DETERMINING INDICATORS AND PREDICTORS OF EFFECTIVE NINTH GRADE ALGEBRA I TEACHERS

by

LAUREN FAY HOLMES

RANDALL E. SCHUMACKER, COMMITTEE CHAIR
STACY HUGHEY-SURMAN
C. JOHN TARTER
SARA TOMEK
STEPHEN TOMLINSON

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in the
Department of Educational Studies in Psychology,
Research Methodology, and Counseling in
the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2016

ABSTRACT

Determining indicators and predictors of effective teachers is not only important to helping current practicing teachers improve, but is also important for the training of the new teaching force. There are now measures of effective teaching, but there are no sets of factors or attributes that define an effective teacher. Increased accountability falls on current educators and drives the questions of not only what makes an effective teacher, but how can teacher effectiveness be reliably measured? As the world looks to the United States to deliver advancements in the areas of science, technology, engineering, and mathematics, the questions remain - What qualities indicate effective Algebra I teachers and can student standardized assessments predict teacher effectiveness? A specific focus on Algebra I teachers is important as it is a required course for all students and it is a course that establishes future success in the areas of Science, Technology, Engineering, and Mathematics (STEM). This study established latent classes of effective and non-effective Algebra I teachers using measurements of teacher promoted engagement, teacher promoted understanding, and classroom management from the MET Project database and supports class identification accuracy with principal effectiveness scores. Indicators of effective teachers are also determined through the Tripod 7Cs Student Perception Survey, the Classroom Assessment Scoring System, and the Framework for Teaching. Established categories of teacher effectiveness helped to identify teacher characteristics that defined effective and non-effective teachers. Teacher effectiveness classification was compared with student achievement data from the ACT QualityCore® Algebra I, and was not able to discriminate established latent classes of teacher effectiveness, suggesting student test scores did not predict effective teachers identified in this study. Knowledge of effective teacher characteristics will help to improve current teacher education programs, establish successful and meaningful professional development, and increase teacher retention rates in the United States, and shift the focus of teacher evaluation away from standardized test scores.

DEDICATION

I would like to dedicate this dissertation to my parents who have supported me through my numerous years in academia. Also, I would like to dedicate this dissertation to my husband and son who have dealt with my absence, constant grumpiness, and neurosis in the last months of my dissertation. I love you all!

LIST OF ABBREVIATIONS

ABIC	Adjusted Bayesian Information Criterion
AERA	American Educational Research Association
AIC	Akaike's Information Criterion
BAM	Balanced Assessment in Mathematics
BIC	Bayesian Information Criterion
CLASS	Classroom Assessment Scoring System
CRP	culturally relevant pedagogy
CTES	comprehensive teacher evaluation system
ESEA	Elementary and Secondary Education Act
FFT	Framework for Teaching
GAO	Government Accountability Office
HLM	hierarchical linear modeling
ICPSR	Inter-university Consortium for Political and Social Research
LCA	latent class analysis
MANOVA	multivariate analysis of variance
MCK	mathematical content knowledge
MET	Measures of Effective Teaching
NCLB	No Child Left Behind
PCK	pedagogical content knowledge
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Programme for International Student Assessment

PLATO	Protocol for Language Arts Teaching Observations
RATE	Rapid Assessment of Teacher Effectiveness
R2T	Race to the Top
SES	socioeconomic status
SPSS©	Statistical Product and Service Solutions, Inc.
STEM	Science, Technology, Engineering, and Mathematics
TEAM	Tennessee Educator Acceleration Model
TPEP	Teacher/Principal Evaluation Project
TRIPOD	Tripod 7Cs Student Perception Survey
TSAS	Texas State Accountability System
TVAAS	Tennessee Value Added Assessment System
VAMs	value-added measures
VDE	virtual data enclave

ACKNOWLEDGEMENTS

I am greatly appreciative for the support of colleagues, friends, and faculty members who enabled me to complete this research project. I am most grateful for the guidance, expertise, and motivation given by my chair, Dr. Randall Schumacker. He forced me to focus my efforts and to accomplish the completion of my dissertation. Dr. Schumacker has gone above and beyond in helping me to achieve my degree. I would also like to thank Dr. Stacy Hughey-Surman for her incredible emotional support and advising through my doctoral degree program. I would like to thank Dr. Sara Tomek for listening to my various methodological concerns, Dr. Stephen Tomlinson for allowing me to view educational research with a different lens, and Dr. C. John Tarter for his enthusiasm and support of my research topic. Again, great thanks to Dr. Schumacker and my committee members for your input, questions, and support throughout this process.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Background

The teaching field is undergoing a renaissance period. We are shifting on a continuum from behavioral learning to cognitive reasoning that begins with the establishment of the Elementary and Secondary Education Act (ESEA), and continues to span the shift from No Child Left Behind (NCLB) standards to the establishment of the Common Core with the nation's competition in Race to the Top (R2T). We have learned our lesson that success on a test is not solely correlated to student knowledge and that without the ability to think critically, analyze, and problem solve we are losing our strength in the fields of Science, Technology, Engineering, and Mathematics (STEM). The nationwide focus has turned to teacher accountability and effectiveness as the keystone for teacher assessment in American education.

A problem when establishing measures for evaluating teacher effectiveness through teacher characteristics resides in the fundamentals of educational philosophy, and seemingly the differences between essentialists and progressives. Essentialists have a traditional view that is teacher-centered, focusing on structure, regimen, and factual learning. Progressives seek to have a student-learning model focusing on connections, adaption, and applied learning. Both views can represent effective teachers and this is an important and under discussed issue in the literature on teacher effectiveness. The instruments used in this study focus on characteristics of effective progressive style teachers.

Even as early as the late 60s, characteristics of good teachers were being studied. Hamachek (1969) examined the relevant literature and concluded that good teachers are flexible, perceptive, democratic, experimental, knowledgeable, and reflective. In the late 1980s, Searles and Kudeki (1987) profiled characteristics of outstanding teachers and concluded that effective teachers are able to maintain a respectful classroom atmosphere, are experts in their subject, are good at presenting content clearly, are concerned with how students are performing, are creative, and are able to provide differentiated instruction. Johnson (1997) described six key indicators of effective teachers including: teachers are content experts, teachers are caring, teachers can control the classroom, teachers interact with students, teachers keep students engaged, and teachers have students with improved achievement. It is clear that effective teachers require much fluidity in all areas of practice. These characteristics are not arbitrary, but until recently there was no comprehensive manner in which to measure teacher characteristics quantitatively.

In order to exhaust related information, it is also important to look to other forms of research. Previous qualitative case studies of different school districts allowed the development of four factors that provided meaningful evaluation of teachers: organizational commitment, evaluator competence, teacher-administrator collaboration, and strategic compatibility (Leinhardt & Wise, 1984). Furthermore, Patrick and Smart (1998) discovered that teacher effectiveness is multidimensional and made up of three factors: respect for students, ability to change students, and organization and presentation skills. One study examining secondary science and math teachers found that teacher dispositions including willingness to collaborate, interactive skills, and the ability to reflect and adjust practice were important for determining effective teachers (Miller & Davison, 2006).

The multidimensionality of teacher effectiveness makes it a difficult topic to research, but the extent of research thus far does reveal some clear themes. Good teachers have strong content knowledge, are great at promoting student understanding, effectively promote student engagement, and can control and manage their classroom environments. The importance of effective teachers is especially relevant in the content area of mathematics. The negative connotations associated with mathematics classes add another level of complexity that magnifies the necessity to research and identify qualities of effective teachers.

Statement of the Problem

Teacher effectiveness is a controversial but not a new topic in the field of educational research. Shifts in the purpose of studying this measure have moved from initial interest in the teaching field, to involvement in civil rights issues of the Coleman report, to quantification of identifiable measures assessing what it means to be an effective teacher. Since the establishment of NCLB, educational research took a quantitative shift in what was deemed as the "era of assessment". Recently, the Student Success Act of 2015 relinquished mandated NCLB testing and allowed local and state governments to decide best practices. However, the heart of the issue, the quality of United States students' STEM ability in the world market, is still the utmost concern. In order to make student ability quantifiable, researchers turn to test scores, but solely using students' standardized test scores is flawed when used to assess teacher effectiveness. Scherrer (2013) argued that teachers cannot be reduced to value-added measures because very different teaching methods can yield student success and that teachers, like students, need individualized interventions. A rebuttal of test score use shifted the research focus to the incorporation of teacher observations in the classroom. Today, researchers have shown the value of using a composite method for evaluating teacher effectiveness that uses multiple measures.

Currently there is no agreed upon successful method for identifying effective teachers based upon teacher quality characteristics. The current push of identifying quality teachers with standardized test scores through value-added measures was recently denounced by The American Educational Research Association (AERA) when the organization stated that there is a severe risk of teacher misclassification if a value added method is used for the evaluation of educators (AERA, 2015). Until teachers can be identified appropriately by the quality of their daily teaching practices, rather than by sole reliance on test results, teachers will not receive merit and recognition for their hard work and dedication to the profession. Also, without a sound basis for identifying characteristics of effective teachers, it will be difficult to improve teacher professional development and teacher training programs.

Purpose of the Study

Research exists on teacher effectiveness, whether qualitatively based on classroom observations or quantitatively based on the analysis of student demographics and school characteristics related to test scores. The naturalistic (qualitative) or the anti-naturalistic (quantitative) approach have both attempted to provide research results related to teacher effectiveness and accountability. In general, naturalists believe that scientific and philosophical theories are not mutually exclusive (Godfrey-Smith, 2003). Science can be a resource of philosophy by contributing answers to philosophical questions. Some anti-naturalists argue that the naturalistic approach is an inappropriate way to resolve questions of the social sciences, while other more moderate anti-naturalists claimed it is an inappropriate way to answer certain kinds of philosophical questions.

Naturalism endorses the methods of the natural sciences to justify meaning and significance in human actions, while anti-naturalism weighs on social sciences for gaining meaning (Rosenberg, 2008). Naturalists posit that the social sciences should mirror or at least mimic the physical sciences to be successful; while, anti-naturalists claim that physical science methods can only be used to study objects and not subjects.

Conventionally, the difference in approaches tends to lend itself in a more quantitative or a more qualitative research study. It is important in this situation to ask: does the research or does the researcher decide? The importance behind this question was considered in the selection of secondary data for this study and for the research methodology used. Regardless of a qualitative or quantitative philosophical belief, the funding of research on teacher effectiveness has been more supported by quantitative research methods.

Using or creating a source of information that takes into account the importance of quantitative and qualitative methods is important in answering questions of teacher effectiveness. The Measures of Effective Teaching Project produced a database demonstrating the initial ability of creating decisive parameters including student surveys, teacher observations, and assessment scores to measure teacher effectiveness (Mihaly, McCaffrey, Staiger, & Lockwood, 2013). Exploration of this database and its varied measurements permitted investigation of important characteristics for studying teacher effectiveness in a way that did not solely commit to quantitative or qualitative approaches, thus a mixed methods approach.

The solution, however, for assessing teacher quality may lie in a multidimensional approach that captures the complexity of the process in education (Graue, M., Delaney, K., & Karch, A., 2013). Halpin and Kieffer (2015) called for "a measurement model that more directly addresses the cross-classified structure of teacher observation data, while also remaining feasible to implement in practice" (p. 271). A novel approach is still desirable for the world of research on teaching and learning because an agreed upon model for successfully identifying effective and non-effective teachers is not prevalent in the research literature.

As policy drives educational change, it is paramount that researchers use measurement instruments that demonstrate score reliability and validity for identifying effective teachers. A missing piece of the puzzle is the impact of the dynamics of teacher performance on student learning (Atteberry, A., Loeb, S., & Wyckoff, J., 2013). The advent of R2T creates a focus on the necessity of teacher quality and the use of identified teacher skills over value added measures (Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J., 2009). Presently the most comprehensive collection of multidimensional data is contained in the Measures of Effective Teaching Project (MET) funded by the Bill and Melinda Gates Foundation. Starting in 2009, the MET project partnered with 3,000 teacher volunteers and numerous independent researchers to find multiple measures that can reliably be used in teacher evaluation to assess teacher effectiveness.

The MET database demonstrated interest in creating decisive measures of effective teachers, including student surveys, teacher observations, and assessment scores (Mihaly, K., McCaffrey, D.F., Staiger, D.O., & Lockwood, J.R., 2013). The data base was developed to help researchers identify characteristics of effective teachers and determine what implications their findings could have for future teacher training and current teacher professional development.

Teacher effectiveness measures using the MET database have predicted student achievement based upon teacher effectiveness in a student-teacher randomized scenario (Kane, T.J., McCaffrey, D.F., Miller, T., & Staiger, D.O., 2013). However, in addition to teacher observations, student opinion, and testing success, MET also included teacher and principal opinions as useful measures of teacher effectiveness. The importance of a principal’s decision about teacher success is determined more by a teacher’s behaviors and skills, including compassion for students, work ethic, and classroom management, rather than their content knowledge or other qualifications (Engel, 2013).

In the past, multiple measurements have been combined in linear models to explore teacher quality (Johnson, Kahle, & Fargo, 2007; Muñoz, Scoskie, & French, 2013; Gargani & Strong, 2014; Harris, Ingle, and Rutledge, 2014). However, there is a gap in the research literature when studying multiple measures assessing teacher effectiveness, especially with

Algebra I teachers. This study therefore focused on the question of how to use multiple measurements to identify characteristics of effective Algebra I teachers.

Study Synopsis and Research Questions

This study used measures of teacher characteristics to determine whether latent classes of effective/non-effective teachers exist, rather than pre-determining and categorizing teachers. The sample represented a diverse geographic and socioeconomic spectrum of school districts, teachers, and students. Data for the MET project database was gathered from 317 schools located in the following school districts: the Charlotte-Mecklenburg Schools, the Dallas Independent Schools, the Denver Public Schools, the Hillsborough County Public Schools, the Memphis Public Schools, the New York City Schools, and the Pittsburgh Public Schools. Data gathered from the MET project encompassed two to three years. This study examined ninth grade Algebra I teachers, which is reflected in the sample size of 152 teachers, and focused on data from the second year of collection. The study used selected second year data for the latent class analysis due to the establishment of the study's randomization and desired teacher commitment in a voluntary data collection approach.

Teacher characteristics classified into the categories of teacher promoted student engagement, teacher promoted student understanding, and classroom environment were used (see Figure 1). Latent class analysis for the data examined seven observed polytomous variables. Characteristics came from data including the Tripod 7Cs Student Perception Survey (TRIPOD), The Classroom Assessment Scoring System (CLASS) observational measure, and The Framework for Teaching (FFT) observational protocol. These characteristics were used in the latent class analysis to determine groups of effective/non-effective teachers. Class differences in characteristics indicating effective/non-effective Algebra I teachers were then compared. Conceptually, student views of teachers, observations of teachers, and principal reviews of teachers are all important teacher qualities that can be used to obtain the latent classes of teachers. Next, decision accuracy was used to validate the latent classes established by conducting a χ^2 test of independence between the established latent classes and the principal rating scores of teacher effectiveness.

Finally, student achievement data on the ACT QualityCore® Algebra I content was used to assess the categorical prediction of effective and non-effective teachers by conducting a discriminant analysis.

