcliffe, 1978). Chronic Stress leads to emotional exhaustion, poor job performance and commitment, teacher burnout, deteriorating health, and mental disorders, especially depression (Bakker & Schaufeli, 2000; Schaufeli & Greenglass, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011). In a study by Hammen and deMayo (1982), one out of five teachers admit experiencing this level of unbearable stress during their career. Although the sources of the teacher stress were numerous, the common experience of lack of control over solutions correlated with the highest depressive symptoms, and teacher depression represents a primary cause of frequent absenteeism and high attrition rates (Hammen & deMayo, 1982; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011).

Multiple research studies of elementary and secondary level teachers report various reasons for teachers stress. The most acute stressor for both is time constraints correlated with work overload (Kyriacou & Sutcliffe, 1978; Kyriacou, 2001). Other primary sources of teacher stress include disruptive student behavior (Borg & Riding, 1991; Kyriacou & Sutcliffe, 1978), and lack of support from the principals, administration, society, and parents (Grayson & Alvarez, 2008; Kyriacou, 2001). Teachers also report poor working relations with colleagues, conflicting job roles, poor working conditions in facilities (Kokkinos, 2007), lack of participation in decision-making (Manthei & Gilmore, 1996), low pay, and low social status (Chaplain, 2007;

Skaalvik, & Skaalvik, 2009) as additional stressors. More recently, high-stakes testing, specifically, student performance data added to teacher effectiveness evaluations, curriculum changes, increased accountability, and strict standards have been identified as new teacher stressors (Hanson, 2006; Kyriacou, 2001; Montgomery & Rupp, 2005). A recent study found teachers, who were exposed to four different accountability models in four different states (including Pennsylvania), all experienced similar high-stakes testing stress as a result of the accountability policies, despite the policy differences. When accountability policies are perceived as threatening, teacher coping resources are depleted (von der Embse, Pendergast, Segool, Saeki & Ryan, 2016).

Another workplace stressor that is common to primary and secondary school employees is emotional labor. In this instance, a teacher tries to regulate feelings to accomplish teaching responsibilities (Grandey, 2000). One employs surface acting, whereby displayed emotions are disassociated from internal emotions (Hochschild, 1979). Teachers who engage more frequently in surface acting or monitoring of their emotions, emotional labor, experience a higher incidence of teacher burnout (Cheung, Tang & Tang, 2011).

The demands for teaching can be overwhelming as they require sustained cognitive, emotional, and physical effort (Feuerhahn, Bellingrath & Kudielka, 2013). Emotional demands, such as conflict or criticism regarding students, parents, colleagues, or administrators contribute to emotional exhaustion, a core component of burnout (Feuerhahn, Bellingrath & Kudielka, 2013). Emotional dissonance also causes great emotional strain on teachers. Emotional dissonance occurs when a teacher performs surface acting or when a teacher displays emotions to parents, students, or colleagues that are not genuinely felt (Cheung, Tang & Tang, 2011; Feuerhahn, Bellingrath & Kudielka, 2013). Cognitive demands such as time constraints and

associated pressures, student disruptions, or any classroom disruptions can directly strain emotional health (Burke & Greenglass, 1989; Kokkinkos, 2007; Feuerhahn, Bellingrath & Kudielka, 2013). The physical demands of teaching include grading, recordkeeping, classroom supervision, instruction, and classroom discipline among others. All of these physical tasks require simultaneous cognitive effort to complete, contributing to more emotional exhaustion (Feuerhahn, Bellingrath & Kudielka, 2013).

Constant or prolonged exposure to stress disrupts a teacher's ability to cope effectively, preventing homeostasis, or a state of well-being. This inability to cope creates feelings of defeat and leads to an unhealthy state of psychosomatic disorders including depression, a predictor of teacher burnout (Bakker & Schaufeli, 2000; Gloria, Faulk & Steinhardt, 2013). Chronic Stress leads to emotional exhaustion and further perpetuates poor job performance and lack of commitment (Bakker & Schaufeli, 2000; Gloria, Faulk & Steinhardt, 2013; Schaufeli & Greenglass, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011). To that end, as emotional exhaustion increases for teachers, so does cognitive stress symptoms and cognitive impairments (Feuerhahn, Bellingrath & Kudielka, 2013). Teachers with extreme emotional exhaustion also have a harder time learning and remembering compared to peers (Feuerhahn, Bellingrath & Kudielka, 2013; Feuerhahn, Stamov-Roßnagel, Wolfram, Bellingrath & Kudielka, 2013). This chronic stress for teachers perpetuates negative consequences for schools and learning environments.

Teacher Burnout

Teacher burnout, a state of mental, physical, and emotional exhaustion, generates from prolonged exposure to intolerable work situations, and emotional and interpersonal stressors, often without sufficient recovery time, leading to teacher disengagement in work (Cherniss,

1980; Hakanen, Bakker, & Schaufeli, 2006; Maslach & Leiter, 2008). Burnout occurs over time, emitting small warning signs such as frustration, headaches, exhaustion, and withdrawal that can escalate into an extreme dread of going to work (Hanson, 2006; Scott, 2006; Omdahl & Fritz, 2006; Zhang & Sapp, 2008). The condition can alleviate if working conditions change. Burnout is not merely stress, but having no support system, no resources, no rewards, or no way out of the current situation. (Blazer & Miami-Dade County Public Schools, 2010; Scott, 2006). Burnout is recognized as a serious problem in education. Studies have found teacher burnout has a negative effect on student learning and it is estimated that 5% to 30% of teachers display burnout symptoms at any particular time (Blazer & Miami-Dade County Public Schools, 2010). A recent study found teacher exhaustion and burnout correlates with high stress, or high recorded cortisol levels, in students (Oberle & Schonert-Reichl, 2016). They do not know if the teacher burnout preceded the student stress or if the student stress preceded the teacher burnout.

A job demands-resources model is often used to interpret two factors, demands and resources, and predict burnout in teachers based on these environmental factors (Maslach, Scaufeli & Leiter, 2001). Job demands of a teacher include work overload, interpersonal conflicts, students' behaviors, role discrepancies, school climate and policies, and crisis control caused by faulty equipment (Byrne 1999; Blazer & Miami-Dade County Public Schools, 2010). Teacher resources include decision-making and latitude, administrative leadership and coaching, support from colleagues and the community, professional development, flexible scheduling, and skill usage (Byrne 1999; Blazer & Miami-Dade County Public Schools, 2010). Prediction of burnout occurs when a teacher's perception encompasses overwhelming demands combined with a lack of resources (Hakanen, Bakker & Schaufeli, 2006). However, other research shows individual motivational factors such as self-esteem, self-efficacy, and optimism can overcome

these perceptions, thereby mediating the connection between workplace factors and burnout (Byrne, 1999; Skaalvik & Skaalvik, 2009). Specifically, autonomous motivation negates burnout (Fernet, Guay, Senecal & Austin, 2012).

Other factors found to precipitate teacher burnout include school reforms and accountability pressure. Studies found schools that were undergoing comprehensive reform contained more teachers experiencing burnout (Hanson, 2006; Kiracou, 2001; Montgomery & Rupp, 2005). Researchers believe this is due to the nature of the top down implementation of most reform policies, whereby the teacher is excluded from decisions and must comply with little or no justification (Blazer & Miami-Dade County Public Schools, 2010). Accountability demands create stress and burnout for teachers in the elementary grades and high-stakes testing subject areas due to test related pressure (Hanson, 2006; Kyriacou, 2001; Montgomery & Rupp, 2005)

Burnout, a psychological syndrome, includes three dimensions that appear sequentially in the following order (Kokkinos, 2007; Malasch & Jackson, 1981): emotional exhaustion, depersonalization, and reduced personal accomplishment (Maslach, 1993; Maslach, Scaufeli & Leiter, 2001). *Emotional exhaustion*, the main identifier associated with burnout, refers to depletion of energy or emotional resources to the point of feeling there is nothing emotionally left to give (Blazer & Miami-Dade County Public Schools, 2010; Fernet, Guay, Senecal & Austin, 2012; Malasch & Jackson, 1981). *Depersonalization* refers to the cynical, detached, or negative attitude that a teacher develops toward their job and students (Blazer & Miami-Dade County Public Schools, 2010; Fernet, Guay, Senecal & Austin, 2012; Malasch & Jackson, 1981). *Reduced personal accomplishment* describes low self-confidence and feelings of dissatisfaction, inadequacy, and ineffectiveness as a teacher (Blazer & Miami-Dade County Public Schools,

2010; Fernet, Guay, Senecal & Austin, 2012; Malasch & Jackson, 1981). Although these burnout components appear sequentially, they can appear at different times and to varying degrees of severity (Kokkinos, 2007; Malasch & Jackson, 1981).

Research shows some teachers possess greater risk factors for burnout. Teachers who experience emotional dissonance or engage more frequently in surface acting, experience a higher incidence of teacher burnout (Cheung, Tang & Tang, 2011). Of all demographic characteristics, age consistently correlates with chronic work stress and burnout. Younger teachers regularly report higher levels of burnout (Malasch, Schaufeli & Leiter, 2001). Some studies have shown no difference in burnout levels when comparing gender. However, other studies found burnout is more prevalent among female teachers (Greenglass, Burke & Ondrack, 1990; Malasch, Schaufeli & Leiter, 2001; Rey, Extremera & Pena, 2012; Schaufeli & Enzmann, 1998). The exception is depersonalization that is more prevalent among male teachers (Anderson & Iwanicki, 1984; Byrne, 1991; Rey, Extremera & Pena, 2012; Van Horn, Scaufeli, Greenglass & Burke, 1997). Some studies have found burnout to be more prevalent among secondary teachers. Secondary female teachers exhibit low levels of perceived personal accomplishment and male secondary teachers display high levels of depersonalization (Anderson & Iwanicki, 1984; Beer & Beer, 1992; Burke & Greenglass, 1989; Rey, Extremera & Pena, 2012; Vandenberghe & Huberman,1999). Others point out the secondary school environment as a factor for increased potential burnout. This inclusive environment includes lack of resources, lack of support, and increased pressure of student performance combined with student issues such as discipline, conflict, apathy, lack of motivation, and poor academic performance (Rey, Extremera & Pena, 2012). High burnout rates are also found in situations where teachers feel overworked, worried about job security, confused about expectations and priorities, and have too

many responsibilities and duties void of pay (Blazer & Miami-Dade County Public Schools, 2010; Scott, 2010).

Symptoms of burnout manifest in psychological, physical, and behavioral forms (Blazer & Miami-Dade County Public Schools, 2010). Psychological symptoms include depression and anxiety. These are continual, prolonged feelings accompanied by loss of joy and interest in work, feeling helpless, feeling resentful, and even feeling dread over going to work (Zhang & Sapp, 2008). Common physical symptoms of exhaustion, digestive disorders, headaches, insomnia, high blood pressure, and heart irregularities appear almost instantly. However, long term physical symptoms of chronic stress and burnout may lead to hypertension, diabetes, or cardiovascular disease (Zhang & Sapp, 2008). Behavioral symptoms manifest in a variety of ways. Lack of interest creates tardiness and absenteeism. These teachers experience a lack of commitment, become less task-oriented, and attend to fewer instructional tasks. The decrease in positive student reinforcement further adds to poor job performance. Emotional distance prevents sympathy for students and provokes less tolerance for student behavioral issues. These teachers often withdraw from colleagues and rely on food, alcohol, caffeine, or drugs to cope with exhaustion or moods (Coman et al., 2013; Firth & Mims, 1985; Malasch & Jackson, 1981; Wisniewski & Garguilo, 1997).

Research suggests that prevention of burnout is easier than reversing it (Maslach, 1993; Maslach, Scaufeli & Leiter, 2001). Activities such as noticing early warning signs, learning stress and time management strategies, being involved in professional development, and joining peer support groups can be helpful. Leadership personnel in schools can participate in supportive training, supply adequate resources, provide feedback, and encourage peer support (Blazer & Miami-Dade County Public Schools, 2010; Maslach, 1993; Maslach, Scaufeli &

Leiter, 2001; Texas Medical Association, 2009). Otherwise, chronic stress can lead to burnout, which has been linked to teacher attrition (Farber, 1991; Malasch & Jackson, 1981).

Teacher Attrition

Research suggests school districts lose billions of dollars each year due to chronic stress and burnout (Omdahl & Fritz, 2006). Money is lost through disability claims, absenteeism, and higher teacher attrition rates. Chronic stress alone causes 75% of doctor's visits, increases the risk of developing cancer or diabetes, and doubles the risk of a heart attack (Chew, 2015; Omdahl & Fritz, 2006). A recent analysis reveals 27% of teachers miss 10 or more days of school during the year, a practice shown to decrease student math achievement and engagement (Sparks, 2016). Reasons given by teachers for leaving the field of education mirror chronic stressors leading to burnout, and attrition has been directly linked to burnout (Farber, 1991). Teachers cite low pay, little respect in society, lack of administrative support, poor parent and student attitudes, increasing school violence, little chance for career advancement, and minimal input into decision-making as main accounts for attrition (Pucella, 2011).

In this study, potential teacher attrition defines the potential or possibility of leaving the field of education. Teachers are more likely to stay if they experience job satisfaction. One study found certain factors contributed to or detracted from teacher stress, burnout, and job dissatisfaction (Otero López, Bolaño, Santiago Mariño & Pol, 2010). Peer support is the best predictor of avoiding stress, burnout, and job dissatisfaction (Otero López, Bolaño, Santiago Mariño & Pol, 2010). A sense of professional accomplishment (hardiness) and optimism are factors that strongly correlated with overall job satisfaction and a decrease in teacher stress and burnout (Otero López, Bolaño, Santiago Mariño & Pol, 2010). However, daily hassles and life events were found to contribute to teacher stress, burnout, and job dissatisfaction (Otero López,

Bolaño, Santiago Mariño & Pol, 2010). A recent teacher poll found teacher job satisfaction dropped from 62% to 39% from 2008 to 2013, and 51% of teachers report feeling extreme stress several days a week (Strauss, 2015).

Demoralization also contributes to the problem of attrition. Santoro (2011) explains that this is not burnout. Instead, although moral rewards of teaching are limitless, many teachers are unable to access these moral rewards in the current climate (Santoro, 2011). Changes have occurred so rapidly that teacher-valued practices, ones viewed as ethical and moral, are questioned. Frustration generates from the changing concept of 'good teaching' and teachers may struggle to remain productive (Santoro, 2011). The intrinsic rewards of having a sense of mission or altruism have been repeatedly found to attract and retain teachers in high-poverty, urban schools for years despite low pay and status (Crocco & Costigan 2007; Freedman & Appleman, 2009; Margolis & Deuel 2009; Ng & Peter 2010; Stotko, Ingram & Beaty-O'Ferrall, 2007). These intrinsic rewards decrease the potential or possibility of teacher attrition. The high-stakes performance and accountability environment has rendered a 'corrosive influence' on the value of teaching and learning (Valli, Croninger, Chambliss, Graeber & Buese, 2008), especially in high-poverty schools. School policy uses terms like 'effectiveness' mixed with other, harsh terms like 'moral obligation' to describe teacher actions (Santoro, 2011). Even the strongest teachers may find little left to sustain them in the profession if the high-stakes accountability atmosphere make moral rewards inaccessible (Santoro, 2011), further increasing the potential of attrition.

Studies and government data have identified specific factors related to increasing the potential or possibility of teacher attrition. The National Center for Education Statistics (2015) reports teacher absenteeism is on the rise for public school teachers and the greatest worry of

public school teachers is job loss. Neither is true for private school teachers, who are not subject to the same teacher effectiveness evaluations that include student performance data (NCES, 2015). Public school teachers also report a steady decrease in feeling they are having a positive influence on teaching in the years following the inception of common core standards (NCES, 2015). At the same time, teacher attrition is steadily on the rise in public schools. The highest rates of attrition occur in city and rural schools, specifically in low-income schools, whereby more than 75% of students register for the free or reduced lunch program (NCES, 2015). The highest attrition rates also occur in public schools comprised of a 50% or more population of multiple races (NCES, 2015). For years prior to teacher evaluations containing student performance data, secondary teachers, more prone to burnout in a less supportive environment, consistently comprise the group with the highest attrition rates (Anderson & Iwanicki, 1984; Beer & Beer, 1992; Burke & Greenglass, 1989; NCES, 2015; Rey, Extremera & Pena, 2012; Vandenberghe & Huberman, 1999). The size of the school and age of the teacher has no strong direct or indirect correlation with attrition rates (NCES, 2015).

Location impacts the possibility of teacher attrition. Certain states and areas experience more attrition than others. For example, Arizona watched 1 to 5 teachers in each district break contracts mid-year to teach across state lines for better pay or to work in other professions. About 1,000 teaching vacancies received no applicants and jobs are filled by non-certified substitutes (Educator Recruitment & Retention Task Force, 2015; Strauss, 2015). These teachers report unfair evaluations, high stress, low pay, little respect, high-stakes testing, increasing mandates, and loss of professional autonomy as reasons for leaving the profession (Educator Recruitment & Retention Task Force, 2015). Kansas and Oklahoma each experienced a similar

teacher exodus across state lines and Oklahoma officials traveled as far as Puerto Rico and Spain to recruit and fill some bilingual education positions (Strauss, 2015).

High-stakes testing is linked to potential attrition and future teacher shortages. In a Nation Education Association survey, 53% of teachers reported a desire to leave the field due to standardized testing, and these teachers also admitted frequently cautioning young adults about entering the profession (Walker, 2014). Another study found 83% of teachers would not recommend teaching as a career (Strauss, 2015; Walker, 2014). Cautionary advice, teacher lay-offs during the recession, and the various reasons cited above are also keeping students from entering the profession, often switching to other majors (Strauss, 2015). Universities in Indiana experienced a 50% decrease in teacher preparation program enrollment, California witnessed a 74% decrease in 10 years, and New York, despite higher salaries, experienced a 22% decrease in 2 years (Strauss, 2015). Indiana reported a 63% decrease in new teacher licenses from 2009-2010 to 2013-2014. Officials claim the decline is due to rigorous testing requirements on teacher evaluations, low teacher wages, limited teacher protection, and the blame-the-teacher mentality touted by politicians in the media (Ladwig, 2015). Officials in California point to high-stakes testing on teacher evaluations, battles over common core, eroding tenure, and a politically hostile anti-teacher environment as reasons for the decline (Strauss, 2015). New York administrators, who are not facing a crisis yet, blame multiple factors resulting from failed reforms (Strauss, 2015). Other states experiencing extreme teacher shortages due to attrition include New Mexico, Nevada, California, New York, Idaho, and Florida (Strauss, 2015). Reasons for the attrition include unfair teacher evaluations based on student test scores in addition to loss of teacher job protection, loss of professional autonomy, increased high-stakes testing, and lack of school resources through under-funding (Strauss, 2015). Thousands of jobs are left unfilled or

are forced to be filled with non-certified substitutes, thereby, ironically decreasing the standard of education in a climate of increased teaching standards.

Self-Determination Theory

The current study reflects the state of education in light of common core standards, high-stakes testing, teacher effectiveness evaluations, and teacher attrition in public schools as viewed through the lens of Self-Determination Theory (SDT). This theory addresses issues of motivation for both students and teachers, whereby a teacher's emotional well-being is affected by external controls that influence both student and teacher motivation (Deci & Ryan, 1985).

Self-Determination Theory (SDT) denotes the study of motivation and personality. This broad framework comprises of a Meta-Theory and a Formal Theory. The meta-theory is used for theoretical framing of motivational studies. The formal theory, within six mini-theories, defines the role of intrinsic and extrinsic motivation on social and cognitive development. It also describes an undermined sense of volition and initiative as a direct result of negative social and cultural factors. However, the highest forms of engagement, performance, volition, persistence, creativity, and initiative are positively correlated with supporting autonomy, competence, and relatedness (Deci & Ryan, 2004; Ryan & Deci, 2015).

Autonomy represents the degree to which an individual perceives the social context in an environment as supportive or controlling. Autonomous individuals prefer creative expression within the construct of defined tasks, that is, generating an expression of self in work or learning (Deci & Ryan, 2004; Ryan & Deci, 2015). *Competence* describes the perceived level of effectiveness an individual feels when interacting with the environment or within a particular behavior domain. SDT stresses that optimal motivation occurs when an individual feels both autonomous and competent in a particular behavior domain (Deci & Ryan, 2004; Ryan & Deci,

2015). *Relatedness* refers to the basic psychological need to be connected, accepted, and cared for by others. SDT reports a strong positive correlation between support of relatedness and competence with volitional or intentional motivation (Deci & Ryan, 2004; Ryan & Deci, 2015).

Meta-Theory: An Organismic Viewpoint

This dialectical approach to predicting human behavior and development begins with active human engagement within social contexts. Individuals have an innate tendency to grow psychologically, master challenges, and integrate new experiences. This development requires ongoing social-contextual support and nurturing. Satisfying the basic need for autonomy, competence, and relatedness promotes healthy psychological development and functioning. Thwarting or hindering these needs disrupts development and can lead to psychopathology and aggression (Deci & Ryan, 2004; Ryan & Deci, 2015).

Formal Theory: Six Mini-Theories

Self-Determination Theory is a broad framework. The Formal Theory of human motivation and personality divides into six mini-theories, each focusing on one aspect of motivation or personality that emerged during research (Deci & Ryan, 2004; Ryan & Deci, 2015).

Cognitive Evaluation Theory (CET)

Intrinsic motivation, the satisfaction of behaving for its own sake, comprises the fundamental basis of Cognitive Evaluation Theory, the first mini-theory. Internal motivation, a lifelong flow of creativity, encourages play and exploration. Research shows supporting basic needs of autonomy and competency within social contexts fosters intrinsic motivation. CET amplifies the importance of intrinsic motivation in education, arts, and sports (Deci & Ryan, 2004; Ryan & Deci, 2015).

Organismic Integration Theory (OIT)

The fundamental element addressed in the second mini-theory is extrinsic motivation that uses outward motivation to control behaviors. The following four subtypes of extrinsic (external) motivation fall along an internalization continuum: External Regulation, Introjection, Identification, and Integration. Social contexts can support autonomy and relatedness to foster internalization. The more internalized the extrinsic motivation becomes, the more autonomous the motivation becomes.

Causality Orientation Theory (COT)

The third mini-theory describes environment orientation and behavior regulation compared to individual differences among people. COT identifies three types of causality orientation. *Autonomy Orientation* describes a person who acts out of concern for what is happening. A person who acts based on rewards, gains, or approval demonstrates *Control Orientation*. Deciding not to act or taking limited action, characterized by competence anxiety, describes *Impersonal* or *Amotivated Orientation* (Deci & Ryan, 2004; Ryan & Deci, 2015).

Basic Psychological Needs Theory (BPNT)

The fourth mini-theory elaborates on the concept of the dependence of psychological and emotional well-being on meeting the basic psychological needs of autonomy, competence, and relatedness. Social-contextual factors that support and do not hinder these needs should positively impact wellness. BPNT views all three as essential, whereby thwarting of one need will have a negative impact on development and well-being (Deci & Ryan, 2004; Ryan & Deci, 2015).

Goal Contents Theory (GCT)

The fifth mini-theory clarifies distinctions between intrinsic (internal) and extrinsic (extrinsic) goals with impact on motivation. Intrinsic goals orient internally such as personal growth, relationships, and community. Extrinsic goals externally orient, such as popularity, appearance, success, or wealth. Extrinsic goals are associated with lower psychological or emotional wellness (Deci & Ryan, 2004; Ryan & Deci, 2015).

Relationships Motivational Theory (RMT)

The sixth mini-theory concerns relatedness, that is, the need to develop, foster, and maintain close relationships. This need for personal connection and acceptance extends to romantic partners, friends, and groups. RMT posits that interaction with others satisfies the need for relatedness and is essential for psychological growth and development (Deci & Ryan, 2004; Ryan & Deci, 2015). Research shows the well-being established with high-quality relationships also placates the need for autonomy and the need for competence to a lesser degree (Ryan & Deci, 2015). The highest quality relationships involve both partners supporting the others needs of autonomy, competence, and relatedness (Deci & Ryan, 2004; Ryan & Deci, 2015).

Self-Determination Theory and Teachers

Deci and Ryan (1985) describe Self-Determination Theory as an approach to motivation, whereby autonomous, or intrinsic, motivation is necessary for optimal functioning. Individuals become autonomously motivated when given a choice to initiate behavior. This intrinsic motivation for teachers transpires by identifying meaning and value in their work. Autonomous motivation, defined as well-internalized values or intrinsic motivation, engenders increased persistence, increased quality of learning, and a positive emotional experience (Grolnick & Ryan, 1987; Ryan & La Guardia, 1999; Ryan, Stiller, & Lynch, 1994). SDT distinguishes

between autonomous motivations (intrinsic) and controlled motivations (external). Controlled motivation, external rewards and punishment, associates with negative emotions, decreased persistence and decreased quality of learning for students (Grolnick & Ryan, 1987; Ryan & La Guardia, 1999; Ryan, Stiller, & Lynch, 1994).

The theory also suggests that internalizing an external control, or extrinsic motivation, can occur by accepting its values and goals, thereby transforming it to an autonomous motivation to exhibit behaviors in line with the acquired values and goals (Deci & Ryan, 2002). However, this internalization creates three other forms of external control: external regulation, introjected regulation, and identified regulation. External regulation describes regulated behavior for the purpose of obtaining rewards or avoiding constraints (Deci & Ryan, 2002). Introjected regulation describes the progression of external demand transforming into an internal representation by way of internal coercion, in the form of anxiety, guilt, or shame. This internal pressure then controls behavior (Deci & Ryan, 2002). Identified regulation involves choice of behavior because of its perceived value, even if the activity is not interesting or engaging. It excludes internal or external pressures and is considered autonomously regulated for that reason (Deci & Ryan, 2002).

Research indicates employees, with more autonomous (intrinsic) motivation, possess a greater sense of well-being than those with more controlled (extrinsic) motivation (Ryan & Deci, 2000). Types of autonomous motivation, such as intrinsic and identified regulation, toward work, have been negatively correlated with burnout. However, types of controlled motivation, such as extrinsic and external regulation, toward work have been positively correlated with burnout, (Fernet, Guay, Senecal & Austin, 2012) which has also been directly linked to attrition (Farber, 1991).

An important aspect of SDT proposes that controlled or external motivations can thwart teacher motivation in the environment. For teachers, work-related factors are considered either autonomy-supportive or controlling (Ryan & Deci, 2000). Autonomy-supportive conditions allow teachers to be involved in decision-making while developing meaningful rationale for those decisions. This environment acknowledges teachers' feelings and views, provides support and feedback, and minimizes pressure (Deci, Eghrari, Patrick & Leone, 1994). Autonomy-supportive conditions also fulfill basic psychological needs that support autonomy, competence, and relatedness. Autonomy reflects the need to feel a sense of self and responsibility for one's behavior. Competence reflects the need to interact and succeed at optimally challenging tasks. Relatedness concerns the degree to which an individual socially integrates and feels accepted (Deci & Ryan, 1985). In turn, teachers experience the highest quality motivational environment and sense of emotional well-being, (Ryan & Deci, 2000) allowing teaching performance, persistence, and creativity to enhance. However, if any of the three conditions is unsupported for a teacher, a detrimental impact to well-being ensues in that environment (Ryan & Deci, 2000). Need thwarting occurs in controlling conditions where teachers are pressured with time constraints, imposed goals, and contingent rewards (Ryan & Deci, 2000).

Self-determination theory predicts highly negative outcomes for student and teacher motivation following attempts to implement external controls to improve student achievement (Deci & Ryan 1985; Ryan & Deci, 2000). As a result, chronic teacher stress will increase, contributing to increased cases of teacher burnout and subsequently attrition (Farber, 1991; Malasch & Jackson, 1981).

Currently, student performance data in part determines teacher effectiveness (PDE, 2015). To that end, teachers focus student attention on this external control. Controlled

motivation decreases student persistence, decreases motivation, creates negative emotions, and decreases quality of learning (Deci & Ryan, 2002). Teachers lose the ability to challenge students of different achievement levels under rigid constraints of common core, and students with disabilities are expected to reach benchmarks simultaneously despite learning differences, backgrounds, and developmental rates (Deci & Ryan 1985; Elliot & Dweck, 2007; Ryan & Deci, 2000). People are optimally challenged, or most intrinsically motivated when tasks are within reach (Elliot & Dweck, 2007). Because high-stakes tests require all students to reach the same benchmark at the same time, it again diminishes any autonomous or internal motivation that may have existed (Elliot & Dweck, 2007).

Two separate studies revealed teachers who were pressured to produce high achieving students became more controlling and less tolerant of student autonomy. Lectures increased along with criticism and praise, all negative impacts on learning. As predicted, in line with SDT, the more the teachers were pressured to perform toward higher standards, the more controlling their instructional behaviors became. As a result, their students scored more poorly on objective based tests than teachers who were not pressured (Deci, Spiegel, Ryan, Koestner & Kauffman, 1982; Flink, Boggiano & Barrett, 1990; Taylor, Ntoumanis, & Standage, 2008). Teachers, pressured to increase student achievement, may be unintentionally controlling students, thereby negatively impacting their learning and ultimately their achievement performance (Elliot & Dweck, 2007).

Feeling external pressure toward an action or outcome, such as pressure to produce student achievement on high-stakes tests, tends to undermine autonomy of the teacher as well (Deci & Ryan, 1985). A recent study identified the perceived thwarting of autonomy, competence, and relatedness as separate and significant indicators of teacher stress and burnout

(Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Skaalvik & Skaalvik, 2009). Because individuals, including teachers, are not motivated by failure, other negative practices have ensued. Curricula narrows to tested items as teachers feel pressure to teach a certain way, thereby decreasing autonomy (Elliot & Dweck, 2007). Teachers use more controlling classroom techniques to meet demands of this high-stakes testing environment and confront increasingly negative experiences daily for teachers and students (Elliot & Dweck, 2007). Teachers report feeling inadequate and incompetent (diminished competence) when repeatedly confronted with evaluations based on student performance data (Elliot & Dweck, 2007). Some teachers report feelings of rejection by peers and colleagues, hindering relatedness (Elliot & Dweck, 2007). Another study reports encroachment on professional autonomy further dismantles professional competence, and ultimately leads to teacher attrition (Bouwma-Gearhart, 2010).

Administrative and managerial behaviors supporting competence include: agreeing on achievable goals, training and preparation for meeting goals, removing physical or procedural barriers, and encouraging regular feedback and discovery of errors while keeping negative feedback to a minimum (Deci & Ryan, 2004). The Pennsylvania Department of Education thwarts these competence-supportive behaviors. Only a small group of teachers are selected to represent all educators in developing assessment questions. Many of these teachers no longer teach in a classroom due to retirement or ascension to administrative or collegiate positions (PDE, 2015). Limited resources are made available for training by the state via the PDE website, however, no formal state initiated training sessions are available for teacher preparation of meeting goals of high-stakes assessments (PDE, 2015). The PSSA tests and Keystone exams do include training sessions for teachers via the web concerning rules for proctoring the exams. This training includes details of fines and other external punishments for failing to follow testing

procedures (PDE, 2015). Teachers, forbidden to look at test questions, cannot discern whether their efforts to teach tested material were on target. Although a few practice questions are released online, teachers cannot discover teaching errors. Feedback is returned at one time and becomes public, open for criticism by the community (PDE, 2015). All of these practices, ensued by the PA Department of Education related to high-stakes testing, thwart the need for competence (Deci & Ryan, 2004), which is linked to teacher stress, burnout, and attrition (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Bouwma-Gearhart, 2010; Skaalvik & Skaalvik, 2009) and negative psychological well-being (Elliot & Dweck, 2007).