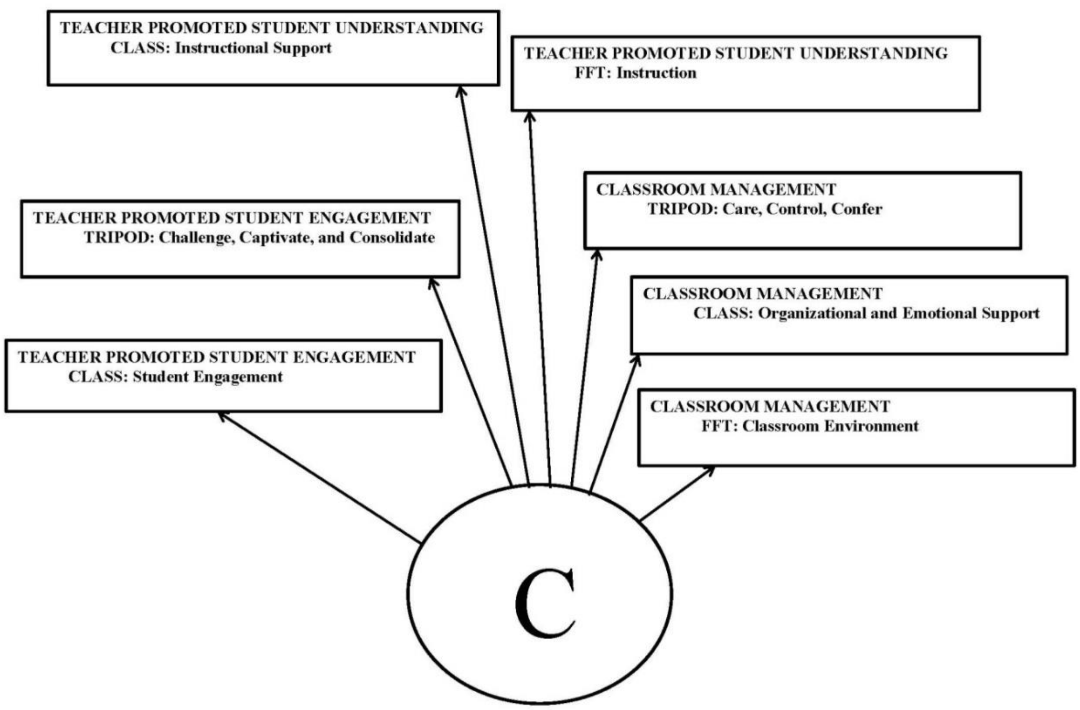

Figure 1. Latent class model for identifying effective/non-effective Algebra I teachers

Research Questions

1. What are the identifiable characteristics that define latent classes of Algebra I teachers?

2. What characteristics indicate effective Algebra I teachers?

3. How do student scores on the ACT QualityCore ® Algebra I standardized exam predict effective/non-effective Algebra I teachers?

Study Significance and Implications

The art of teaching is now a science that contains identifiable parameters necessary for the establishment of successful teachers. These universal parameters are not definitive in scope but can be identifiable through the examination of a diverse group of educators. The implementation of current policy is striving for applicable measures of teacher effectiveness through qualitative and quantitative characteristics. Evidence of success in teacher evaluation is arguably directly realized in the Tennessee Educator Acceleration Model (TEAM) and indirectly realized in the Texas State Accountability System (TSAS). Many states nationwide are implementing teacher evaluation practices including Washington State's Teacher/Principal Evaluation Project (TPEP), which uses the Marzano Teacher Evaluation Model, and New York's Teacher Effectiveness program that uses Danielson's updated *Framework for Teaching* (FFT). As R2T searches for measures of teacher effectiveness to improve the quality of student learning, all involved stakeholders desire to define and evaluate the outgrowth of value added models in establishing an evaluation model for teacher accountability.

The MET project's initial findings have established that effective teaching can be measured through a system of balanced weights that indicate multiple measurements including the following: classroom observations, student surveys, and student assessment gains (Mihaly, McCaffrey, Staiger, & Lockwood, 2013). Information on teacher effectiveness measures is highly pertinent to the Government Accountability Office (GAO) for the development and coordination process to improve teacher quality in higher education institutions and state authorized programs (Teacher Quality, 2009).

Many obstacles are faced in the training and development of effective teachers beginning with poor undergraduate programs that fail to merge content knowledge with pedagogy, that do not prepare students for urban teaching, that lack instruction on the integration of technology and inquiry, and that fail to stress the importance of reflection and collaboration practices for self-efficacy (Kahle & Kronebusch, 2003). Kahle and Kronebusch (2003) further argue that current professional development is driven by capitalism and is not meaningful, but rather disjointed and unconnected from individual teaching goals. Although a certain amount of professional development is required, it is often not in school budgets and teachers select meaningless topics to fulfill requirements, or even worse, move up the salary scale (Kahle & Kronebusch, 2003). With regard to mathematics teachers, the fracture between state level education professionals and classroom teachers leads to a missed opportunity for increasing teacher effectiveness both in the field of undergraduate teacher training programs and in professional development opportunities.

Although there are now measures of effective teaching, there are no sets of factors or attributes that define an effective teacher. In order to identify effective teachers it was important to learn not only how to measure the qualities of a good teacher, but also how to define these qualities. The MET project provides invaluable data that allows the identification of latent classes of effective/non-effective teachers, which can then be used to test the efficacy of predicting effective/non-effective teachers using standardized test scores. The rationale is that standardized test scores used in value added models are not a sound predictor of effective/non-effective teachers.

Limitations and Delimitations of the Study

Using secondary data has its advantages and disadvantages. Data collection from the MET study provided increased diagnostic information on the multiple measurements used for identifying classes of teacher effectiveness. Purely causal interpretation of teacher characteristics and student outcomes was avoided. This study used item sections from the complete instruments to create the seven observed variables used as measures of teacher characteristics. No prior student achievement data was included in this study as it was not relevant to the model explored. The MET study's adapted FFT and CLASS instruments were not provided for release and there is no known data on the principal rating score used in decision accuracy.

Also, restriction put on data use and extraction from the MET study virtual data enclave played a role in this study. A researcher is restricted to using the MET data online via a virtual data enclave, thus they are not able to download the secondary data. Variable selection and data missingness were also a delimiting concern in the study. Furthermore, the researcher must use the MET software programs and request approval to download any computer output files generated from computer analyses. Consequently, the access to data, variable selection, and statistical analysis were governed by the University of Michigan who controlled the data set and process.

CHAPTER 2: REVIEW OF THE LITERATURE

Teacher Evaluation

Historically, the concept of teacher evaluation is not novel. The Journal for Science and Mathematics was created for secondary teachers in 1905. As early as the 1920s, The Learned and Bagley report "insisted that student learning should be the guiding principle for teaching, and that teachers should be judged on their effectiveness in promoting student learning" (Imig & Imig, 2006, p 168). Teacher evaluation methods were discussed before the establishment of the United States Department of Education in 1979. The method of teacher judgment is what is up for debate. There are two methodological approaches for studying teacher effectiveness: "evaluating the process and evaluating the product" (Medley, 1971). Separation of the two approaches leaves critical information out of the equation. The split view also fractionalizes the world of educational research into taking different qualitative and quantitative standpoints relating to measuring teacher personality characteristics and relating to measuring standardized test scores. The implementation of teacher evaluation systems is not uniformly dependable (Glazerman, et al, 2011). Even by teaching district, it is not possible to recognize the top 25% of teachers with confidence (Glazerman, et al, 2010). This separation of teacher evaluative methods leads to the collection of different data, including survey, observational, and test score data collection. Instead of a piecemeal exploration of teachers, why not look at a more holistic view of teacher quality? In practice, this is not a totally new idea; however, a successful methodology for such examination of data has not yet been employed.

A common theme materializes in the literature: that of differing constructs for establishing the characteristics of effective teachers. In examination of the National Board for Professional Teaching Standards, teachers are assessed using attributes, behaviors, and knowledge, and evaluated on value added models, achievement, quality of learning, and normative reference points (Hattie & Clinton, 2001). Current models tend to classify the majority of teachers as adequate and are not effective at establishing typographies of effective teaching. Hattie and Clinton (2001) argued that because of the inability to establish distinct categories of teaching related to levels of effectiveness, more emphasis should rely on fostering and rewarding excellence in the teaching profession. This may explain why current teacher evaluation methods do not successfully divide teachers into effectiveness categories.

Beller (1971) proposed three aspects necessary for complete teacher evaluation: role, style, and technique. Teacher roles included the duties, responsibilities and function of the teacher, teacher style referred to the teacher's personality and attitudes, and lastly teacher technique established the specific strategies teachers used to accomplish their teaching objectives. Teacher style as defined by Beller (1971) is more difficult to evaluate. Beliefs and teaching style seem to have a clear connection and student's rate teachers as more effective when they utilize divergent teaching styles. In Beller's three aspect model, he indicated that the greatest strides in evaluating effectiveness are in the areas of technique because there is a measurable outcome or product of teaching. The quantifiable nature of this data is what proves as an attractive methodology today. Goe, Bell and Little (2008) claimed that teacher evaluation can be approached by three different angles: measurements of inputs, measurements of processes, or measurements of outputs. This is a very similar organizational structure as the model Beller described almost 40 years earlier that identified inputs, process, and outputs of teaching. Inputs mirror teacher roles, and encompass functions a teacher must put in to perform his or her job; process refers more to the teacher's personal style, and outputs are analogous to accomplishing teaching objectives. The main difference between the two is the word choice.

Goe, et al. (2008) used a teacher evaluation structure centered on "measurements," that highlighted a quantitative shift in educational research.

The three levels deemed important to teacher evaluation have been studied in the order that both Beller (1971) and Goe, et al. (2008) presented them. The teacher input or teacher role keeps demonstrating that teacher credentials, years of experience, and past years of novice teaching, do not play an important role in teacher quality (Jepsen, 2005; Rice, 2010). Now researchers look to the process or the product. Teacher processes or teacher style are explored with surveys and teacher observations. Teacher output or teacher technique saw the height of popularity with exploring high stakes testing and value added models. Now researchers are scrambling to support their preferred measurements and results.

Arne Duncan stated that current measures of teacher evaluation do little to establish effective teachers; all teachers regardless of skill are rated basically the same (Gottlieb, 2013). Duncan's suggested solution for fairly assessing teacher effectiveness requires the inclusion of student test scores to rate and classify teachers (Gottlieb, 2013). The essential problem with this approach was that it reduced a teacher to a number, and did not take into account numerous factors that describe successful teaching and that delineate effective teachers. Gottlieb (2013) stated that "Duncan's policy proposals seems to describe a situation in which a teacher's "effectiveness" is simply equated with a teacher's "quality," such that a teacher's impact on students' achievement scores will function as the "preponderant criterion" for making termination and remuneration determinations" (pp. 16-17). This proposed a problem in the many faceted definition of what it means to be a teacher. A new model for evaluating teacher effectiveness is necessary if teacher preparation programs and professional development activities are to improve.

Evaluation decisions based solely upon test scores do not fairly judge teacher quality and effectiveness. A new method of evaluating teachers is necessary, especially since test score effects are poor measurements of teacher quality at the high school level for Algebra I teachers (Jackson, 2014). Jackson (2014) examined data on over 3,500 Algebra I teachers in North Carolina and revealed that high school teachers positively impact test scores, and experienced teachers tend to have higher test scores, but that overall test scores are inadequate for determining Algebra I teacher effectiveness.

The best approach to teacher evaluation is the inclusion of multiple measures. Gottlieb (2013) claimed that the use of multiple measures proposed is merely an illusion for a holistic evaluation process because items are said to be included according to their agreement with test scores. A lack of understanding of quantitative methods seems to be responsible for this thought process. Gottlieb (2013) used the arguments of Elliot Eisner to caution that current ideas of "assessment" are merging with "evaluation," creating dangerous perceptions for teacher effectiveness. The seemingly quantitative view presupposes " that those conducting the assessment already know everything that will be relevant to a judgment of quality, and that the means of assessment are in every case adequate to revealing what Eisner elsewhere calls the "universe of particulars" which refers to the numerical test score itself (Gottlieb, 2013, p. 20). There are indeed some researchers and now some states that are moving forward with a one-dimensional approach leading many stakeholders to question the impartiality of using teacher assessment strategies based on test scores for evaluation.

Eisner (1976) revealed the fear of qualitative researchers when he stated "quality becomes converted to quantity and then summed and averaged as a way of standing for the particular quality from which the quantities were initially derived" (p. 137). Eisner's advocacy for an age of the evaluation of teacher effectiveness rooted in connoisseurship and criticism is the qualitative mirror image to the quantitative style of evaluation argued by (R2T). However, the argument developed by Gottlieb (2013) falls back onto dichotomizing the world of educational research into the camps of qualitative and quantitative methods even though he argues both sides ultimately share the same resolution. In this study, multiple measures are used, and a mixed methods research approach is used to resolve this divided view of research methodology.

Initial quantification of the relationship between teacher effects and student outcomes were examined by economists in the 1970s (Hanushek, 1971). Teachers as an economic unit are not found to be easily comparable on merits of effectiveness used in the past and so, in addition to teacher effectiveness, teacher effectiveness information is becoming important in describing educational outcomes to be used in policy and practice (Jackson, Rockoff, & Staiger, 2014). Information gained from such studies should be used for improving teacher education programs and improving professional development opportunities. Even though teaching is said to have undergone a progressive movement, Imig and Imig (2006) argued that essentialists are responsible for the control of educational constructs like teacher effectiveness today due to standards established by educational law. This lack of uniformity adds another concern when examining teacher effectiveness in evaluation models across the United States.

Why have teacher evaluation systems moved away from using a holistic approach? Just as in industry, the process and the product are connected. Harris, Ingle, and Rutledge (2014) used a mixed methods approach that examined principal interviews and teacher ratings along with ratings on value added measures to address discrepancies in the two evaluation methods. There is a modest overlap of value added measures and the principal ratings, but it is more clearly distinguished for the highest performing teachers. Their most interesting finding claimed "various measures differ not just in their validity, but in the construct they measure" (p 105). Some observational instruments positively correlated with teachers' value added measures (VAMs), student's self-reported engagement, and social and emotional development (Kane & Staiger, 2012). The components comprising effective teachers are not defined by a single measurement or even a single construct as there is inherent multidimensionality in social science research. Multiple measures and multiple ratings are a few key standards in a system important to meaningful teacher evaluation (The New Teacher Project, 2010). When multidimensionality is applied to characteristics of effective teachers, the psychometric usage supports the identification of components of effective teachers (Grossman, Cohen, Ronfeldt, & Brown, 2014; Lasarev & Newman, 2014).

Another issue in teacher evaluation is the tie of student test scores to teacher merit. Educational reforms are currently pushing merit pay by increasing the importance of student test scores (Brewer, Myers, & Zhang, 2015). There is little to show that student test scores demonstrate student gains and therefore would not be connected to the identification of effective teachers. Levine and Levine (2013) examined longitudinal data for high-stakes testing used in English and mathematics data from New York State. They concluded that high-stakes testing is not useful as an assessment of improved learning, and that the product of what achievement scores actually measure is unknown, not to mention the high cost of providing teacher training and student score analysis shows little change in reducing the minority gap. Analysis of a pay for performance program in the Austin Independent School District in Texas showed Mathematics test scores only improved the first year of the program and that student learning objectives were not significantly correlated with the test scores (Balch & Springer, 2015). These teacher merit pay programs tied to student test scores disassociate quality teacher characteristics defining teacher effectiveness.