Administrative or managerial behaviors that support relatedness include: meeting regularly, supporting cooperation, keeping certain information confidential, and refraining from negatively representing a party not present (Deci & Ryan, 2004). In contrast, the Pennsylvania Department of Education hinders this support. The PDE does not meet with teachers. However, the PDE will send representatives to deliver presentations to certain events such as the Pennsylvania Council of Teachers of Mathematics annual conference (PCTM, 2015; PDE, 2015). PDE thwarts efforts of confidentiality and cooperation when district test results are publicly posted, and districts are ranked in order of performance (PDE, 2015). The recent attachment of student performance data to individual teacher's effectiveness evaluations casts a negative stigma on some teachers who may have inaccurate ratings due to PVAAS (Darling-Hammond, 2013; Sentell, 2012; Watanabe, 2012). This evaluation may negatively represent the teacher in Harrisburg despite inaccuracies. Continual hindering of relatedness-supportive behaviors related to high-stakes testing by PDE further contributes to teacher stress, burnout, and attrition (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Bouwma-Gearhart, 2010; Skaalvik & Skaalvik, 2009).

This study contributes to the emergent body of work relating teacher stress, burnout, and attrition related to student performance data on teacher effectiveness evaluations through the lens of SDT. Research indicates external pressure on teachers to improve student performance decreases autonomy for students. The more pressure increases for teachers to improve achievement, the more controlling teachers become, and the more poorly their students perform, potentially generating more teacher stress (Deci, Spiegel, Ryan, Koestner & Kauffman, 1982; Flink, Boggiano & Barrett, 1990). Teachers are also at risk for loss of autonomy, diminished competence, and diminished relatedness as several studies link these to the increasing demands of repeated effectiveness evaluations that include student performance data (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Fernet, Guay, Senecal & Austin, 2012; Skaalvik & Skaalvik, 2009). This thwarting of autonomy, competence, and relatedness, as described in SDT, further correlates with increased teacher stress, burnout, and attrition (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Bouwma-Gearhart, 2010; Skaalvik & Skaalvik, 2009).

Summary

Poor working conditions can create stress and weaken the effectiveness of any teacher (Darling-Hammond, 2013). Chronic work stress, burnout syndrome, and subsequent teacher attrition have been problematic issues in education for four decades (Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011; Tye & O'Brien, 2002). Chronic work stress and burnout occur when a teacher experiences the inability to cope with unbearable levels of stress and demands of work (Hamman & deMayo, 1982; Kyriacou & Sutcliffe, 1978). Both negatively affect mental and physical health, whereby suffering teachers report experiencing exhaustion, poor performance, emotional exhaustion, emotional detachment, discouragement, reduced productivity and achievement (Ahola, et al., 2005; Malasch, Schaufeli, & Leiter, 2001; Steinhardt, Smith-Jaggars,

Faulk & Gloria, 2011). Chronic teacher stress, an independent predictor of burnout (Chan, 2003), can result in depression (Bakker & Schaufeli, 2000), a primary cause of increased absenteeism and high attrition rates (Hammen & deMayo, 1982; Hastings & Bahm, 2003). Some studies reveal increased external pressure on teachers to improve student performance diminishes autonomy for students (Deci & Ryan, 1985) and teachers (Elliot & Dweck, 2007). In turn, students perform more poorly, generating more teacher stress, and teachers experience diminished competence and diminished relatedness in light of repeated performance evaluations (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Fernet, Guay, Senecal & Austin, 2012; Skaalvik & Skaalvik, 2009). Self-determination Theory explains how this thwarting of autonomy, competence, and relatedness in the workplace correlates with increased teacher stress, burnout, and attrition in research studies (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Bouwma-Gearhart, 2010; Skaalvik & Skaalvik, 2009).

CHAPTER 3

PROCEDURES

The purpose of this quantitative study was to examine the impact that stress related to the inclusion of student standardized testing data in teacher evaluation has on burnout and potential attrition rates among current Pennsylvania math teachers who identify as either Teachers of Record (TOR) or Non-Teachers of Record (NTOR). In Pennsylvania, a TOR receives 15% of scores on their teacher effectiveness evaluations directly from student performance data and an additional 15% from the School Performance Profile (SPP) score. A NTOR only receives a 15% score from the SPP. Grades and courses scheduled for standardized assessment of students, and Teacher of Record parameters were set by the Pennsylvania Department of Education (2015).

This study was important for (a) identifying teacher stress levels related to high-stakes testing, (b) comparing stress levels between TOR and NTOR, (c) identifying teacher burnout levels related to high-stakes testing, (d) comparing burnout levels between TOR and NTOR, ® comparing potential teacher attrition levels between TOR and NTOR, (f) identifying any correlation among teacher stress, teacher burnout, and teacher attrition rates, and (g) identifying any correlation of effects of other factors (gender, school location, school socio-economic status, years of teaching experience, grade level of students) with stress, burnout, and attrition. A more intricate understanding of the relationship of teacher stress, teacher burnout, and potential attrition rates related to high-stakes testing may influence government policy related to mathematics or possibly STEM (Science, Technology, Engineering, and Math) fields regarding high-stakes testing of these content areas.

Chapter 1 contained an introduction to the study, including a theory, hypothesis, terms, and variables. Chapter 2 detailed a review of the literature, including teacher chronic stress,

burnout, and attrition through the lens of Self-Determination Theory. Additional teacher effectiveness evaluation practices and policies were also meticulously reviewed. Chapter 3 discusses research methodology including: research method and appropriateness of research design, population and sampling, informed consent, confidentiality, location, data collection, and instrumentation with validity and reliability. The chapter concludes with data analysis and a summary.

Research Methods and Design Appropriateness

A quantitative research method, which involves collecting numerical data for the purpose of statistical analysis to answer specific research questions, was used to conduct the study (Creswell, 2008; Vogt, 2007). More specifically, the causal-comparative/quasi-experimental design was used. This design of research attempts to establish cause-effect relationships among the variables. The term *quasi*-experiment means to have some resemblance to a real experiment (Vogt, 2007). Analogous to actual experiments, the experimenter identifies an independent variable but does not manipulate it. The researcher measures effects of the independent variable on the dependent variable. Groups form naturally or pre-exist as no random assignment process is necessary for groups to occur. The researcher identifies control groups and treatment groups, and studies and compares both groups (Green, Camilli, & Elmore, 2006). Typical design of quasi-experiments calls for pretests and posttests, however, this is not required when two groups are considered to be probabilistically equivalent (Vogt, 2007). The current study received IRB approval (see Appendix A) on March 23, 2016.

Advantages to quasi-experiments, or natural experiments, include the elimination of ethical concerns since groups are already formed and treatment is decided by bureaucrats, legislators, physicians, or therapists (Creswell, 2008; Vogt, 2007). These experiments are

typically robust on internal validity (Vogt, 2007). However, the inability to match groups based on similarities, threatens external validity. Therefore, it is critical to match as many similarities as possible among groups allowing for generalizations to be made about a population (Creswell, 2008; Vogt, 2007). Participants were placed in two groups by the legislators and PDE (2015). These groups are Teacher of Record (TOR) and Non-Teacher of Record (NTOR). The TOR group consisted of public school math teachers in grades 4 through 8 and Algebra 1. The NTOR group includes all public school math teachers in grades K through 3 and any high school math other than Algebra 1. Classification of a math teacher as belonging to the TOR (treatment group) or NTOR (control group) follows specifications defined by PDE (2015). Concerning group matching, similarities between the groups included a history of at least a bachelor's degree in education or courses beyond a bachelor's degree in education, a semester of student teaching experience, a valid PA teaching certificate, and math certification for teachers certified in grades 7-12. Teachers were also required to be currently teaching math at a public school either full-time or teach math as a part of their elementary schedule. Similarities between the groups were considered probabilistically equivalent, thereby maximizing external validity (Creswell, 2008; Vogt, 2007). An attempt was made to establish a cause and effect relationship among variables (Green, Camilli, & Elmore, 2006). All participants were considered to have an equal likelihood of exposure to everyday life events, or external variables, and thereby increased internal validity.

Another concern, confounding variables, or lack of control over extraneous variables, creates problems with internal validity, or issues concluding causal relationships among the variables (Creswell, 2008). All participants were Pennsylvania certified Math teachers who are currently teaching math full-time in a PA public school or as a part of their elementary teaching schedule. Math teachers who only work in substitute and long-term substitute positions were

excluded to limit confounding variables as teacher evaluations would differ and they would not have enough time to experience the teacher effectiveness evaluation system. Pre-K teachers were excluded because the ETSI was constructed for use with educators of grades K through 12 only (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015).

All TOR and NTOR completed demographic questions, the Educators Test Stress Inventory (ETSI) (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015), and the Malasch Burnout Inventory Educators Survey (MBI-ES) (Malasch, Jackson & Schwab, 1986). All surveys were distributed via the PSEA email listserv of public school teachers and administered via Qualtrics. Site approval was requested by the researcher (see Appendix B) and permission was granted by the PSEA on March 7, 2016 to distribute a link to the survey via the listserv to a stratified sample of high school, middle school, and elementary school math teachers (See Appendix C). The purpose of the current study was to examine the relationships among the dependent variables of high-stakes testing teacher stress, teacher burnout, and potential teacher attrition plans with the independent variable of direct pressure of high-stakes testing results on the new teacher evaluation system by being a teacher of record.

Research Questions

Research Question 1

Does high-stakes testing as a component of teacher evaluation lead to higher levels of stress, leading to burnout, and eventually attrition.

Research Question 2

Does high-stakes testing as a component of teacher evaluation lead to a correlation among stress, burnout, and attrition?

Research Question 3

Are other factors (gender, school location, school socio-economic status, years of teaching experience, grade level of students) effecting stress, burnout, and teacher attrition?

Hypothesis

To thoroughly compare levels of teacher stress, teacher burnout, and attrition between TOR and NTOR, the following hypotheses were developed and tested. The first three null hypothesis ($H1_0 – H3_0$) were written to answer Research Question 1. The fourth null hypothesis ($H4_0$) was written to answer Research Question 2. The last three null hypotheses ($H5_0 – H7_0$) were written to answer Research Question 3.

$H1_0$: There is no statistically significant difference in stress levels between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

$H1_A$: There is a statistically significant difference in stress levels between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

$H2_0$: There is no statistically significant difference in levels of teacher burnout between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

$H2_A$: There is a statistically significant difference in levels of teacher burnout between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

$H3_0$: There is no statistically significant difference in levels of potential teacher attrition between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

$H3_A$: There is a statistically significant difference in levels of potential teacher attrition between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

$H4_0$: There is no correlation among high-stakes testing teacher stress, teacher burnout, and potential teacher attrition rates.

$H4_A$: There is a correlation among high-stakes testing teacher stress, teacher burnout, and potential teacher attrition rates.

$H5_0$: No difference exists among demographic factors (gender, school location, school socio-economic status, years of teaching experience, grade level of students) and teacher stress.

$H5_A$: A difference exists among demographic factors (gender, school location, school socio-economic status, years of teaching experience, grade level of students) and teacher stress.

$H6_0$: No difference exists among demographic factors (gender, school location, school socio-economic status, years of teaching experience, grade level of students) and teacher burnout.

$H6_A$: A difference exists among demographic factors (gender, school location, school socio-economic status, years of teaching experience, grade level of students) and teacher burnout.

$H7_0$: No difference exists among demographic factors (gender, school location, school socio-economic status, years of teaching experience, grade level of students) and potential teacher attrition rates.

$H7_A$: A difference exists among demographic factors (gender, school location, school socio-economic status, years of teaching experience, grade level of students) and potential teacher attrition rates.

Population

Pennsylvania public school teachers were chosen because math teachers are evaluated differently within the same state. In Pennsylvania, an evaluation for a TOR includes individual student high-stakes testing data and SPP data (PDE, 2015). An evaluation for a NTOR includes only SPP data (PDE, 2015). According to the National Center for Education Statistics (2015), about 250,000 or 7.4% of teachers in the United States are math teachers in public and private schools. Although it appears that 10,104 middle level and high school math teachers were employed in PA public schools during year 2014-2015, an additional group of special education teachers were also reported to teach math increasing the total to 12,359 (7-12) teachers (PDE, 2015). Another misrepresentation to the total reported number of math teachers by the file is the classification of elementary teachers who are not reported as math teachers. Any of the primary, intermediate, or special education elementary teachers could be teaching math as a part of their schedule. This includes an additional 49,048 teachers of whom the majority most likely teach math and the majority of the 20,232 intermediate teachers are most likely TOR for math. Table 1 illustrates the approximate number of current TOR and NTOR math teachers in the state of Pennsylvania as described above.

For the purpose of the study, only current, full-time public school math teachers or elementary teachers scheduled to teach math were invited to participate. Math teachers who only work in substitute and long-term substitute positions were excluded as teacher evaluations would

differ. Pre-K teachers were excluded because the ETSI was constructed for use with educators of grades K through 12 only (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015).

Table 1

Number of TOR and NTOR Math Teachers in PA 2014-2015

	NTOR Elementary Primary (K-3)	TOR Elementary Intermediate (4-6)	**TOR Middle Level Math (7-9)	**NTOR High School Math (10-12)	Total
Elementary and Math Teachers	25,276	17,772	5,069	5,235	53,352
Special Education Math Teachers	*3,540	*2,460	1,172	883	8,055
Total	28,816	20,232	6,241	6,118	61,407

Note. *Amount is estimated. PDE reports 6,075 (PK-6) Special Education elementary teachers who could possibly teach math (2015). The above number of special education teachers assigned to each elementary level was based on the frequency of elementary teachers in the above grade categories minus the frequency of PK (K4) teachers.

**Most Middle Level Math teachers would be classified as TOR except for grade 9 teachers who do not teach Algebra 1. Likewise, most Math (10-12) teachers would be NTOR unless students were enrolled in the course titled Algebra 1.

Sampling

A sample is a representative subset of a larger population (Brase & Brase, 2013). When conducting quantitative research, different methods of sampling are used based on the research circumstances. Sampling methods include stratified sampling, cluster sampling, systematic sampling, simple random sampling, and multistage sampling, among others (Brase & Brase, 2013). These represent types of probability sampling, whereby each item in the population has an equal likelihood of being selected to participate in the sample (Brase & Brase, 2013). For this study, the researcher requested permission from the program director of the Pennsylvania State Education Association (PSEA) to utilize the email listserv of registered public school math teachers in Pennsylvania. The researcher requested to send the survey link that included the

Informed Consent letter to the research department at PSEA who transmits messages via the email listserv. Transmission of surveys through the listserv attendant maintained anonymity of participants and security of listserv email accounts through PSEA. Approval to utilize the listserv as stated was granted by PSEA officers and a permission letter was obtained. The researcher chose stratified random sampling in this circumstance to maximize external validity (Brase & Brase, 2013; Vogt, 2007). The research department was able to use their computer system to extract 4 groups, or strata, of math teachers which included: (10-12) high school, (7-9) middle school, (4-6) elementary intermediate, and (K-3) primary teachers. Equal groups of participants representing the TOR group and the NTOR group of the various 4 strata are desired but unnecessary. The same computer system was able to randomly select equal amounts of participants from within each of the 4 groups, or strata, to survey as shown in Figure 1.

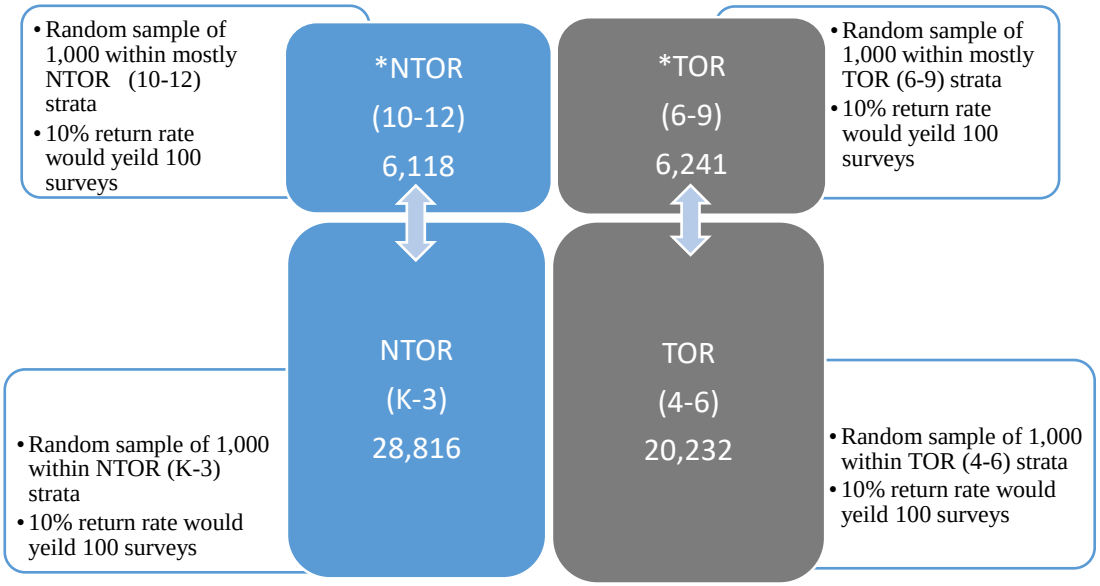


Figure 1. Stratified sampling method of survey population.

The survey was distributed to all randomly selected members on the listserv at the same time, whereby each member of the sample had an equal likelihood of participating in the study (Brase & Brase, 2013). The Informed Consent letter preceded the survey within the link and was

immediately followed by the electronic consent signature question. The second confirmation question stated, "To continue with the survey, you must be currently teaching full-time at a public school as a math teacher or as a middle school or elementary teacher who teaches math as a part of schedule. (Substitute and long-term substitute positions are excluded.) Do the above statements describe you?" If a participant selects 'yes,' the survey continues, however, if the participant selects 'no,' the survey ends.

Most researchers suggest a 95% confidence level and a 5% confidence interval (margin of error) for the social sciences (Creswell, 2008; Shaughnessy, Zechmeister & Jeanne, 2011; Vogt, 2007) when determining sample size. Because there exists possibly 61,407 public school math teachers in Pennsylvania (PDE, 2015) including elementary and special education teachers, a sample size of 382 teacher responses is required for a 95% confidence level and a 5% confidence interval (The Research Advisors, 2015). However, the current quantitative study compares two independent samples whose primary endpoints consist of averages. The following study parameters were used to calculate a sample size using power analysis. The researcher tested various anticipated means and standard deviations for the MBI-ES and the ETSI from prior studies with $\alpha = 0.05$ and $\beta = 0.2$ (Gold, 1984; Iwanicki & Schwab, 1981; Malasch & Jackson, 1986; Malasch, Jackson & Leiter, 1996; Malasch, Jackson & Schwab, 1986; von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015). The estimated enrollment ratio of the groups used was set equal to one as equal groups were desired. The alpha $\alpha = 0.05$ represents the probability of (a false positive) a Type-I error and $\beta = 0.2$, or the power $(1 - \beta) = 0.8$, represents the risk of (a false negative) a Type-II error (Brase & Brase, 2013). The power $(1 - \beta)$ represents the ability to detect actual differences between groups (Brase & Brase, 2013). The various calculations frequently yielded a total sample size less than 400 of equal groups. The researcher

concluded that both the above sample size of 382 and the power analysis calculations were similar and comparable values. Therefore, 382 was accepted as the total sample size. For the purpose of survey collection, this projected approximately 400 surveys, 200 within each TOR and NTOR group. PSEA (2015) reports a 10% return rate for digital surveys. For this reason, 1,000 individuals were randomly selected via the PSEA computer listserv system within each of the four strata. This process of stratified sampling attempted to obtain 100 participants from within each of the four strata, thereby obtaining 200 in each group of TOR and NTOR with varying age ranges within each group.

Informed Consent

According to governing rules set forth by the United States government, humans participating in any research must give their informed consent (Office for Human Research Protections, 2009). The Code of Federal Regulations (CFR) for the protection of human subjects (Title 45, Part 46) requires specific information be available to subjects before they agree to become participants (OHRP, 2009). In compliance with CFR, the researcher constructed a letter of informed consent (see Appendix D) that preceded the survey within the link. If subjects chose to participate, active consent was given electronically by selecting a radio button at the beginning of the Qualtrics survey (see Appendix E). The first question in the survey stated, "By clicking on 'I agree,' you acknowledge that you read and understand the Informed Consent Form, and you are consenting to participate in this research study. Clicking 'I agree' on this survey represents your electronic signature of consent to participate." If a participant selects 'I agree,' the survey commences. If a participant selects "I do not agree," the survey is ended.

The Notice of Informed Consent stated that the study involved research and detailed who was eligible to participate. The letter contained a statement that participation was voluntary, and

the participant could withdraw at any time without penalty (OHRP, 2009). It also explained the purposes of the research, identified the expected duration of participation, and described procedures. The letter indicated the research involved "minimal risk," whereby no foreseeable risks, or discomforts to participants were anticipated. Possible benefits of the research to the subject or others were described (OHRP, 2009). Contact information was given if participants had questions or wanted to withdraw. Potential subjects were informed that their identity would remain anonymous to the researcher and notified that research information would be kept on file, saved to a thumb drive, and stored in a safe at the researcher's home for three years before being destroyed in accordance with federal guidelines (OHRP, 2009).

Anonymity

Maintaining anonymity among research participants increases the likelihood of obtaining truthful responses on surveys (Brase & Brase, 2013). In this study, participants were asked to enter two pieces of identifying information, known only to the participants, for the purpose of withdrawal. First, participants were asked to enter a pseudonym or code name. The pseudonym could be any combination of letters, numbers, or symbols that could be used to identify participants for withdrawal purposes. The second requested piece of information was a date of birth in the event that two individuals provided the same pseudonym. The researcher did not know actual names of participants, only code names and dates of birth ensuring anonymity and no coercion. Surveys were forwarded by the PSEA listserv attendant at the request of the researcher to protect anonymity and privacy of membership email addresses.

Geographic Location

The purpose of this quantitative study was to examine the impact that stress related to the inclusion of student standardized testing data in teacher evaluation has on burnout and potential

attrition rates among current Pennsylvania math teachers who identify as either Teachers of Record (TOR) or Non-Teachers of Record (NTOR). The study participants were current PA public school math teachers, grades K through 12, who were located throughout the state of Pennsylvania. Survey links were distributed via an email listserv of math teachers through PSEA using Qualtrics. Site approval was requested by the researcher. Approval was granted and a permission letter was obtained by the elected officers at the Pennsylvania State Education Association (PSEA) to utilize the email listserv to distribute surveys to public school math teachers registered with PSEA.

Data Collection

Data collection in educational, quantitative research often involves the use of surveys (Creswell, 2008). Three general types of survey research design include cross-sectional studies, successive independent samples, and longitudinal studies (Creswell, 2008; Shaughnessy, Zechmeister & Jeanne, 2011). A cross-sectional survey collects data from participants at one time (Creswell, 2008), that is, participants answer questions and click submit to send survey data to the researcher. The cross-sectional survey design is a type used with causal-comparative/quasi-experimental design (Creswell, 2008), and the survey design chosen for the study.

The current study collected data using three cross-sectional surveys. The first survey was a demographic survey, designed to classify participants for the purpose of the study (see Appendix F). The second and third studies were previously developed. The ETSI was developed by von der Embse, Kilgus, Solomon, Bowler, and Curtiss in 2015 (see Appendix G), and the MBI-ES was developed by Malasch, Jackson, and Schwab in 1986 (see Appendix H). All surveys utilized closed-ended questions, standardizing data (Shaughnessy, Zechmeister &

Jeanne, 2011). The previously developed surveys were constructed in a Likert-type format. Before the onset of data collection, permission was obtained to use the ETSI and the MBI-ES from the developers of each (see Appendices I-J). Approval was also granted by the Internal Review Board at Indiana University of Pennsylvania prior to data collection. On average, participants spent a total of 5 minutes reading the letter of Informed Consent and 9 minutes completing the demographic survey, ETSI, and MBI-ES. All surveys were distributed via the Pennsylvania State Education Association (PSEA) email listserv using a link to Qualtrics. Data collection began on May 26, 2016 and was set to cease upon receiving 400 survey submissions or July 8, 2016. Three email reminders were sent following the original email. Survey collection was set to suspend electronically upon receiving the 400th survey as 400 license agreements (see Appendix H) were purchased by the researcher for use of the MBI-ES for $288. Although a sample size of 382 was required for a 95% confidence level and a 5% confidence interval (Creswell, 2008; Shaughnessy, Zechmeister & Jeanne, 2011; Vogt, 2007), this number is based on an estimated population. A total of 293 complete surveys were collected.

Instrumentation

Quantitative research attempts to use numerical data and statistical analysis to reach unbiased conclusions about a population (Brase & Brase, 2013). In the current study, closed-question surveys were used to collect numerical data. A review of the literature revealed two previously constructed instruments used for teacher stress and burnout. The Educator Test Stress Inventory (ETSI) was used to collect data about teacher stress relevant to the educator effectiveness evaluation that included student performance data (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015). The Malasch Burnout Inventory Educators Survey (MBI-ES) was used to collect data relevant to teacher burnout (Malasch, Jackson & Leiter, 1996;

Malasch, Jackson & Schwab, 1986). Another survey of the cross sectional study contained only demographic questions for the purpose of classifying participants into TOR and NTOR groups and obtaining information regarding potential teacher attrition plans in a Likert-type format question.

The *Educator Test Stress Inventory (ETSI)* (see Appendix G) is a new multidimensional instrument developed to measure teacher stress and anxiety related to policy change and educator effectiveness evaluations that include student performance data from high-stakes tests (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015). The ETSI is developed for public school teachers in grades K through 12. The final version contains 11 items measuring Source of Stress (ETSI-S) and Manifestations of Stress (ETSI-M). The ETSI-S subscale consists of 5 items measuring sources of stress and anxiety. The score range is 5-25 with a mean of 17.25, and a standard deviation (SD) of 4.43. The ETSI-M subscale measures psychological and emotional symptoms. The score from these six items ranges from 6-30 with a mean of 14.94, and a SD of 5.01. The total score (ETSI-T) ranges from 11-55 with a mean of 32.18, and a SD of 8.51. Teachers self-report on this Likert-type scale. Responses worth 1 to 5 points range from 1 indicating "very much disagree" to 5 indicating "very much agree" (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015).

The *Malasch Burnout Inventory Educators Survey (MBI-ES)* (see Appendix H), recognized as the leading measure of burnout, derived from the original Malasch Burnout Inventory (Malasch & Jackson, 1986; Malasch, Jackson & Leiter, 1996; Malasch, Jackson & Schwab, 1986). Incorporating 25 years of research, the MBI-ES includes three scales: emotional exhaustion, depersonalization, and personal accomplishment. The format consists of a 22 item survey constructed of a 7 point Likert-type responses ranging from 1 indicating "never" to 7

indicating “every day.” 9 of 22 items measure emotional exhaustion. Five items measure depersonalization and eight items measure personal accomplishment. High subscale scores on emotional exhaustion and depersonalization, combined with low subscale scores on personal accomplishment, indicate burnout (Malasch, Jackson & Leiter, 1996).

Validity

Validity refers to the truth or accuracy of the research and falls under three main types: content validity, criterion-related validity, and construct validity (Vogt, 2007). Content validity refers to whether the instrument is measuring what it is supposed to measure. Criterion-related validity refers to usefulness of predictive validity or concurrent validity of an external criterion (Vogt, 2007). Construct validity refers to how well an instrument corresponds to other variables as predicted by a theory or rationale (Chalghaf, Azaiez, & Elarbi, 2014; Vogt, 2007). Threats to validity can occur through design, measurement, or analysis.

A systematic content validation process was used to create the *Educator Test Stress Inventory (ETSI)*. Initially, 20 items were constructed based on intensive review of literature. These items were sent to eight content experts and rated for readability, clarity, and importance. Revisions were sent for re-evaluation by experts until 15 items remained. The STAI, an assessment measuring both state and general anxiety, was used to assess construct-related validity. The ETSI received a median alpha coefficient of .90 for general anxiety and a median alpha coefficient of .93 for state anxiety (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015). The ETSI is the first assessment of teacher stress related to teacher evaluation and high-stakes testing. Evidence is limited to confirm factorial stability, face validity, and convergent validity. Predictive validity and concurrent validity will need to be examined over time with future research (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015).

Validity for the *Malasch Burnout Inventory Educators Survey (MBI-ES)* has been authenticated similarly to the original Malasch Burnout Inventory. Scores for individuals were correlated with job characteristics associated with burnout and with measures conjectured to be burnout-related to confirm convergent validity. Scores of the individuals were also correlated with ratings derived from persons who knew the individuals well (Malasch & Jackson, 1986; Malasch, Jackson & Leiter, 1996; Malasch, Jackson & Schwab, 1986). Regarding validity of instrument, data further demonstrated high workload correlated with high emotional exhaustion scores, high depersonalization scores, and low personal accomplishment scores (Malasch & Jackson, 1981). Gold also used a factor analytic (1984) in a study of 462 teachers to support using the 3 factor structure. The MBI-ES was also validated in study of 469 teachers by Iwanicki and Schwab (1981).

Reliability

Reliability refers to the consistency of research design or measurement over time, which is consistency with replication (Vogt, 2007). There are four classes of reliability: inter-observer reliability, test-retest reliability, parallel forms reliability, and internal consistency reliability (Vogt, 2007). Inner-observer reliability, or inner-rater reliability, refers to the consistency of estimates of the same phenomena given by different raters or observers (Vogt, 2007). Test-retest reliability describes the measurement consistency from one study to another (Vogt, 2007). Parallel-forms reliability is used to distinguish consistency between results of two tests created to measure the same content in similar ways (Vogt, 2007). Internal consistency reliability evaluates consistency of items within a test by calculating correlation coefficients of internal items (Vogt, 2007). Reliability coefficients, or correlation coefficients, range from 0.0 (completely inconsistent) to 1.0 (entirely consistent) reliability (Vogt, 2007).

For the *Educator Test Stress Inventory (ETSI)* the researchers calculated alpha coefficients to evaluate internal consistency and convergent reliability (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015). The following results were obtained: ETSI-M = .85, ETSI-S = .82, and ETSI-T = .89. Bivariate correlation was used to compare each ETSI with all STAI assessments. All ETSI-STAI correlations were statistically significant. Correlations were highest for the ETSI-T subscale with all STAI subscales (general, state, and total anxiety). This total ETSI was also found to be the best predictor of all three STAI subscales. The ETSI-M and ETSI-S were equally predictive and correlated with the three STAI subscales.