Teacher Effectiveness

The current idea of what makes an effective teacher is a measurement related to an outcome. A narrow definition of teacher effectiveness “refers to a teacher’s ability to improve student learning as measured by student gains on standardized achievement tests” (Little, Goe, & Bell, 2009). The idea of what makes a good teacher is generally defined by teacher characteristics. Both are important factors for analysis of effective teachers. The theory and evidence behind effective teachers must harmonize multiple practices (Darling-Hammond & Bransford, 2005). Because of multiple practices, new measurements of teacher effectiveness must be established to reveal identifiable characteristics of effective and non-effective teachers. Current Federal R2T initiatives do not explore both the process and the product, but rather focus on a one-sided definition of effective teachers that only examines the factor of increased rates in student growth. R2T’s emphasis on value added measures (VAMs) leaves out characteristics of effective teachers. This is analogous to treating a patient’s symptoms without diagnosing the patient. It is exploring the effect without examining the causes.

The importance of effective teaching is directly related to a student’s academic success. Student growth or lack of growth is directly related to the strength of the teacher (Hanushek, Piopiunik, & Wiederhold, 2014). A teacher’s ability to engage students and to promote understanding should in turn improve a student’s understanding of the content. The importance of this issue is not only relevant federally or locally, but is critical for the students involved. Due to student inclination toward aversion of mathematics content, secondary mathematics teachers rely heavily on the importance of student engagement to promote learning and understanding of mathematics. Engagement can take several forms, for example, a promoted form of whole class discourse that promotes student thinking (Bennett, 2013). The use of teacher promoted student engagement through whole class discourse in the mathematics classroom gives a broader formative assessment of student learning and effective teaching than test scores which only provide a finite example of student understanding (Bennett, 2013). Good teachers have strong content knowledge, are great at promoting student understanding, effectively promote student engagement, and can control and manage their classroom environments.

The most important players in the equation, the students, agree with the themes above. Though often disregarded as an unimportant or biased measure of teacher effectiveness, student opinion plays an important role in establishing characteristics of effective mathematics teachers. A study of 73 secondary mathematics teachers used multiple classroom observations focusing on emotional, organizational, and instructional support to determine the quality of teacher-student interactions (Martin, Rimm-Kaufman, 2015). Students were surveyed on their feelings of engagement in their mathematics classes three times during the school year after the lessons in which teachers were observed (Martin & Rimm-Kaufman, 2015). Results revealed that effective teachers gave significantly more emotional support in their classroom environments. Student and teacher relationships are critical to establishing trust, and respect in the classroom which directly relates to more successful classroom management.

The exploration of effective STEM teachers is an area that had a close examination during the Cold War, and then another surge of interest shortly after the institution of NCLB. The current importance of STEM demands a reexamination of Algebra I teacher effectiveness as it is important to the future of establishing student interest in mathematic and scientific endeavors. In an examination of international STEM, teachers teaching in the United States were asked to rate effective teacher qualities. These teachers reported that effective teachers clearly explain the content, enjoy teaching, and are fair in their grading practices (O'Donnell, 2009).

Another issue of concern in the area of STEM is the lack of minority presence and the existing achievement gap between majority and minority races in the United States. One aspect that is often overlooked is the opinion of students on culturally relevant pedagogy (CRP). CRP is important in determining the construct of teacher promoted student engagement as it directly relates with a teacher's ability to make mathematics content relevant to students, therefore forming an investment in content and an establishment of strong student and teacher relationships. A study of African American students found that by increasing CRP, it in turn increased student motivation and engagement, and in conjunction, student understanding (Hubert, 2014).

Effective Algebra I teachers are able to make adaptions in the classroom and this is important to convey in teacher education programs. Pedagogical content knowledge (PCK) has a greater influence on student achievement than mathematical content knowledge (MCK); however, one study found that even though MCK and PCK are strongly correlated in predicting students' mathematical achievement gains, MCK does not significantly influence PCK and is not adequate on its own for creating effective secondary mathematics teachers (Goos, 2013). Teacher pedagogy, including the importance of engaging students and promoting student understanding, is important for creating skillful and effective teachers.

A qualitative study examining a collaboration of mathematics education specialists and researchers supported that there are three key indicators of effective teaching and therefore effective teachers (Nebesniak, 2012). These indicators include: teaching conceptually for understanding, making connections with content, and directing attention through student engagement. The themes reoccur today. Effective teachers are flexible, knowledgeable, and successful at creating and maintaining a positive classroom climate.

After NCLB, studies into what defined “highly qualified” became the new political prompt for definitions of good teachers. Characteristics of effective mathematics teachers were seen by pre-service and in-service teachers as multifaceted, containing aspects related to MCK, PCK, and concern for students (Strickland & Page, 1991). Specific characteristics identified were effective teachers’ abilities to explain mathematics clearly, work with students, and express interest in students (Strickland & Page, 1991). Pedagogical problems seemed to be the most related to teacher effectiveness, but for many teachers, content knowledge was also important because teachers are expected and required to teach subjects in which they have little to no backgrounds due to teacher shortages.

One study examining secondary science and math teachers found that teacher dispositions including willingness to collaborate, interactive skills, and the ability to reflect and adjust practice were important for determining effective teachers (Miller & Davison, 2006). Beishuizen, Hof, van Putten, Bouwmeester, and Asscher (2001) revealed that secondary students could identify two dimensions of good teachers: personality and ability. Both teachers and secondary students emphasized the importance of teacher-student relationships for successful teaching. Student views of engagement in mathematics showed dissatisfaction with mathematics teaching and a lack of enjoyment of learning the content due to ineffective teaching methods (Murray, 2011). A survey conducted by Learning Point Associates and Public Agenda revealed 72 % of teachers’ rate student engagement in coursework as a good or excellent indicator of teacher effectiveness (Rebora, 2010).

Overall, years of teaching experience has proven to be a poor indicator of teacher quality; however, if years of experience are related to novice versus veteran teachers it becomes a relevant indicator of teacher effectiveness. A 40 year longitudinal study of elementary and middle school teachers revealed that experience does have an impact for novice teachers who have taught five years or less compared to veteran teachers (Rice, 2010). In a study by Ng, Nicholas, and Williams (2010) novice teachers relayed belief systems of effective teaching pre and post experience in the classroom. Initially teachers felt effectiveness would be related to the strength of their expertise and training, while in the end, they established effectiveness as more related to charisma and their ability to build relationships with their students. Another study of novice teachers revealed their concern with a lack of preparation from teacher education programs to prepare them for common problems in the classroom (Watson, 2006). Novice teachers expressed concern in a variety of areas including lack of preparation to deal with endless paperwork, changes in schedules, non-existent mentors, and an inability to deal with classroom management and discipline issues (Watson, 2006). In urban schools, it is difficult to attract and retain effective teachers. Two characteristics that stand out include the differences between novice (1-3 years) and experienced (greater than 3 years) teachers and also differences in teacher cognitive ability (Jacob, 2007) that determine teacher effectiveness. Teacher cognitive skills are an important determinant in student achievement differences between the United States and other countries.

A comparison of 23 developed economies revealed differences in teacher numeracy and literacy scores from the Programme for the International Assessment of Adult Competencies (PIAAC) (Hanushek, Piopiunik, & Wiederhold, 2014). The PIAAC along with the student data from the Programme for International Student Assessment (PISA) determined that differences in student performance rested on the cognitive ability of the teachers: smarter teachers, higher performing students. Background into the discrepancy of teacher cognitive ability by country revealed the pool from which teachers are selected and recruited. Countries with teachers scoring higher on the PIAAC, recruit and select from the top tier of college students and are able to do so because of the high pay and prestige attached to the profession in said country. The authors suggest teacher cognitive ability as the missing factor in identifying effective teachers (Hanushek, Piopiunik, &Wiederhold, 2014). Goldhaber (2007) revealed a positive correlation between teacher licensure scores and student achievement. Although content knowledge is important and used as a determinant for Algebra I teachers to be highly qualified, some other aspects are being overlooked including the teachers ability to communicate effectively with students (Miller & Davison, 2006). Much conversation about what it means to be an effective Algebra I teacher sometimes overlooks what some of the difficulties are for students to learn Algebra I content effectively.

Teacher Evaluation Methods

Teachers report that current evaluation systems do not lead to actualizing areas needed for improvement and therefore they cannot illicit meaningful professional development (Weisberg, Sexton, Mulhern & Keeling, 2009). Until there is a fair method for teacher evaluation, teachers will not be on board and teacher improvement through professional development and college instruction will be stymied. There are currently three utilized methods of evaluation in practice including observations, surveys, and value added measures.

Current teacher observational rankings reflect tiers of teacher performance resulting in a majority of teacher rankings in the satisfactory category (Gordon, Kane, & Staiger, 2006; Weisberg, Sexton, Mulhern & Keeling, 2009). This most commonly used methodology for teacher evaluation is not successful at differentiating teacher quality and also is not helpful for educators to assess areas of self-improvement. The MET project found that when 127 administrators and teachers conducted 24 teacher observations using Charlotte Daniel's FFT, a majority of teachers scored in the middle two categories of "basic" and "proficient" (Ho & Kane, 2012). High principal evaluation scores generally represent teachers with both strong classroom management and strong teacher student relationships (Stronge, Ward, & Grant, 2011). Principals have also been found to distinguish among teachers who can produce the smallest and the greatest standardized achievement gains though they cannot distinguish middle performing teachers (Jacob & Lefgren, 2008). Ho and Kane (2012) found a Spearman rank-order correlation of 0.87 between administrator observations and outsider observations for the MET data. The reliability between principal and outsider observations implies limited personal bias in the principal's evaluations. However, solely using these principal observations for evaluations is not enough to get a total picture of effective teacher characteristics.

Principal observations, while holding some criticism, seem to be a reliable measure of establishing teacher effectiveness. One study reviewing principal observations found that ineffective teachers seem to have the most problems in areas of pedagogical content knowledge including: lesson-implementation skills, ability to establish rapport with students, and classroom-management skills (Torff & Sessions, 2009). These skills relate to teacher and student interaction skill sets. This finding held for schools from a variety of economic backgrounds suggesting that teacher effectiveness was not related to the socioeconomic status of students.

Although a useful tool, principal observations cannot be the total factor in measuring teacher quality due to the nature of their compressing rating scales. In order to improve this form of measurement, the MET study included the use of increased observations by outside trained persons in order to “provide an ongoing check against ‘in-school bias’ (Bill and Melinda Gates Foundation, 2013, p 5). A concern with outsider teacher observation for the evaluation of teacher effectiveness is validity and accuracy. In regards to the validity of observations, it has been established that expert panels are able to rate an instrument’s content validity. In one study, Van Tassel-Baska, Quek, and Feng (2006) demonstrated expert agreement could be validated for individual items and for the clarity of language.

Different sites employ different methods for consistent and fair assessment, including evaluator training, using third party evaluators, using evaluator tests, rating evaluators, as well as using data checks, reviewing student performance and including superintendent reviewed, principal discretion supplemental values (Doyle & Han, 2012). Effective training of outside observers for the MET project found that less than 10% of the variance in teacher observation scores was a result of inconsistency in observer scoring (Bill and Melinda Gates Foundation, 2012, p 2). Also, observer reliability increased from 0.66 to 0.69 when multiple shorter observations were employed (Bill and Melinda Gates Foundation, 2013, p 5). In general, other studies assessing inter-rater reliability for teacher observations found a reliability between 0.55-0.89 (Gersten, Baker, Haager, & Graves, 2005; Van Tassel-Baska, Quek, & Feng, 2006). The better trained the observer the more consistent and accurate the observation. However teacher observation methods are not without criticism. Critics argue that inconsistency and poor inter-rater reliabilities make observational instruments unlikely to be able to provide the only measurement for effective teaching.

The least utilized in secondary education, surveys are often employed to garner a student perspective on teacher performance. There is not much historical use of surveys for this age range and therefore there is not much data to discuss (Peterson, Wahlquist, & Bone, 2000).

There is, however, evidence of the stability of secondary student surveys over time with correlations of multi-year evaluations, 0.7-.0.85, for the TRIPOD used in the MET study after correcting for measurement error (Bill and Melinda Gates Foundations, 2010). Further research by Hajdin and PaŽur (2012) examined high school students' abilities to differentiate between evaluating teachers versus evaluating teaching. There are not clear differences between the two categories, but students did identify teaching methodology, teacher personal characteristics, elements of teaching, and teacher student relationships as being the best indicators of effectiveness. One study related high school student evaluations of teacher skills and characteristics using the Student Perception Questionnaire. Student scores on the questionnaire were compared to evaluations of the same teachers by university and school supervisors and the results suggest agreement between the two data sets supporting students' ability to fairly assess teacher skills and characteristics (Tairab & Wilkinson, 1991). Overall the use of study surveys may be more stable over time than the use of observations (Polikoff, 2015).

Value added measures are growth models that claim to assess a teacher's impact on students while controlling for factors outside a teacher's control, namely, student and school characteristics. Beginning in the 1990s, statistical methodologies like value added measures revealed an incredible variation in teacher performance with regards to student output and teacher product (Rowan, Correnti, & Miller, 2002; Sanders & Rivers, 1996; Webster, Mendro, Orsak, & Weerasinghe, 1998; Wright, Horn, & Sanders, 1997). The use of only objective statistical evaluations to measure teacher quality is not a consistent method. For example, the "lagged score" VAM model below identifies the general measures used:

$$Y_{isjt} = \beta_0 + Y_{isjt-1}\beta_1 + X_{isjt}\beta_2 + S_{isjt}\beta_3 + T_{isjt}\theta + \varepsilon_{isjt}$$

where Y_{isjt} is a test score for student *i* at school *s* with teacher *j* in year *t*, X_{isjt} serves as a vector of student characteristics, S_{isjt} serves as a vector of school and/or classroom characteristics, T_{isjt} serves as a vector of teacher indicator variables and ε_{isjt} is the error term (Koedel, Mihaly, & Rockoff, 2015).

The Tennessee Value Added Assessment System (TVAAS) is a value added approach developed by William Sanders at the University of Tennessee to implement a mixed-effects model for longitudinal standardized test score data across multiple subject areas (Kupermintz, Shepard, & Linn, 2001). The TVAAS was one of the first statistical methodologies employed claiming to fairly assess indicators of teacher quality while accounting for influential factors of student achievement like ethnicity, SES, prior achievement, and general ability (Kupermintz, Shepard, & Linn, 2001). They suggested that further examination of value added measures is necessary due to

> the potential of confounding of teacher effects and other independent factors contributing to student academic progress, the dependency of estimates of teacher effects on model assumptions and on the context of their school systems, the explicit links between student score gains and instructional practices, and the generalizability of multiple choice test results as indicators of instructional impact on student progress toward desirable educational goals (p. 19).

The use of value added measures as a finite methodology for evaluating teacher quality cannot be used without a more comprehensive view of effective teacher characteristics and its relevance to current educational policy cannot be ignored. There are typically three clusters of growth models used: simple change, residualized change, and value tables (Schafer, et al., 2012). These value added models cannot solely be used for evaluating high stakes decision making with regards to teacher evaluation. Even when accounting for control variables like gender, SES, ethnicity, and English proficiency, there are too many confounding variables with regards to classroom and learning dynamics (Schafer, et al., 2012). Some states like Colorado, Louisiana, and Tennessee have current evaluation programs that specify 50% of a teacher's performance rating is determined by student growth data (Green, Baker, & Oluwole, 2012). Laws in Maine plan to implement the use of teacher evaluation based solely on standardized test scores in 2015-2016, however Representative MacDonald from Maine wishes to propose an act to limit the effects of standardized tests on teacher evaluations (Cousins, 2013). Nationwide, the debate continues to ensue over the use of standardized test scores for evaluating teacher performance.