The three *Malasch Burnout Inventory Educators Survey (MBI-ES)* subscales: emotional exhaustion, depersonalization, and personal accomplishment, were evaluated for internal consistency reliability in three separate studies (Malasch, Jackson & Leiter, 1996; Wheeler, Vassar, Worley & Barnes, 2011). The initial Cronbach alpha scores, or correlational coefficients, are as follows: emotional exhaustion was .90, depersonalization was .76, and personal accomplishment was .76 (Iwanicki & Schwab, 1981). The second study reported Cronbach alpha scores of .88 for emotional exhaustion, .74 for depersonalization, and .72 for personal accomplishment (Gold, 1984). A third study by Wheeler, Vassar, Worley, and Barnes (2011) calculated Cronbach alpha scores of .90 for emotional exhaustion, .79 for depersonalization, and .71 for personal accomplishment.

Data Analysis

The current study used the Statistical Package for the Social Sciences (SPSS) software to calculate all statistical analysis. Descriptive statistics, including frequencies and percentages, were computed for the following demographic variables: (a) gender, (b) years of math teaching experience (public or private school), (c) level of students taught (elementary, middle, or

secondary), (d) school classification (rural, suburban, city) with predominant income levels, and (e) teacher classification (TOR or NTOR). Ranges, means, and standard deviations were calculated for the Educator Test Stress Inventory, the Malasch Burnout Inventory – Educator Survey, and the one question regarding potential teacher attrition (Brase & Brase, 2013; Lund & Lund, 2015).

The first three null hypotheses ($H1_0$ – $H3_0$) were developed to compare Teachers of Record with Non-Teachers of Record on the three interval-level composite scales. To that end, independent samples *t*-tests were calculated unless assumptions were violated for tests of normality and outliers were identified. Otherwise, Mann-Whitney U, non-parametric tests, were used for testing the null hypotheses because the surveys contained Likert scale items. The use of independent samples *t*-tests allows for more powerful analysis of ordinal data as if it were ratio data while comparing means, and the Mann-Whitney U compares medians, a different measure of central tendency (Brase & Brase, 2013; Lund & Lund, 2015).

The fourth null hypothesis ($H4_0$) was developed to identify if any correlation exists among high-stakes testing stress, burnout, and potential attrition of teachers. The Pearson product-moment correlation generated a correlation coefficient *r* between two of the three variables at a time. This design allowed for use of the Spearman Correlation coefficient r_s because the conditions of linearity and normality were not met (Lund & Lund, 2015).

The last three null hypotheses ($H5_0$ – $H7_0$) were written to determined how much the variation of the dependent variable can be explained by the factors described in the literature review. For one evaluation, the Independent samples *t*-tests could not be calculated because Levene's test for equality of variances assessed a violation of the assumption of homogeneity of variances. In that case, a subsequent Welch t-test for unequal variances was run (Lund & Lund,

2015). In all other cases, an attempt was made to conduct a one-way ANOVA. If statistical assumptions were not met, meaning outliers were present and data was not normally distributed, a subsequent non-parametric test, the Kruskal-Wallis H test, was run to detect statistically significant differences (Lund & Lund, 2015)

Summary

In the current education environment, teacher effectiveness evaluations contain student performance data. Research discloses external pressure on teachers to improve student performance diminishes autonomy for both students and teachers (Deci & Ryan, 1985). In turn, these students perform more poorly on high-stakes tests (Elliot & Dweck, 2007). These external demands of high-stakes testing are also linked to loss of autonomy, diminished competence, and diminished relatedness for teachers as effectiveness evaluations depend on improved student performance data (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Fernet, Guay, Senecal & Austin, 2012; Skaalvik & Skaalvik, 2009). Several studies reveal thwarting of autonomy, competence, and relatedness further correlates with increased teacher stress, burnout, and attrition (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Bouwma-Gearhart, 2010; Skaalvik & Skaalvik, 2009).

This study compared teacher high-stakes testing stress, burnout, and potential plans of attrition between Teachers of Record (TOR) and Non-Teachers of Record (NTOR) among the population of public school math teachers, grades K through 12, in Pennsylvania. Teachers of Record (TOR) are those math teachers who are accountable for individual standardized test score results of their students, with the high-stakes testing scores counting as 15% of their teacher evaluation (Doherty & Jacobs, 2013; Pennsylvania Department of Education, 2014). These teachers differ from math teachers classified as Non-Teachers of Record (NTOR) who are not

accountable for standardized testing scores of their students on teacher evaluations. TOR evaluations can be lowered when student test results are recorded, potentially resulting in dismissal from teaching and subsequent loss of income (Doherty & Jacobs, 2013; Pennsylvania Department of Education, 2014). This threat to only the TOR can increase chronic teacher stress. These threats generate feelings of discouragement and inequity (Malasch, Schaufeli, & Leiter, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011; Tye & O'Brien, 2002), while also increasing potential for greater rates of burnout and teacher attrition (Chang, 2009). This study is important because classification as a TOR or NTOR occurs randomly by teaching assignment. Comparison of both groups may reveal differences in high-stakes testing stress, burnout, and attrition relative to student performance data used to evaluate teacher effectiveness in an environment of a growing teacher shortage.

CHAPTER 4

DATA AND ANALYSIS

The purpose of this quantitative study was to examine the impact that stress related to the inclusion of student performance data in teacher evaluations has on burnout and potential attrition rates among current Pennsylvania math teachers. Chapter 1 contains an introduction to the study, Chapter 2 details a review of the literature, and Chapter 3 outlines research procedures. Chapter 4 includes results of the data analyses of the three research questions by examination of the seven hypotheses. The chapter begins with descriptive statistics of the demographic information and follows with results for research question 1, research question 2, and research question 3. The chapter concludes with a summary.

Descriptive Statistics

Total responses included 330 with 37 individuals exiting the study before completing the survey information. A total of 293 math teachers participated in the study for a 91.35% confidence level and a margin of error of ± 5.71%, indicating the data can be trusted 91.35% of the time within ±5.71% of the response. Table 2 contains descriptive statistics and demographic information of participants. Of these, 70.6% were female, 28.7% were male, 0.3% responded other, and 0.3% declined to answer. Most participants had 11 to 20 years of teaching experience (46.1%), followed by similar years of experience in the 6 to 10 (18.8%) and 11 to 20 (18.1%) year categories. Only 11.9% of participants teach at high-income schools while the others are split fairly evenly with 46.1% at middle-income schools and 42% at low-income schools. The majority of participants teach in suburban schools (51.5%) followed by rural schools (34.8%) and urban schools (13.7%). Grade levels were somewhat evenly represented with 27.3% elementary teachers (K-5), 38.2% middle school teachers (6-8), and 34.5% secondary teachers

(9-12). The majority of participants, 70.3% reported to be a Teacher of Record (TOR), and only 29.7% reported to be a Non-Teacher of Record (NTOR).

Table 2

Descriptive Statistics for Sample Demographic (N = 293)

	Frequency	Percent
Gender		
Female	207	70.6
Male	84	28.7
Other	1	.3
Decline to answer	1	.3
Total	293	100.0
Years of Teaching Experience		
1 to 5	38	13.0
6 to 10	55	18.8
11 to 20	135	46.1
21 to 30	53	18.1
31 +	12	4.1
Total	293	100.0
School Classification		
Low-income	123	42.0
Middle-income	135	46.1
High-income	35	11.9
Total	293	100.0
School Location		
Rural	102	34.8
Suburban	151	51.5
Urban	40	13.7
Total	293	100.0
Grade Level		
Elementary (K-5)	80	27.3
Middle School (6-8)	112	38.2
Secondary (9-12)	101	34.5
Total	293	100.0
Teacher of Record		
Yes	206	70.3
No	87	29.7
Total	293	100.0

Research Question 1

Does high-stakes testing as a component of teacher evaluation lead to higher levels of stress, leading to burnout, and eventually attrition?

Hypothesis 1: Stress Levels between TOR and NTOR

The first null hypothesis and alternate hypothesis were established for examination of the following piece of the first research question. Is there a statistically significant difference in teacher stress levels between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system?

$H1_0$: There is no statistically significant difference in stress levels between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

$H1_A$: There is a statistically significant difference in stress levels between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

There were 203 TOR and 87 NTOR participants. The ETSI consists of 3 scores: the Manifestations of Stress (ETSI-M) subscale, the Sources of Stress (ETSI-S) subscale, and the Total (ETSI-T) subscale (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015). The study design included a continuous dependent variable, a two-group categorical independent variable, and independent observations.

The visual boxplot assessment (see Figure 2) of the ETSI-M subscale scores revealed an outlier for the NTOR group and the assumption of normality was violated for the TOR group. For these reasons the Mann-Whitney U, a non-parametric test, was used to determine if statistically significant median differences existed between the two groups (Lund & Lund, 2015). Distributions of ETSI-M subscale median scores for TOR and NTOR were similar, as assessed by visual inspection. The difference of median ETSI-M subscale scores for TOR (15.5) and

NTOR (16.0) was not statistically significant, $U = 9{,}465$, $p = .445$. Based on this distribution, the researcher failed to reject the null hypothesis and failed to accept the alternate hypothesis. Data indicate the two groups experience no real difference in Manifestations of Stress (e.g., feeling anxious, increased heart rate, sweating) as it is not statistically significant (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015).

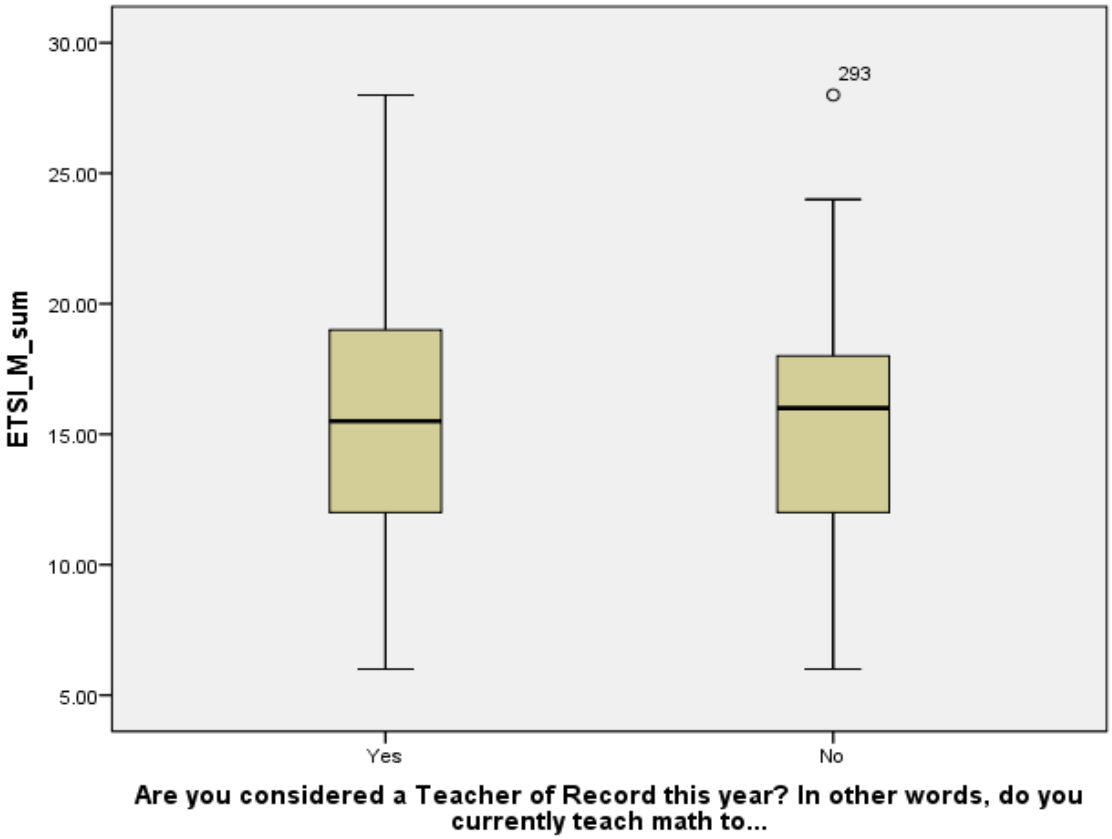


Figure 2. Boxplots of ETSI-M scores for TOR and NTOR. An outlier is visible for the NTOR group (IBM Corp., 2015).

A visual inspection of boxplots for the ETSI-S subscale scores revealed outliers for both the TOR and NTOR groups (see Figure 3). The assumption of normality was also violated for both groups, and the Mann-Whitney U was run to determine if statistically significant median differences existed between the two groups (Lund & Lund, 2015). The visual assessment

revealed similar median ETSI-S subscale scores for TOR and NTOR. The difference of median ETSI-S subscale scores for TOR (19.0) and NTOR (19.0) was not statistically significant, U = 8,024, p = .156. Based on this distribution, the researcher failed to reject the null hypothesis and failed to accept the alternate hypothesis. Data indicate the two groups experience no real difference in Sources of Stress (i.e., pressure from the administration, pressure from parents) as it is not statistically significant (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015).

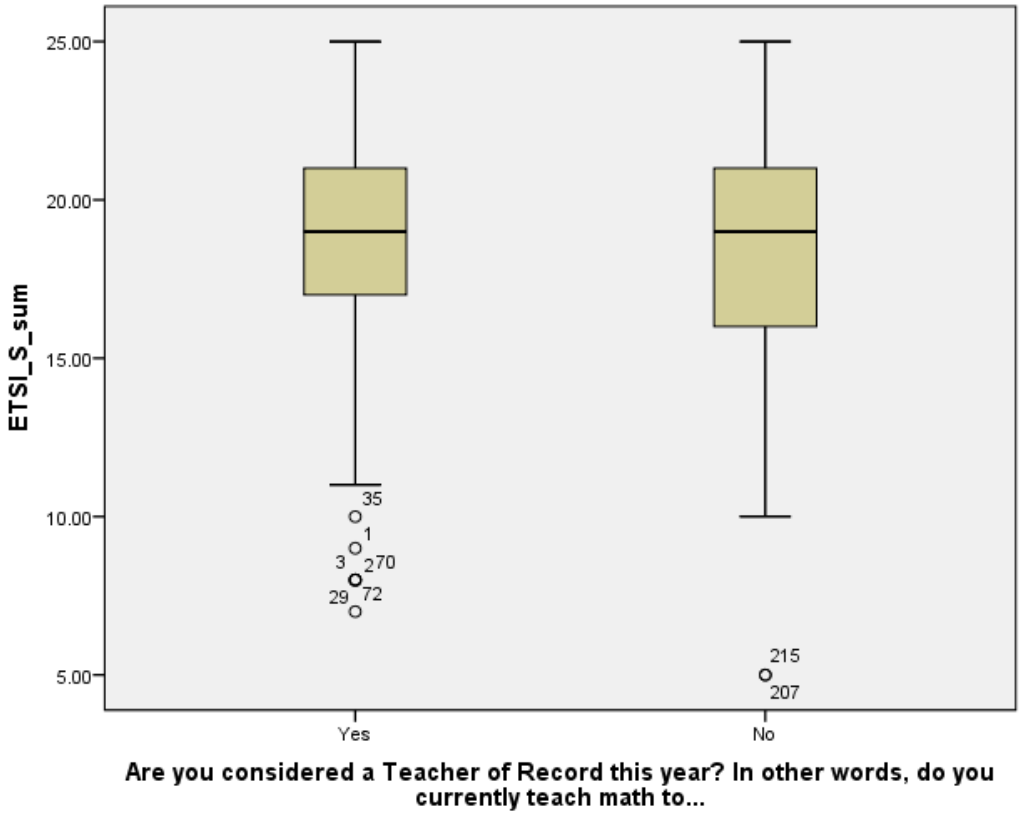


Figure 3. Boxplots of ETSI-S for TOR and NTOR. Outliers are visible on both boxplots for both groups (IBM Corp., 2015).

A visual assessment of the ETSI-T boxplots in Figure 4 revealed no outliers. ETSI-T scores were normally distributed for both TOR and NTOR, as assessed by the Shapiro-Wilk's test (p > .05). Homogeneity of variance was assessed by Levene's test for equality of variances (p = .961). Therefore, assumptions were met, and an independent-samples t-test was run to

determine if a statistically significant difference existed between TOR and NTOR regarding average educator test stress levels (Lund & Lund, 2015). The ETSI-T mean ± standard deviation was TOR (34.06 ± 7.26) and NTOR (33.79 ± 7.24). The level of significance (p = .771) indicates a statistically significant mean difference does not exist. The researcher, thereby, failed to reject the null hypothesis and rejected the alternate hypothesis. The data show no real statistically significant difference in higher levels of total high-stakes testing stress scores for either group.

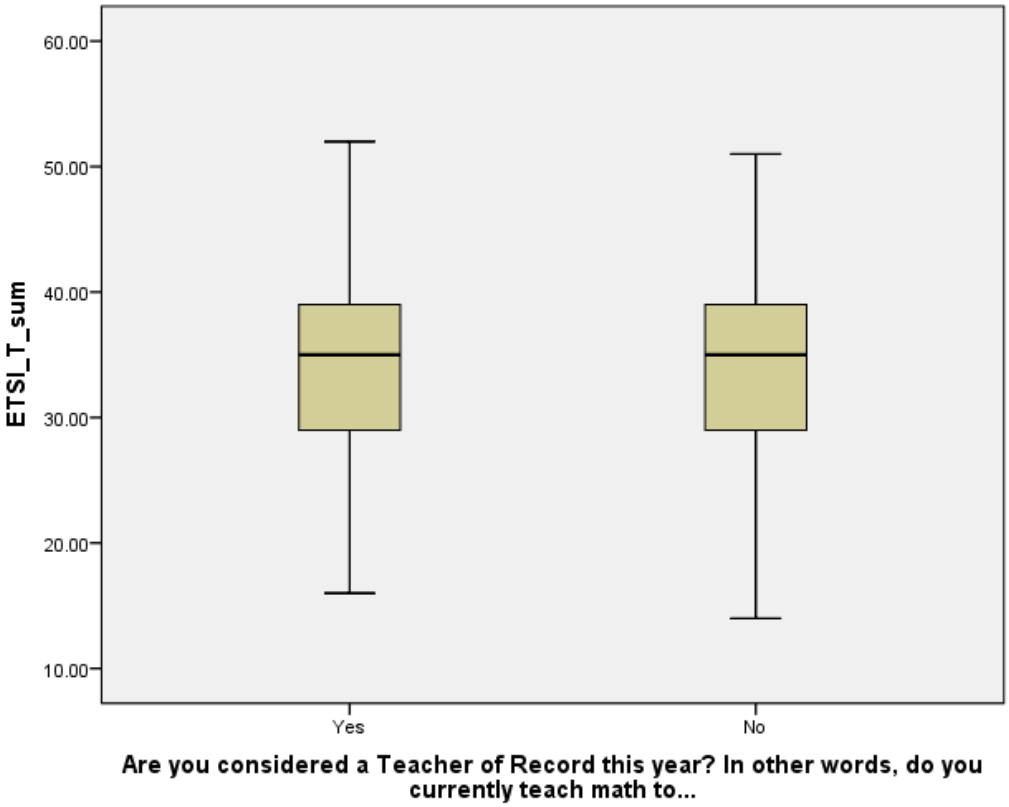


Figure 4. Boxplots of ETSI-T scores for TOR and NTOR. No outliers are visible (IBM Corp., 2015).

Hypothesis 2: Levels of Teacher Burnout between TOR and NTOR

The second null hypothesis and alternate hypothesis were written to evaluate this part of the first research question. Is there a statistically significant difference in levels of teacher

burnout between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system?

$H2_0$: There is no statistically significant difference in levels of teacher burnout between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

$H2_A$: There is a statistically significant difference in levels of teacher burnout between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

According to the US National Library of Medicine (2016), no precise scientific definition or diagnosis of burnout exists. Likewise, no valid, standardized procedure or instrument has been internationally accepted to diagnose burnout from a medical perspective. However, the MBI remains the most common self-assessment questionnaire for scientific research, not a clinical diagnosis. The MBI-ES consists of three subscales: Emotional Exhaustion (EE), Depersonalization (DP), and Personal Accomplishment (PA). High scores in EE and DP combined with low scores in PA indicate burnout (Malasch & Jackson, 1986; Malasch, Jackson & Leiter, 1996; Malasch, Jackson & Schwab, 1986). Most researchers agree on burnout consisting of three distinct constructs with continuous measures. However, calculating one total score into a single variable is deemed unreliable (Malasch, Leiter & Scaufeli, 2008). Still, comparing individual scales between groups yields an incomplete picture as burnout is considered a construct of all three. For these reasons, statistical cut-off points from the normative distribution in thirds range from "low," "moderate" and "high" to classify scores (Malasch, Jackson & Leiter, 1996). Total Emotional Exhaustion (EE) scores range from Low (0-16), Moderate (17-26), and High (27 and over). Scores for Depersonalization (DP) rate Low (0-

8), Moderate (9-13), and High (14 and over). Personal Accomplishment (PA) scores interpret in reverse and scale as follows: Low (0-30), Moderate (31-36), and High (37 or over) as noted in Table 3 (Malasch, Leiter & Scaufeli, 2008).

Table 3

Frequency Categorization: Malasch Burnout Inventory - Educators Survey

Frequency Category	Emotional Exhaustion (EE)	Depersonalization (DP)	*Personal Accomplishment (PA)
High	27 and over	14 and over	37 and over (48)
Moderate	17-26	9-13	31-36
Low	0-16	0-8	0-30

Note. *Scores for Personal Accomplishment are interpreted in reverse. "High" PA counteracts burnout, whereas "high" EE correlates with burnout.

In the clinical realm, no exact code or standard of diagnosis exists, leaving this up to the discretion of the clinician to distinguish between burnout and other disorders such as depression or anxiety. Despite the MBI's acceptance for scientific research, not a clinical diagnosis, some health professionals have adopted its use as accepted practice to help distinguish burnout from other diagnoses (Malasch, Leiter & Scaufeli, 2008). A recent study by Kokkinos (2006) found EE to be the most robust measure of burnout and the first dimension to appear. Coping with excessive demands then leads to DP to survive the situation, followed by recognition of the detachment and feelings of inadequacy concerning job performance or decreased PA (Kokkinos, 2006).

The Emotional Exhaustion (EE) subscale boxplots in Figure 5 revealed an outlier for the NTOR group upon visual inspection. The TOR and NTOR groups both violated the assumption of normality and the Mann-Whitney U test was used to determine if statistically significant median differences existed (Lund & Lund, 2015). The visual assessment revealed similar graph and median EE subscale scores for TOR and NTOR. The difference of median EE subscale

scores for TOR (33) and NTOR (37) was not statistically significant, $U = 9{,}948$, $p = .136$. Based on this distribution, the researcher failed to reject the null hypothesis and failed to accept the alternate hypothesis. Thereby, the two groups experience no real difference in emotional exhaustion scores from a statistically significant comparison of values.

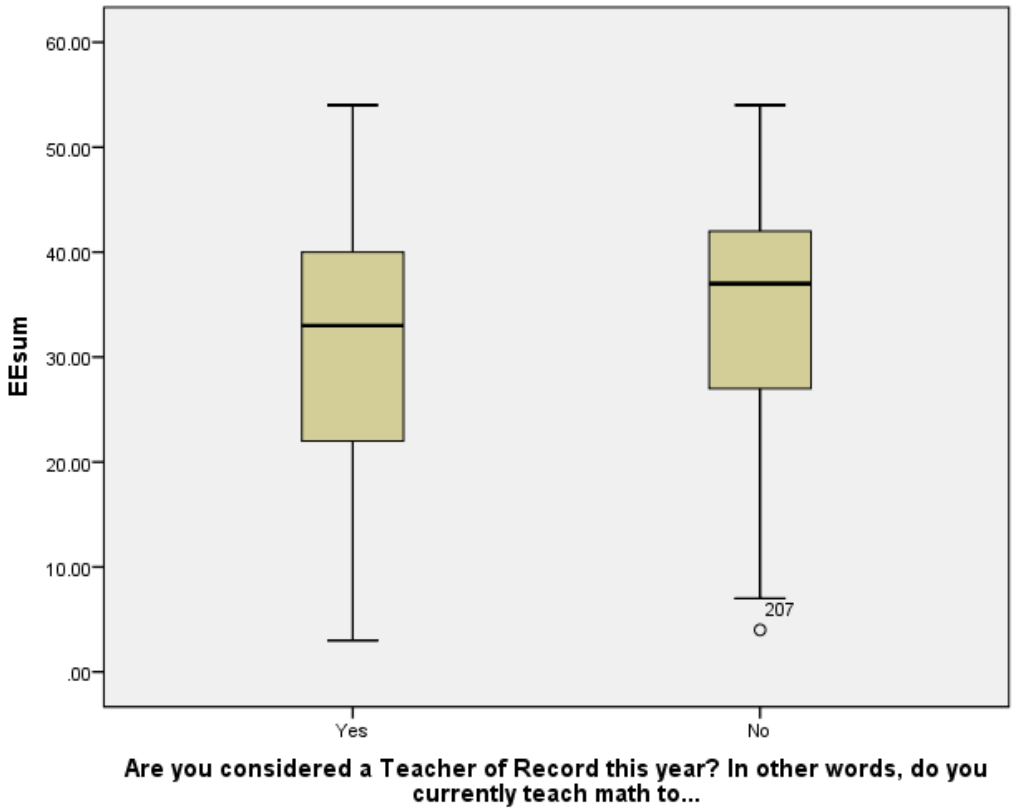


Figure 5. Boxplots of EE subscale scores for TOR and NTOR. An outlier is visible for the NTOR group (IBM Corp., 2015).

A visual inspection of boxplots for the Depersonalization (DP) subscale scores revealed an outlier for the TOR group as seen in Figure 6. The assumption of normality was also violated for both groups, and the Mann-Whitney U test was run (Lund & Lund, 2015). Visual inspection showed similar medians and graphs of DP subscale scores for both groups. The difference of median DP subscale scores for TOR (9) and NTOR (8) was not statistically significant, $U = 9{,}094$, $p = .841$. Based on this distribution, the researcher failed to reject the null hypothesis and

failed to accept the alternate hypothesis. There exists no real statistically significant difference in the median depersonalization scores for both groups.

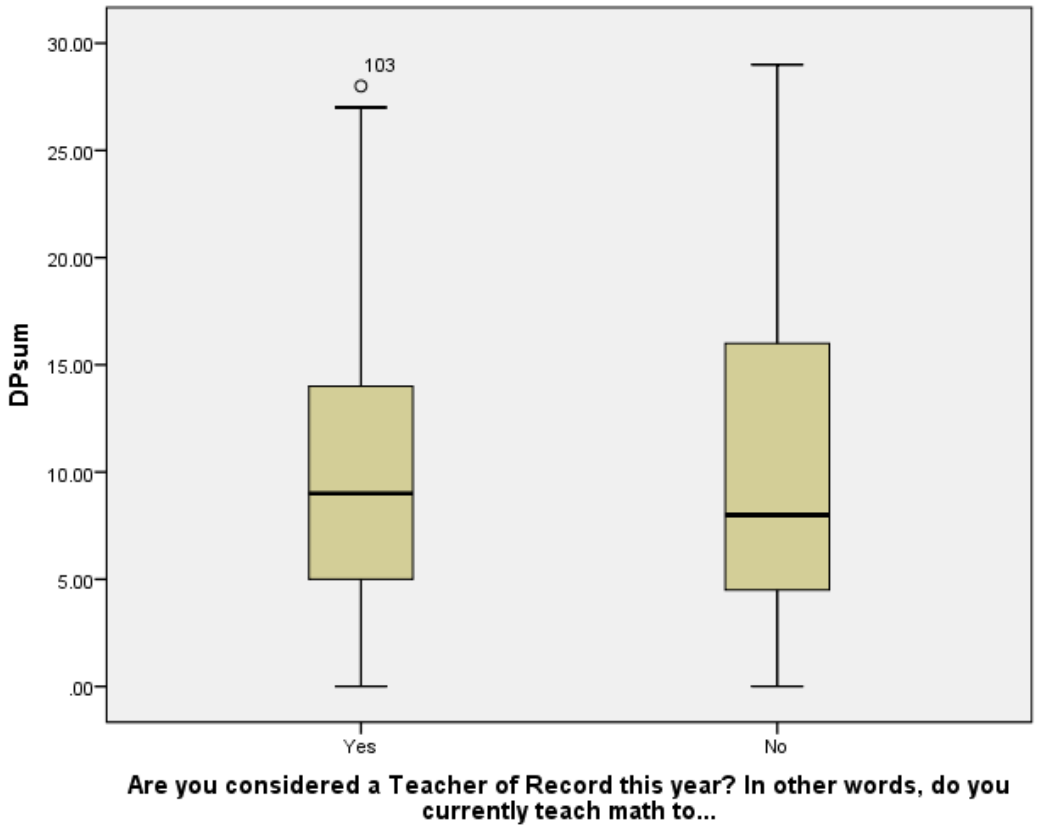


Figure 6. Boxplots of DP subscale scores for TOR and NTOR. A visible outlier exists for the TOR group (IBM Corp., 2015).

A visual inspection of boxplots for the Personal Accomplishment (PA) subscale scores showed an outlier for the TOR group (see Figure 7) and the assumption of normality was violated for both groups. The Mann-Whitney U was run to determine if statistically, significant median differences existed (Lund & Lund, 2015). Similar median PA subscale scores were visually assessed for the groups. The difference of median PA subscale scores for TOR (37.5) and NTOR (38.0) was not statistically significant, $U = 9{,}263$, $p = .648$. Based on this distribution, the researcher failed to reject the null hypothesis and failed to accept the alternative

hypothesis. There exists no statistically significant difference, or no real difference, in personal achievement scores between the groups.

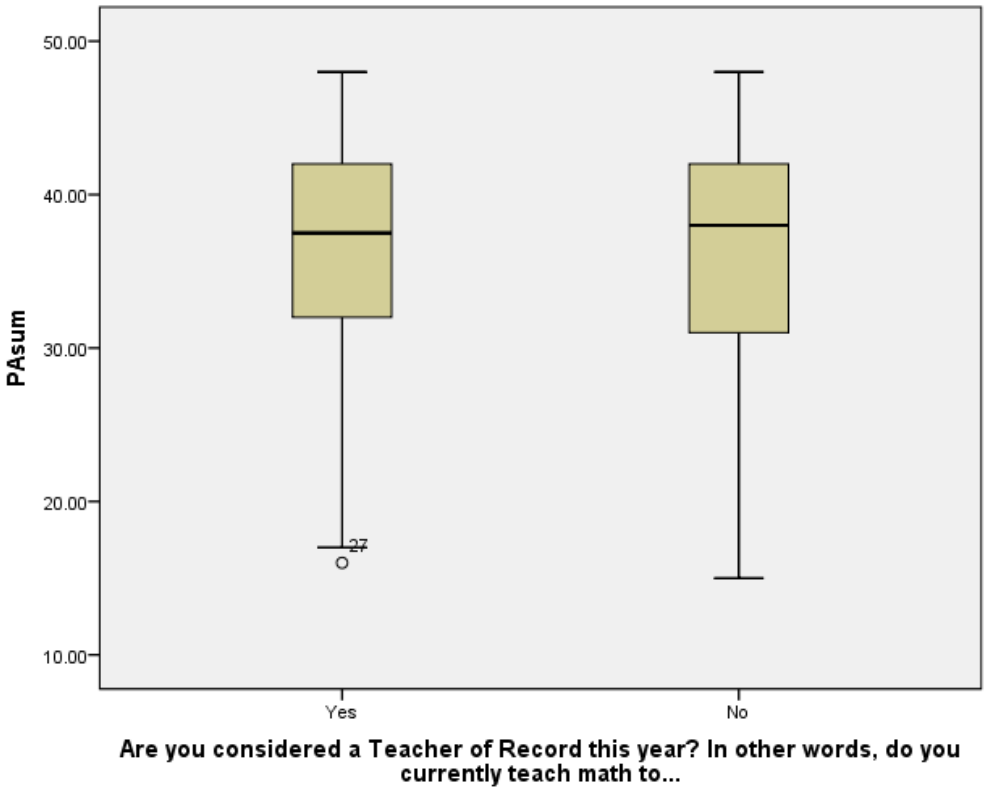


Figure 7. Boxplots of PA subscale scores for TOR and NTOR. A visible outlier exists for the TOR group (IBM Corp., 2015).