Koedel, Mihaly, and Rockoff (2015) provided a comprehensive review of VAMs and concluded that students will benefit from VAMs being included in personnel decisions for teachers; however there are few studies reported for high schools and it would be of greatest value for VAMs to validate "the relationships between value added and alternative evaluation metrics, like those used in combined measures of teaching effectiveness (e.g., classroom observations, student surveys, etc.)" (p. 192). Di Carlo (2012) stated that VAMs just have a bad reputation and are neither good nor bad. Rather than banishing their use in teacher effectiveness models, VAMs should be used cautiously through avoiding mandating universally high weights for value added measures, paying attention to all components of evaluation, addressing error, and continually monitoring results and evaluating evaluations (Di Carlo, 2012). There is not a single method previously discussed, that alone, demonstrates a way to identify effective and non-effective teachers. Newer approaches suggest composite measures that incorporate additional information like observation and student survey data into value added measures. Unfortunately, it still remains whether this is a better approach.

Combining Measures of Teacher Effectiveness

It is important for the evaluation of teacher roles to encompass all stakeholders. Beller (1979) argued "future evaluation of teacher roles should never rely on one source alone but should include the perception and judgment of all groups involved in the teaching of the child" (p. 128). Although there is not one instrument in use that identifies effective teachers, Gordon, Kane, and Staiger (2006) propose a combination of methods including teacher observations, principal evaluations, parent evaluations, teacher absences, and the use of value added measures. This composite evaluation method is also referred to as the "portfolio view" of teachers or as a comprehensive teacher evaluation system (Gordon, Kane, & Staiger, 2006; Moss, Schutz, & Collins, 1998). Comprehensive teacher evaluation systems (CTESs) are systems that use multiple measures of teacher effectiveness including student achievement, student feedback, teacher peer reviews, teacher self-evaluation and personnel review. MacCalla (2014) used secondary data to evaluate the use of teacher surveys, expert assessments, and classroom observations to determine their value in teacher evaluation. Data was taken from the California state-funded Improving Teacher Quality Program which was a voluntary professional development program implemented in the Los Angeles school district. In this study, expert assessments were found to be the most reliable measurement, followed by classroom observations, and finally teacher surveys (MacCalla, 2014).

In a study of New York City public school teachers, subjective evaluations of teacher effectiveness are shown to complement objective measures of teacher effectiveness when used to predict future student achievement gains (Rockoff, & Speroni, 2011). Using multiple measures including observations, portfolios of teacher work, student learning objectives, and student surveys as well as VAMs are important to creating increased reliability and validity if valued aspects of teacher are targeted (Harris, 2012). Blending VAMs and teacher observation based evaluations is a better assessment of teacher effectiveness for teachers in low-income, minority schools (Mangiante, 2011). The MET project is the most recent and accessible data set to review multiple measures and their relationship to determining effective and non-effective teachers.

More recently researchers have been advocating the importance of multiple measures for the evaluation of teacher effectiveness. Ten sites including: Delaware, Tennessee, Rhode Island, Hillsborough County, FL, Houston, TX, New Haven, CT, Pittsburgh, PA, Washington, D.C., Achievement First (CMO), and Relay Graduate School of Education were used in a study examining methods for measuring teacher effectiveness (Doyle & Han, 2012). It was found that several sites are using multiple measures of student achievement even when student testing data exists in a method to strengthen evaluation due to the potential weakness of any one individual measurement. Student performance data in these sites was typically calculated using value added models, student growth percentiles, pre/post-test growth, or mastery of standards. The weighting of student performance varied from using flat percentages, matrices, and rating ceilings. Additionally, classroom observations were taken at all sites. The use of multiple measures indicated a shift in teacher evaluation that took into account a more holistic approach.

The most common method of observation was use of Charlotte Danielson's FFT including observational scoring for four defined domains of teaching: planning and preparation, the classroom environment, instruction, and professional responsibilities. Chosen observers at different institutions included: administrators, coaches, third party evaluators, and teacher peers. Observational frequency also differed by individual site often based on the type of teacher, low-performing, high-performing and novice. Additional non-academic measures taken into account at some sites included student perceptions, student character, peer ratings, and teacher commitments and contributions to the school.

Teacher quality indicators like the FFT are now being used in conjunction with value added measures. A study using 222 middle school math teachers compared value added scores to other teacher quality indicating instruments including surveys and observations and found a significant correlation (Hill, Kapitula, & Umland, 2011). A recent study of the Pittsburg Public Schools teacher evaluation system examined a multi-composite measure including the Research-based Inclusive System of Evaluation based on Danielson's FFT, TRIPOD based upon Fergusons' 7Cs student survey, and a value added measure developed by Johnson, Lipscomb, Gill, Booker, and Bruch in 2012 (Chaplin, Gill, Thompkins, & Miller, 2014). The components of the three measures, FFT, TRIPOD, and the VAM were moderately and significantly correlated, suggesting acceptable levels of internal consistency (Chaplin, Gill, Thompkins, & Miller, 2014). Correlations across the measures "are capturing various aspects of effective teaching" in complementary ways (Chaplin, Gill, Thompkins, & Miller, 2014, p. 3).

Since the MET study, a new teacher observation tool called the Rapid Assessment of Teacher Effectiveness (RATE) has been developed. RATE is a fast and cheap 6 item rubric using raters that is comparable or better than longer popular observation tools used in the original MET study (Gargani & Strong, 2014). The importance of teacher observational data in gauging teacher effectiveness is now clearly established and RATE provides a new tool for future data quantification; although, the largest current source of data for analysis on teacher effectiveness is in the MET study.

The Measures of Effective Teaching Project (MET) was funded by the Bill and Melinda Gates Foundation in 2009 and was proposed to identify qualities of teacher effectiveness to help improve teacher quality. Six school districts with over 3000 teachers participated in the MET project. It is the most comprehensive data collection effort exploring measures of effective teaching and used seven observational instruments, student and teacher perception surveys, and student achievement gains on the SAT-9, the ACT, and the Balanced Assessment in Mathematics (BAM).

The MET project purports the more recent idea of composite measurements, which is a variation of value added measures that depended on only growth models. Composite measures include multiple, different types of instruments, with weighted total scores to determine their ability to accurately assess effective teachers in relation to the student growth measures. A composite measure of effective teaching using the MET project data including student survey responses, classroom observations, and student achievement growth was implemented to explore a combined methodology for using multiple measures to assess teacher effectiveness (Mihaly, McCaffrey, Douglas, & Lockwood, 2013). In the composite measure employed, components of all indicators were positively correlated, suggesting the ability to find a common factor related to effective teaching (Mihaly, et al., 2013). However, the use of equally weighted indicators will not lead to identifying characteristics of effective teachers, but may give a more accurate prediction, and will be more stable over time (Mihaly, et al, 2013). The measures of effective teaching were able to predict which teachers had the greatest amount of student achievement growth (Kane, McCaffrey, Miller, & Staiger, 2013).

Critics of the MET project claim that it found supporting measurements that align with and strengthen the use of VAMs, and therefore did not provide a new way to explore teacher effectiveness (Gabriel & Allington, 2012). However, this study is not proposing a VAM growth model, and VAMs should not be solely used in high stakes assessment of effective teachers. Just as the topic of what makes an effective teacher is complex and multifaceted, so should be the measurements used to assess teacher effectiveness. The idea of using a different methodology in this study, namely latent class analysis, will better support the realization of characteristics that categorize teachers as effective and non-effective.

Toward Improving the Future

Determining indicators and predictors of effective teachers is not only important to helping current practicing teachers improve, but it is also important for the training and professional development of the new teaching force. Right now, teacher effectiveness is very relevant to the federal government and the push for quantitative VAMs is gathering interest in the field of educational research. Although VAMs may not be the best approach for determining teacher effectiveness, the issue and its relation to the current problems of teacher attrition is very important.

A study of North Carolina novice teachers, or teachers having five or less years of experience, revealed some interesting trends about teacher effectiveness. Henry, Bastian, and Fortner (2011) found that novice teachers increased in effectiveness until year 3, in which they reached a plateau. Teachers who left after 5 years were on average less effective than teachers who continued teaching. The on the job training required to reach greater effectiveness takes about a year. This first year struggle could be ameliorated or maybe even surpassed with proper professional development including mentor programs or specialized professional development as these skills are not often or easily taught in teacher education programs. Another idea that increased teacher effectiveness examined the method in which teachers were instructed pedagogically. It is argued that instructing STEM teachers using metacognition enables them to engaging in the learning process and influences them to require the same metacognitive approach from future students, in turn, increasing their teaching effectiveness (Eldar & Miedijensky, 2015).

Examining teacher preparation programs revealed that the multiple rating instruments measured only one rather than multiple constructs of future teacher performance (Henry, et al., 2013). The designed rating instruments and standardized teacher tests could not predict teacher effectiveness in the classroom (Henry, et al., 2013). It was problematic and revealed that teacher preparation programs are not able to predict teacher effectiveness and are not providing accurate education in developing strategies for novice teachers to have skills sets necessary for successful teaching.

Not only is it important for teacher preparation programs to develop the best teachers possible, but it is essential that teachers are adequately prepared to reduce nationwide problems of attrition. Teacher preparation programs are coming under fire and are being included in the accountability orders of NCLB; however, this accountability is also linked to gains in student test scores.

Summary

The separation of teacher evaluation methods leads to the collection of different data, including survey, observational, and test score data collection. Instead of piecemeal exploration of teachers, why not look at a more holistic view of teacher quality? This is not a totally new idea; however, a successful methodology for such examination of data has not yet been employed. The components comprising effective teachers are not defined by a single measurement or even a single construct, as there is inherent complexity in social science research.

Although assessing and interpreting achievement is an important aspect when assessing teacher effectiveness, it is just as important to know combined characteristics of effective teachers. The multidimensionality of teacher effectiveness makes it a difficult topic to research, but the extent of research thus far does reveal some clear themes. Good teachers have strong content knowledge, are great at promoting student understanding, effectively promote student engagement, and can control and manage their classroom environments. Until there is a fair method for teacher evaluation, teachers will not be on board and teacher improvement through professional development and college instruction will be stymied.

The MET project is the most recent and accessible data set to review multiple measures and their relationship to determining latent classes of effective/non-effective teachers. Using multiple measures, including observations, surveys, and teacher background information to identify groups of teachers through latent class analysis will provide a new approach in educational research.

CHAPTER 3: METHODOLOGY

This study used a quantitative design with secondary data from the MET Project in order to find characteristics of teachers that would categorize classes of teacher effectiveness. The study used a combination of measures including observational measures (Framework for Teaching, and Classroom Assessment Scoring System), survey measures (TRIPOD Survey, and MET Principal Rating Score), and student ACT Quality Core® Algebra I scores. An overview of the sample including the participants, a discussion of the setting, and procedures of data collection are presented. Afterward, a detailed description of all instruments used in the study will follow, and complete original surveys that were accessible are listed in Appendices D and E. Finally, the research questions, data analysis procedures, and the purpose for selection of various methodologies will be explained.

Description of the Site and Participants

The sample represents a diverse geographic and socioeconomic spectrum of school districts, teachers, and students. The MET project volunteers are from 317 schools and 6 school districts including the Charlotte-Mecklenburg Schools, the Dallas Independent Schools, the Denver Public Schools, the Hillsborough County Public Schools, the Memphis Public Schools, the New York City Schools, and the Pittsburg Public Schools. The MET project collected multiple years of data. The quantitative analysis described used selected second year data for the latent class analysis (LCA) due to the establishment of the study's randomization and desired teacher commitment in a voluntary data collection approach.

In order to narrow the sample from the approximately 3,000 teachers involved in the study, a subsample of Algebra I teachers was selected. This study will only examine ninth grade Algebra I teachers, which is reflected in the sample size of 152 teachers. SAS 9.4 software

(2012) was used to determine the power given an approximate non-centrality parameter. The critical chi-square value with one degree of freedom is 3.84 for $p = .05$, and the approximate non-centrality parameter is 9.286. Power results indicated that the sample of 152 teachers is sufficient for the data analysis with power = 0.98. Chi-square values for other sample sizes can be calculated based on chi-square being equal to 2*sample size*F. Multiplying chi-square by the ratio of the new sample size to the sample size for the chi-square gives a chi-square for the new sample size that can used to obtain the power for that sample size. For example, the chi-square for sample size 100 is equal to 100/50*9.286 or 18.572. *Mplus* 7.4© (2015) power analysis indicated sufficient power given the sample size for this analysis. This study will not analyze all data from the MET Project database, but rather the files listed in Appendix C.

When examining the demographics of the Algebra I teachers in the MET study project, approximately 63% of the teachers were female and 37% were male. Around 65% of the teachers were considered experienced teachers with greater than or equal to 4 years of teaching experience. The least amount of experience was 0 years and the greatest amount of experience was 37 years with a standard deviation of 6.97. Ethnically, teachers were 61% white, 29% black, 6% Hispanic, and 4% other. For the FFT and CLASS instruments, teachers provided on average 4 to 6 lessons. Tripod student perception surveys were also used for the specific classrooms; however, since it was a classroom specific instrument, average class sizes were not reported.

Data Acquisition

Data acquisition involved a two-step process including an application proposal and an Agreement for Use of Confidential Data Form (see Appendix B) signed jointly by The University of Michigan and The University of Alabama. Once cleared, data access licensure for this dissertation required the purchase of two licenses at $350.00 apiece through The University of Michigan; licensure was paid for privately. Access was coordinated via The Inter-University Consortium for Political and Social Research (ICPSR), a part of the Institute for Social Research at The University of Michigan, and is an international group that includes more than 700 institutions and research organizations including The University of Alabama (ICPSR, 2016). The ICPSR is responsible for activities including data access training, leadership training, and methods training for the social science research community (ICPSR, 2016).

Access to MET data files required an expedited IRB review (see Appendix A). Specific data files applying to the study were requested and only access to certain files were permitted (see Appendix C). The data requested identified participants in the study by coded IDs and data users were not aware of any identity details. The six school districts involved were also given coded IDs. No individual student level data was requested because increased restrictions by the MET study on data release were of concern. Due to the sensitive nature of the data included, the ICPSR required data access through a virtual data enclave management system (VDE). A VDE is a virtual machine accessed from a personal computer that uses a remote server to allow limited and controlled data manipulation (ICPSR, 2016). No data outputs, graphs or tables were allowed to be extracted without a formal permission request and no raw data file extraction was allowed. No internet access or access to personal computer functions was allowed during data cleanup, exploration, or data analysis.

After access to the VDE was granted, a user account was set up with The University of Michigan. Then a token account linked with the MET VDE was established on a personal Apple iPhone 6® using the downloaded application RSA SecurID®. Tokens are used to sign into the VDE and are eight digit numbers that are generated every 60 seconds. Finally, the MET VDE was downloaded on a private password protected laptop computer. Data access required both a username, token input, and a secure password. Once logged into the VDE there was no internet access and work was restricted to the data files and programs set up on the database. Any data manipulation was saved to a user linked proxy H drive and all analyses were approved by the ICPSR before results were extracted and published. Details of the online security and data extraction are included in Appendix B.

Instrumentation

Observational instruments.

Observational measures for the MET project were selected or created to focus an observer's attention on the teaching practice, and to establish evidence of good teaching practices (Kane & Staiger, 2012). Some instruments were preexisting with validated measures, while others were adapted for the project.