Mean scores and the standard deviation for each subscale category were calculated for the TOR group, NTOR group, and combined group of PA public school math teacher participants for comparison with the U.S. teacher sample from 1996. The information can be found in Table 4. Average Emotional Exhaustion (EE) scores for all groups of PA teacher participants fell in the "high" category compared to the 1996 sample of U.S. teachers that averaged scores classified in the "moderate" category. Of all subscales, Emotional Exhaustion has been most closely associated with burnout.

Table 4

MBI-ES: PA Teacher Sample Mean Data Comparison with U.S. Teacher Sample of 1996

	Emotional Exhaustion	Depersonalization	Personal Accomplishment
TOR Mean	31.58	9.85	36.15
SD (N=206)	11.43	6.43	7.14
NTOR Mean	33.68	10.23	36.52
SD (N=87)	11.47	7.10	7.31
Total PA Mean	32.20	9.96	36.26
SD (N=293)	11.46	6.63	7.18
U.S. Sample Mean	*21.25	11.00	33.54
SD (N=4163)	11.01	6.19	6.89

Note. All mean EE scores for PA math teacher participants fell in the high range compared to the U.S. sample of 1996 in the *moderate range.

Hypothesis 3: Levels of Potential Teacher Attrition between TOR and NTOR

The third null hypothesis and alternate hypothesis were designed to analyze this piece of the first research question. Is there a statistically significant difference in levels of potential teacher attrition between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system?

$H3_0$: There is no statistically significant difference in levels of potential teacher attrition between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

$H3_A$: There is a statistically significant difference in levels of potential teacher attrition between TOR and NTOR regarding high-stakes testing and the new teacher evaluation system.

The design included a continuous dependent variable, a two-group categorical independent variable, and independent observations. The assumption of normality was violated, so a non-parametric test was used. A Mann-Whitney U test was run to determine if there were

differences in potential attrition median scores between TOR and NTOR (Lund & Lund, 2015). Distribution of potential attrition scores between TOR and NTOR were similar, as assessed by visual inspection of the histogram in Figure 8. The difference of median potential attrition scores for TOR (29.50) and NTOR (30.00) was not statistically significant, $U = 9{,}257$, $p = .655$. Based on this distribution, the researcher failed to reject the null hypothesis and failed to accept the alternative hypothesis. In conclusion, the data show no statistically significant differences in potential attrition rates between TOR and NTOR. Both groups consist of teachers responding similarly. Responses ranged between thinking of leaving 0% of the time to 100% of the time in both groups.

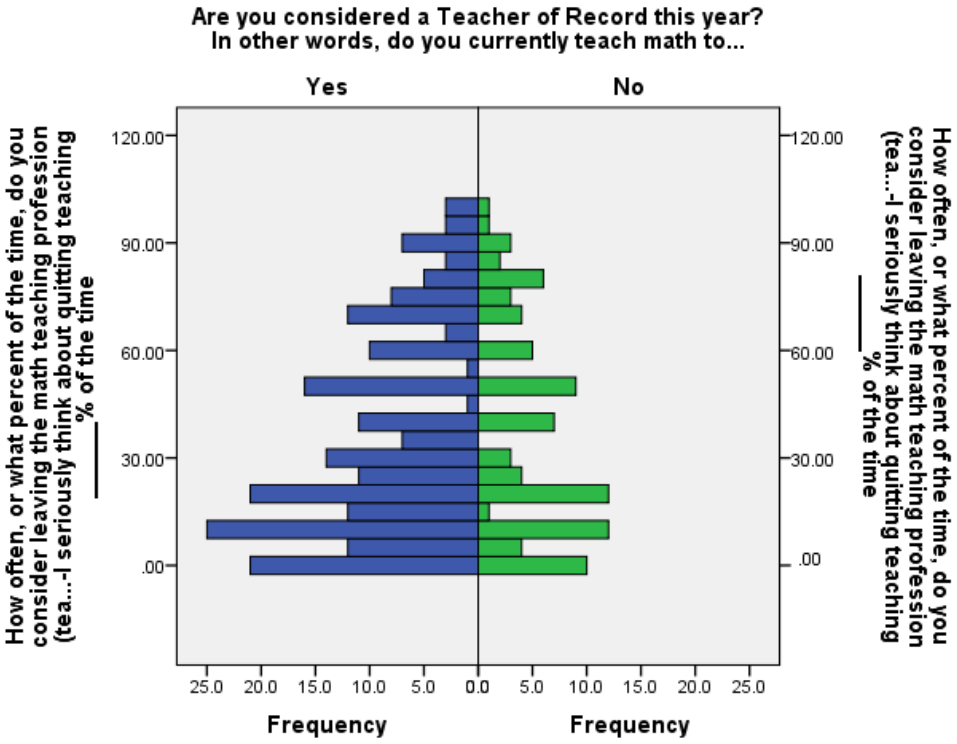


Figure 8. Histogram of potential attrition rates for TOR and NTOR. Although the TOR group in this study consists of more teachers, the responses are similarly distributed on each side (IBM Corp., 2015).

Research Question 2

Does high-stakes testing as a component of teacher evaluation lead to a correlation among stress, burnout, and attrition?

Hypothesis 4: Correlation among Testing Stress, Burnout, and Teacher Attrition?

The null hypothesis and alternate hypothesis below were created to examine research question two. Specifically, is there a correlation among high-stakes testing teacher stress, teacher burnout, and potential teacher attrition?

$H4_0$: There is no correlation among high-stakes testing teacher stress, teacher burnout, and potential teacher attrition rates.

$H4_A$: There is a correlation among high-stakes testing teacher stress, teacher burnout, and potential teacher attrition rates.

A Spearman's rank-order correlation was run to assess the relationships between testing-stress/burnout, testing-stress/attrition, and burnout/attrition. EE, ETSI-T, and the reported rates of potential attrition were used. EE was selected to represent burnout as this construct has repeatedly been correlated most closely with burnout (Kokkinos, 2006; Malasch, Leiter & Schaufeli, 2008; Schaufeli, Daamen & van Mierlo, 1994). Preliminary analysis by visual inspection of scatterplots showed each of the three relationships to be monotonic. A monotonic relationship between two variables describes the existence of either a positive or negative correlation. A positive correlation shows one variable increasing as the other increases, and a negative correlation shows one variable increasing as the other decreases. The relationship must be consistently positive or negative, but does not necessarily need to be linear. A moderate positive correlation was found between potential attrition and ETSI-T, $r_s(291) = .335$, $p < .0005$. A moderate positive correlation was found between ETSI-T and EE, $r_s(291) = .380$, $p < .0005$.

A strong positive correlation was identified between EE and potential attrition, $r_s(291) = .552$, $p < .0005$. All three correlations were statistically significant, meaning the correlations exist. Two are moderately correlated, and one is a statistically strong correlation. All three are positive or direct correlations, meaning as one factor increases the other also increases. In this case, an increase in EE (emotional exhaustion) is strongly correlated with an increase in potential attrition. Likewise, a decrease in EE is strongly correlated with a decrease in potential attrition. The researcher thereby rejected the null hypothesis and accepted the alternate hypothesis.

Research Question 3

Are other factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) affecting stress, burnout, and teacher attrition?

Hypothesis 5: Which Demographic Factors have an Effect on Teacher Stress?

The fifth null and alternate hypotheses were designed to study the following part of the third research question. Which of the following demographic factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) have an effect on teacher stress?

$H5_0$: No difference exists among demographic factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) and teacher stress.

$H5_A$: A difference exists among demographic factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) and teacher stress.

An independent samples t-test was run to assess if statistically significant differences exist in high-stakes teacher stress (ETSI-T) between “females” (n = 207) and “males” (n = 84).

Visual inspection of the boxplots revealed no outliers more than 1.5 box lengths from the edge of the box (see Figure 9). ETSI-T scores were normally distributed for each gender, as assessed by Shapiro-Wilk's test ($p > .05$). Levene's test for equality of variances ($p = .002$) assessed a violation of the assumption of homogeneity of variances. A subsequent Welch t-test for unequal variances was run (Lund & Lund, 2015). The mean ± standard deviation ETSI-T score for females was 34.46 ± 6.65 and for males was 32.70 ± 8.48. The mean difference between groups failed to be statistically significant ($p = .092$), meaning there is no real difference between male and female participants regarding high-stakes testing stress. The researcher failed to reject the null hypothesis and rejected the alternate hypothesis regarding high-stakes testing teacher stress and gender.

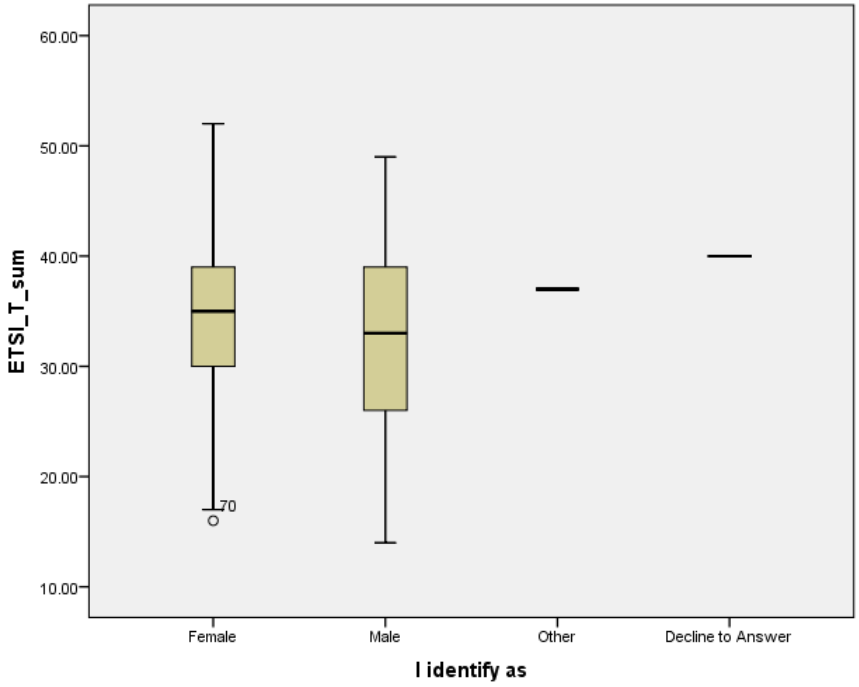


Figure 9. Boxplots of ETSI-T scores for males and female. No outliers are more than 1.5 box-lengths from the edge of the box (IBM Corp., 2015).

Statistical assumptions were not met for a one-way ANOVA. The Kruskal-Wallis H test, a non-parametric test, was run to detect statistically significant differences (Lund & Lund, 2015) in high-stakes testing stress concerning school location as follows: "rural" (n = 102), "suburban" (n = 151), and "urban" (n = 40). Visual assessment of the boxplots displayed similar distributions (see Figure 10). The difference in mean rank values for rural (155.07), suburban (139.62), and urban (154.29) locations was not statistically significant, $\chi^2(2) = 2.372$, $p = .305$. There exists no statistically significant difference, or no real difference, in high-stakes testing stress among public school teacher participants in rural, suburban, and urban locations. For this distribution, the researcher failed to reject the null hypothesis and rejected the alternate hypothesis for the school location.

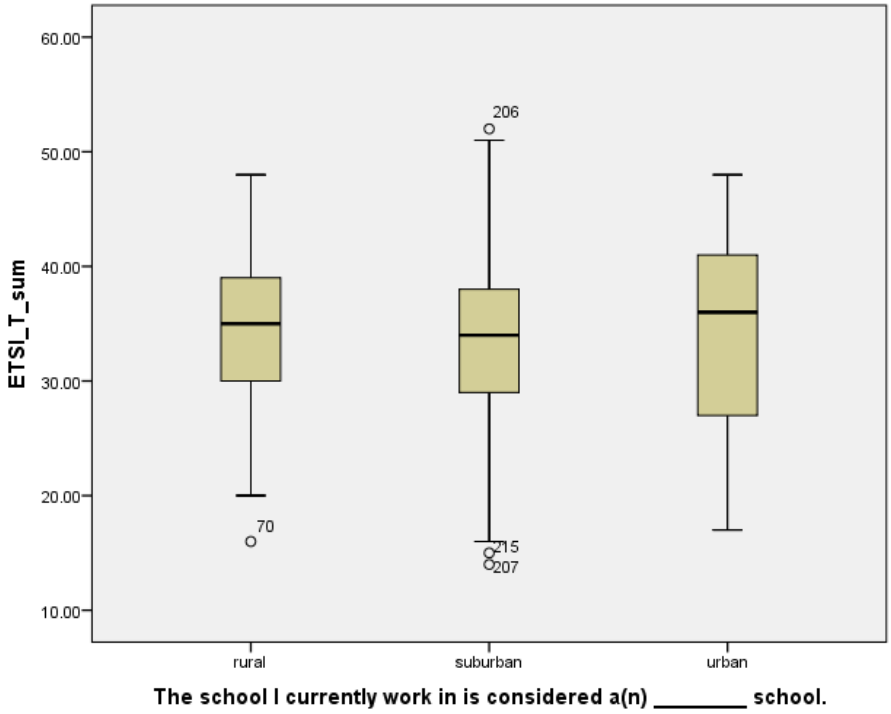


Figure 10. Boxplots of ETSI-T scores for school location. Similar distributions are displayed for rural, suburban, and urban schools (IBM Corp., 2015).

Assumptions were met to conduct a one-way ANOVA (Lund & Lund, 2015) to evaluate if statistically significant differences exist in high-stakes teaching stress levels among public school teachers based on the socioeconomic status of the district. Participants were classified as teaching in the following type of district: "low-income" (n = 123), "middle-income" (n = 135), and "high-income" (n = 35) schools. There were no outliers, as assessed by the boxplots in Figure 11, and data was normally distributed, as assessed by the Shapiro-Wilks test ($p > .05$). Levene's test of homogeneity of variances ($p = .810$) confirmed homogeneity of variance. Mean ± standard deviation ETSI-T scores were low-income (35.02 ± 7.35), middle-income (33.67 ± 6.98), and high-income (31.54 ± 7.35). The difference in ETSI-T scores was statistically significant between the teachers at low-income and high-income schools, $F(2, 290) = 3.442$, $p = .033$, $\omega^2 = 0.02$. Tukey post hoc analysis revealed the mean difference ETSI-T scores between teachers at high-income and low-income schools was 3.48 (95% CI (0.2 to 6.7)), which is also statistically significant ($p = .032$). A statistically significant difference, or a real difference, existed between the high-stakes testing stress levels of math teacher participants at low-income and high-income schools, whereby participants at low-income schools reported higher levels of high-stakes testing stress. For this comparison group, between participants at low-income schools and participants at high-income schools, the researcher rejected the null hypothesis and accepted the alternate hypothesis. For other comparisons of high-stakes testing stress between teacher participants at schools of different income levels, the researcher failed to reject the null hypothesis and rejected the alternate hypothesis.

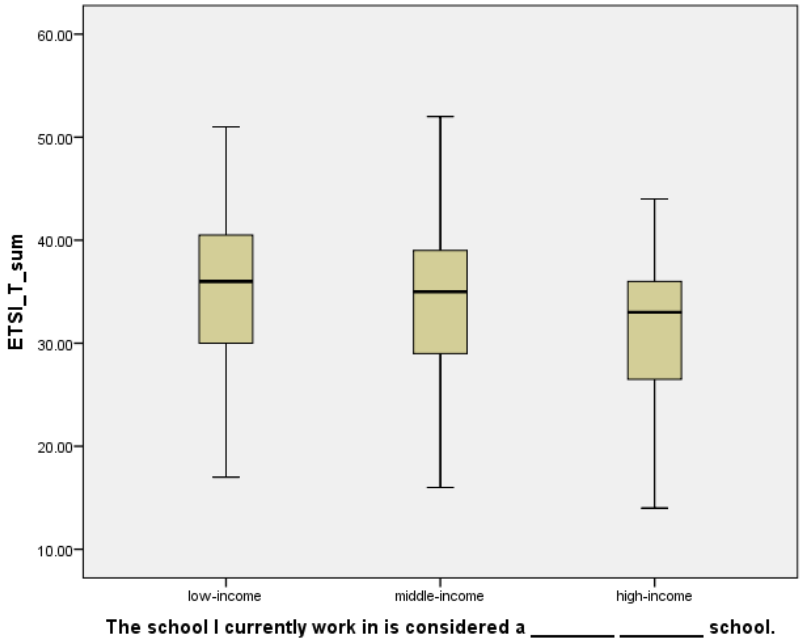


Figure 11. Boxplots of ETSI-T scores and socioeconomic status. No visible outliers exist (IBM Corp., 2015).

Two outliers were present on the boxplots for years of teaching experience, as assessed by visual inspection (see Figure 12). Despite the data being normally distributed, a one-way ANOVA was not used. The Kruskal-Wallis H test (Lund & Lund, 2015), a non-parametric test, was run to detect statistically significant differences in ETSI-T scores regarding years of teaching experience. The groups included: "1 to 5" (n = 38), "6 to 10" (n = 55), "11 to 20" (n = 135), "21 to 30" (n = 53), and "31+" (n = 12). Visual evaluation of boxplots revealed different distributions as seen in Figure 15. The mean rank values calculate as follows: "1 to 5" (147.79), "6 to 10" (126.65), "11 to 20" (161.19), "21 to 30" (145.94), and "31+" (82.83). The difference of ETSI-T scores was statistically significant between the "11 to 20" and "31+" groups, $\chi^2(4) = 13.884$, $p = .008$. Pairwise comparison post hoc tests were performed in SPSS using Dunn's procedure. The difference of mean rank values for "11 to 20" (161.19) and "31+" (82.83) was statistically significant with an adjusted *p*-value ($p = .021$). There exists a statistically significant difference, or real difference, in high-stakes teacher stress for participants in the "11 to 20"

years-experience group than participants in the “31+” years-experience group, whereby the median ETSI-T scores are higher for the “11 to 20” group. For this comparison group (“11 to 20” and “31+”), the researcher rejected the null hypothesis and accepted the alternate hypothesis for high-stakes test stress levels regarding years of teaching experience. The researcher failed to reject the null hypothesis and rejected the alternate hypothesis for high-stakes test stress levels regarding years of teaching experience for all other comparison groups of ETSI-T scores and years of teaching experience.

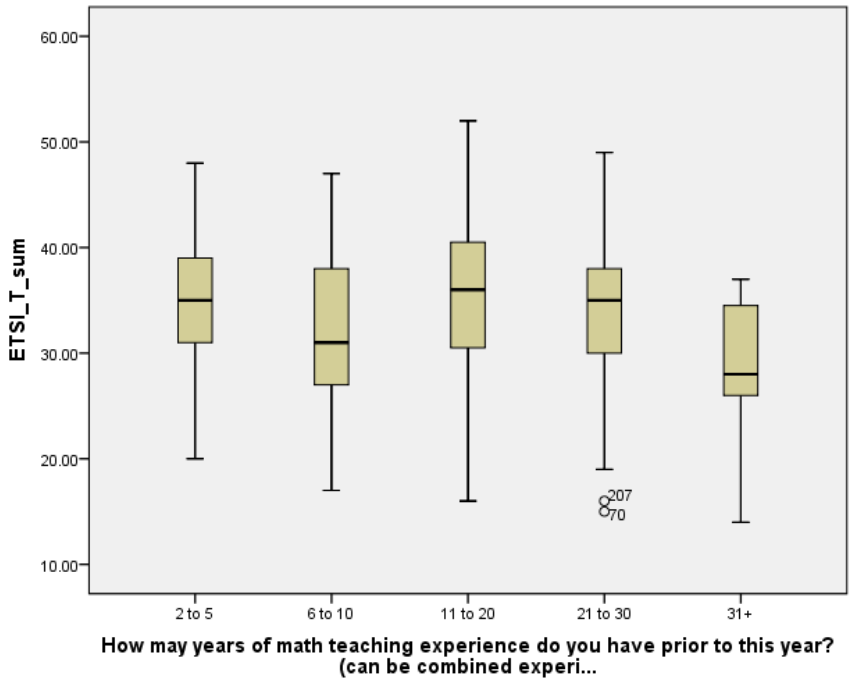


Figure 12. Boxplots of ETSI-T scores and years of teaching experience. Two outliers are visible (IBM Corp., 2015).

Statistical assumptions were not met for a one-way ANOVA. The Kruskal-Wallis H test was run (Lund & Lund, 2015) to detect statistically significant differences in ETSI-T scores compared to grade level taught. The grade levels included: “elementary grades k - 5” (n = 80), “middle school grades 6 - 8” (n = 112), and “secondary grades 9 -12” (n = 101). Visual assessment of the boxplots displayed an outlier and unrelated distributions as seen in Figure 13.

The difference of mean rank values for elementary (166.73), middle school (147.64), and secondary (130.66) levels was statistically significant, $\chi^2(2) = 8.120$, $p = .017$. Pairwise comparison post-hoc tests were performed in SPSS using Dunn's procedure, a non-parametric procedure created to follow rejection of the Kruskal-Wallis H test (Lund & Lund, 2015). The difference of mean rank values for "elementary" (166.73) and "secondary" (130.66) was statistically significant with an adjusted *p*-value ($p = .013$). There exists a statistically significant difference in ETSI-T scores between elementary and secondary public school teacher participants, whereby elementary participants report higher levels of high-stakes testing stress. For this distribution, the researcher rejected the null hypothesis and accepted the alternate hypothesis for the elementary-secondary comparison. The researcher failed to reject the null hypothesis and rejected the alternate hypothesis for all other ETSI-T scores and grade level comparisons.

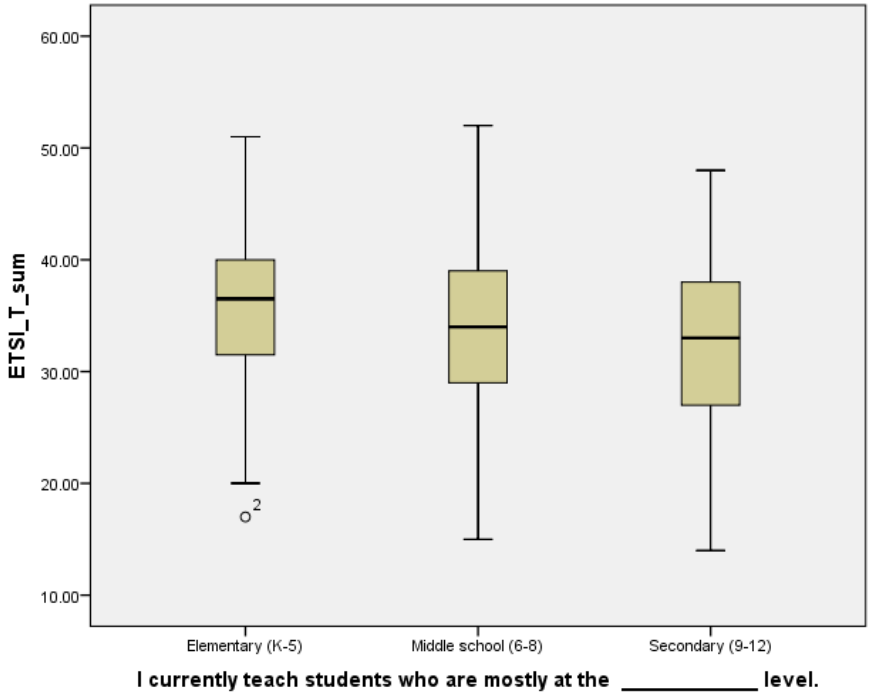


Figure 13. Boxplots of ETSI-T for grade level taught. The distributions are visibly unrelated and display a visible an outlier (IBM Corp., 2015).

Hypothesis 6: Which Demographic Factors have an Effect on Teacher Burnout?

The sixth null hypothesis and alternate hypothesis were derived to examine the following piece of the third research question. Which of the following demographic factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) have an effect on teacher burnout?

$H6_0$: No difference exists among demographic factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) and teacher burnout.

$H6_A$: A difference exists among demographic factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) and teacher burnout.

Assumptions were not met for an independent samples t-test. The Kruskal-Wallis H test was run (Lund & Lund, 2015) to determine if statistically significant differences exist in teacher burnout (EE) subscale scores between "females" (n = 207) and "males" (n = 84). Visual inspection of the boxplots in Figure 14 revealed distributions that were not similar. The difference of EE (emotional exhaustion) subscale scores were statistically significant between females and males, $\chi^2(3) = 10.232$, $p = .017$. Pairwise comparison post hoc tests were performed in SPSS using Dunn's procedure, the procedure designed to follow the rejection of the Kruskal-Wallis H test (Lund & Lund, 2015). The difference of mean rank values for female (156.95) and male (123.15) participants was statistically significant with an adjusted *p*-value ($p = .012$). There exists a statistically significant difference in higher burnout (EE) subscale scores for female participants than male participants. The researcher rejected the null hypothesis and accepted the alternate hypothesis for burnout levels regarding gender.

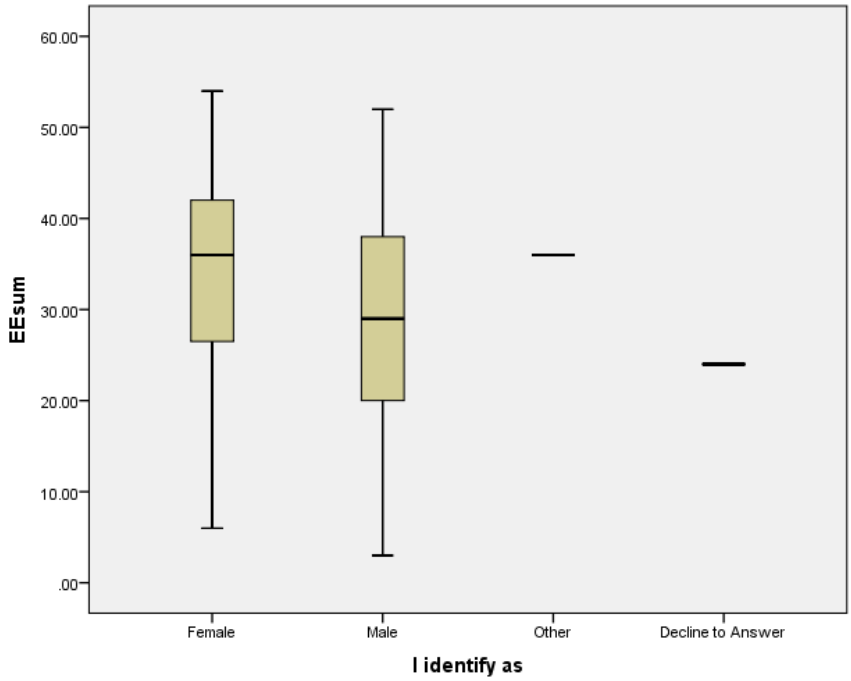


Figure 14. Boxplots of EE subscale scores for males and females. Distributions are not similar (IBM Corp., 2015).

Statistical assumptions were not met to run the one-way ANOVA for school location data. The independent samples Kruskal-Wallis test was run (Lund & Lund, 2015) to determine if statistically significant differences in teacher burnout exist regarding school location: "rural" (n = 102), "suburban" (n = 151), and "urban" (n = 40). Visual inspection of the boxplots in Figure 15 revealed similar distributions. The difference of mean rank values for rural (141.72), suburban (146.86), and urban (161.01) locations was not statistically significant, $\chi^2(2) = 1.493$, $p = .474$. There exists no statistically significant difference in teacher burnout among public school teacher participants in rural, suburban, and urban locations. For this distribution, the researcher failed to reject the null hypothesis and rejected the alternate hypothesis for the school location.

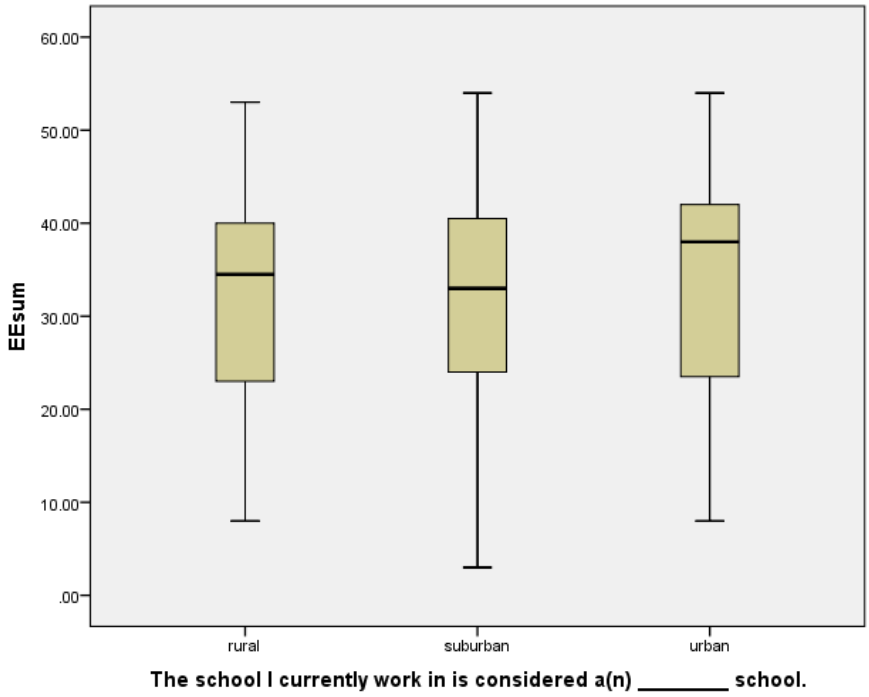


Figure 15. Boxplots of EE subscale scores for school location. Distributions appear similar (IBM Corp., 2015).