The premise for Charlotte Danielson's Framework for Teaching (FFT) Evaluation Instrument started in 1996 and soon turned into a usable instrument in 2007 that was used by the MET project in 2009. It has since been updated in 2011 and in 2013. The FFT is based in constructivist theory and correlates with the Interstate Teacher Assessment and Support Consortium (InTASC) standards that champion progressive teaching. It is made up of four domains: planning and preparation, classroom environment, instruction, and professional responsibilities, which in turn represent 22 components, for a 76 item instrument (The Danielson Group, 2015). The domains of planning and preparation and professional responsibilities are assessed not by classroom observations, but rather by gathering artifacts and evidence and were not explored in this study. However, classroom environment and instruction were observable domains. Both observable domains contain five components coded on a four point scale ranging from unsatisfactory, to basic, to proficient, and to distinguished (see Table 1). A study by

Milanowski, Heneman, and Kimball (2011), reviewing multiple implementations of the Framework for Teaching revealed a correlation between the FFT score with value added estimates of teacher effectiveness to be in the 0.2-0.3 range (p. 11). When multiple observers and multiple observations of teachers are made using the FFT, the ratings were reliable (Milanowski, Heneman, and Kimball 2011).

Table 1.

Summary of FFT Reliability Findings

Instrument	Domain	Component	Inter-rater Reliability
FFT			**0.67**
	Classroom Environment	a. Creating and environment of respect and rapport	0.51-0.60
		b. Establishing a culture for learning	
		c. Managing classroom procedures	
		d. Managing student behavior	
		e. Organizing physical space	
	Instruction	a. Communicating with students	0.47-0.50
		b. Using questioning and discussion techniques	
		c. Engaging students in learning	
		d. Using assessment in instruction	
		e. Demonstrating flexibility and responsiveness	

Note. Reliabilities were reported in Kane & Staiger (2012), Table 11

The Classroom Assessment Scoring System (CLASS) was originally developed as an evaluation tool for elementary classrooms (Pianta, Hamre, Hayes, Mintz, & LaParo, 2008). CLASS was developed in the Curry School of Education Center for Advanced Study of Teaching and Learning at The University of Virginia. There are multiple versions based upon level of instruction ranging from preschool to high school. The goal of the CLASS is to focus on observing student and teacher interactions in classrooms. The Classroom Assessment Scoring System –Secondary focuses on what teaching elements support student cognitive and social development (Hamre & Pianta, 2010). The CLASS-S used in the MET study contained four domains: emotional support, classroom organization, instructional support, and student engagement, which used a total of 12 items each scored on a 7 point scale (see Table 2).

Table 2.

Summary of CLASS Reliability Findings

Instrument	Domain	Component	Inter-rater Reliability
CLASS			**0.63**
	Emotional support	a. Positive climate	0.61
		b. Negative climate	
		c. Teacher sensitivity	
		d. Regard for student perspectives	
	Classroom organization	a. Behavior management	0.62
		b. Productivity	
		c. Instructional learning formats	
	Instructional support	a. Quality of feedback	0.55
		b. Content understanding	
		c. Analysis and problem solving	
		d. Instructional dialogue	
	Student engagement	a. Student engagement	--

Note. Reliabilities were reported in Kane & Staiger (2012), Table 11.

CLASS has been validated in over 2000 classrooms, is multidimensional, and describes aspects of teaching linking student achievement and development (Hamre & Pianta, 2010). It was designed as a method to provide feedback to novice and experienced teachers with regards to their personal effectiveness (Hamre & Pianta, 2010).

Survey instruments.

The Tripod 7Cs survey (TRIPOD) is a survey that is used to measure student perceptions of teaching quality and learning engagement (Ferguson & Danielson, 2014). Developed by Ferguson in 2000, TRIPOD is based upon Erik Erikson's five stages of life-cycle identity development (Ferguson & Danielson, 2014, p 104). It is adapted to a 7Cs framework including:

challenge, control, care, confer, captivate, clarify, and consolidate. The MET researchers grouped the first two Cs, challenge and control into a construct called “press” that most strongly predict student achievement gains (Ferguson & Danielson, 2014). “Press” includes challenge and control and addresses effort and rigor, while control concerns class behavior. The remaining Cs are grouped into “support” which focus more on the quality of teaching and the student experience (Ferguson & Danielson, 2014). “Support” includes teacher’s supportive relationships, student interactions concerning discussion, engagement, clarification, and summarization. The MET project provided evidence of the instrument’s score validity and reliability.

The principal survey was developed by the MET study. It is a 4 question survey that first asked the number of teacher’s formal observations, and the number of teacher’s informal observations. The third question on the principal rating scores asked: Among teachers you have known who taught the same grade/subject, how would you rate this teacher overall? It is scored on a Likert type scale ranging from 1 to 6 with 1 representing exceptional teachers (top 5%), 2 representing very good teachers (top 25%), 3 representing good teachers (top 50%), 4 representing fair teachers (top 75%), 5 representing poor teachers (bottom 25%), 6 representing very poor teachers (bottom 5%). The fourth questions asked the principal’s confidence in the rating they assigned the teacher. There are no known published results using this measure.

Student growth measurement.

The ACT QualityCore® Algebra I is an end of course assessment. It is a component of ACT’s College and Career Readiness System and is designed to measure the learning outcomes necessary for student success in college and for career readiness. The ACT QualityCore® Algebra I is a standardized 70 item multiple choice test. The ACT worked with The Education

Trust in the 2003-2004 school year to generate the “On Course for Success Study”. “On Course for Success” examined 10 schools with significant minority and low-income populations in order to create standards for the QualityCore® set of exams that would be successful in a more diverse population (ACT & The Educational Trust, 2004). The mathematics standards test development required that 35% or greater of the student population scored a 22 or greater on the ACT Mathematics Assessment Test (ACT & The Educational Trust, 2004, p 7); 7 out of the 10 schools met the requirement (ACT & The Educational Trust, 2004, p 37). The seven high-performing minority and low-income schools that demonstrated a strong ability for college preparatory in mathematics were examined qualitatively to create the derived course standards for the ACT QualityCore® Algebra I (see Appendix D). This initial validity assessment examined content, clarity and specificity, and organization of the standards created (see Appendix D). A validity study was conducted by a survey given to high school teachers in 2005 that addressed the rigorous nature of the created ACT QualityCore ® Standards from the 2003-2004 school year (ACT QualityCore®, 2014). Additional validity was determined by a content experts' panel review. A combination of the expert panel review and the high school teachers' determination of essential skills determined the validity of the standards (ACT QualityCore®, 2014).

Item development began with a diverse group of high school teachers, was refined by ACT test development specialists, and reviewed by external content experts. New test items are field tested with no fewer than 500 student responses per item (ACT QualityCore®, 2014). Raw scores are scaled into a range of 125-175 with different value meanings per content area. Constructed-response items are reviewed by trained outside raters. Initial QualityCore® End of Course Assessments were given nationally in the 2008-2009 school year using a stratified random design with a minimum of 50-60 schools per subject to determine base data (ACT QualityCore®, 2014). The ACT QualityCore® Algebra I demonstrated a mean score of 144.85 ± 4.31, with the standard error of the mean equaling 2.22, and the test reliability equaling 0.75 (ACT QualityCore®, 2014, p. 31).

Connections between instruments.

The correlation between the FFT and the CLASS for cross-subject curriculum is 0.88 even though the two instruments measure distinct competencies (Kane & Staiger, 2012, p 32). In the MET Project both the FFT and the Tripod produced valid and reliable score indicators of teaching quality at the classroom level (Cantrell & Kane, 2013; Ferguson & Danielson, 2014; Ho & Kane, 2012; Kane, McCaffrey, & Staiger, 2010; Kane, et al., 2012).

The FFT and TRIPOD are conceptually compatible and empirically compatible with all of the 7Cs components, and are statistically significantly correlated with FFT components with 95 percent confidence levels (Ferguson & Danielson, 2014). Furthermore, a crosswalk study of FFT and TRIPOD demonstrated that teacher observations and student opinions of teacher effectiveness in classroom environments showed Control and Challenge from the TRIPOD, as well as any components measuring Classroom Environment from the FFT were stronger and more consistently positive predictors of SAT-9 and BAM scores (Ferguson & Danielson, 2014, p 129).

Research Questions

1. What are the identifiable characteristics that define latent classes of Algebra I teachers?

2. What characteristics indicate effective Algebra I teachers?

3. How do student scores on the ACT QualityCore ® Algebra I standardized exam predict effective/non-effective Algebra I teachers?

Data Analysis Procedures

Unlike survey instruments, observational measurements are often examined as whole score measures; however, using observational measurements as a single construct does not explore the multiple dimensions potentially captured. Psychometric research proposes that single construct measurements may in fact fall under this multidimensional idea (Grossman, Cohen, Ronfeldt, & Brown, 2014; Lazarev & Newman, 2014). Also, the evaluation of individual items is an approach only taken in professional development situations where individual items might be linked to a specific teacher characteristic (Danielson, 2012).

Data from the MET project was organized for desired variables by teacher IDs of ninth grade Algebra I teachers using IBM SPSS Statistics 23®. Table 5 lists the observable variables, their scales and their scale values, and the domain that supported effective teacher quality.

Latent class analysis (LCA) was used to categorize teachers as effective or non-effective based upon seven continuous observed polytomous variables hypothesized as a latent class model, M_0 (null) and M_1 (dichotomous or polytomous). The model in which there are seven observed variables and one unobserved latent dichotomous or polytomous variable (X) can be expressed in equation form as:

$$\pi_{ijklmn}^{ABCDEFX} = \pi_r^X \pi_{ir}^{\bar{A}X} \pi_{jr}^{\bar{B}X} \pi_{kr}^{\bar{C}X} \pi_{lr}^{\bar{D}X} \pi_{mr}^{\bar{E}X} \pi_{nr}^{\bar{F}X}$$

for i = 1, … , I; j = 1, … , J: k = 1, … , K; l = 1, … , where $\pi_{ijklmn}^{ABCDEFX}$ has the shared probability that an observation of class i is on variable A, class j is on variable B, class k is on variable C, et cetera, and that A, B, C, and D are mutually independent (Hagenaars & McCutcheon, 2002). The observed variables reflect the teacher characteristics in the LCA model (Figure 1). Gathering information about strengths and weaknesses in the measurements will sort teachers into latent classes. The use of LCA will not only serve to explore categories of teacher

Table 3.

Items and Scales of Instruments Used in the MET Study

Domain	Measures (MET file)	Scale	Values	Meaning
Teacher Promoted Student Engagement	TRIPOD: Challenge, Captivate, & Consolidate	5	1	Totally untrue
	(34414-0004)		2	Mostly untrue
			3	Somewhat untrue/true
			4	Mostly true
			5	Totally true
	CLASS: Student Engagement	7	1 & 2	low
	(34346-0002)		3, 4, & 5	mid
			6 & 7	high
Teacher Promoted Student Understanding	CLASS: Instructional Support	7	1 & 2	low
	(34346-0002)		3, 4, & 5	mid
			6 & 7	high
	FFT: Instruction	4	1	Unsatisfactory
	(34346-0004)		2	Basic
			3	Proficient
			4	Distinguished
Classroom Management	TRIPOD: Care, Control, Confer	5	1	Totally untrue
	(34414-0004)		2	Mostly untrue
			3	Somewhat untrue/true
			4	Mostly true
			5	Totally true
	CLASS: Organization & Emotional	7	1 & 2	low
	(34346-0002)		3, 4, & 5	mid
			6 & 7	high
	FFT: Classroom Environment	4	1	Unsatisfactory
	(34346-0004)		2	Basic
			3	Proficient
			4	Distinguished

effectiveness, but will also be an easily translatable method “because teachers’ scores and their measurement errors can be easily computed, once the parameters of the measurement model have been estimated (Halpin & Kieffer, 2015, p 264).

LCA uses measured (observed) variables to indicate latent (unobserved) classes which are categorical in nature. Using LCA allowed the clustering of teachers based upon similarities in their scores. The goal of the clustering is to explore latent classes of teachers from a spectrum of non-effective to effective. Similarly scoring teachers demonstrated differences in teacher characteristics because the observed variables are not correlated with the unobserved variables. Then, from the latent classes, to use the terms of Halpin and Kieffer (2015), a kind of “profile membership score” versus a total score becomes a summary of the measurement. LCA was performed using *Mplus* 7.4© (Muthén & Muthén, 2014) with two or three fixed classes selected. The *Mplus* 7.4© program is in Appendix F. The numbers of classes were examined for the goodness of fit for different models (Lubke & Neale, 2008; Schumacker & Pugh, 2013; Schumacker & Lomax, 2016). It was important to use robust standard errors to address model misspecification (Halpin & Kieffer, 2015).

After the latent class analysis, the decision accuracy for validity was determined using the principal rating score of teacher effectiveness. The chi-square statistic was used to determine if the latent class categorization and the principals’ categorization were independent or dependently related.

Principal observation evaluation instruments are often used when evaluating teacher effectiveness. High principal evaluation scores generally represent teachers with both strong classroom management and strong teacher student relationships (Stronge, Ward, & Grant, 2011).

Table 4. Decision Accuracy Using Principal Effectiveness Scores

		Decision made on teacher effectiveness based upon latent class analysis	
		Effective Teacher	**Non-effective Teacher**
Principal assessment of teacher effectiveness	**Effective Teacher**	Correct Classification	Misclassification
	Non-effective Teacher	Misclassification	Correct Classification

A multivariate analysis of variance (MANOVA) was also conducted to determine differences between teachers' mean scores on the seven observed variables based on the two and three latent class models. The MANOVA results were then compared to the results from the two and three latent class models using discriminant analysis.

Discriminant analysis estimates relationships between one or more categorical variables with a set of quantitative variables. In order to examine the ability of standardized exam scores to predict effective and non-effective Algebra I teachers, the association of LCA scores with ACT QualityCore® Algebra I scores was compared using discriminant analysis. Discriminant analysis used student success by score on the Algebra I portion of the ACT QualityCore® series to predict effective and non-effective teacher latent group membership.

There are two types of discriminant analysis: simple discriminant analysis which uses the theory of maximum group differences to determine classification functions and canonical discriminant analysis (Klecka, 1980). Simple classification functions are used in the analysis as no tests of significant dimensionality were performed. The simple discriminant analysis uses the following general formula: $D = v_1X_1 + v_2X_2 + v_3X_3 = \ldots\ldots\ldots v_iX_i + a$, in which the discriminant function is represented by D, the discriminant coefficient is represented by v, the respondent's score for that variable is represented by X, the number of predictor variables is i, and a is a constant (Burns & Burns, 2008). Variables are assigned to groups with the highest discriminant score taking the cost of misclassification into account (Klecka, 1980). Determining prediction accuracy requires a classification matrix in which an analysis sample and the latent class results are used to compute the discriminant function and a validation sample, the ACT QualityCore® Algebra I scores were used to develop the classification matrix.

Summary

First, latent classes of effective/non-effective Algebra I teachers were established by selecting categories of teacher skills for the observed variables. These observed variable categories were generated from the student survey (TRIPOD), the CLASS, and FFT, observational scores. *Mplus* 7.4© was used to perform the latent class analysis resulting in an undetermined number of latent classes (null model). Next, decision accuracy was used to validate the latent classes established by conducting a chi-square test of independence between the established latent classes and the principal survey ratings of teacher effectiveness. MANOVA was used to indicate mean differences in the continuous observational variable for two class and three latent class models. Finally, student achievement data on the ACT QualityCore ® Algebra I content was used to assess the categorical prediction of effective/non-effective teachers by conducting a discriminant analysis.

CHAPTER 4: RESULTS

Introduction

The measurements supporting teacher promoted student engagement included the TRIPOD Challenge, Captivate, and Consolidate, and the CLASS Student Engagement. The measurements supporting teacher promoted student understanding included the FFT Instruction and CLASS Instructional Support. The measurements supporting classroom management included FFT measures. The observed variables were selected to represent the most important constructs in the literature and were explained in Chapter 3.

Table 5. Observed Variable Names for LCA

Domain	Measurement	Variable Name
TEACHER PROMOTED STUDENT ENGAGEMENT	TRIPOD: Challenge, Captivate, and Consolidate	T_ENGAGE
	CLASS: Student Engagement	C_ENGAGE
TEACHER PROMOTED STUDENT UNDERSTANDING	CLASS: Instructional Support	C_UNDER
	FFT: Instruction	F_UNDER
CLASSROOM MANAGEMENT	TRIPOD: Care, Control, and Confer	T_MANAGE
	CLASS: Organizational and Emotional Support	C_MANAGE
	FFT: Classroom Environment	F_MANAGE

Note. *n* = 143

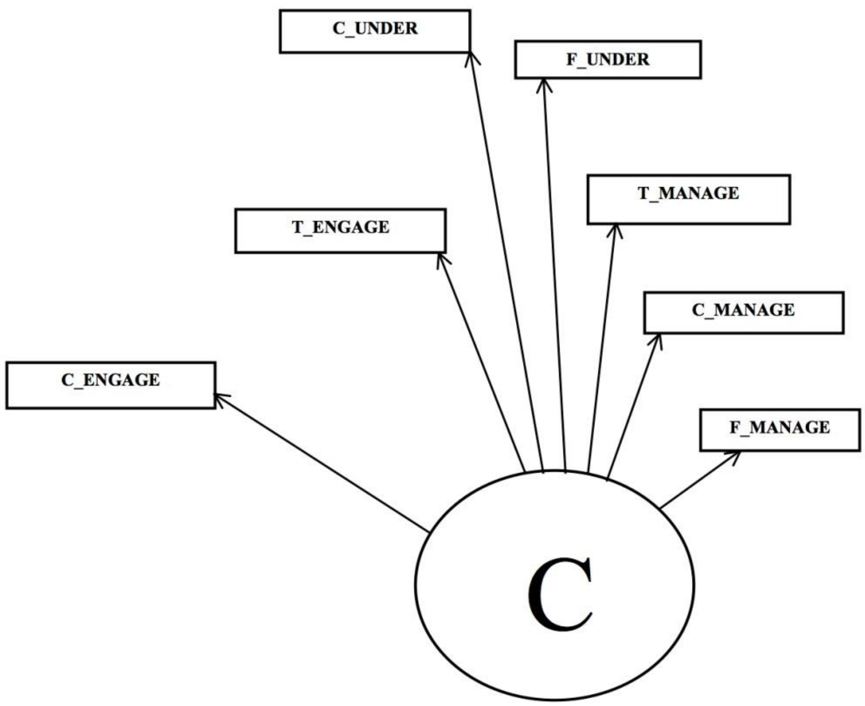

Figure 2. Latent Class Model with Variable Names.

Usage of the MET study VDE and its corresponding data proved more restrictive than anticipated. The first and most dramatic alteration in data usage was due to data missingness. Continuous observed variables were selected from the requested files (see Appendix C) for ninth grade Algebra I teachers (*n* =152) that represented the proposed instrument domains (see Table 5) using *IBM SPSS Statistics* 22®. The creation of a master file that organized desired variables by teacher IDs resulted in the realization of missing completely at random data for 28 percent of the seven observed variables used in the LCA. *IBM SPSS Statistics* 22® was used to impute missing values for these teachers. Further missingness was a result of non-existent data values for nine teachers on the variables T_ENGAGE and F_UNDER. Dropping variables with high non-response was necessary in this context (Rubin, 1976). Adaptations due to severe missingness for these observed variables per teacher reduced the size to $n = 143$, a nine participant loss of 6% from original sample size of $n = 152$. Although listwise deletion is often deemed a last resort measure, complete case analysis was more powerful and appropriate for this study (Little & Schemker, 1995).

Research Questions

Research question 1.

What are the identifiable characteristics that define latent classes of Algebra I teachers?

Using the observed variables from Figure 2, and *Mplus* 7.4© by Muthén & Muthén, exploratory LCA was run using two and three latent class models to sort Algebra I teachers into categories of effectiveness. (For details of the *Mplus* 7.4© programs see Appendix F.)

LCA with two classes.

Results of the two class LCA solution indicated that the model estimation terminated normally with 143 observations, seven dependent variables, and one categorical latent variable. See Table 6 for mean differences in the two latent class model. Model fit information identified 22 free parameters with a loglikelihood value of -1516.44, a sample-size adjusted BIC of 3072.46, and a classification quality, or entropy, of .93. Final class counts and proportions of latent classes resulted in two latent classes, the first class identified 40 teachers or 28% of the teachers with a probability of .95 as belonging to a distinct Class 1 and the second class identified 103 teachers or 72% of teachers with a probability of .98 of belonging to a distinct Class 2 membership. These were subjectively labeled non-effective ($n = 40$) and effective teachers ($n = 103$), respectively. The Vuong, Lo, Mendell, Rubin likelihood ratio test for one versus two latent class models was 1681.24 with a $p < .00001$.

Table 6. Two Class Solution for Teacher Effectiveness

Variable		*Two Class Solution*	
	Overall Item Means	*Class 1*	*Class 2*
T_ENGAGE	59.89	57.17	60.94
C_ENGAGE	4.14	3.52	4.38
C_UNDER	2.78	2.42	2.92
F_UNDER	8.64	7.42	9.12
T_MANAGE	50.19	47.26	51.33
C_MANAGE	4.35	3.81	4.55
F_MANAGE	9.34	7.66	9.99
n	143	40	103

Figure 3 represents the sample means of the established two latent class solution in *Mplus* 7.4©. The x-axis represents the seven observed variables and the y-axis shows the mean values for the sorted teacher classes. The trend lines represent Class 1 or non-effective teachers, 28%, of the population and Class 2 effective teachers, 72%, of the population. A greater separation between trend lines of Class 1 and Class 2 shows the significant difference in the two latent class solution.

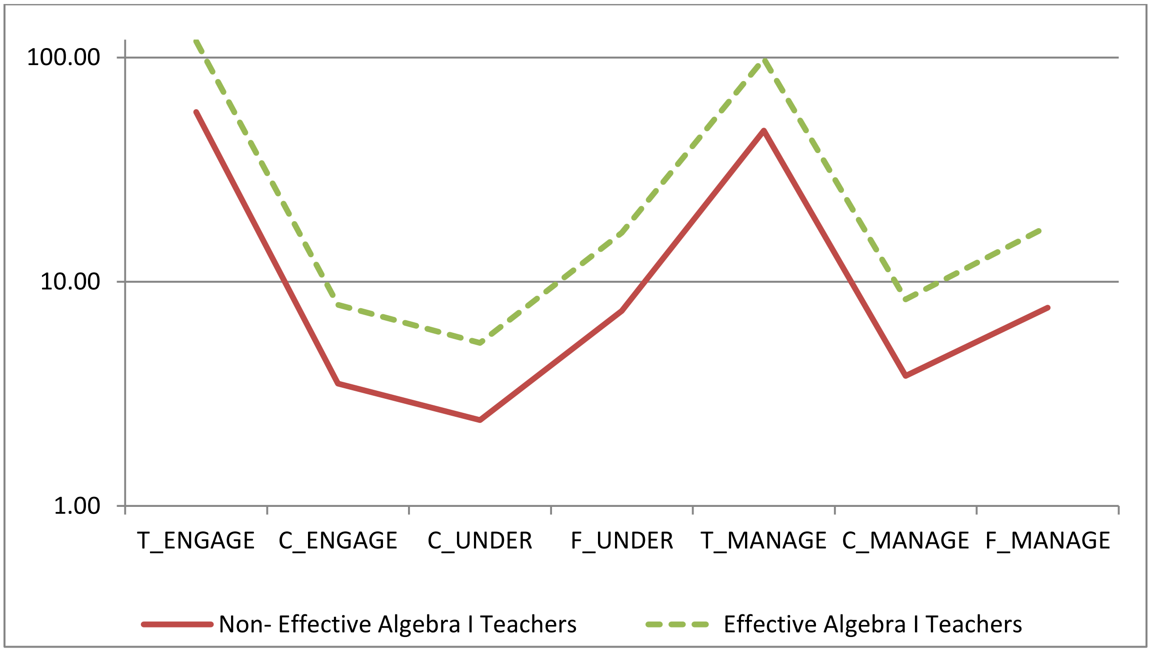

Figure 3. Log transformed mean scores for two class LCA solution

LCA with three classes.

Results of the three class LCA indicated that the model estimation terminated normally with 143 observations, seven dependent variables, and one categorical latent variable. See Table 10 for mean differences in the three latent class model. Model fit information identified 30 free parameters with a loglikelihood value of -1475.79, a sample-size adjusted BIC of 3005.54, and a classification quality, or entropy, of .82. Final class counts and proportions of latent classes resulted in three latent classes; the first class identified 28 teachers or 19% of the teachers with a probability of .96 as belonging to a distinct class, a second class identified with 61 teachers or 43% of teachers with a probability of .90 of belonging to a distinct second class membership, and a third class identified 54 teachers or 38% of the teachers with a probability of .92. These were subjectively labeled non-effective (n = 28), moderately effective (n = 61), and effective (n = 54) teachers. The Vuong, Lo, Mendell, Rubin likelihood ratio test for two versus three latent classes was -1516.44 with p = .604.

Table 7. Three Class Solution for Teacher Effectiveness

Variable		*Three Class Solution*		
	Overall Item Means	*Class 1*	*Class 2*	Class 3
T_ENGAGE	59.89	56.29	60.15	61.46
C_ENGAGE	4.14	3.42	4.07	4.56
C_UNDER	2.78	2.32	2.70	3.11
F_UNDER	8.64	7.13	8.62	9.46
T_MANAGE	50.19	46.65	50.30	51.90
C_MANAGE	4.35	3.69	4.30	4.74
F_MANAGE	9.34	7.36	9.46	10.23
n	143	28	61	54

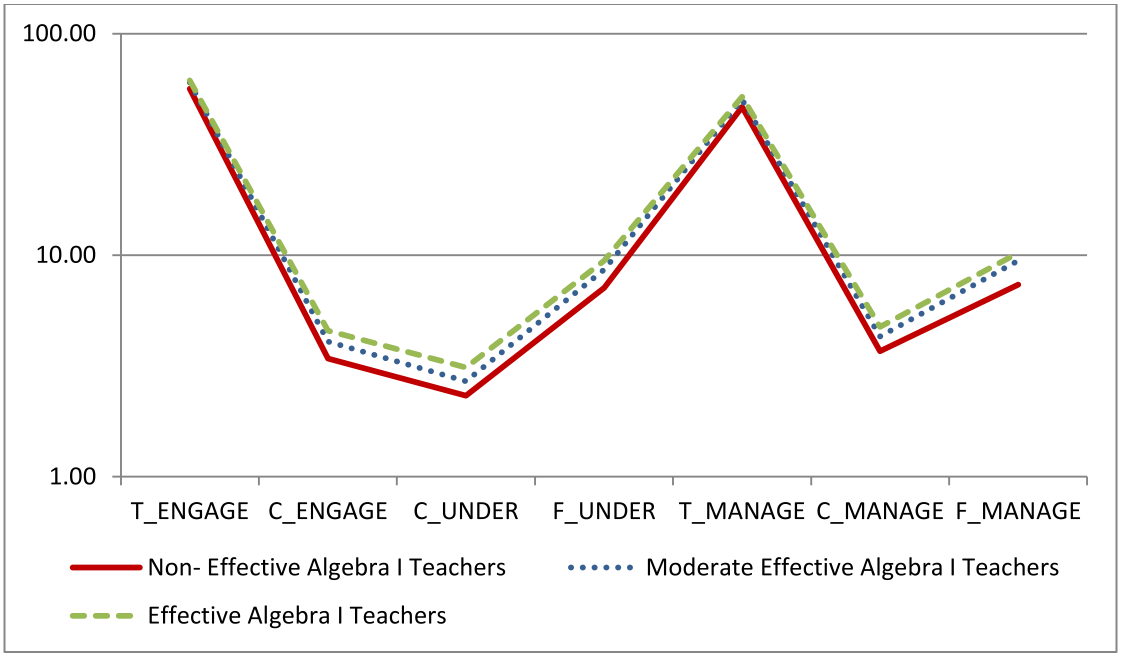

Figure 4. Log transformed mean scores for three class LCA solution

Figure 4 represents the sample means of the established three latent classes in *Mplus* 7.4©. The x-axis represents the seven observed variables and the y-axis shows the mean values for sorted teacher classes. The trend lines represent Class 1 or non-effective teachers, 19%, of the population, Class 2 moderate effective teachers, 43%, of the population, and Class 3 effective teachers, 38% of the population. The close proximity of the trend lines supports a poor class fit for the three LCA solution.

Model fit.

Evaluating the two latent class and the three latent class solutions revealed that the two latent class solution had a better model fit (see Table 8). The Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are penalized likelihood criteria used for selecting best predictor subsets and for comparing non-nested models, with BIC representing a stricter measurement (Dziak, Coffman, Lanza, & Li, 2012). Generally, a lower value for AIC and BIC represents a better model fit (Dziak, Coffman, Lanza, & Li, 2012). Adjusted BIC (ABIC) depends on sample size (Dziak, Coffman, Lanza, & Li, 2012). In LCA, entropy or classification certainty was greater for the two latent class model than for the three latent class solution (see Table 8), and entropy values greater than .80 designate a good separation of identified latent classes (Ramaswamy, et al, 1993). The sample size, n = 143, and the number of items, seven, suggested that the best method for examining model fit was the Lo, Mendell, Rubin Test, which is based upon bootstrapping (Dziak, Coffman, Lanza, & Li, 2012).

Table 8. Assessing Fit for Different Numbers of Classes

Fit Indices	*1 Class*	*2 Classes*	*3 Classes*
AIC	3390.47	3076.88	3011.58
BIC	3431.95	3142.07	3100.47
Sample Size Adjusted BIC	3387.66	3072.46	3005.54
Entropy	na	0.93	0.82
Lo, Mendell, Rubin Test	na	2 v 1	3 v 2
		-1681.24	-1516.44
		$p < .00001$*	$p = .604$
N for each class	*C*1 = 143	*C*1 = 40 *C*2 = 103	*C*1 = 28 *C*2 = 61 *C*3 = 54

Note. All fit statistics were calculated using *Mplus* 7.4®
*Significant at $\alpha = .05$.

The Lo, Mendell, Rubin Test revealed a significant difference between the two latent class solution versus the three latent class solution (see Table 8), revealing that the two latent class solution had the best model fit. Figures 3 and 4 also support these conclusions.

Decision accuracy for the validity of the two latent class solution was determined by comparing results with the principal effectiveness score from the short MET Principal Rating Survey. The principal score was available for 100 out of the 143 Algebra I teachers (70%). No principals rated teachers as a 6 or 7, the lowest 25% and 5% of teachers rated, respectively. The scale of remaining scores, 1 through 5, created categories of teachers that were smaller than n =

10, so scores were recoded into 1 and 2, with scores of 1 representing the “worst” teachers and scores of 2 representing the “best” teachers. A chi-square statistic revealed no significant difference between the two latent class solution for LCA and the principal’s rating of effective teachers with $\chi^2(1) = .512$, $p = .474$.