For school socioeconomic status, boxplots in Figure 16 revealed an outlier and data was not normally distributed. For these reasons, the Kruskal-Wallis H test was run (Lund & Lund, 2015) to determine if statistically significant differences in teacher burnout (EE) subscale scores exist regarding school socioeconomic status: the "low-income" (n = 123), "middle-income" (n = 135), and "high-income" (n = 35) schools. Visual assessment of the boxplots also revealed somewhat similar distributions. The difference of mean rank values for low-income (156.96), medium-income (136.16), and high-income (153.81) schools was not statistically significant, $\chi^2(2) = 4.141$, $p = .126$. There exists no statistically significant difference in teacher burnout among public school teacher participants in schools of different socioeconomic status. For this distribution, the researcher failed to reject the null hypothesis and rejected the alternate hypothesis for school socioeconomic status.

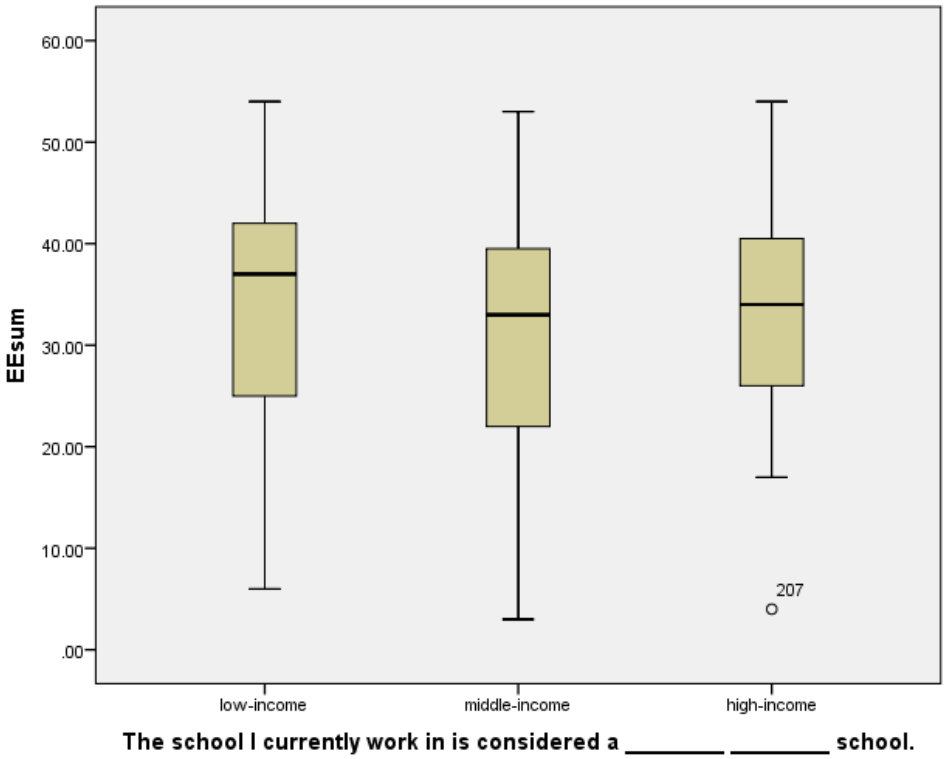


Figure 16. Boxplots of EE subscale scores for socioeconomic status. Visual assessment reveals an outlier and somewhat similar distributions (IBM Corp., 2015).

The visual assessment of the boxplots in Figure 17 for years of teaching experience revealed several outliers and data was not normally distributed. Hence, the Kruskal-Wallis H test was conducted (Lund & Lund, 2015) to determine if statistically significant differences in EE (emotional exhaustion) subscale scores exist among teachers with different years of teaching experience. The groups included the following years: "1 to 5" (n = 38), "6 to 10" (n = 55), "11 to 20" (n = 135), "21 to 30" (n = 53), and "31+" (n = 12). The boxplot distributions also appeared dissimilar. The mean rank values calculate as follows: "1 to 5" (108.20), "6 to 10" (137.67), "11 to 20" (164.84), "21 to 30" (152.90), and "31+" (85.83). The difference of EE subscale scores was statistically significant between the "11 to 20" and "31+" groups and the "11 to 20" and the "1 to 5" groups, $\chi^2(4) = 21.159$, $p < .0005$. Pairwise comparison post hoc tests

were performed in SPSS using Dunn's procedure. The difference of mean rank values for "11 to 20" (164.84) and "31+" (85.83) was statistically significant with an adjusted *p*-value ($p = .020$). The difference of mean rank values for "11 to 20" (164.84) and "1 to 5" (108.20) was statistically significant with an adjusted *p*-value ($p = .003$). There exists a statistically significant difference in burnout, whereby teacher participants in the "11 to 20" years-experience group report higher levels of burnout than teacher participants in the "1 to 5" and "31+" years-experience group. For these comparison groups ("11 to 20" compared with "1 to 5" and "31+" years of teaching experience), the researcher rejected the null hypothesis and accepted the alternate hypothesis for burnout (EE) subscale scores. The researcher failed to reject the null hypothesis and rejected the alternate hypothesis for burnout (EE) subscale scores regarding years of teaching experience for all other comparison groups.

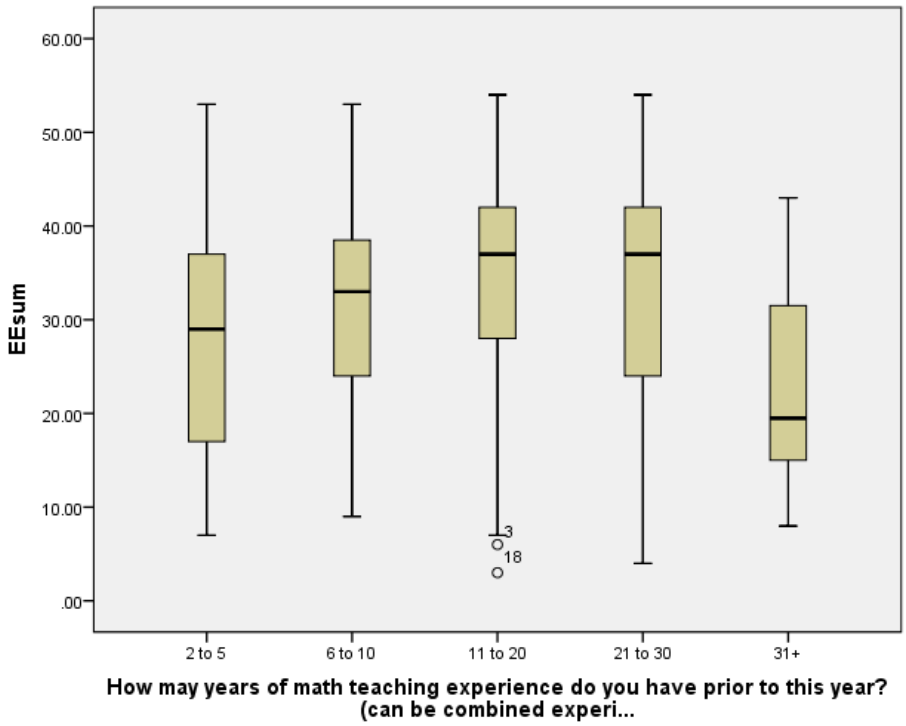


Figure 17. Boxplots of EE subscale scores for years of teaching experience. Several outliers are visible and distributions appear dissimilar (IBM Corp., 2015).

Statistical assumptions were not met for a one-way ANOVA. The Kruskal-Wallis H test was run (Lund & Lund, 2015) to detect statistically significant differences in burnout (EE) subscale scores compared to grade level taught. The groups included: "elementary grades k - 5" (n = 80), "middle school grades 6 - 8" (n = 112), and "secondary grades 9 -12" (n = 101). Visual inspection of the boxplots in Figure 18 revealed similar distributions. The difference of mean rank values for elementary (151.92), middle school (147.93), and secondary (142.07) levels was not statistically significant, $\chi^2(2) = 0.626$, $p = .731$. There exists no statistically significant difference in burnout (EE) subscale scores among elementary, middle school, and secondary level public school teacher participants. For this distribution, the researcher failed to reject the null hypothesis and rejected the alternate hypothesis for all EE subscale scores and grade level comparisons.

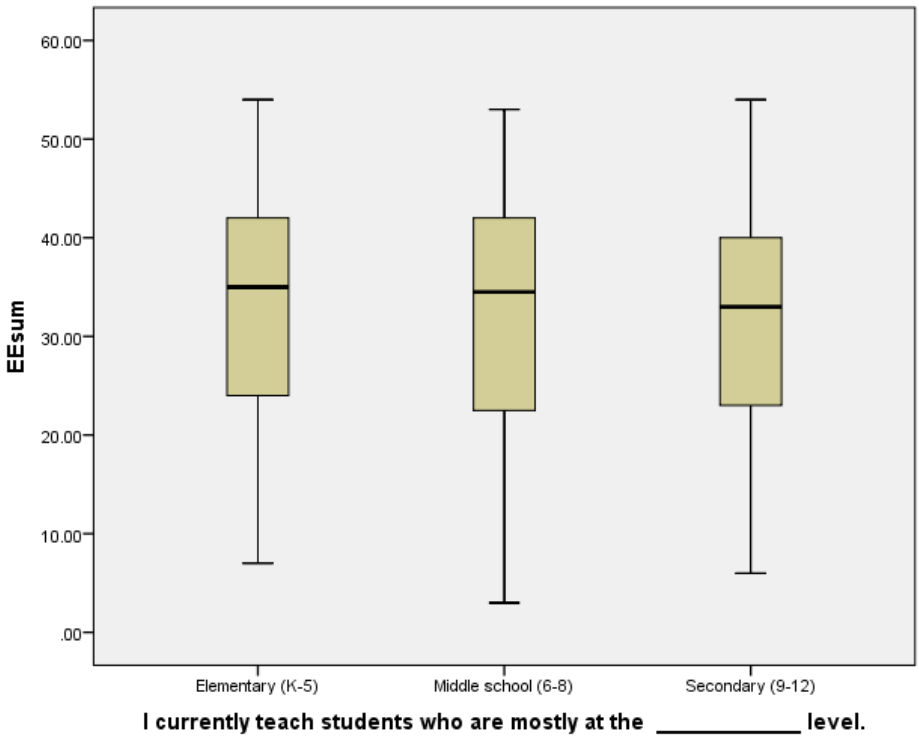


Figure 18. Boxplots of EE subscale scores for grade level taught. Similar distributions are visible (IBM Corp., 2015).

Hypothesis 7: Which Demographic Factors have an Effect on Teacher Attrition Rates?

This final null hypothesis and alternate hypothesis were generated to investigate the last part of the third research question. Which of the following demographic factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) have an effect on potential teacher attrition rates?

$H7_0$: No difference exists among demographic factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) and potential teacher attrition rates.

$H7_A$: A difference exists among demographic factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) and potential teacher attrition rates.

Statistical assumptions were not met, and a non-parametric test was selected. The Kruskal-Wallis H test was run (Lund & Lund, 2015) to detect statistically significant differences of teacher potential teacher attrition rates between the following groups: "female" (n = 207) and "male" (n = 84). Visual assessment of the boxplots in Figure 19 displayed similar distributions. The difference of mean rank values for female (150.85) and male (136.22) participants was not statistically significant, $\chi^2(3) = 2.609$, $p = .456$. There exists no statistically significant difference of potential attrition rates between male and female participants. For this distribution, the researcher failed to reject the null hypothesis and rejected the alternate hypothesis for gender.

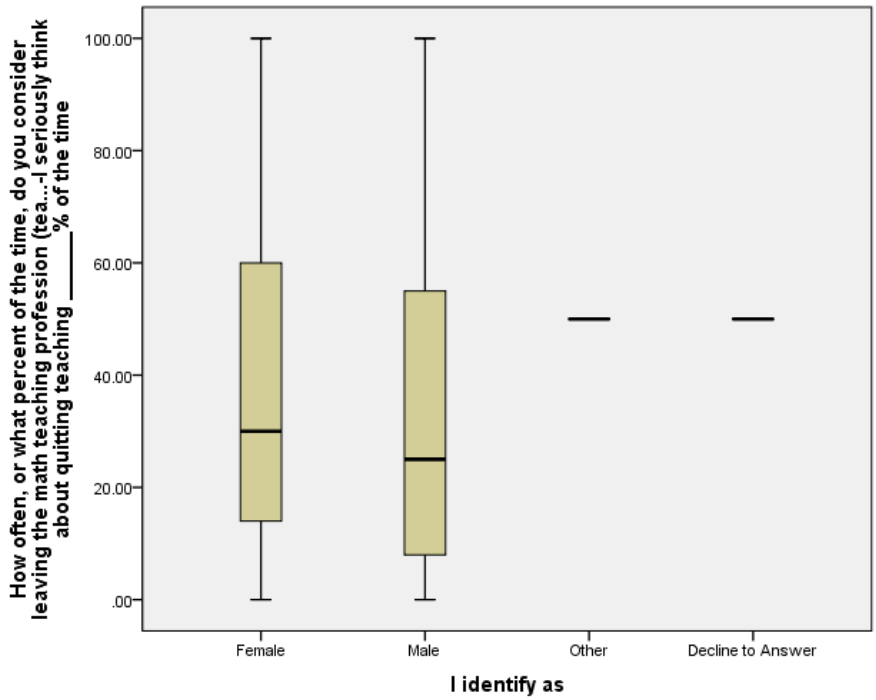


Figure 19. Boxplots of potential attrition rates for males and females. Distributions appear similar (IBM Corp., 2015).

Assumptions were not met to run the one-way ANOVA for school location data, and the Kruskal-Wallis H test was run (Lund & Lund, 2015) to detect if statistically significant differences of potential teacher attrition rates exist among locations. These locations were grouped as follows: "rural" (n = 102), "suburban" (n = 151), and "urban" (n = 40) school locations. Visual inspection of the boxplots in Figure 20 revealed mostly similar distributions. The difference of mean rank values for rural (157.80), suburban (137.91), and urban (153.79) locations was not statistically significant, $\chi^2(2) = 3.667$, $p = .160$. There exists no statistically significant difference of potential attrition rates among public school teacher participants in rural, suburban, and urban locations. For this distribution, the researcher failed to reject the null hypothesis and rejected the alternate hypothesis for the school location.

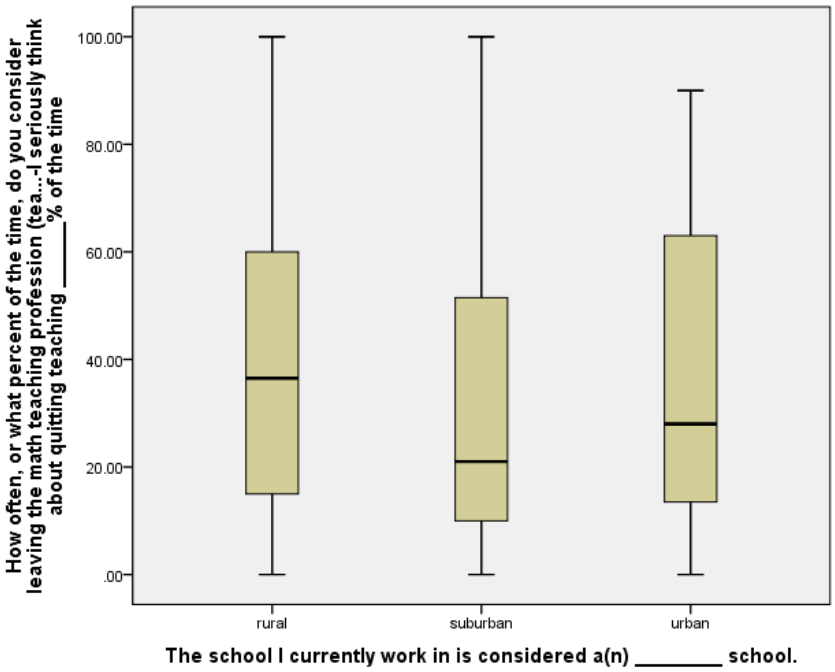


Figure 20. Boxplots of potential attrition rates for school location. Distributions appear similar (IBM Corp., 2015).

For socioeconomic status, no outliers were present in the boxplots (see Figure 21), as assessed by visual inspection, but data was not normally distributed. For this reason, the independent samples Kruskal-Wallis H test was run (Lund & Lund, 2015) to detect if statistically significant differences exist for potential teacher attrition rates among "low-income" (n = 123), "middle-income" (n = 135), and "high-income" (n = 35) schools. Visual assessment of the boxplots revealed similar distributions. The difference of mean rank values for low-income (155.96), middle-income (139.49), and high-income (144.50) schools was not statistically significant, $\chi^2(2) = 2.474$, $p = .290$. There exists no statistically significant difference of potential attrition rates among public school teachers in low-income, middle-income, and high-income schools. For this distribution, the researcher failed to reject the null hypothesis and rejected the alternate hypothesis for socioeconomic status.

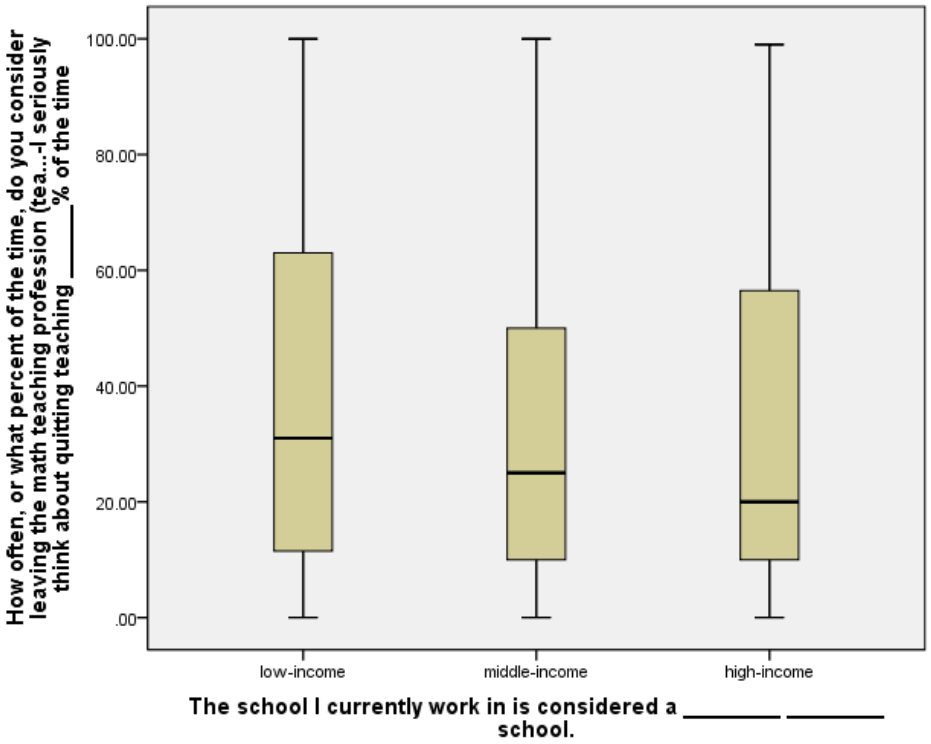


Figure 21. Boxplots of potential attrition rates for socioeconomic status. Distributions appear similar (IBM Corp., 2015).

Assumptions were violated for parametric tests, thereby the Kruskal-Wallis H test was executed (Lund & Lund, 2015) to determine if statistically significant differences in potential attrition rates exist among teacher participants with different years of teaching experience. The groups were divided in years as follows: "1 to 5" (n = 38), "6 to 10" (n = 55), "11 to 20" (n = 135), "21 to 30" (n = 53), and "31+" (n = 12). Boxplot distributions appeared dissimilar upon visual inspection (see Figure 22). The mean rank values calculate as follows: "1 to 5" (109.84), "6 to 10" (144.62), "11 to 20" (161.75), "21 to 30" (137.25), and "31+" (152.71). The difference of potential attrition rates was statistically significant between the "11 to 20" and the "1 to 5" group, $\chi^2(4) = 12.242$, $p = .016$. Pairwise comparison post hoc tests were performed in SPSS using Dunn's procedure. The difference of mean rank values for "11 to 20" (161.75) and "1 to 5" (109.84) was statistically significant with an adjusted *p*-value ($p = .008$). There exists a

statistically significant difference in potential teacher attrition, whereby participants in the “11 to 20” years-experience group report higher rates of potential attrition than participants in the “1 to 5” years-experience group. For the comparison group (“11 to 20” compared to “1 to 5” years of teaching experience), the researcher rejected the null hypothesis and accepted the alternate hypothesis for potential teacher attrition rates. The researcher failed to reject the null hypothesis and rejected the alternate hypothesis for potential teacher attrition rates regarding years of teaching experience for all other comparison groups.

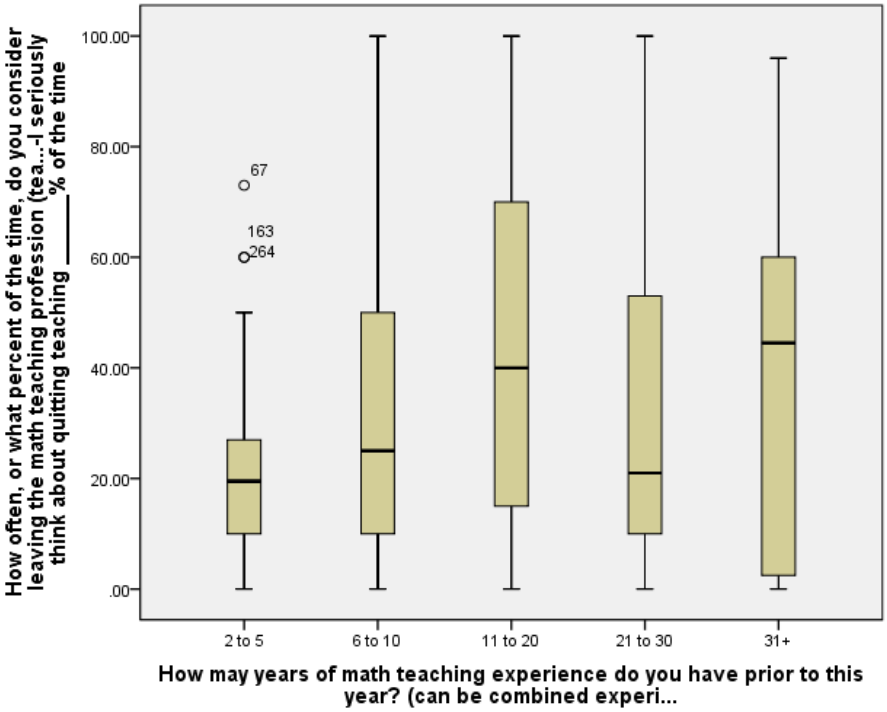


Figure 22. Boxplots of potential attrition rates for years of teaching experience. Distributions appear dissimilar (IBM Corp., 2015).

Statistical assumptions were not met for a one-way ANOVA. The Kruskal-Wallis H test was run (Lund & Lund, 2015) to detect statistically significant differences in potential teacher attrition rates compared to grade level taught. Groups were divided as follows: “elementary grades k - 5” (n = 80), “middle school grades 6 - 8” (n = 112), and “secondary grades 9 -12” (n =

101). Visual assessment of the boxplots in Figure 23 revealed similar distributions. The difference of mean rank values for elementary (143.53), middle school (147.89), and secondary (148.76) levels was not statistically significant, $\chi^2(2) = 0.191$, $p = .909$. There exists no statistically significant difference in potential attrition rates among elementary, middle school, and secondary level public school teacher participants. For this distribution, the researcher failed to reject the null hypothesis and rejected the alternate hypothesis for all potential attrition rates and grade level comparisons.

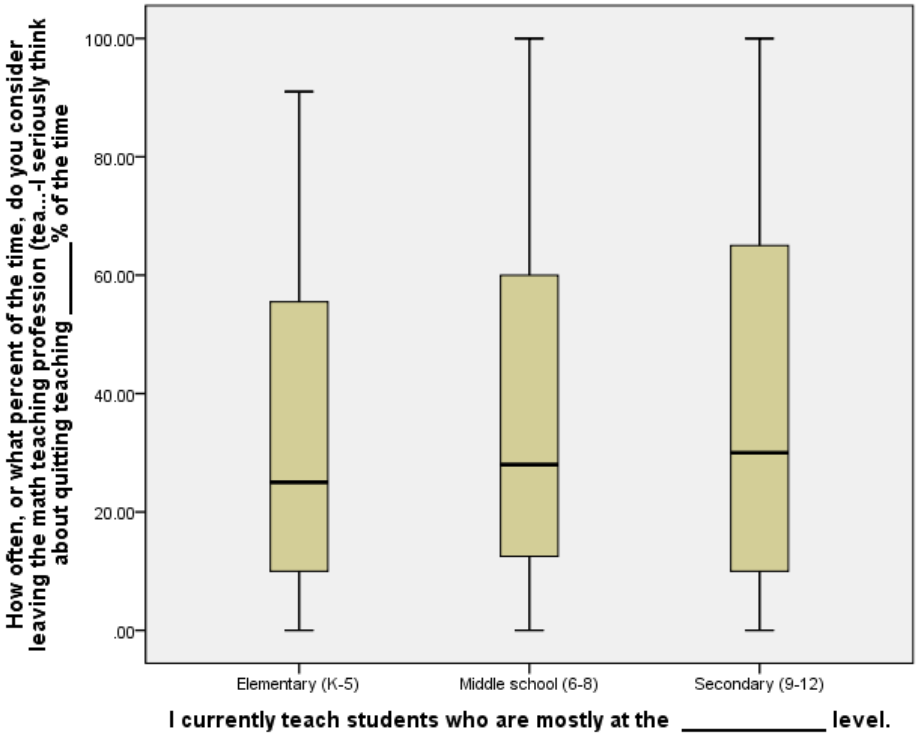


Figure 23. Boxplots of potential attrition rates for grade level taught. Distributions appear similar (IBM Corp., 2015).

The 11-20 years of experience group showed statistically, significant differences on the (high-stakes testing stress) ETSI-T scores, the MBI-ES (emotional exhaustion or burnout) EE subscale scores, and the potential teacher attrition rates. Also, all math teacher participant data show moderate, positive correlations between high-stakes testing stress and burnout and high-

stakes testing stress and attrition. A strong, positive correlation was identified between burnout and attrition. For this reason, a separate data set of the 11-20 years of experience group was extracted to compare the correlation differences with the entire sample of math teacher participants. EE, ETSI-T, and the reported rate of potential attrition were again used for equivalent comparisons.

Table 5

Spearman's Rho Correlations for Math Teacher Sample (N = 293)

			ETSI_T_sum	EEsum	How often, or what percent of the time, do you consider leaving the math teaching profession (tea...-I seriously think about quitting teaching _____% of the time
Spearman's rho	ETSI_T_sum	Correlation Coefficient	1.000	.380**	.335**
		Sig. (2-tailed)	.	.000	.000
		N	293	293	293
	EEsum	Correlation Coefficient	.380**	1.000	.552**
		Sig. (2-tailed)	.000	.	.000
		N	293	293	293
	How often, or what percent of the time, do you consider leaving the math teaching profession (tea...-I seriously think about quitting teaching _____% of the time	Correlation Coefficient	.335**	.552**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	293	293	293

**. Correlation is significant at the 0.01 level (2-tailed).

Preliminary analysis by visual inspection of scatterplots showed each of the three relationships to be monotonic without outliers. However, not all data sets were normally distributed, therefore, a Spearman's rank-order correlation was run to assess the relationships between testing-stress/burnout, testing-stress/attrition, and burnout/attrition. Spearman's rho

correlations for the math teacher sample are presented in Table 5 (IBM Corp., 2015). A moderate positive correlation was found between ETSI-T and EE, $r_s(135) = .394$, $p < .0005$. Another moderate positive correlation was found between ETSI-T and potential attrition rates, $r_s(135) = .365$, $p < .0005$. A strong positive correlation was identified between EE and potential attrition rates, $r_s(135) = .598$, $p < .0005$. All three correlations were statistically significant or real. Two are moderate, positive correlations, and one is a strong, positive correlation. All three pairings of the 11-20 years of experience group are respectively stronger correlations compared to those of the entire math teacher sample.

Table 6

Spearman's Rho Correlation for the 11-20 Years of Experience Group (N = 135)

			ETSI_T_sum	EEsum	How often, or what percent of the time, do you consider leaving the math teaching profession (tea...-I seriously think about quitting teaching _____% of the time
Spearman's rho	ETSI_T_sum	Correlation Coefficient	1.000	.394**	.363**
		Sig. (2-tailed)	.	.000	.000
		N	135	135	135
	EEsum	Correlation Coefficient	.394**	1.000	.598**
		Sig. (2-tailed)	.000	.	.000
		N	135	135	135
	How often, or what percent of the time, do you consider leaving the math teaching profession (tea...-I seriously think about quitting teaching _____% of the time	Correlation Coefficient	.363**	.598**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	135	135	135

**. Correlation is significant at the 0.01 level (2-tailed).

Spearman's rho correlations for the 11-20 years of experience group are represented in Table 6 (IBM Corp., 2015). Values for statistically significant differences previously detected by SPSS are highlighted in the table. The correlation coefficient indicates how close the points correlate to a consistently increasing (positive coefficient) or decreasing (negative coefficient) function. A higher decimal indicates a stronger correlation. For example, for the ETSI-T/EE pairing, the correlation coefficient for the entire sample of math teacher participants is $r_s(293) = .380$ and the correlation coefficient for the 11-20 years of experience group is $r_s(135) = .394$. Because $0.394 > 0.380$, the group with 0.394 has the stronger correlation. In this case, the 11-20 years group has the stronger correlation coefficient for all pairings.

Summary

Chapter 4 contains the findings of the study. Research questions 1, 2, and 3 were evaluated by examination of the seven hypotheses. Chapter 5 follows with a summary of results, conclusions, and recommendations for future study.

Research question 1 asked: does high-stakes testing as a component of teacher evaluation lead to higher levels of stress, leading to burnout, and eventually attrition? Both the Teachers of Record (TOR) and Non-Teachers of Record (NTOR) experience similar levels of high-stakes testing stress on all scales. Both groups also experience similar levels of burnout on all three EE, DP, and PA subscales. Both TOR and NTOR report similar rates of potential attrition and 69% of all participants scored "high" for burnout on the emotional exhaustion (EE subscale).

Research question 2 posed the following question: does high-stakes testing as a component of teacher evaluation lead to a correlation among stress, burnout, and attrition? There is a moderate positive correlation between high-stakes testing stress and teacher burnout and between high-stakes testing stress and potential teacher attrition. There is a strong positive

correlation between teacher burnout and potential teacher attrition. Similar and slightly stronger correlations were found for all three comparisons for the 11-20 years-experience teaching group.

Research question 3 asked the following: are other factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) affecting high-stakes testing stress, teacher burnout, and teacher attrition? Teacher participants from low-income schools reported increased levels of high-stakes testing stress compared to those from high-income schools. Teacher participants at the elementary level reported higher levels of high-stakes testing stress compared to reported levels of secondary teacher participants. Female teacher participants reported higher levels of burnout compared to male participants.

Participants with 11-20 years of teaching experience reported the highest levels of stress compared to the lowest levels reported by the 31+ years of teaching experience group. Teacher participants with 11-20 years of experience teaching also report higher levels of burnout, especially compared to the 31+ group and the 1-5 year group. Additionally, this 11-20 years of experience teaching participant group reports the highest level of potential attrition rates compared to the lowest level reported by the 1-5 year participant group.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The current quantitative study examined the impact of high-stakes testing stress related to the inclusion of student standardized testing data in teacher evaluations on burnout and potential attrition rates among current Pennsylvania math teachers who identify as either Teachers of Record (TOR) or Non-Teachers of Record (NTOR). Chapter 1 introduced the study; Chapter 2 reviewed the literature; Chapter 3 outlined research procedures, and Chapter 4 reported results of the data analyses. Chapter 5 includes a summary and conclusions and follows with recommendations.

Summary

This causal-comparative/quasi-experimental study attempted to establish a cause and effect relationship among high-stakes testing stress, teacher burnout, and potential teacher attrition rates between PA public school math TOR and NTOR through observation of an ongoing natural experimental-type situation. The problem is that high-stakes tests measure student performance outcomes and not teacher behaviors that are controllable, thereby creating a copious amount of chronic stress for teachers (Elliot & Dweck, 2007). In turn, prolonged or chronic teacher stress can lead to potential emotional exhaustion, otherwise known as burnout (Malasch, Schaufeli, & Leiter, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011), and increased rates of teacher attrition (Chang, 2009). Several studies link the demands of high-stakes testing to loss of autonomy for teachers, diminished competence, and relatedness concerning performance evaluations. These same factors (loss of autonomy, diminished competence, and relatedness) correlate with increased teacher stress, burnout, and attrition as viewed through the lens of Self-Determination Theory (Bartholomew, Ntoumanis, Cuevas & Lonsdale, 2014; Bouwma-Gearhart,

2010; Deci & Ryan, 1985; Fernet, Guay, Senecal & Austin, 2012; Skaalvik & Skaalvik, 2009). There exists a small, but growing body of new research that in some way relates teacher stress, burnout, and/or attrition to the use of student performance data on teacher effectiveness evaluations (Baker, Oluwole & Green, 2013; Bouwma-Gearhart, 2010; Hanson, 2006; Hewitt, 2015; von der Embse, Pendergast, Segool, Saeki & Ryan, 2016). However, a gap exists, specifically among math teachers regarding high-stakes testing stress, burnout, and attrition, related to the use of student performance data on teacher effectiveness evaluations.

The independent variable of the study was Teacher of Record (TOR) or Non-Teacher of Record (NTOR) participants among the population of public school math teachers in Pennsylvania. The dependent variables of the study were Educators Test Stress Inventory (ETSI) scores, Malasch Burnout Inventory - Educators Survey (MBI-ES) scores, and potential attrition rates. The investigation included evaluation of the effect of individual factors on the dependent variables as observed in prior studies including gender, school location, school socioeconomic status, years of teaching experience, and grade level of students. Self-Determination Theory provided the theoretical framework for the study (Deci & Ryan, 1985). The scope of the survey was limited to evaluating factors including high-stakes testing stress, teacher burnout, and potential teacher attrition rates. The study was limited by self-reported data from participants and quantitative data, although objective, does not allow in-depth analysis of responses.

The current study collected data in Qualtrics format using three cross-sectional surveys. A demographic survey (see Appendix F), the ETSI (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015) (see Appendix G), and the MBI-ES (Malasch, Jackson & Schwab, 1986) (see Appendix H) were utilized. Permission was granted, and transmission of surveys occurred

through the Pennsylvania State Education Association (PSEA) listserv attendant, maintaining the anonymity of participants and security of listserv email accounts. The researcher used stratified random sampling in this circumstance to maximize external validity (Brase & Brase, 2013; Vogt, 2007). An approximate population 61,407 PA public school math teachers yielded a desired sample size 382 for a 95% Confidence Level and ±5% Margin of Error. The PSEA research department (which yields about a 10% email survey return) was able to use their computer system to randomly sample 4,000 math teachers. A total of 1,000 were randomly selected from within each of 4 extracted groups, or strata, which included: (10-12) high school math teachers, (7-9) middle school math teachers, (4-6) intermediate elementary teachers, and (K-3) primary teachers. Surveys were distributed via email one time followed by three reminder emails beginning one week after the Pennsylvania state testing window closed from May 26, 2016, to July 8, 2016. Initially, 330 potential participants responded. However, 293 participants completed the survey without exiting and dropping out of the study. The final sample size calculates to a 93.1% Confidence Level and a ±5.71 Margin of Error, meaning the data is true 93.1% of the time within plus or minus 5.71%.

SPSS software was used to evaluate data for the three surveys to test the seven null hypotheses. Three research questions were posed in the current study. Research question 1 asked: does high-stakes testing as a component of teacher evaluation lead to higher levels of stress, leading to burnout, and eventually attrition? A comparison of both participant groups of TOR and NTOR was completed to determine if statistically significant differences in the values

of the specific variables exist as follows: high-stakes testing stress levels (3 scales), burnout levels (3 scales), and potential attrition rate.

The ETSI was scored to evaluate high-stakes testing stress, and it consists of 3 scores: the Manifestations of Stress (ETSI-M) subscale, the Sources of Stress (ETSI-S) subscale, and the Total (ETSI-T) subscale (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015). No statistically significant differences were detected between the TOR and NTOR groups of participants for any of the three scales of ETSI. Therefore, $H1_0$ fails to be rejected. Statistical evaluation in SPSS also determined no statistically significant differences exist between TOR and NTOR participants regarding all three scales of burnout (Emotional Exhaustion, Depersonalization, and Personal Accomplishment). Hence, $H2_0$ fails to be also rejected. No statistically significant difference was detected for potential attrition rates between TOR and NTOR participant groups. Additionally, $H3_0$ fails to be rejected. Despite the attempt to establish a cause and effect relationship among high-stakes testing stress, teacher burnout, and potential teacher attrition by choosing a causal-comparative/quasi-experimental study, the researcher was unable to definitively accomplish this since no statistically significant differences existed and qualitative data was not obtained.

However, upon review of the data, the researcher noticed the scores for both groups of TOR and NTOR participants might not have shown a statistical difference because the ETSI scores, MBI-ES scores, and potential attrition rates were high for both groups. The creators of the ETSI caution its limitations for interpreting a total stress score and suggest its utilization for identifying sources and manifestations of stress (von der Embse, Kilgus, Solomon, Bowler & Curtiss, 2015). It should be noted upon further review that all questions in the Sources of Stress (ETSI-S) portion of the survey revealed mean scores greater than 3.5 to 4. A mean score of this

value on a 1 to 5 scale indicates many participants rated the each stressor as a 4 (agree) or a 5 (strongly agree) as a source of stress. A rating of a 3 (neither agree nor disagree) indicates a neutral response. A rating of a 2 (disagree) or a 1 (strongly disagree) indicates the item is not a stressor. The TOR participants report 76% are anxious before standardized testing and 50% are anxious after the testing. All opposing percentages reveal either the item is neutral or not a stressor. The NTOR participants report 62% are anxious before standardized testing and 44% are anxious following the testing. Both participant groups also report feeling as though they are being evaluated by standardized testing (82% TOR and 75% NTOR). Numbers reveal 45% of TOR and 48% of NTOR feel pressured by parents to raise student test scores, also indicated by a response of 4 or 5. However, a total of 89% of all participants (89% TOR and 89% NTOR) report feeling pressured by the administration to raise student test scores, suggesting the participants do not merely hear about test scores. These participants are being pressured to have students perform well.

For the MBI-ES, mean Emotional Exhaustion (EE) scores for both groups fell in the "high" category, while the U.S. mean scores in 1996 (before high-stakes testing evaluations) were in the "moderate" category (Malasch, Leiter & Schaufeli, 2008). Out of the three MBI-ES scales, EE is the scale most closely associated with burnout (Malasch, Leiter & Schaufeli, 2008). Increased Depersonalization (DP) scales tend to follow prolonged EE scores, as teachers begin to distance themselves from students after prolonged feelings of emotional exhaustion (Malasch, Leiter & Schaufeli, 2008). A decrease in Personal Accomplishment (PA) scores typically happens last. After an increase in depersonalization, this loss of personal connection to students can diminish a sense of personal accomplishment in teaching (Malasch, Leiter & Schaufeli, 2008). At this point, 69% of Pennsylvania math teacher participants (from both groups) are

experiencing “high” burnout (EE scores). However, the burnout has not yet progressed to “high” DP or “low” PA scores, indicating increased burnout progression.

Potential teacher attrition rates, or the amount of potential for leaving the field of education, were also similar for both the TOR and NTOR participant groups. Because differences failed to be statistically significant, $H3_0$ fails to be rejected. Both displayed a range of responses from 0% to 100% when asked the following: “How serious, or what percent of the time, do you consider leaving the teaching profession (teacher attrition)? I think about quitting _____% of the time.” Only 10% responded that they never consider leaving. Half of the participants responded that they think about quitting 33% or a third of the time or more. About 30% of participants acknowledged that they think about quitting 50% of the time or more.

Research question 2 posed the following question: does high-stakes testing as a component of teacher evaluation lead to a correlation among stress, burnout, and attrition? Data suggest there exists a moderate, positive correlation between high-stakes testing stress and teacher burnout, and between high-stakes testing stress and potential teacher attrition. There exists a strong, positive correlation between teacher burnout and potential teacher attrition. In all three cases there exists a positive correlation. Hence, $H4_0$ is rejected. All three are positive correlations meaning as one variable (i.e. burnout) increases, the other (i.e. potential attrition) also increases. The variables also decrease together. For example, as burnout decreases, potential attrition also decreases. The first two correlations as mentioned above were of moderate degree, meaning the data fit a continuously increasing function pattern, but the points were somewhat scattered around the graph pattern with a correlation coefficient between 0.3 and 0.5 (Brase & Brase, 2015). The strong correlation indicates the points were scattered closer to the function pattern, with some possibly on the graph, and a correlation coefficient between 0.5

and 0.7 (Brase & Brase, 2015). For reference, a very strong correlation would be points even closer to the function with more points on the graph, and a correlation coefficient greater than 0.7 or 0.8 (Brase & Brase, 2015). A perfect correlation finds all points on the graph with a correlation coefficient of 1 or a -1 for a perfect negative correlation for a decreasing function pattern (Brase & Brase, 2015). The data show, as high-stakes testing stress increases, burnout increases, and as high-stakes testing stress increases, potential attrition increases. However, the data show a stronger correlation exists between burnout and potential attrition, meaning teacher burnout is a stronger predictor of potential teacher attrition.

Research Question 3 asked the following: are other factors (gender, school location, school socioeconomic status, years of teaching experience, grade level of students) affecting stress, burnout, and teacher attrition? This question was separated into three parts (stress, burnout, and attrition) and each dependent variable was individually evaluated with each of the above factors.

In the case of high-stakes testing stress, participants who were grouped by gender and school location (rural, suburban, urban), were found to have no statistically significant differences. However, math teacher participants from low-income schools reported increased levels of high-stakes testing stress compared to math teacher participants from high-income schools. Participants with 11-20 years of experience reported the highest levels of high-stakes testing stress compared to the lowest levels communicated by the 31+ years of experience group. Participants at the elementary level reported higher levels of high-stakes testing stress compared to secondary participants reporting lower levels with middle-level teachers reporting levels in between the two. In the case of $H5_{0,}$ this hypothesis is rejected for the case of participants from

low-income schools, participants with 11-20 years of experience and participants at the elementary level. $H5_0$ fails to be rejected for all other factors evaluated in the current study.

In the case of burnout, participants grouped by school location, school socioeconomic status, and grade level taught reported similar levels of burnout. Recall 69% of the participants scored "high" for burnout (EE). However, female participants reported higher levels of burnout than males. Also, participants with 11-20 years of experience reported higher levels of burnout (the same group identified for high-stakes testing stress), especially compared to the 31+ group and the 1-5 years of experience group. In the case of $H6_{0,}$ this hypothesis is rejected for the event of participants with 11-20 years of experience and female participants. $H6_0$ fails to be rejected for all other factors evaluated in the current study.

When evaluating the individual factors with potential teacher attrition, groups compared by gender, school location, school socioeconomic status, and grade level of students taught contained similar ranges of responses. The only exception was the 11-20 years of experience group that reported the highest level of potential attrition compared to the lowest reported level by the 1-5 year group. In the case of $H7_{0,}$ this hypothesis is rejected for the case of participants with 11-20 years of experience. $H7_0$ fails to be rejected for all other factors evaluated in the current study.

Conclusions

Several key findings emerged from data analysis. The first research question focused on possible differences in levels of high-stakes testing stress, teacher burnout, and potential attrition between Teachers of Record (TOR) and Non-Teachers of Record (NTOR) and noted similarities. Research Question 2 focused on correlations among high-stakes testing stress, burnout, and potential teacher attrition. Research Question 3 evaluated factors and explored mixed results.

Research Question 1: Finding More Similarities

Despite these teacher participants being evaluated differently on effectiveness evaluations by the Pennsylvania Department of Education (PDE), TOR and NTOR participants experience similar levels of high-stakes testing stress, teacher burnout, and potential teacher attrition. Therefore, the researcher was unable to establish a definitive cause and effect relationship that directly links high-stakes testing stress to teacher burnout and potential teacher attrition rates for PA public school math participants without qualitative data.

However, several unexpected key findings regarding similarities were identified. Both groups of participants experience the same sources of stress to similarly high levels as revealed by the Educators Stress Test Inventory (ETSI-S) subscale question scores. The TOR/NTOR participants report 76%/62% are anxious before standardized testing and 50%/44% after the testing. Both participant groups also report feeling as though they are being evaluated by standardized testing (82% TOR and 75% NTOR), most likely because the math score is a part of the School Performance Profile (SPP) that appears on all teachers' evaluations in a Pennsylvania school district. Also, K-3 NTOR prepare students for future testing years, and secondary NTOR often teach remediation classes or students in other courses who have not yet passed the Algebra 1 Keystone Exam. These students have until grade 11 to pass for scores to positively reflect on the SPP. Those who teach subjects without standardized tests would not have this same responsibility. Numbers also reveal TOR/NTOR participants feel pressured by some parents (45%/48%) and most administrators (89%/89%) to raise student test scores. The data suggest high-stakes testing stress levels score high for both TOR and NTOR participants. What is unclear is how much other teacher stressors may simultaneously impact high chronic stress levels that contribute to burnout.

The most alarming statistic uncovered in the data remains the significant number of math teacher participants in both groups, 69% of the total participants, who scored in the "high" category for burnout. This score significantly differs from the 1996 sample group of over four thousand U.S. teachers that established the normal curve whereby the mean fell in the "moderate" category (Maslach, Jackson & Leiter, 1996). It is important to note that the 1996 sample was taken before high-stakes testing scores were factored into teacher evaluations. Other prior studies revealed that 5% to 30% of teachers across the U.S. were expected to suffer from burnout at any given time (Blazer & Miami-Dade County Public Schools, 2010), well below 69%. Earlier reported high-stakes testing sources of stress scores from participants (ETSI-S) indicate that conflict or criticism from parents and administrators contributes to the emotional exhaustion component of teacher burnout. These ETSI-S scores also indicate the presence of anxiety from high-stakes evaluation and comprehensive reform that contributes to burnout. Despite these findings, additional factors contributing to the high percentage of burnout cannot be excluded without a qualitative component. Other factors not addressed include work overload, interpersonal conflicts, students' behaviors, role discrepancies, school climate, policies, loss of autonomy, and crisis control caused by faulty equipment (Bakker & Schaufeli, 2000; Burke & Greenglass, 1989; Gloria, Faulk & Steinhardt, 2013; Kokkinkos, 2007; Feuerhahn, Bellingrath & Kudielka, 2013). Furthermore, the data from 1996 is 20 years old, and many aspects of a school, technology, teaching, and society have changed.

Another finding indicated that about half of all participants from both groups of participants responded that they seriously think about leaving the profession 33% of the time or more. The finding equates to thinking about leaving at least 1.6 days or more in a 5 day week. There exists a well-documented link whereby high-stakes testing stress contributes to burnout

which further contributes to attrition (Chang, 2009; Elliot & Dweck, 2007; Malasch, Schaufeli, & Leiter, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011). However, there exist other factors that also contribute to attrition such as low pay, little respect in society, lack of administrative support, poor parent and student attitudes, increasing school violence, little chance for career advancement, and minimal input into decision-making (Pucella, 2011). A qualitative component is needed to determine which factors, if any, also contribute to attrition.

Research Question 2: Correlations

The second research question tested if a correlation existed among high-stakes testing stress, teacher burnout, and potential attrition. Data suggest there exists a moderate, positive correlation between high-stakes testing stress and teacher burnout, and between high-stakes testing stress and potential teacher attrition. There exists a strong, positive correlation between teacher burnout and potential teacher attrition. It is well documented that chronic stress, including high-stakes testing stress, can lead to burnout which in turn contributes to teacher attrition (Chang, 2009; Elliot & Dweck, 2007; Malasch, Schaufeli, & Leiter, 2001; Steinhardt, Smith-Jaggars, Faulk & Gloria, 2011). The data show, as high-stakes testing stress increases, burnout increases and attrition increases. However, the strongest positive correlation of the three for the Pennsylvania math teacher participants in both groups is between burnout and potential attrition. Burnout is a strong predictor of potential attrition, and currently, 69% of all math teachers in the study scored "high" on the EE (burnout) scale of the Malasch Burnout Inventory - Educators Survey (MBI-ES). However, without the qualitative component, the researcher was unable to determine if high-stakes testing stress was the only factor contributing to burnout and attrition.

Research Question 3: Mixed Results

The third research question evaluated the influence of other factors on high-stakes testing stress, teacher burnout, and potential teacher attrition. Findings concerning these other factors are mixed. High-stakes testing stress was higher for teachers at low-income schools, for teachers with 11-20 years of experience, and for teachers at the elementary level. Although high-stakes testing stress levels are statistically significantly higher for these factors, this study cannot exclude other factors that may contribute to the overall stress impacting these participants, and potentially contributing to elevated high-stakes testing stress levels. Contributing factors other than accountability policies may include time constraints, work overload, disruptive student behavior, lack of participation in decision-making, poor working relations with colleagues, conflicting job roles, poor working conditions in facilities, low pay, low social status, and lack of support from the principals, administration, society, and parents (Borg & Riding, 1991; Chaplain, 2007; Grayson & Alvarez, 2008; Kokkinos, 2007; Kyriacou & Sutcliffe, 1978; Kyriacou, 2001; Manthei & Gilmore, 1996; Skaalvik, & Skaalvik, 2009). An additional qualitative component may help explain the increased response from teachers at low-income schools where the faculty is often spread thin and retiring employee work is divided and distributed among remaining teachers, increasing the workload. Other stressors that may impact teachers in low-income districts include poor working conditions in crumbling facilities and low-paying jobs offering low social status in the community. Teachers at low-income and elementary schools may experience conflicting job roles more often as part-time caregivers to some students, ensuring those children's basic needs of food, clothes, and shelter are met along with educational needs. Elementary teachers are more aware of their students' personal needs as they typically tend to fewer numbers of the same students each day and teachers in low-income

districts have greater numbers of students in need. Some elementary level teachers may also feel inadequately prepared to teach math under the high-stakes accountability policies. Despite all of the above factors discussed, teachers must continue to teach students and prepare them for high-stakes testing.

Findings concerning other factors affecting teacher burnout revealed higher burnout scores for teachers with 11-20 years of experience and female teachers. Constant or prolonged exposure to stress disrupts a teacher's ability to cope effectively, preventing a state of well-being, leading to burnout. The physical demands of teaching combined with the simultaneous cognitive demands contribute to increased emotional exhaustion (Feuerhahn, Bellingrath & Kudielka, 2013). Burnout studies are mixed comparing gender. However, this study found that burnout is more prevalent among female teachers and teachers in the 11-20 years experience group. The differences were statistically significant; however, there is a possibility of some response bias as these two groups disproportionally outnumber other groups. Additionally, those suffering from burnout may have been more likely to complete the survey. It is also unclear how other known teacher stressors may contribute to burnout without a qualitative component.

A distinct pattern emerged with the 11-20 years of experience group of math teacher participants. The 11-20 year of experience group has shown statistically significant differences in higher levels of high-stakes testing stress, higher scores for teacher burnout, and higher rates of potential attrition compared to all other groups in the current study. These findings are supported by correlations found within the current study as high-stakes testing stress, burnout, and potential attrition were all found to be positively correlated. The added conflict of criticism or lack of support from parents and administrators, as confirmed by the ETSI-S (Sources of Stress) survey, further contributes evidence to the prolonged high-stakes testing stress that

contributes to the emotional exhaustion sub-scale of burnout. Scores for 69% of math teacher participants rated high for burnout. However, there is a possibility of response bias as the 11-20 years of experience group may be overrepresented and not necessarily reflective of the teaching population. Also, those suffering from burnout symptoms may have been more likely to complete the survey.

Although the above information supports reasons for high levels for both groups of TOR and NTOR participants, it does not adequately explain the reason for the 11-20 years of experience group facing statistically significantly higher levels in all three areas compared to the other groups. Especially when prior studies repeatedly found new teachers to report higher levels of stress, burnout, and attrition. Self-Determination Theory, the theoretical framework for this study, may help clarify some circumstances surrounding the findings. For teachers, work-related factors are considered either autonomy-supportive or controlling (Ryan & Deci, 2000). The teachers in the 11-20 years of experience category, who began their careers in an era of a student-centered focus, are now being told to adapt to a data-driven focus, or an accountability system similar to a business model. The shift has created a newly identified group of teacher stressors including high-stakes testing, student performance data added to teacher effectiveness evaluations, curriculum changes, increased accountability, and strict standards (Hanson, 2006; Kyriacou, 2001; Montgomery & Rupp, 2005). In autonomy supportive conditions, teachers experience the highest quality motivation and emotional well-being, allowing for optimal teaching performance and creativity. However, in an autonomy controlling condition, need thwarting of any of the three conditions for a teacher, has a detrimental impact on well-being (Ryan & Deci, 2000). The new accountability system removes participation in decision-making from the skill set of the once "professionally" respected opinion of the teacher, creating a loss of

autonomy which contributes to burnout (Fernet, Guay, Senecal & Austin, 2012). Additionally, this group has endured longevity in the profession and may have outlasted others who have already left the profession. A qualitative component could strengthen this conjecture and help identify other factors creating higher levels of high-stakes testing stress, burnout, and potential attrition for this group of participants.

The general deduction of this study is that most math teacher participants in Pennsylvania in both groups, the TOR and NTOR, experience high levels of high-stakes testing stress from similar sources, score in the "high" range for burnout, and report high rates of possible teacher attrition. Each of the three variables including, high-stakes testing stress, teacher burnout, and potential teacher attrition, were found to be positively correlated, whereby as one variable increases, the other variable also increases. Burnout and potential attrition had the strongest positive correlation of all pairings, meaning burnout is a strong or reliable predictor of potential attrition. Other factors identified as contributing to high-stakes testing stress included: teaching at a low-income school, teaching in an elementary school, and teaching for 11-20 years. Other factors identified as contributing to teacher burnout include: teaching for 11-20 years and identifying as a female teacher. The only other factor contributing to potential teacher attrition is teaching for 11-20 years.

Recommendations

A primary goal of this investigation was to determine if accountability policies attaching high-stakes testing data to teacher evaluations impacts levels of stress, burnout, and attrition rates for groups of math teachers. Findings suggest many math PA teachers are impacted by these policies. Results of the present study indicate more research is needed in the area of high-stakes testing stress, teacher burnout, and attrition to support these findings. Similar studies should be

expanded to include other tested subject areas such as English Language Arts and Science. Additional longitudinal studies that follow teachers over time could reveal how many teachers leave the field of education, contributing to the validity of the long range predictions. The addition of qualitative data for a mixed methods study would also contribute to the unanswered questions that remain from obtaining only quantitative data in this study. Additional research should investigate ways to support teachers through stress, provide quality professional development, and return autonomy supportive environments to classroom teachers.

Based on the findings and conclusions of this study, lawmakers should consider the larger ramifications of the data-driven, accountability policy system that is currently in place. These education policies force a top-down implementation with little or no input from the majority of actively teaching public school math educators. The accountability business model replaced an autonomy-supportive environment with an autonomy-controlling environment. Once effective, student-centered teachers are now labeled as failures based on the high-stakes testing Value-Added Measures (VAM) system of scoring; the same system found to be 'capricious' and 'arbitrary' by a New York court (Strauss, 2016). Federal and state lawmakers are responsible for the environment that has 69% of Pennsylvania public school math teacher participants scoring 'high' for burnout. Because preventing burnout is easier than reversing it, and burnout was found to have a strong, positive correlation with potential teacher attrition, there exists a strong possibility that Pennsylvania could lose and not replace a large number of currently employed public school math teachers to attrition if the current high-stakes testing policies remain.

To add to the crisis, Pennsylvania, one of the top five teacher producing states in the nation, suffers from a drastic decline in education majors similar to the other top four teacher producing states (Malmont, 2016). A new report found PA enrollment in education majors

decreased by 62%, similar to major enrollment declines in other states across the nation (Malmont, 2016). Due to negative political attacks to pensions, past wage freezes, school funding cuts, larger class sizes, and continued political attacks on teachers, STEM majors opt out of education careers for those with better pay and more stability. In 2013, Pennsylvania awarded 558 math teaching certificates, in 2015, this decreased to 204 (Malmont, 2016). The trend is similar for all other fields of education in Pennsylvania, not just STEM subjects. Hence, if the burnout of the 69% translates to attrition, Pennsylvania may not have enough math teachers to fill the open positions. In turn, the state might need to resort to emergency teaching certifications as Arizona has in the midst of its attrition crisis (Education Recruitment & Retention Task Force, 2015). These inadequately trained teachers have higher rates of attrition than traditionally trained teachers, causing more educational disruption with higher turnover rates (Education Recruitment & Retention Task Force, 2015). Lawmakers must concede that this policy no longer classifies as placing highly qualified teachers in the classroom in an optimal learning environment. It seems as though lawmakers, through federal mandates, are removing the best research-based asset to any classroom, its teachers.

The high levels of sources of high-stakes testing stress could disappear with the removal of the high-stakes tests as a part of teacher evaluation. Reversing the threat of a looming attrition crisis may require the removal of harmful, 'failing' labels, VAM scores, and a data-driven environment. A return to an autonomy-supportive environment and a student-centered focus may be enough to reverse burnout for some educators, thereby reversing their thoughts on attrition and maintaining their public school math teacher status. High-stakes teacher stress, teacher burnout, and potential attrition rates of math teachers could undermine desired progress in Science, Technology, Engineering, and Math (STEM) fields in this country. As lawmakers

continue to engage in creating educational policy changes, research studies such as this can provide a more intricate understanding of the unintended consequences of proposed changes.

References

Ahola, K., Honkonen, T., Isometsa, E., Kalimo, R., Nykyri, E., Aromaa, A., & Lonnqvist, J. (2005). The relationship between job-related burnout and depressive disorders-results from the Finnish Health 2000 Study. *Journal of Affective Disorders, 88*, 55-62. doi: 10.1016/j.jad.2005.06.004

Alliance for Excellent Education (2014). *On the path to equity: Improving the effectiveness of beginning teachers.* Retrieved from http://all4ed.org/wp-content/uploads/2014/07/PathToEquity.pdf

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS[R] EVAAS[R]) in the Houston Independent school District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives, 20*(12), 1-31.

Anderson, M. B. G., & Iwanicki, E. F. (1984). Teacher motivation and its relationship to burnout. *Education Administration Quarterly, 20,* 109-132.

Baker, B. D., Oluwole, J. O., & Green, P. C. (2013). The legal consequences of mandating high stakes decisions on low-quality information: Teacher evaluation in the race-to-the-top era. *Education Policy Analysis Archives, 21*(5), 1-23.

Bakker, A. B., & Schaufeli, W. B. (2000). Burnout contagion processes among teachers. *Journal of Applied Social Psychology, 30*(11), 2289-2308.

Bartholomew, K. J., Ntoumanis, N., Cuevas, R., & Lonsdale, C. (2014). Job pressure and ill-health in physical education teachers: The mediating role of psychological need thwarting. *Teaching and Teacher Education, 37,* 101-107. doi:10.1016/j.tate.2013.10.006

Beer, J., & Beer, J. (1992). Burnout and stress, depression, and self-esteem of teachers. *Psychological Reports, 71,* 1331-1336.

Bill & Melinda Gates Foundation. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching Project*. Seattle, WA: Author.

Blazer, C., & Miami-Dade County Public Schools, R. S. (2010). Teacher burnout. *Information Capsule, 1003*. Research Services, Miami-Dade County Public Schools. Retrieved from http://files.eric.ed.gov/fulltext/ED536515.pdf

Borg, M. G., & Riding, R. J. (1991). Occupational stress and satisfaction in teaching. *British Educational Research Journal, 17*(3), 263-281.

Bouwma-Gearhart, J. (2010). Pre-service educator attrition informed by self-determination theory: Autonomy loss in high-stakes education environments. *Problems of Education in the 21St Century, 26,* 30-41.

Bracey, G. W. (2007). The Sputnik effect: Why it endures, 50 years later. *Education Week, 27.* Retrieved from http://www.edweek.org/ew/articles/2007/10/02/06bracey_web.h27.htm

Brase, C. H., & Brase, C. P. (2013). *Understanding basic statistics* (6th ed.). Boston, MA: Brooks/Cole.

Burke, R. J., & Greenglass, E. R. (1989). The clients' role in psychological burnout in teachers and administrators. *Psychological Reports, 64,* 1299-1306.

Byrne, B. M. (1991). Burnout: Investigating the impact of background variables for elementary, intermediate, secondary and university educators. *Teaching and Teacher Education: An International Journal of Research, 7,* 197-209.

Carroll, T. G., & Foster, E. (2008). *Learning teams: Creating what's next*. Washington, DC: National Commission on Teaching and America's Future.

Chalghaf, N., Azaiez, F., & Elarbi, B. (2014). Triggers of burnout among Tunisian teachers at higher institutes of sport and physical education. *IOSR Journal of Humanities and Social Science, 19*(7), 70-77.

Chan, D. W. (2003). Hardiness and its role in the stress burnout relationship among prospective Chinese teachers in Hong Kong. *Teaching and Teacher Education*, 19, 381-395. doi: 10.1016/j.tate.2006.04.005

Chan, A. H., Chen, K., & Chong, E. Y. (Eds.). (2010). Work stress of teachers from primary and secondary schools in Hong Kong. In *Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol III*.

Chang, M. (2009). An appraisal perspective of teacher burnout: Examining the emotional work of teachers. *Educational Psychology Review, 21*, 193-218.

Chaplain, R. P. (2001). Stress and job satisfaction among primary head teachers: A question of balance? *Educational Management and Administration, 29*(2), 197-215.

Cherniss, C. (1980). *Staff burnout: Job stress in the human services*. Beverly Hills, CA: Sage.