Research question 2.

What characteristics indicate effective Algebra I teachers?

Means of the seven continuous variables were compared with a MANOVA for the two latent class and the three latent class solutions using *IBM SPSS Statistics* 22®. Normality of the two latent class solution was satisfied with Box’s Test ($M = 34.25$; $F = 1.14$, $p = .282$). The MANOVA revealed a statistically significant difference in teacher effectiveness measures based upon the two latent class solution, $F(1,141) = 45.45$, $p < .001$; Wilks’ $\Lambda = .30$, partial η^2 of .70. Univariate tests supported that all of the seven observed variables were significant, see Table 12 for individual observed variable scores. All variables had significantly higher average scores for LCA with two latent classes of teachers: non-effective teacher ($n = 40$) and effective teacher ($n = 103$). Partial eta squared was based upon Cohen’s (1969) derivation of values of f resulting in effect sizes of small (.0099), medium (.0588), and large (.1379) (pp. 278-280). Effect sizes with regards to the observed variables had medium to large effect sizes (see Table 9). Although the variables T_ENGAGE and T_MANAGE were significant in the two latent class solution MANOVA, the η^2 values are much lower than for the other observed variables.

Normality of the three latent class solution was not satisfied with a Box’s M test ($M = 105.48$; $F = 1.73$, $p = .001$). The MANOVA revealed a statistically significant difference in teacher effectiveness measures based upon the three latent class solution, $F(2,140) = 30.30$, $p < .001$; Wilks’ $\Lambda = .15$, partial η^2 of .61. Univariate tests supported that all of the seven observed variables were significant, see Table 13 for individual observed variable scores. Effect sizes with regards to the observed variables had medium to large effect sizes (see Table 10). Pairwise comparisons with a Dunn-Bonferroni adjustment revealed significant differences between the three latent classes for all of the seven observed variables except for T_MANAGE. Variables T_ENGAGE, C_ENGAGE, C_UNDER, F_UNDER, C_MANAGE, and F_MANAGE showed latent Class 3 effective teachers having a significantly higher mean than latent Class 2 moderate effective teachers, and latent Class 1 non-effective teachers. The variable T_MANAGE demonstrated no significant difference between latent Class 3 and latent Class 2 groups in the three class solution.

The two latent class and three latent class solution MANOVA results were followed up with discriminant analysis to examine group differences using latent class probability of group membership, rather than purely exploring a linear combination of outcome variables (Klecka, 1980). For the two latent class solution, the discriminant function explained canonical $R^2 = .70$. The function was significant with Wilks' $\Lambda = .30$, χ^2 (7), p <.001. The three latent class solution revealed two discriminant functions. The first explained 99% of the variance, canonical $R^2 = .85$, while the second function explained 1% of the variance, canonical $R^2 = .05$. In combination these discriminant functions significantly differentiated classes, Wilks' $\Lambda = .15$, χ^2 (14) = 259.98, $p < .001$, but after removing the first function, the second function did not significantly differentiate the classes, Wilks' $\Lambda = .95$, χ^2 (6) = 7.54, $p = .274$. Discriminant analysis showed no difference between the two latent class and three latent class solutions established in the LCA on the observed variables T_ENGAGE, C_ENGAGE, C_UNDER, F_UNDER, C_MANAGE, T_MANAGE, and F_MANAGE.

Table 9. Univariate Tests Two LCA Model

	CLASS 1			CLASS 2					
Variable	*M*	*SD*	*SE*	*M*	*SD*	*SE*	*F*(1,141)	*p**	η^2
T_ENGAGE	57.17	6.55	1.03	60.94	5.60	0.55	11.84	0.001	0.08
C_ENGAGE	3.52	0.38	0.06	4.38	0.32	0.03	184.23	< .001	0.57
C_UNDER	2.42	0.34	0.05	2.92	0.35	0.03	60.74	< .001	0.30
F_UNDER	7.42	0.94	0.15	9.12	0.83	0.08	110.80	< .001	0.44
T_MANAGE	47.26	5.78	0.91	51.33	6.16	0.61	13.04	< .001	0.09
C_MANAGE	3.81	0.32	0.05	4.55	0.27	0.03	193.31	< .001	0.56
F_MANAGE	7.66	1.30	0.20	9.99	1.04	0.10	126.27	< .001	0.47

Note. Class 1, n = 40; Class 2, n = 103; * $\alpha = .05$

Table 10. Univariate Tests for Three LCA Model

	CLASS 1			CLASS 2			CLASS 3					
Variable	*M*	*SD*	*SE*	*M*	*SD*	*SE*	*M*	*SD*	*SE*	*F*(2,140)	*p*	η^2
T_ENGAGE	56.29	7.19	1.36	60.15	5.65	0.72	61.46	5.26	0.72	7.32	0.001	0.09
C_ENGAGE	3.42	0.40	0.07	4.07	0.28	0.04	4.56	0.23	0.03	154.83	< .001	0.68
C_UNDER	2.32	0.32	0.06	2.70	0.22	0.03	3.11	0.36	0.05	66.02	< .001	0.48
F_UNDER	7.13	0.84	0.16	8.62	0.68	0.09	9.46	0.88	0.12	79.80	< .001	0.53
T_MANAGE	46.65	6.32	1.20	50.30	5.62	0.72	51.90	6.40	0.87	6.93	0.001	0.09
C_MANAGE	3.69	0.28	0.05	4.30	0.28	0.02	4.74	0.22	0.03	211.44	< .001	0.75
F_MANAGE	7.36	1.30	0.25	9.46	1.24	0.16	10.23	0.88	0.12	60.12	< .001	0.46

Note. Class 1, n = 28 ; Class 2, n = 61 ; Class 3, n= 54

Research question 3.

How do student scores on the ACT QualityCore ® Algebra I standardized exam predict effective/non-effective Algebra I teachers?

Discriminant analysis was used to categorically predict effective/non-effective teachers based on student achievement data from the ACT QualityCore ® Algebra I. Again missingness of data limited the analysis to *n* = 138 rather than the *n* = 143 Algebra I teachers used in the LCA.

Final ACT QualityCore® scale scores were used with established latent classes from the LCA two latent class solution. The overall sample mean of the ACT QualityCore® Algebra I scores was 140.25 (*SD* = 3.18), the mean for the non-effective teachers was 139.19 (*SD* = 3.02), and the mean for the effective teachers was 140.67 (*SD* = 3.16). The ACT QualityCore® Algebra I scores for this student sample was lower than the normalized values given by the ACT®, *M* = 144.85, *SD* = 4.31). Students in this study averaged one standard deviation below the normalized ACT raw score, thus below average students. For the two latent class solution, the discriminant function explained, canonical $R^2 = .04$. The function was significant with Wilks' $\Lambda = .96$, $\chi^2(1) = 6.08$, $p = .014$. The three latent class solution revealed one discriminant function. The function explained, canonical $R^2 = .04$. The three latent class discriminant function did not significantly differentiate classes, Wilks' $\Lambda = .96$, $\chi^2(2) = 5.75$, $p = .06$.

Discriminant analysis was followed by chi-square analysis using predicted group membership from the discriminate analysis and class membership designated from the LCA. For the two latent class LCA, there was no significant difference between groups, $\chi^2(1) = .473$, $p = .492$. There is no relationship between student achievement data from the ACT QualityCore ® Algebra I and the latent classes of effective and non-effective Algebra I teachers.

Summary

Selected observed variables from the MET study including components of TRIPOD, CLASS, and FFT were selected for Algebra I teachers to represent the teacher qualities of student engagement, the promotion of student understanding and classroom management. These seven observed variables were used in LCA to successfully categorize teachers into two latent classes; non-effective teachers and effective teachers. This classification was supported by a comparison of principal ratings of teacher effectiveness with the two class solution in LCA.

Observed variable mean scores were then compared for the characteristics of teachers for the two and three latent class solutions. The two latent class solution was determined to have the best model fit. MANOVA and discriminant analysis was used to further investigate the viability of the two latent class solution. Finally, discriminant analysis was used to test whether student standardized test scores on the ACT QualityCore® Algebra I predict effective and non-effective teachers. Results indicated that the test scores do not predict the teacher effective/non-effective group membership.

The following chapter presents a more in depth discussion of the findings and relates possible conclusions for using a LCA methodology to identify effective and non-effective teachers rather than using VAMs and composite measures that rely on student standardized test scores. These methods and the results of these methods are compared to the current literature. Strengths and weakness of the study are examined. Future implications with regard to comprehensive teacher evaluation systems and improving teacher education programs and professional development opportunities are discussed.

CHAPTER 5: DISCUSSION

Introduction

The purpose of this study was to use multidimensional observed variables as characteristics of teachers to determine if they identified latent classes of teachers (effective/non-effective), and to determine if student standardized test scores on the ACT QualityCore® Algebra I can predict teacher effective/non-effective group membership.

One distinguishing characteristic of the LCA methodology used in this study was to classify effective and non-effective teachers, which is a departure from VAM models using covariates such as SES, gender, ethnicity, and English language proficiency for high school students when exploring the results of standardized test scores. Controlling for these variables is traditional in the hierarchical linear modeling (HLM) approach, but was not regarded as having methodological importance in this study.

Research question 1 explored the identification of characteristics for establishing latent classes of Algebra I teachers. The teacher characteristics were represented by components of two observational measures (CLASS and FFT), and one student survey measure (TRIPOD), which served to give separate identities to classes of Algebra I teachers. Characteristics used to establish latent classes of teachers included teacher promoted student engagement, teacher promoted student understanding, and classroom management. Results identified a two latent class solution, with groups labeled non-effective and effective teachers. A three latent class solution was determined to not have a good model fit. Research question 2 identified characteristics indicating effective Algebra I teachers. Since all of the continuous observed variables had statistical significance between the two latent class solution, each measure supported one or more characteristic to successfully establish latent classes of Algebra I teachers including teacher promoted student engagement, teacher promoted student understanding, and classroom management. The second research question further explored the individual continuous observed variables and their scoring scales relative to the two latent class solution. The characteristic representing teacher promoted student engagement was supported by both TRIPOD and CLASS measurements, indicating teachers with higher scores were more successful in demonstrating this quality of teacher effectiveness. The same holds true for the use of the relevant components of CLASS and FFT for representing teacher promoted student understanding, and the relevant components of CLASS, FFT, and TRIPOD for assessing classroom management. Research question 3 investigated latent class membership with ACT QualityCore® Algebra I test scores revealing that student scores on the standardized test did not predict effective/ non-effective group membership of teachers determined by the LCA.

Summary of Results

The multidimensionality of teaching practice deserves a multidimensional approach for the classification of teacher effectiveness. However, previous complications with a multidimensional approach encountered problems with indeterminacy (Grossman, et al., 2014; Lasarev & Newman, 2014). LCA used in this study helped to solve the problem of psychometric indeterminacy by using seven polytomous variables to represent components of larger measurements. Constructs of quality teacher characteristics posits a different and determinant methodology for classifying teachers. The novelty of using LCA in determining two latent classes is important because it differs from current HLM models that use nested effects and covariates in traditional VAM modeling (Gottlieb, 2013). Teachers regardless of skill were being rating basically the same (Gottlieb, 2013). The MET project (2013) reported that a composite measure comprising equally weighted whole versions of the measurements (TRIPOD, FFT, and CLASS) could be used along with achievement growth (ACT scores) to predict teacher-student achievement growth; which they indicated as a measure of effective teaching, but cautioned that using these equally weighted indicators would not be able to identify characteristics of effective teachers (Mihaly, McCaffrey, Douglas, & Lockwood, 2013).

The two latent class solution established two distinctly different groups of Algebra I teachers using the MET data. Harris, Ingle, and Rutledge (2014) have made clear distinctions for the highest of performing teachers, but many approaches are unable to separate the commonly acceptable classification of teachers when using mixed rating instruments; including teacher observations, principal opinions, and student surveys. Results of this study demonstrated that teacher characteristics do indicate effective and non-effective teachers on key measures when using latent class analysis.

The use of latent class analysis to identify quality teachers is novel, and especially when using observable characteristics from both trained outside evaluators and student viewpoints. This gives an important perspective for examining quality teaching (Bill and Melinda Gates Foundation, 2013). Student opinion surveys have been proposed as more stable over time than other measurements (Polikoff, 2013) and have been regarded as able to fairly assess teacher skills and characteristics when compared with supervisor evaluations (Tairab & Wilkinson, 1991). However, although the student opinions from TRIPOD did indicate significance, the TRIPOD variables for teacher promoted student engagement and classroom management had much smaller effect sizes, suggesting that an emphasis on student evaluations is not as important for determining effective and non-effective teachers. Outsider observations of teacher performance and student opinion have both been determined as reliable teacher evaluation measurements (Gersten, Baker, Haager, & Graves, 2005; Van Tassel-Baska, Quek, & Feng, 2006); however, there are fewer studies revealing the combination of using these tools for evaluating teacher effectiveness, and the majority of them are using the MET database. Halpin and Kieffer (2015) used the MET study to examine profiles of instructional practice for middle school English language arts teachers. Using the application of LCA they used classroom observational instruments including the FFT, CLASS and the Protocol for Language Arts Teaching Observation (PLATO) scores to compare mean scores that classified teachers into four profiles (Halpin & Kieffer, 2015). Profile association was then discussed with regards to teacher strengths and weaknesses in areas of discourse, modeling, instruction, time, and behavior management revealing only slight teacher nuances and variability in teacher strengths and weaknesses within group profile membership (Halpin & Kieffer, 2015). In contrast, these results indicated two latent classes of teachers based on teacher observations and student opinions, which is unique in this study of MET Algebra I teachers.

The two latent class model fit was supported by principal evaluations of their teachers. High principal evaluation scores are shown to represent teachers with strong classroom management scores and for teachers that build strong relationships with students (Stronge, Ward, & Grant, 2011). Principals are able to pinpoint poor PCK skills including problems with classroom management, and developing a rapport with students by creating engaging lessons (Torff & Sessions, 2009). The lack of a significant difference between the two latent class solution of effective and non-effective teachers based upon principal effectiveness scores, and on the use of observed variable mean scores to classify teachers, further supported previous research. The principal score evaluating effective teachers may be a subjective measurement; however, it stands as a successful predictor of effective and non-effective latent classes of teachers in this study.

Characteristics of effective teachers indicated by the observed variables in this study were designed to support themes in current and past literature that indicated quality or teacher effectiveness. Regardless of scale, the TRIPOD, FFT, and CLASS measurements used for the data analysis indicated a higher score as representing a more effective teacher. The components of Ronald Ferguson's TRIPOD instrument were found to yield reliable and valid scores in the MET study and student opinions of teachers (Bill and Melinda Gates Foundation, 2012; Fergurson & Danielson, 2014). It is important to mention that the TRIPOD scores though significant, did not have as great of an effect size as the CLASS and FFT scores in determining the two class solution for effective and non-effective teachers, suggesting that student opinion is not as important as teacher observations when determining characteristics of quality teachers. FFT has been established to also yield valid and reliable scores in both the MET study and through the creator Charlotte Danielson (Bill and Melinda Gates Foundation, 2012, Danielson, 2012; Milanowski, Heneman, & Kimball 2011). Finally, the CLASS measurement by Ronald Pianta has also established reliable and valid scores in the MET study (Bill and Melinda Gates Foundation, 2012). The teacher quality components of the measurements used are what is important and relatable to the literature.