Cheung, F., Tang, C. S.-K., & Tang, S. (2011). Psychological capital as a moderator between emotional labor, burnout, and job satisfaction among school teachers in China. *International Journal of Stress Management, 18,* 348–371. doi:10.1037/a0025787

P Chew. (2015, May 18). Start with family heart history to uncover FH. [Web log comment]. Retrieved from http://blog.sanofi.us/2015/05/18/start-with-family-heart-history-to-uncover-fh/

Civic Impulse. (2014). *H.R. 1 -111th Congress: American Recovery and Reinvestment Act of 2009*. Retrieved from https://www.govtrack.us/congress/bills/111/hr1

Coman, D., Alessandri, M., Gutierrez, A., Novotny, S., Boyd, B., Hume, K., Sperry, L., & Odom, S. (2013). Commitment to classroom model philosophy and burnout symptoms among high fidelity teachers implementing preschool programs for children with autism spectrum disorders. *Journal of Autism & Developmental Disorders, 43*(2), 345-360. doi:10.1007/s10803-012-1573-1

Corcoran, S. P., Jennings, J. L., & Beveridge, A. A. (2011). Teacher effectiveness on high and low-stakes tests. Working paper, New York University.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Crocco, M. S., & Costigan, A. T. (2007). The narrowing of curriculum and pedagogy in the age of accountability: Urban educators speak out. *Urban Education, 42*(6), 512-535.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement.* New York, NY: Teachers College Press.

Darling-Hammond, L., & Richardson, N. (2009, February). Teacher learning: What matters? *Educational Leadership, 5*(66), 46-53.

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality, 62*(1), 119-142.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York, NY: Plenum Press.

Deci, E. L., & Ryan, R. M. (2000). The "what" and the "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11,* 227-268.

Deci, E. L., & Ryan, R. M. (2002). The paradox of achievement: The harder you push, the worse it gets. In D. I. Cordova & J. Aronson (Eds.), *Improving academic achievement: Contributions of social psychology* (pp. 61-87). New York, NY: Academic Press.

Deci, E. L., & Ryan, R. M. (2004). *Handbook of self-determination research.* Rochester, NY: University of Rochester Press.

Deci, E. L., Spiegel, N. H., Ryan, R. M., Koestner, R., & Kauffman, M. (1982). The effects of performance standards on teaching styles: The behavior of controlling teachers. *Journal of Educational Psychology, 74*, 852-859.

Doherty, K. M., & Jacobs, S. (2013). *State of the states 2013 connect the dots: Using evaluations of teacher effectiveness to inform policy and practice.* Retrieved from http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

Duffet, A., Farkas, S. Rothertham, A. J., & Silva, E. (2008). *Waiting to be won over: Teachers speak on the profession, unions, and reform.* Washington, DC: Education Sector.

Education Commission of the States (2014). *Blueprint for college readiness: A 50-state policy analysis.* Retrieved from: http://www.ecs.org/docs/ECSblueprint.pdf

Educator Recruitment & Retention Task Force (2015). *Education recruitment and retention report*. Prepared for Arizona Department of Education. Retrieved from http://www.azed.gov/wp-content/uploads/2015/02/err-initial-report-final.pdf

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2013). *Selecting growth models for school and teacher evaluations.* National Center for Analysis of Longitudinal Data in Education Research. Retrieved from http://www.caldercenter.org/publications/calder-working-paper-80.cfm

Elliot, A. J., & Dweck, C. S. (Eds.). (2007). *Handbook of competence and motivation.* New York, NY: Guilford Press.

Farber, B. A. (1991). *Crisis in education.* San Francisco, CA: Jossey-Bass.

Fernet, C., Guay, F., Senecal, C., & Austin, S. (2012). Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teaching and Teacher Education: An International Journal of Research and Studies, 28*(4), 514-525.

Feuerhahn, N., Bellingrath, S., & Kudielka, B. M. (2013). The interplay of matching and non-matching job demands and resources on emotional exhaustion among teachers. *Applied Psychology: Health & Well-Being, 5*(2), 171-192. doi:10.1111/aphw.12002

Feuerhahn, N., Stamov-Roßnagel, C., Wolfram, M., Bellingrath, S., & Kudielka, B. M. (2013). Emotional exhaustion and cognitive performance in apparently healthy teachers: A longitudinal multi-source study. *Stress & Health: Journal of The International Society for The Investigation of Stress, 29*(4), 297-306. doi:10.1002/smi.2467

Firth, G. H., & Mims, A. (1985). Burnout among special education paraprofessionals. *Teaching Exceptional Children, 17,* 225-227.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th eds.). Boston, MA: Pearson.

Flink, C., Boggiano, A. K., & Barrett, M. (1990). Controlling teaching strategies: Undermining children's self-determination and performance. *Journal of Personality and Social Psychology, 59,* 916-924.

Freedman, S. W., & Appleman, D. (2009). In it for the long haul: How teacher education can contribute to teacher retention in high-poverty, urban schools. *Journal of Teacher Education, 60*(3), 323-337.

Gardner, H., Csikszentmihalyi, M., & Damon, W. (2001). *Good work: When excellence and ethics meet*. New York, NY: Basic.

Gloria, C., Faulk, K., & Steinhardt, M. (2013). Positive affectivity predicts successful and unsuccessful adaptation to stress. *Motivation & Emotion, 37*(1), 185-193. doi:10.1007/s11031-012-9291-8

Gold, Y. (1984). The factorial validity of the Malasch Burnout Inventory in a sample of California elementary and junior high school classroom teachers. *Educational and Psychological Measurement, 44,* 1009-1016.

Grandey, A. A. (2000). Emotion regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology, 5*, 95–110. doi:10.1037/1076-8998.5.1.95

Grayson, J. L., & Alvarez, H. K. (2008). School climate factors relating to teacher burnout: A mediator model. *Teaching and Teacher Education, 24,* 1349-1363.

Green, J. L., Camilli, G., & Elmore, P. B. (Eds.). (2006). *Handbook of complementary methods in education research.* New York, NY: Routledge.

Greenglass, E., Burke, R. J., & Ondrack, M. (1990). A gender role perspective of coping and burnout. *Applied Psychology: An International Review, 39,* 5-27.

Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology, 52,* 890-898.

Hakanen, J. J., Bakker, A. B., & Schaufeli, W. B. (2006). Burnout and work engagement among teachers. *Journal of School Psychology, 43,* 495-513.

Hall, S. H. (2016, March 27). Growing number of parents opting children out of PSSA tests. *The Times-Tribume.com.* Retrieved from http://thetimes-tribune.com/news/growing-number-of-parents-opting-children-out-of-pssa-tests-1.2023749

Hammen, C., & deMayo, R. (1982). Cognitive correlates for teacher stress and depressive symptoms: Implications for attributional models of depression. *Journal of Abnormal Psychology, 91*(2), 96-101.

Hanby, D. (2011, October). *Pioneer reflections from Ohio's 2010-11 journey: Pilot year insights from Leg I and Leg II travelers.* Presentation at the Ohio Council of Teacher Educators Conference, Columbus, OH.

Hanson, A. M. (2006). *No child left behind: High-stakes testing and teacher burnout in urban elementary schools.* Dissertation submitted to the University of Phoenix, Phoenix, AZ. ERIC Document Reproduction Service No. ED493443.

Hastings, R. P., & Bham, M.S. (2003). The relationship between student behaviour patterns and teacher burnout. *School Psychology International*, *24*(1), 115-127.

Hewitt, K. K. (2015). Educator evaluation policy that incorporates EVAAS value-added measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives, 23*(76), 1-49.

Hochschild, A. R. (1979). Emotion work, feeling rules, and social structure. *American Journal of Sociology, 85,* 551–575. doi:10.1086/227049

Howard, T. C. (2003). Who receives the short end of the shortage? Implications of the U.S. teacher shortage on urban schools. *Journal of Curriculum and Supervision, 18*(2), 142-160.

Hull, J. (2013). *Trends in teacher evaluation: How states are measuring teacher performance.* Retrieved from http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-Evaluation-Full-Report-PDF.pdf

IBM Corp. (2015). IBM SPSS: Statistics for Windows (Version 23.0) [software]. Armonk, NY: IBM Corp.

Iwanicki, E. F., & Schwab, R. L. (1981). A cross validation study of the Malasch Burnout Inventory. *Educational and Psychological Measurement, 41*, 1167-1174.

Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational Leadership, 60*(8), 30-33.

King, J. A., & Rohmer-Hirt, J. A. (2011). Internal evaluation in American public school districts: The importance of externally driven accountability mandates. In B. B. Volkov & M. E. Baron (Eds.), *Internal evaluation in the 21st century: New Directions for Evaluation, 132,* 73-86.

Kokkinos, C. M. (2006). Factor, structure and psychometric properties of the Malasch Burnout Inventory-Educators Survey among elementary and secondary school teachers in Cyprus. *Journal of the International Society for the Investigation of Stress, 22*(1), 25-33.

Kokkinos, C. M. (2007). Job stressors, personality and burnout in primary school teachers. *British Journal of Education Psychology, 77,* 229-243.

Kress, S., Zechmann, S., & Schmitten, J. M. (2011). When performance matters: The past, present, and future of consequential accountability in public education. *Harvard Journal on Legislation, 48,* 185-234.

Kyriacou, C. (2001). Teacher stress: Directions for future research. *Educational Review, 53*(1), 27-35. doi: 10.1080/00131910120033628

Kyriacou, C., & Sutcliffe, J. (1978). A model of teacher stress. *Educational Studies, 4*(1), 1-6. doi: 10.1080/ 0305569780040101

Ladwig, B. (2015, July 3). Schools struggling to find teachers. *The Greensburg Daily News.* Retrieved from http://www.greensburgdailynews.com/news/local_news/schools-struggling-to-find-teachers/article_b273ddbc-a74a-5c7a-b659-3977f29cdd7d.html

Leedy, P. D., & Ormrod, J. E. (2010). *Practical research: Planning and design* (9th ed.). Upper Saddle River, NJ: Pearson.

Lund, A., & Lund, M. (2015). *Laerd Statistics.* Retrieved from https://statistics.laerd.com/aboutus.php

Maisel, M. (2016, February 3). Gov. Wolf delays keystone exams, wants to add $200M to education budget. *Fox 43 News.* Retrieved from http://fox43.com/2016/02/03/gov-wolf-delays-keystone-exams-wants-to-add-200m-to-education-budget/

Malmont, J. (2016, June 4). Closer look: Teacher shortage looms on horizon. *The Sentinel.* Retrieved from http://cumberlink.com/news/local/closer_look/closer-look-teacher-shortage-looms-on-horizon/article_f2ebf75f-bfb6-5b8a-a704-0bd51316b895.html

Manthei, R., & Gilmore, A. (1996). Teacher stress in intermediate schools. *Educational Research, 38*(1), 3-19.

Maslach, C. (1993). Burnout: A multidimensional perspective. In W.B. Schaufeli, C. Maslach, & T. Marek (Eds.), *Professional burnout: Recent development in theory and research* (pp. 19-32). Washington, DC: Taylor & Francis.

Malasch, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior, 2,* 99-113.

Maslach, C., & Jackson, S. E. (1986). *MBI: The Maslach Burnout Inventory: Manual research edition.* Palo Alto, CA: Consulting Psychologists Press.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach Burnout Inventory manual* (3rd ed.). Palo Alto, CA: Consulting Psychologist Press.

Maslach, C., & Leiter, M. P. (2008). Early predictors of job burnout and engagement. *The Journal of Applied Psychology, 93,* 498-512.

Malasch, C., Leiter, M. P., & Schaufeli, W. B. (2008). Measuring burnout. In C. L. Cooper & S. Cartwright (Eds.), *The Oxford Handbook of Organizational Wellbeing* (86-108). Oxford: Oxford University Press.

Malasch, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology, 52,* 397-422.

Malasch, C., Jackson, S. E., & Schwab, R. L. (1986). *Malasch Burnout Inventory – educators survey*. Retrieved from http://www.mindgarden.com/117-maslach-burnout-inventory

Margolis, J., & Deuel, A. (2009). Teacher leaders in action: Motivation, morality, and money. *Leadership and Policy in Schools, 8*(3), 264-286.

McNamara, G., & O'Hara, J. (2008). The importance of the concept of self-evaluation in the changing landscape of educational policy. *Studies in Educational Evaluation, 34,* 173-179.

MET (2013). Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study. Retrieved from http://www.metproject.org/downloads/MET_Ensuring_Fair_ and_Reliable_Measures_Practitioner_Brief.pdf

Montgomery, C., & Rupp, A.A. (2005). A meta-analysis for exploring the diverse causes and effects of stress in teachers. *Canadian Journal of Education, 28*(3), 458-486.

Morse, R. (2015). How states compare in the 2015 best high schools rankings. *U.S. News & World Report.* Retrieved from: http://www.usnews.com/education/best-high-schools/articles/how-states-compare

Muijs, D. (2011). *Doing quantitative research in education with SPSS* (2nd ed.). Thousand Oaks, CA: Sage.

National Assessment of Educational Progress (2013). Retrieved from http://nces.ed.gov/nationsreportcard/about/

National Center for Education Statistics (2015). *Digest of education statistics.* Retrieved from https://nces.ed.gov/programs/digest/current_tables.asp

National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform.* Retrieved from http://www2.ed.gov/pubs/NatAtRisk/index.html

National Council on Teacher Quality. (2003). SREB focus on retention. *Teacher Quality Bulletin* (January 31).

Nevo, D. (2001). School evaluation: Internal or external? *Studies in Educational Evaluation, 27,* 95-106.

Newhouse, K. (2015, August 10). PSSA math scores drop sharply in Lancaster county schools. *Lancaster Online.* Retrieved from http://lancasteronline.com/news/local/pssa-math-

scores-drop-sharply-in-lancaster-county-schools/article_726ec29e-3f8d-11e5-b507-1b86474a56ba.html

Newhouse, K. (2015, November 24). House approves two year keystone exam delay; Senate must ok bill amendment. *Lancaster Online.* Retreived from http://lancasteronline.com/news/local/house-approves-two-year-keystone-exam-delay-senate-must-ok/article_542e4dd4-92aa-11e5-a98c-77c013ea29c0.html

Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives, 18*(23), 1-27.

Ng, J. C., & Peter, L. (2010). Should I stay or should I go? Examining the career choices of alternatively licensed teachers in urban schools. *Urban Review, 42*(2), 123-142.

Niederberger, M., & Crompton, J. (2016, January 20). Keystone exams delayed until 2019. *Pittsburgh Post-Gazette.* Retrieved from http://www.post-gazette.com/news/state/2016/01/20/State-moves-ahead-with-plan-to-delay-Keystone-Exams-as-graduation-requirement/stories/201601200211

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

Oberle, E., & Sconert-Reichl, K. A. (2016). Stress contagion in the classroom? The link between classroom teacher burnout and morning cortisol in elementary school students. *Social Science & Medicine, 159,* 30-37.

Office for Human Research Protections (2009). *Title 45 code of federal regulations part 46: protection of human subjects.* Prepared for Department of Health and Human Services. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html

Omdahl, B. L., & Fritz, J. M. H. (2006). *Coping with problematic relationships in the workplace: Strategies that reduce burnout.* Paper presented at the annual meeting of the International Communication Association, Dresden Germany, June 2006. Retrieved from http://www.allacademic.com/meta/p93215_index.html

O'Shaughnessy, l. (2013, October 31). More U.S. states leaning on teachers. Retrieved from http://www.cbsnews.com/news/more-us-states-leaning-on-teachers/

Otero López, J. M., Bolaño, C. C., Santiago Mariño, M. J., & Pol, E. V. (2010). Exploring stress, burnout, and job dissatisfaction in secondary school teachers. *International Journal of Psychology & Psychological Therapy, 10*(1), 107-123.

Palochko, J. (2015, June 16). Superintendents call the push to delay the state's graduation requirements a "good first step". *The Morning Call.* Retrieved from http://www.mcall.com/news/breaking/mc-lehigh-valley-keystone-exams-changes-20150616-story.html

Pathe, S., & Choe, J. (2013). *A brief overview of teacher evaluation controversies.* Retrieved from http://www.pbs.org/newshour/rundown/teacher-evaluation-controversies/

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Pecheone, R. L., & Chung, R. R. (2006). Evidence in teacher education: The Performance Assessment for California Teachers (PACT). *Journal of Teacher Education, 57*(1), 22-36.

Pecheone, R. L., & Stansbury, K. (1996). Connecting teacher assessment and school reform. *Elementary School Journal, 97,* 163-177.

Pennsylvania Department of Education (2015). *2014-2015 Professional personnel individual staff report* [Data file]. Retrieved from http://www.education.pa.gov/Data-and-Statistics/Pages/Professional-and-Support-Personnel.aspx#.VoXHLxUrLIU

Pennsylvania Department of Education (2015). *Educator effectiveness systems in Pennsylvania.* Retrieved from http://www.portal.state.pa.us/portal/server.pt/community/educator_effectiveness_project/20903

Pennsylvania Department of Education (2015). *Frequently asked questions: Act 48 continuing professional education requirements. Retrieved from* file:///C:/Users/bpisc_000/Downloads/FAQ%20Act%2048%20Continuing%20Professional%20Education%20Requirements.pdf

Pennsylvania Department of Education (2015). *Pennsylvania Value Added Assessment System (PVASS).* Retrieved from http://www.portal.state.pa.us/portal/server.pt/community/state_assessment_system/20965/pennsylvania_value_added_assessment_system_%28pvaas%29/1426500

Pennsylvania Department of Education (2015). *Pennsylvania Value Added Assessment System (PVASS): Guide to PVASS teacher specific reporting.* Retrieved from: https://pvaas.sas.com/

Pennsylvania Department of Education (2015). *Pennsylvania Value Added Assessment System (PVASS): SY 13-14 teacher specific reporting.* Retrieved from https://pvaas.sas.com/

Pennsylvania Department of Education (2015, April). *PVAAS and the transition of PA's assessment system*. Retrieved from file:///C:/Users/bpisc_000/Downloads/PVAAS_Transition_Assessment%20System.pdf

Pennsylvania Department of Education (2015). *PVAAS teacher-specific reporting* [Training video slides]. Retrieved from https://evaas.sas.com/videos/PA/TeacherSpecificReports_ForTeachers/story.html

Pennsylvania Department of Education (2015). *PVAAS teacher value added summary: A* [Data file]. Retrieved from http://www.education.pa.gov/Pages/default.aspx#.VoIEfhUrLIU

Pennsylvania Department of Education (2015). *PVAAS teacher value added summary: B* [Data file]. Retrieved from http://www.education.pa.gov/Pages/default.aspx#.VoIEfhUrLIU

Pennsylvania State Education Association (2015, March). Toxic Testing. *PSEA Voice,* 8-12.

Pucella, T. J. (2011). The impact of national board certification on burnout levels in educators. *Clearing House, 84*(2), 52-58. doi:10.1080/00098655.2010.511306

Rey, L., Extremera, N., & Pena, M. (2012). Burnout and work engagement in teachers: Are sex and level taught important?. *Ansiedad Y Estrés, 18*(2/3), 119-129.

Rothstein, J. (2011). Review of learning about teaching. *Boulder, CO: National Education Policy Center, 27,* 1-12.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist, 55,* 68-78.

Ryan, R. M., & Deci, E. L. (2015). Self-determination theory: An approach to human motivation and personality. Retrieved from http://www.selfdeterminationtheory.org/

Ryan, R. M., & La Guardia, J. G. (1999). Achievement motivation within a pressured society: Intrinsic and extrinsic motivations to learn and the politics of school reform. In T. Urdan (Ed.), *Advances in motivation and achievement* (Vol. 11, pp. 45-85). Greenwich, CT: JAI Press.

Ryan, R. M., Stiller, J., & Lynch, J. H. (1994). Representations of relationships to teachers, parents, and friends as predictors of academic motivation and self-esteem. *Journal of Early Adolescence, 14,* 226-249.

Santoro, D. A. (2011). Good teaching in difficult times: Demoralization in the pursuit of good work. *American Journal of Education,118*(1), 1-23. doi:10.1086/662010

Schaufeli, W. B., Daamen, J., & van Mierlo, H. (1994). Burnout among Dutch teachers: An MBI validity study. *Educational and Psychological Measurement, 54,* 803-812.

Schaufeli, W. B., & Greenglass, E. R. (2001). Introduction to special issue on burnout and health. *Psychology & Health, 16,* 501-510.

Schaufeli, W.B., & Enzmann, D. (1998). *The burnout companion to study and practice: A critical analysis.* Philadelphia, PA: Taylor & Francis, Ltd.

Scott, E. (2006). *Stress and burnout: Burnout symptoms and causes.* Retrieved from http://stress.about.com/od/burnout/a/stressn_burnout.htm

Scott, E. (2010). *Job burnout: Job factors that contribute to employee burnout.* Retrieved from http://stress.about.com/od/burnout/a/job_burnout.htm

Sentell, W. (2012, October, 16). Teachers: Educators at top schools fear reviews will cost their jobs. *The Baton Rouge Advocate.* Retrieved from http://theadvocate.com/home/4161514-125/teachers-reviews-threaten-their-jobs

Shaughnessy, J., Zechmeister, E., & Jeanne, Z. (2011). *Research methods in psychology* (9th ed.). New York, NY: McGraw-Hill.

Skaalvik, E. M., & Skaalvik, S. (2009). Does school context matter? Relations with teacher burnout and job satisfaction. *Teaching & Teacher Education, 25*(3), 518-524. doi:10.1016/j.tate.2008.12.006

Smith, A., Brice, C., Collins, A., Matthews, V., & McNamara, R. (2000). *The scale of occupational stress: A further analysis of the impact of demographic factors and type of job.* Sudbury: Health and Safety Executive.

Sparks, S. D. (2016, June 27). 1 in 4 teachers miss 10 or more school days, analysis finds. *Education Week.* Retreived from http://www.edweek.org/ew/articles/2016/06/27/1-in-4-teachers-miss-10-or.html

Stecher, B., Garet, M., Holtzman, D., & Hamilton, L. (2012). Implementing measures of teacher effectiveness: Teachers and leaders in pilot sites say the evaluation systems are beneficial, and leaders say they're well-prepared for their new responsibilities, even as they worry about increased workload and stress from the new expectations. *Phi Delta Kappan, 94*(3), 39-43.

Steinhardt, M. A., Smith-Jaggars, S. E., Faulk, K. E., & Gloria, C. T. (2011). Chronic work stress and depressive symptoms: Assessing the mediating role of teacher burnout. *Stress & Health: Journal of the International Society for the Investigation of Stress, 27*(5), 420-429. doi:10.1002/smi.1394

Stotko, E. M., Ingram, R., & Beaty-O'Ferrall, M. E. (2007). Promising strategies for attracting and retaining successful urban teachers. *Urban Education, 42*(1), 30-51.

Strauss, V. (2016, May 10). Judge calls evaluation of N.Y. teacher 'arbitrary' and 'capricious' in case against new U.S. secretary of education. *Washington Post.* Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2016/05/10/judge-calls-evaluation-of-n-y-teacher-arbitrary-and-capricious-in-case-against-new-u-s-secretary-of-education/

Strauss, V. (2015, August 24). The real reason behind the U. S. teacher shortage. *The Washington Post.* Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/wp/2015/08/24/the-real-reasons-behind-the-u-s-teacher-shortage/

Taylor, I. M., Ntoumanis, N., & Standage, N. (2008). A self-determination theory approach to understanding the antecedents of teachers' motivational strategies in physical education. *Journal of Sports and Exercise Psychology, 30,* 75-94.

The Research Advisors (2015). *Sample size table.* Retrieved from http://www.research-advisors.com/tools/SampleSize.htm

Tye, B. B., & O'Brien, L. (2002). Why are experienced teachers leaving the profession? *Phi Delta Kappan, 84*(1), 24. Retrieved September 29, 2004, from the ProQuest database.

U.S. Department of Education (2015). *Every Student Succeeds Act (ESSA).* Retrieved from http://www.ed.gov/essa

U.S. National Library of Medicine (2016). *Depression: What is burnout syndrome?* Retrieved from http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0072470/

Valli, L., Croninger, R. G., Chambliss, M. J., Graeber, A. O., & Buese, D. (2008). *Test driven: High-stakes accountability in elementary schools*. New York, NY: Teachers College Press.

Van Horn, J. E., Schaufeli, W. B., Greenglass, E., & Burke, R. J. (1997). A Canadian-Dutch comparison of teachers' burnout. *Psychological Reports, 81,* 371-382.

Vandenberghe, R., & Huberman, A. M. (1999). *Understanding and preventing teacher burnout.* New York, NY: Cambridge University Press.

Vogt, W. P. (2007). *Quantitative research methods for professionals.* Boston, MA: Pearson.

von der Embse, N. P., Kilgus, S. P., Solomon, H. J., Bowler, M., & Curtiss, C. (2015). Initial development and factor structure of the educator test stress inventory. *Journal of Psychoeducational Assessment, 33*(3), 223-237. doi:10.1177/0734282914548329

von der Embse, N. P., Pendergast, L. L., Segool, N., Saeki, E., & Ryan, S. (2016). The influence of test-based accountability policies on school climate and teacher stress across four states. *Teaching & Teacher Education, 59,* 492-502. doi:10.1016/j.tate.2016.07.013

Walker, T. (2014, November 2). NEA survey: Nearly half of teachers consider leaving profession due to standardized testing. Retrieved from http://neatoday.org/2014/11/02/nea-survey-nearly-half-of-teachers-consider-leaving-profession-due-to-standardized-testing-2/

Watanabe, T. (2012, October 28). Measuring the worth of a teacher? L. A. Unified school district's academic growth over time measurement system, based on students' progress on standardized tests, spurs debate over fairness, accuracy. Retrieved from http://articles.latimes.com/2012/oct/28/local/la-me-teacher-evals-20121029

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad.* Dallas, TX: National Staff Development Council and Standford, CA: Stanford Center for Opportunity Policy and Education.

Wheeler, D. L., Vassar, M., Worley, J. A., & Barnes, L. L. B. (2011). A reliability generalization meta-analysis of coefficient alpha for the Malasch Burnout Inventory. *Educational and Psychological Measurement, 71*(1), 231-244.

Wisniewski, L., & Gargiulo, R. M. (1997). Occupational stress and burnout among special educators: A review of the literature. *The Journal of Special Education, 31*(3), 325-346.

Zhang, Q., & Sapp, D. A., (2008). A burning issue in teaching: The impact of teacher burnout and nonverbal immediacy on student motivation and effective learning. *Journal of Communication Studies, 1*(2), 152-168.

Appendix A

IRB Approval Letter and Email Addendum

Indiana University of Pennsylvania
www.iup.edu

Institutional Review Board for the
Protection of Human Subjects
School of Graduate Studies and Research
Stright Hall, Room 113
210 South Tenth Street
Indiana, Pennsylvania 15705-1048

P 724-357-7730
F 724-357-2715
irb-research@iup.edu
www.iup.edu/irb

March 23, 2016

Becky L. Piscitella
3741 Graham Avenue
Windber, PA 15963

Dear Ms. Piscitella:

Your proposed research project, "Stress, Burnout, and Attrition: Implications of Student Performance Data on Math Teacher Effectiveness Evaluations," (Log No. 16-086) has been reviewed by the IRB and is approved. In accordance with 45CFR46.101 and IUP Policy, your project is exempt from continuing review. This approval does not supersede or obviate compliance with any other University requirements, including, but not limited to, enrollment, degree completion deadlines, topic approval, and conduct of university-affiliated activities.

You should read all of this letter, as it contains important information about conducting your study.

Now that your project has been approved by the IRB, there are elements of the Federal Regulations to which you must attend. IUP adheres to these regulations strictly:

1. You must conduct your study exactly as it was approved by the IRB.
2. Any additions or changes in procedures must be approved by the IRB before they are implemented.
3. You must notify the IRB promptly of any events that affect the safety or well-being of subjects.
4. You must notify the IRB promptly of any modifications of your study or other responses that are necessitated by any events reported in items 2 or 3.

The IRB may review or audit your project at random *or* for cause. In accordance with IUP Policy and Federal Regulation (45CFR46.113), the Board may suspend or terminate your project if your project has not been conducted as approved or if other difficulties are detected

Although your human subjects review process is complete, the School of Graduate Studies and Research requires submission and approval of a Research Topic Approval Form (RTAF) before you can begin your research. If you have not yet submitted your RTAF, the form can be found at http://www.iup.edu/page.aspx?id=91683 .

IRB to Becky L. Piscitella, March 23, 2016

While not under the purview of the IRB, researchers are responsible for adhering to US copyright law when using existing scales, survey items, or other works in the conduct of research. Information regarding copyright law and compliance at IUP, including links to sample permission request letters, can be found at http://www.iup.edu/page.aspx?id=165526.

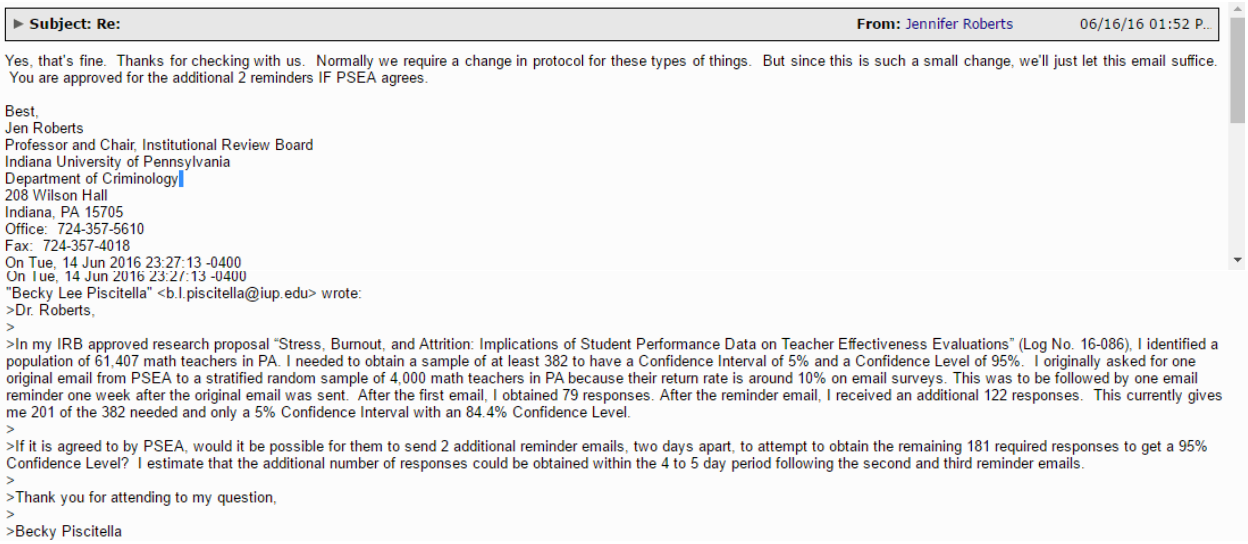
I wish you success as you pursue this important endeavor.

Sincerely,

Jennifer Roberts, Ph.D.
Chairperson, Institutional Review Board for the Protection of Human Subjects
Professor of Criminology

JLR:jeb

Cc: Dr. Valeri Helterbran, Dissertation Advisor
Dr. Kelli Jo Kerry-Moran, Graduate Coordinator
Ms. Brenda Boal, Secretary



▶ **Subject: Re:** **From:** Jennifer Roberts 06/16/16 01:52 P..

Yes, that's fine. Thanks for checking with us. Normally we require a change in protocol for these types of things. But since this is such a small change, we'll just let this email suffice. You are approved for the additional 2 reminders IF PSEA agrees.

Best,
Jen Roberts
Professor and Chair, Institutional Review Board
Indiana University of Pennsylvania
Department of Criminology
208 Wilson Hall
Indiana, PA 15705
Office: 724-357-5610
Fax: 724-357-4018
On Tue, 14 Jun 2016 23:27:13 -0400
On Tue, 14 Jun 2016 23:27:13 -0400
"Becky Lee Piscitella" <b.l.piscitella@iup.edu> wrote:
>Dr. Roberts,
>
>In my IRB approved research proposal "Stress, Burnout, and Attrition: Implications of Student Performance Data on Teacher Effectiveness Evaluations" (Log No. 16-086), I identified a population of 61,407 math teachers in PA. I needed to obtain a sample of at least 382 to have a Confidence Interval of 5% and a Confidence Level of 95%. I originally asked for one original email from PSEA to a stratified random sample of 4,000 math teachers in PA because their return rate is around 10% on email surveys. This was to be followed by one email reminder one week after the original email was sent. After the first email, I obtained 79 responses. After the reminder email, I received an additional 122 responses. This currently gives me 201 of the 382 needed and only a 5% Confidence Interval with an 84.4% Confidence Level.
>
>If it is agreed to by PSEA, would it be possible for them to send 2 additional reminder emails, two days apart, to attempt to obtain the remaining 181 required responses to get a 95% Confidence Level? I estimate that the additional number of responses could be obtained within the 4 to 5 day period following the second and third reminder emails.
>
>Thank you for attending to my question,
>
>Becky Piscitella

Appendix B

Site Approval Request

Indiana University of Pennsylvania
Department of Professional Studies in Education

Project: Stress, Burnout, and Attrition: Implications of Student Performance Data on Math Teacher Effectiveness Evaluations

January 23, 2016

Dear PSEA Officers, Mr. Wazeter, and Mr. Mercer,

I, Becky Piscitella, the researcher, am formally requesting site approval from PSEA as an email distributer of my dissertation research survey link to elementary, middle grade, and high school math teachers. As per our recent phone conversations and emails, I understand it is your policy to obtain approval from certain members of PSEA prior to granting this request. If approved by PSEA, the Indiana University of Pennsylvania (IUP) Internal Review Board (IRB) requires official approval correspondence to be on PSEA letterhead for me to include in my IRB application. This letter can be sent to me via email at cbfj@iup.edu. Once approved by IUP and the IRB Committee, I will forward the IRB approval to PSEA before any distribution of emails can begin.

Summary of Site Approval:

1. PSEA will agree to distribute the dissertation survey link via the PSEA listserv and email an approval letter on PSEA letterhead to Becky Piscitella at cbfj@iup.edu.
2. Upon receiving IUP and IRB Committee approval, Becky Piscitella will email a copy of the IRB approval and the survey link to Mr. Wazeter and Mr. Mercer at PSEA. No individuals or districts will be known to me, so anonymity is assured and no coercion can be applied to any teacher agreeing to participate. Procedures for participants to withdraw from the study will be included in the informed consent letter sent to them through PSEA.
3. The PSEA research department will extract math teachers from the listserv and random sample 1,000 teachers from within each group: Elementary (K-3), Elementary (4-6), Middle Level Math (7-9), HS Math (10-12)
    a. Surveys will be sent to all members in a group of 1,000 or fewer teachers.
    b. The elementary level may be one group (K-6) with a random sample of 2,000 if it cannot be separated. The research department will notify Becky Piscitella of the outcome after extracting math teacher data.
4. One follow-up reminder email, authored and provided by Becky Piscitella, will be sent to original recipients by PSEA one week following the original email with the link to participate.
5. Becky Piscitella is responsible for collecting and analyzing data, and reporting findings.

Seeking Institutional Approval: If you agree to participate in this study as described and indicate your willingness to participate on PSEA letterhead, the researcher, Becky Piscitella, will seek approval from Indiana University of Pennsylvania Institutional Review Board for the Protection of Human Subjects. Phone: (724) 357-7730

Thank you. Please feel free to contact me if you have any questions about this study.

Respectfully,
Becky Piscitella

Principal Investigator
Becky L. Piscitella
Principal Investigator and Doctoral Candidate
Professional Studies in Education
Curriculum & Instruction
Indiana University of Pennsylvania
3741 Graham Ave.
Windber, PA 15963
Phone: (814) 467-0295
cbfj@iup.edu

Faculty Sponsor
Dr. Valeri. R. Helterbran
Faculty Sponsor
Professional Studies in Education
Indiana University of
Pennsylvania
Davis Hall Room 323
Indiana, PA 15705
Phone: (724) 357-2400
vhelter@iup.edu

Appendix C

PSEA Site Approval Permission

PSEA PENNSYLVANIA STATE EDUCATION ASSOCIATION

400 North Third Street
P.O. Box 1724
Harrisburg, PA 17105-1724

(717) 255-7000 • (800) 944-PSEA (7732)
Fax: (717) 255-7128 • (717) 255-7124
www.psea.org

W. Gerard Oleksiak, *President*
Dolores M. McCracken, *Vice President*
Richard W. Askey, *Treasurer*
James G. Vaughan, *Executive Director*

March 7, 2016

Project Director for
Mrs. Becky Piscitella, Doctoral Student
Indiana University of Pennsylvania
Department of Professional Studies in Education
Davis Hall
Indiana, PA 15705

To whom this matter concerns:

The purpose of this letter is to notify you that the officers and executive director of PSEA have approved Mrs. Piscitella's survey research project, *A Quantitative Study: Stress, Burnout, and Attrition: Implications of Student Performance Data on Math Teacher Effectiveness Evaluations*. The letter serves as recognition that PSEA has given official permission and site approval for this project, which is Mrs. Piscitella's dissertation research.

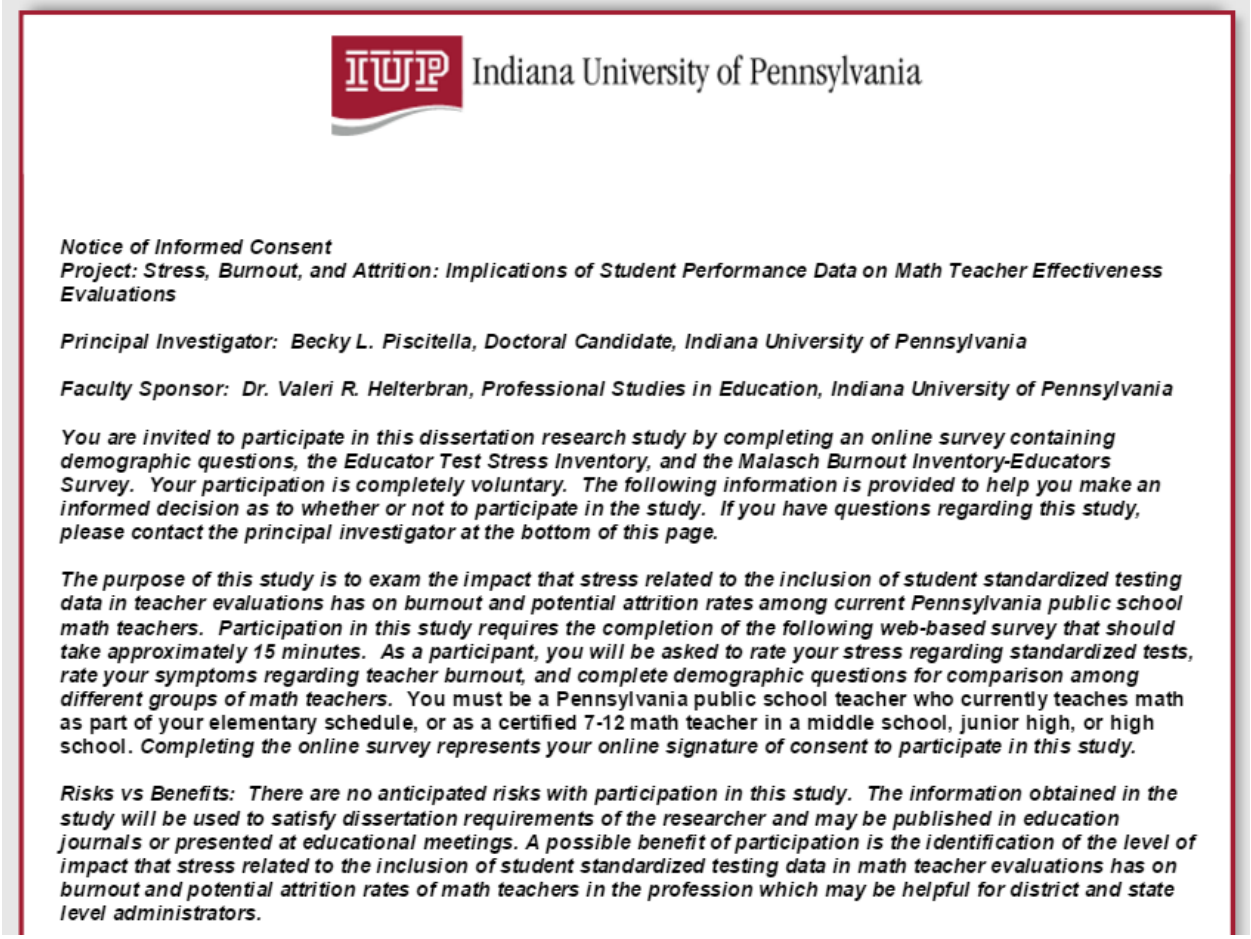
Their approval means that PSEA staff will send PSEA members e-mail communications that contain links to Mrs. Piscitella's survey. Elementary, middle grade, and high school math teachers will receive the e-mails with the survey links. Mrs. Piscitella's survey will proceed later this winter and in the spring. The exact schedule has yet to be determined.

Sincerely,

Daniel C. Mercer, Ph.D.
Assistant Director of Research, Survey Director
PSEA
400 North Third Street
Harrisburg, PA 17105
(717) 255-7038
dmercer@psea.org

Appendix D

Notice of Informed Consent



Indiana University of Pennsylvania

Notice of Informed Consent
Project: Stress, Burnout, and Attrition: Implications of Student Performance Data on Math Teacher Effectiveness Evaluations

Principal Investigator: Becky L. Piscitella, Doctoral Candidate, Indiana University of Pennsylvania

Faculty Sponsor: Dr. Valeri R. Helterbran, Professional Studies in Education, Indiana University of Pennsylvania

You are invited to participate in this dissertation research study by completing an online survey containing demographic questions, the Educator Test Stress Inventory, and the Malasch Burnout Inventory-Educators Survey. Your participation is completely voluntary. The following information is provided to help you make an informed decision as to whether or not to participate in the study. If you have questions regarding this study, please contact the principal investigator at the bottom of this page.

The purpose of this study is to exam the impact that stress related to the inclusion of student standardized testing data in teacher evaluations has on burnout and potential attrition rates among current Pennsylvania public school math teachers. Participation in this study requires the completion of the following web-based survey that should take approximately 15 minutes. As a participant, you will be asked to rate your stress regarding standardized tests, rate your symptoms regarding teacher burnout, and complete demographic questions for comparison among different groups of math teachers. You must be a Pennsylvania public school teacher who currently teaches math as part of your elementary schedule, or as a certified 7-12 math teacher in a middle school, junior high, or high school. *Completing the online survey represents your online signature of consent to participate in this study.*

Risks vs Benefits: There are no anticipated risks with participation in this study. The information obtained in the study will be used to satisfy dissertation requirements of the researcher and may be published in education journals or presented at educational meetings. A possible benefit of participation is the identification of the level of impact that stress related to the inclusion of student standardized testing data in math teacher evaluations has on burnout and potential attrition rates of math teachers in the profession which may be helpful for district and state level administrators.

Confidentiality: Your identity will remain anonymous. You will be asked to enter a pseudonym and date of birth for the purpose of locating and deleting your data if you choose to withdraw from the study. Data from the study will remain on a thumb drive in a locked safe in the researcher's home for three years before being destroyed in accordance with federal guidelines.

Compensation: Participants will not be compensated for participation in this study. Similarly, no penalty shall occur for opting not to participate.

Voluntary Participation: Participation in this study is voluntary. You may withdraw at any time by notifying the researcher by mail or email containing your pseudonym and date of birth for data identification. Your information will be removed and not used in the final study. If you have not finished the survey and wish to withdraw, you may do this by closing the web browser without clicking submit.

Questions Regarding Study Participation: If you have questions regarding this study, please contact the principal investigator at the bottom of this page. Please print this page for your records.

Statement of Institutional Approval: This study has been approved by the Indiana University of Pennsylvania Institutional Review Board for the Protection of Human Subjects. Phone: (724) 357-7730

Becky L. Piscitella
Principal Investigator and Doctoral Candidate
Professional Studies in Education
Curriculum & Instruction
Indiana University of Pennsylvania
3741 Graham Ave.
Windber, PA 15963
Phone: (814) 467-0295
cbfj@iup.edu

Dr. Valeri. R. Helterbran
Faculty Sponsor
Professional Studies in Education
Indiana University of Pennsylvania
Davis Hall Room 323
Indiana, PA 15705
Phone: (724) 357-2400
vhelter@iup.ed

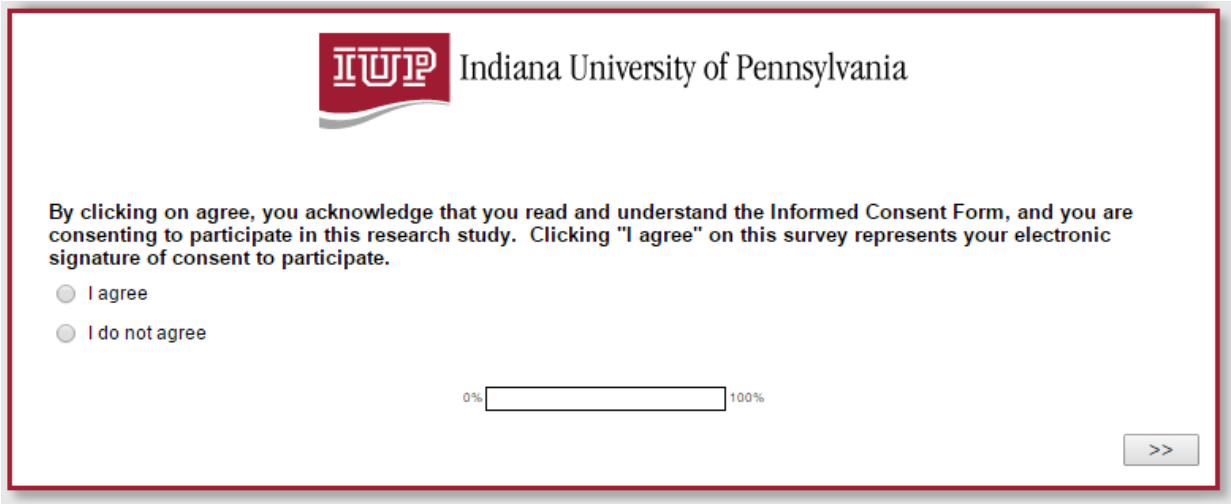
- ○ Continue
- ○ Exit

0% 100%

>>

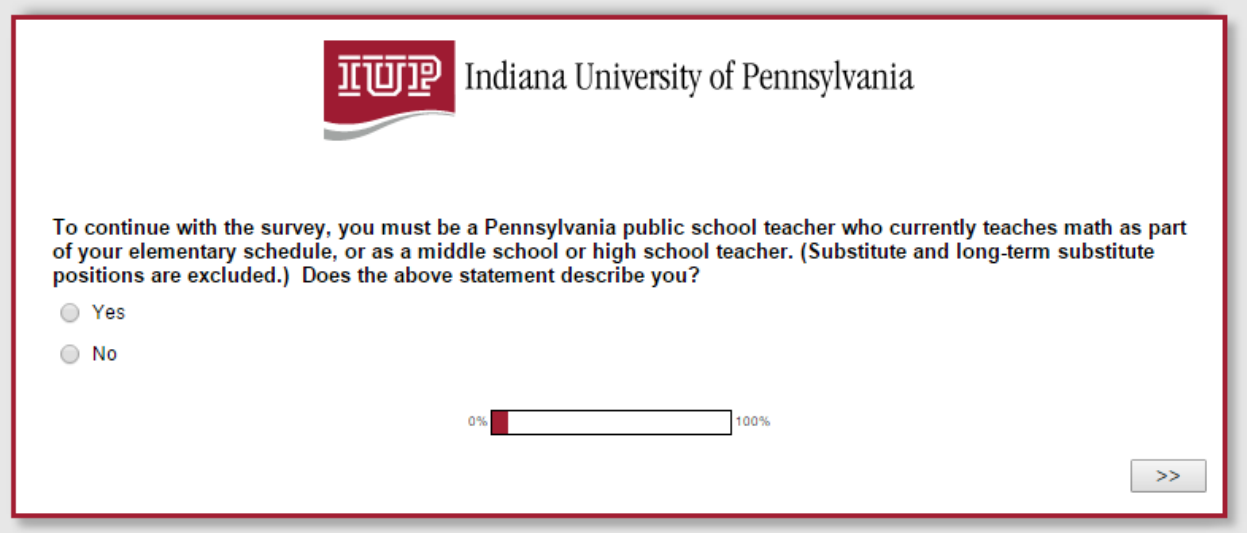
Appendix E

Informed Consent Digital Signature



If the participant selects “I do not agree,” the survey ends.
If the participant selects “I agree,” the survey progresses to the next question.

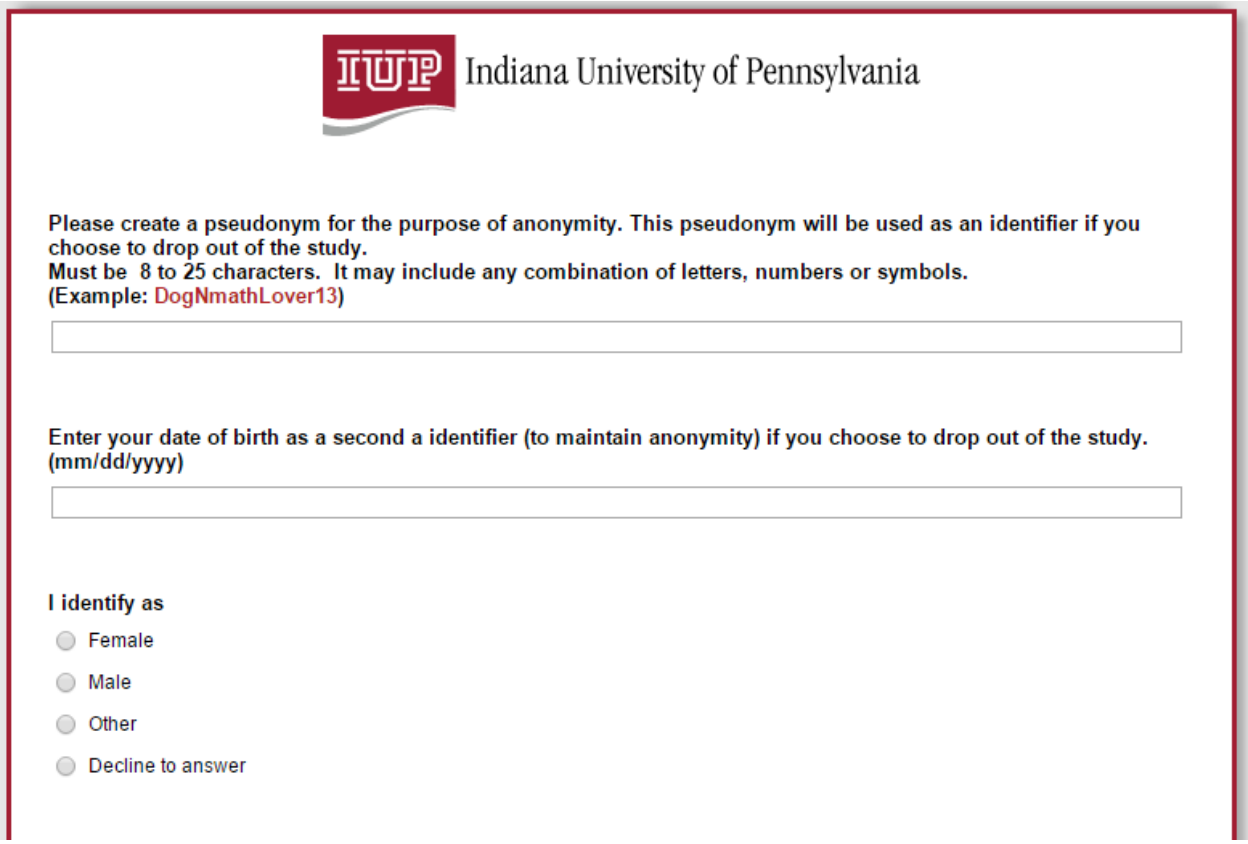
Confirmation of Requirements to Participate



If the participant selects “No,” the survey ends.
If the participant selects “Yes,” the survey opens to the three main parts that includes: the demographic questions, ESTI, and MBI-ES.

Appendix F

Survey Demographic Questions



How may years of math teaching experience do you have prior to this year? (can be combined experience from public and private schools as well as experience teaching in states other than PA)

- 2 to 5
- 6 to 10
- 11 to 20
- 21 to 30
- 31+

The school I currently work in is considered a _____ _____ school.

- low-income
- middle-income
- high-income

The school I currently work in is considered a(n) _____ school.

- rural
- suburban
- urban

I currently teach students who are *mostly* at the _____ level.

- Elementary (K-5)
- Middle school (6-8)
- Secondary (9-12)

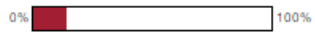
Are you considered a Teacher of Record this year? In other words, do you currently teach math to grades 4, 5, 6, 7, or 8, or Algebra 1, while verifying a roster of students with the state?

- Yes
- No

How often, or what percent of the time, do you consider leaving the math teaching profession (teacher attrition)?

0 10 20 30 40 50 60 70 80 90 100

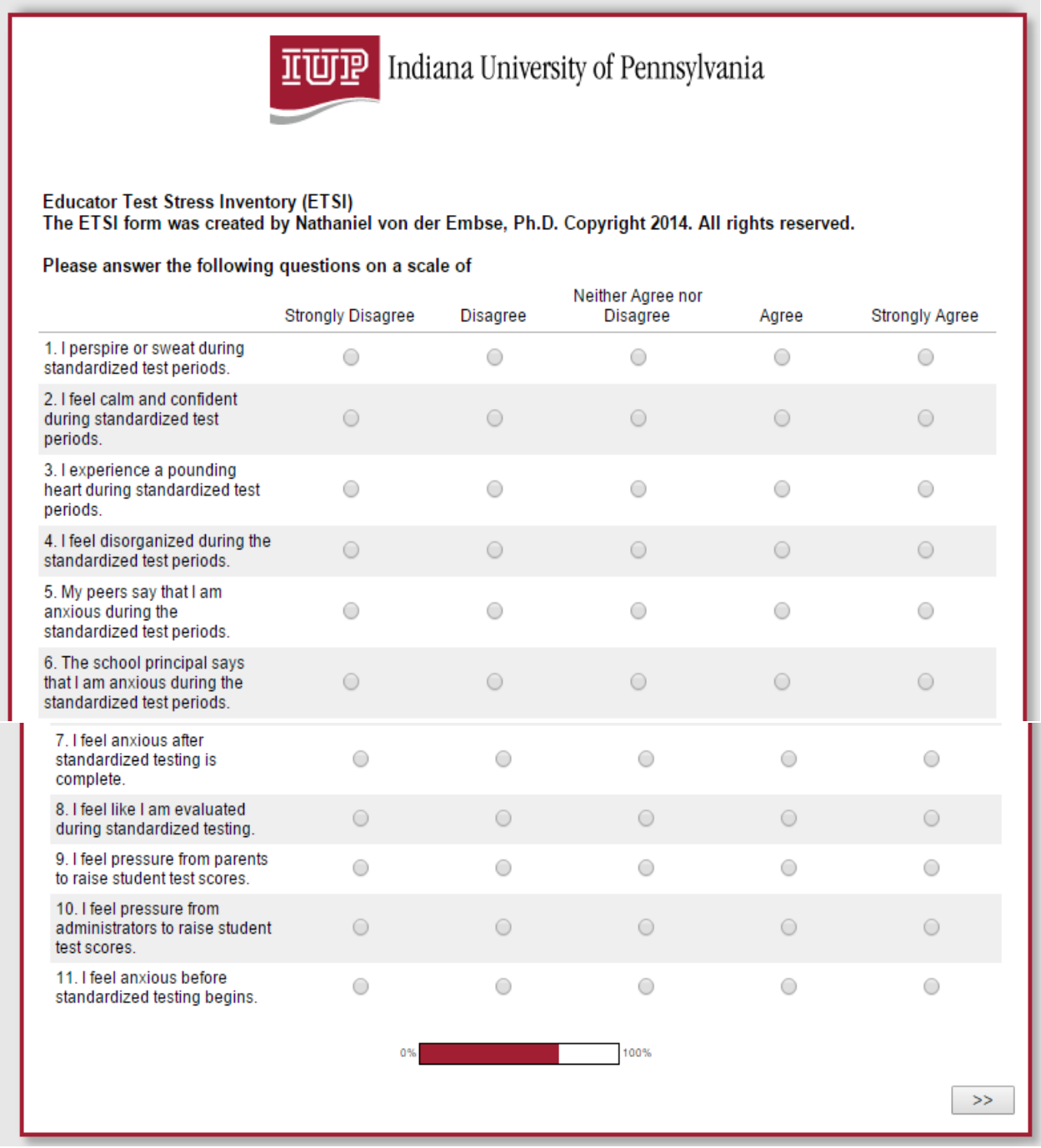
I seriously think about quitting teaching _____% of the time



>>

Appendix G

Educator Test Stress Survey (ETSI) in Qualtrics Format



Indiana University of Pennsylvania

Educator Test Stress Inventory (ETSI)
The ETSI form was created by Nathaniel von der Embse, Ph.D. Copyright 2014. All rights reserved.

Please answer the following questions on a scale of

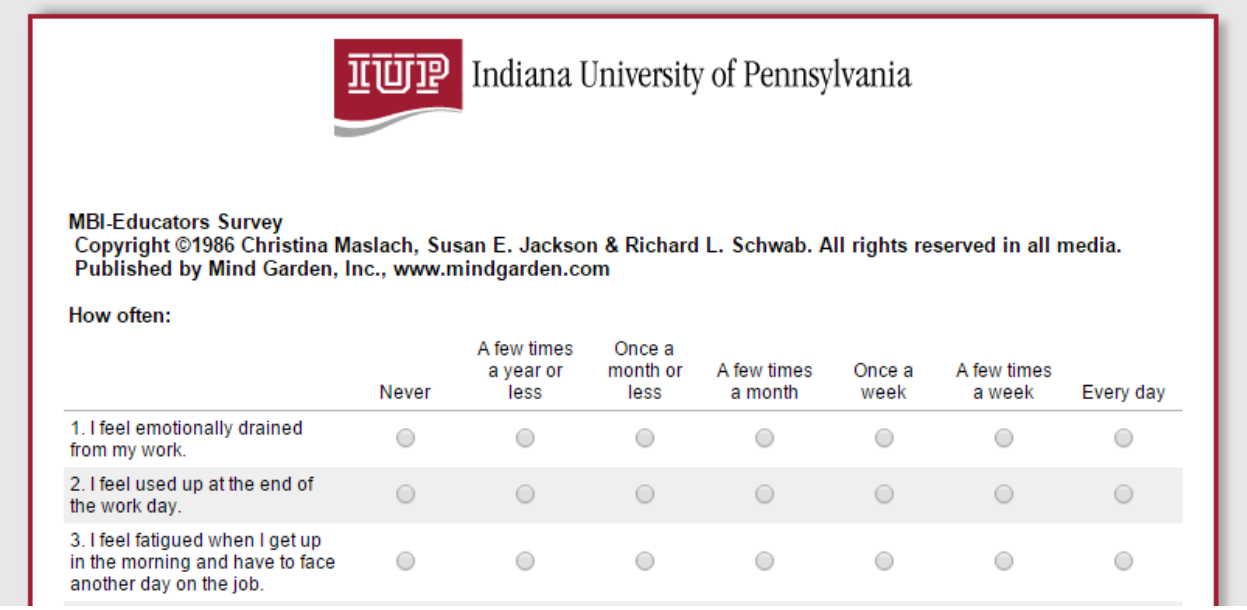
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
1. I perspire or sweat during standardized test periods.					
2. I feel calm and confident during standardized test periods.					
3. I experience a pounding heart during standardized test periods.					
4. I feel disorganized during the standardized test periods.					
5. My peers say that I am anxious during the standardized test periods.					
6. The school principal says that I am anxious during the standardized test periods.					
7. I feel anxious after standardized testing is complete.					
8. I feel like I am evaluated during standardized testing.					
9. I feel pressure from parents to raise student test scores.					
10. I feel pressure from administrators to raise student test scores.					
11. I feel anxious before standardized testing begins.					

0% 100%

>>

Appendix H

Malasch Burnout Inventory - Educators Survey (MBI-ES) in Qualtrics Format



Indiana University of Pennsylvania

MBI-Educators Survey

How often:

	Never	A few times a year or less	Once a month or less	A few times a month	Once a week	A few times a week	Every day
1. I feel emotionally drained from my work.	○	○	○	○	○	○	○
2. I feel used up at the end of the work day.	○	○	○	○	○	○	○
3. I feel fatigued when I get up in the morning and have to face another day on the job.	○	○	○	○	○	○	○

Only 3 questions of 22 are permitted for viewing or reproduction in a proposal or dissertation per permission (see Appendix J).

Appendix I

Educator Test Stress Survey (ETSI) Permission

Appendix J

Malasch Burnout Inventory-Educators Survey (MBI-ES) Permission

mind garden

www.mindgarden.com

To whom it may concern,

This letter is to grant permission for the above named person to use the following copyright material for his/her thesis or dissertation research:

Instrument: ***Maslach Burnout Inventory, Forms: General Survey, Human Services Survey & Educators Survey***

Copyrights:

MBI-General Survey (MBI-GS): Copyright ©1996 Wilmar B. Schaufeli, Michael P. Leiter, Christina Maslach & Susan E. Jackson. All rights reserved in all media. Published by Mind Garden, Inc., www.mindgarden.com

MBI-Human Services Survey (MBI-HSS): Copyright ©1981 Christina Maslach & Susan E. Jackson. All rights reserved in all media. Published by Mind Garden, Inc., www.mindgarden.com

MBI-Educators Survey (MBI-ES): Copyright ©1986 Christina Maslach, Susan E. Jackson & Richard L. Schwab. All rights reserved in all media. Published by Mind Garden, Inc., www.mindgarden.com

Three sample items from a single form of this instrument may be reproduced for inclusion in a proposal, thesis, or dissertation.

The entire instrument may not be included or reproduced at any time in any published material.

Sincerely,

Robert Most
Mind Garden, Inc.
www.mindgarden.com