Effective Algebra I teachers in the two latent class solution demonstrated significantly higher scores on the TRIPOD for areas of both student engagement and classroom management, significantly higher scores on the FFT for teacher promoted understanding and classroom management, and significantly higher scores for CLASS in areas of teacher promoted understanding and classroom management for effective teachers. Characteristics of quality teachers reflect effective teachers and promote student success. The strength (effectiveness) of a teacher was directly related to student growth (Hanushek, Piopiunik, & Wiederhold, 2014). The importance of student engagement is amplified in the Algebra I classroom as student dissatisfaction with mathematics content engagement due to ineffective teaching is a major complaint in their ability to learn content (Murray, 2011).

Effective teachers in this study that scored higher on engagement measurements including components of the TRIPOD and CLASS instruments supported the importance of student opinion in classifying effective teachers, which is also supported by teacher observations. Mathematics students feel more engaged with the content due to the level of emotional support teachers exhibited (Martin & Rimm-Kaufman, 2015), a strength measured both by the TRIPOD and CLASS classroom management components. Interestingly the areas of engagement and classroom management demonstrated the greatest effect sizes for Algebra I teachers in this study. Furthermore, the MET study linked FFT components with TRIPOD components demonstrating statistically significant relationships between conceptually compatible areas including the area of classroom management (Ferguson & Danielson, 2014). A study examining 82 secondary Algebra I teachers used the CLASS instrument to evaluate teacher effectiveness (Gitomer, et al., 2014). Results of the observational instrument revealed teachers had the lowest scores in the domain of Instructional Support, middle scores in Emotional Support, and the highest scores in Classroom Management (Gitomer, et al., 2014).

CLASS component scores of Algebra I teachers in this study mimicked results of the Gitomer et. al. (2014) study which showed the understanding component did not play as large a role as the engagement and classroom management components. Clearly, components of classroom management are extremely important in determining teacher effectiveness based upon the literature and based upon the observation that all three measurements used in this study contain sections to gauge the quality of effective teaching through classroom management of which two measures had significant effect sizes.

So how do qualities of effective Algebra I teachers relate to student standardized test scores? They do not. An examination of the Algebra I teachers in this study revealed similar ACT QualityCore ® Algebra I test scores for teachers regardless of latent class assignment and predicted group membership. Characteristics of teacher effectiveness related to practice are what is important. This finding both confirms and opposes different views in the research literature. An examination of three large metropolitan school districts comprising 1,291 ninth grade Algebra I students demonstrated that teachers' practices and collaboration with colleagues had the greatest impact on student achievement on the ACT Quality Core® in Algebra I (Ndum, Allen, & Fisher, 2012). Although Ndum, Allen and Fisher (2012) did find student growth with the student measure used in their study, they did not compare classes of teacher effectiveness; and it is important to note that their study was performed by ACT to gauge a teacher ACT preparation program. Unfortunately teacher merit pay schemes can be responsible for undermining successful teacher collaboration practices (Brewer, Myers, & Zhang, 2015). Different research from a larger study of over 3,500 Algebra I teachers established a relationship between teachers' positively impacting test scores, but also concluded that the test scores were inadequate to determining the effectiveness of Algebra I teachers (Jackson, 2014). The core purpose of examining raw ACT QualityCore® Algebra I test scores for ninth grade students was to remove the use of standardized test scores as a main evaluator for determining effective teachers, here specifically Algebra I teachers.

Implications for Comprehensive Teacher Evaluation Systems

Evaluating teachers is still a controversial but important topic affecting numerous stakeholders and multiple levels in the education field from research to classroom application. Using quality teacher characteristics to implement decisions of teacher effectiveness serves to give a more holistic approach to teacher evaluation and is based upon a more complete picture of the examination of teaching practice. Numerous limitations of using standardized test scores paint an incomplete picture of teacher effectiveness by evaluating teachers on a finite measurement. Although this testing approach is more convenient for researchers and policy makers, the implications of a multidimensional evaluation system for determining indicators and predictors of effective Algebra I teachers would have a greater impact on the field of teaching. The implications of using a latent class model to identify non-effective/effective teachers based upon quality teacher characteristics could greatly impact the field of teacher evaluation. Furthermore, the principal evaluations are strong predictors of determining effective and non-effective teachers. Principals are doing a good job of evaluating their teachers even though a subjective measure. The exploration of quality principal evaluation instruments could play an important future role in teacher evaluation, especially if they are created to target the evaluation of quality teacher characteristics.

A shift from using student standardized test data could adapt the field of teacher evaluation. Using quality teacher characteristics that focus on daily teaching practices such as teacher promoted student engagement, teacher promoted student understanding, and classroom management gives a better picture of what effective teachers do daily. This study indicated that quality teacher characteristics should be used instead of student achievement data, and that to determine classes of effective and non-effective teachers these quality characteristics must be quantifiable. This would create a successful evaluative method pleasing to all stakeholders.

Implications for Improving Teacher Education Programs and Professional Development

Determining indicators and predictors of effective Algebra I teachers is critical for the improvement of teacher education programs and meaningful teacher professional development opportunities. A summary study of the usefulness of professional development approaches for the promotion of effective mathematics teachers revealed that only 2 out of 910 showed positive effects on students' math proficiency (Gersten, Taylor, Keys, Rolfhus, & Newman-Gonchar, 2014). One of the two workshops focused on strengthening teacher content knowledge in mathematics and the other focused on teacher collaboration to create engaging materials to promote student understanding (Gersten, et al., 2014). Both of these topics, increased content knowledge and student engagement are important to the quality teacher characteristics explored in this study. Identifying characteristics of effective Algebra I teachers and effective math teachers will be instrumental in establishing useful professional development workshops.

In regards to the concern of unsuccessful engagement in mathematics content, Algebra I teachers novice or experienced would benefit from learning more engagement strategies to incorporate in their curriculum. Algebra I content by nature is not always exciting to students, and manipulating Algebra I content to make it more engaging would in turn promote student understanding, and perhaps improve classroom management issues. Results from this study indicated the importance of classroom management for effective teaching, which is also an important component of quality teaching that needs to be addressed in both teacher education programs and professional development.

Study Limitations

Limitations of the study relate to the instruments used for analysis, the restrictive nature of the MET data files, and the subset of Algebra I teachers. Psychometric properties of the instruments could use a firmer establishment in the literature and results are not generalizable due to the subset of participants analyzed. The study delimitations also pointed out the restrictions placed on computer analysis and obtaining results. Although much thanks is given to *The Bill and Melinda Gates Foundation* for supporting the MET Study, numerous data restrictions added difficulty to conducting this study.

Directions for Future Research

Interest in other student performance data would add richness to the LCA approach. Student GPAs or summative Algebra I scores might help to support the relationship between the classes of effective teacher's identified with another important measurement. Also, the availability of these measurements for teachers could allow teacher self-assessment strategies for improving teacher performance, if found to be significant.

Follow up examination of student performance data with the ACT QualityCore® Algebra I test could reverse exploration of teacher effectiveness by examining the raw test scores. A reverse approach of examining the top and bottom 25% of student standardized scores and then comparing them with the two LCA solution using chi-square might further reveal information about the use of standardized test scores to predict teacher effectiveness.

It would also be interesting to develop and assess different measurements for evaluating the areas of engagement, student understanding, and classroom management. There seems to be a dynamic relationship between these dimensions which warrants additional exploration.

Exploration of a larger sample of teachers would be of great interest. Testing the LCA model for secondary teachers in different subject areas and then comparing effectiveness of the model across subject areas could reveal important information related to general characteristics of quality teachers, and also perhaps differences amongst teachers with different background content.

In addition to looking at differences in secondary teachers based upon content, it would be interesting to explore teacher effectiveness across student education levels including primary, secondary, and even post-secondary instructors.

Conclusion

This study was limited to the MET study data and Algebra I teachers. The chosen observed variables from the TRIPOD, FFT, and CLASS measurements were used to represent the quality teaching practices of successfully engaging students, promoting understanding of content, and managing a positive classroom environment. These measurements were used to promote a potential teacher evaluation system that dispels the fears of teacher evaluation argued by many opponents including Eisner's (1976) foreboding condemnation of reducing the quality of teaching to a quantity. Even though principal scores are viewed as a subjective measurement in teacher evaluation, this study suggests that principal scores are more evaluative than student test scores. Findings from this study of Algebra I teachers in the MET study, suggests that successfully evaluating teacher effectiveness using teacher characteristics rather than reducing them to student associated test scores, gives hope for the improvement of teacher professional development and teacher education programs.

Current education policy is searching to expand regulations on public education and local school systems are trying to develop ways to increase student achievement by evaluating teacher effectiveness. This is an important issue; however, current evaluation methods are relying on the use of student standardized test scores as the predominant measurement of student growth to indicate effective teachers. This is problematic as it does not provide a holistic evaluation monitoring the multidimensionality of quality teacher characteristics and therefore effective teachers. Fortunately leading organizations, like the *American Educational Research Association* (AERA, 2015), are speaking out to warn of the deleterious effects of teacher evaluation strategies involving high stakes decisions for teacher tenure and employment based upon these types of student growth measures traditionally assessed by HLM. In order to warn the government, AERA issued a statement on the use of VAM models for the evaluation of educators and educator programs cautioning against the use of such models due to psychometric problems and study validity issues (AERA, 2015). The desire for methods of evaluation for teachers and teacher education programs is not subsiding. A hopeful new approach using quality teacher characteristics for evaluating teacher effectiveness might satisfy all stakeholders.

REFERENCES

ACT QualityCore®. (2014). *Technical manual.* v 3.0, Iowa City, IA: Author.

ACT & The Education Trust. (2004). On course for success: A close look at selected high school courses that prepare all students for college. Iowa City, IA: Authors.

AERA (2015). Statement of use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher, 44*(8), 448-452.

Atteberry, Allison, Susanna Loeb, and James Wyckoff. "Do first impressions matter? Improvement in early career teacher effectiveness." *NBER Working Papers* (June 2013):1-53.

Balch, R., & Springer, M. G. (2015). Performance pay, test scores, and student learning objectives. *Economics Of Education Review*, *44,* 114-125.

Beishuizen, J. J., Hof, E., van Putten, C. M., Bouwmeester, S., & Asscher, J. J. (2001). Students' and teachers' cognitions about good teachers. *British Journal of Educational Psychology, 71*(2), 185-202.

Beller, E. K. (1971). Teacher evaluation: Why, what, and how! *Peabody Journal of Education 48*(2), 125-140.

Bennett, C. A. (2013). Teachers' perspectives of whole-class discourse: Focusing on effective instruction to improve student learning. *Action in Teacher Education, 35*(5-6), 475-488.

Bill and Melinda, Gates Foundation. (2010). *Working with teachers to develop fair and reliable measures of effective teaching. MET project.*

Bill and Melinda Gates Foundation. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Policy and Practice Brief. *MET Project.*

Bill and Melinda Gates Foundation. (2013). Ensuring Fair and Reliable Measures of Effective Teaching: Culminating from the MET Project’s Three-Year Study. Policy and Practice Brief. *MET Project.*

Brewer, T. J., Myers, P. S., & Zhang, M. (2015). Islands unto themselves: How merit pay schemes may undermine positive teacher collaboration. *Critical Questions In Education*, *6*(2), 45-54.

Burns, R.B. & Burns, R. S. (2008). *Business Research Methods and Statistics Using SPSS.* Thousand Oaks, CA: Sage Publications Inc.

Cantrell, S. & Kane, T.J. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-study.* Seattle, WA: Bill & Melinda Gates Foundation.

Chaplin, D., Gill, B., Thompkins, A., Miller, H., & Regional Educational, L. M. (2014). *Professional practice, student surveys, and value-added: Multiple measures of teacher effectiveness in the Pittsburgh public schools.* (Research Report REL 2014-024). Retrieved from ERIC database.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cousins, C. (2013, April 24). Debate continues over weight of student test scores in teacher evaluations. *Bangor Daily News*. Retrieved from http://bangordailynews.com/2013/04/24/news/state/debate-continues-over-weight-of-student-test-scores-in-teacher-evaluations/

Danielson, C. (2012). Observing Classroom Practice. *Educational Leadership*, *70*(3), 32-37.

Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA, US: Jossey-Bass.

Di Carlo, M. (2012). How to use value-added measures right. *Educational Leadership, 70*(3), 38-42.

Doyle, D., & Han, J.G. (2012). *Measuring teacher effectiveness: A look "under the hood" of teacher evaluation in 10 sites*. New York: 50CAN; New Haven, CT: ConnCAN; and Chapel Hill, NC: Public Impact. Retrieved from http://www.conncan.org/learn/research/teachers/measuring-teacher-effectiveness

Dziak, J.J, Coffman, D.L., Lanza, S.T., and Li, R. (2012). Sensitivity and specificity of information criteria. (Research Report No 12.29). Retrieved from The Methodology Center at Pennsylvania State University website: https://methodology.psu.edu/media/techreports/12-119.pdf

Eisner, E. W. (1976). Educational connoisseurship and criticism: Their form and functions in educational evaluation. *Journal of Aesthetic Education,* 10(3/4), 135- 150.

Eldar, O., & Miedijensky, S. (2014). Designing a metacognitive approach to the professional development of experienced science teachers. In Alejandro Peña-Ayala (Ed.), *Metacognition: Fundaments, Applications & Trends,* (pp. 299-319). Switzerland: Springer International Publishing.

Engel, M. (2013). Problematic Preferences? A mixed method examination of principals' preferences for teacher characteristics in Chicago. *Educational Administration Quarterly*, *49*(1), 52-91.

Ferguson, R.F & Danielson, C. (2014). How Framework for Teaching and Tripod 7Cs evidence distinguish key components of effective teaching. In T.J.Kane, K.A. Kerr, and R.C. Pianta (Eds.), *Designing teacher evaluation systems.* (pp. 98-143). San Francisco: John Wiley & Sons, Inc.

Gabriel, R., & Allington, R. (2012). THE MET PROJECT: The wrong $45 million question. *Educational Leadership, 70*(3), 44-49.

Gargani, J., & Strong, M. (2014). Can we identify a successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal Of Teacher Education*, (5), 389-402.

Gersten, R., Baker, S.K., Haager, D., and Graves, A. W. (2005). Exploring the role of teacher quality in predicting reading outcomes for first-grade English learners: An observational study. *Remedial & Special Education, 2* (4), 197-206.

Gitomer, D., Bell, C., Yi, Q. I., McCaffrey, D., Hamre, B. K. 2., & Pianta, R. C. (2014). The instructional challenge in improving teaching quality: Lessons from a classroom observation protocol. *Teachers College Record, 116*(6), 1-32.

Glazerman, S., Loeb, S. Goldhaver, D., Staiger, D., Raudenbush, S. and Whitehurst, G. (2010). Evaluating Teachers: The important role of value-added. (Research Report November 17). Retrieved from the Brown Center on Education Policy at Brookings website: http://www.brookings.edu/research/reports/2010/11/17-evaluating-teachers

Godfrey-Smith, P. (2003). *Theory and Reality: An introduction to the philosophy of science.* Chicago, IL: The University of Chicago Press.

Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effective-ness: A research synthesis. (Research Report S283B050051). Retrieved from National Comprehensive Center for Teacher Quality website: http://files.eric.ed.gov/fulltext/ED521228.pdf

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, *42*(4), 765-794.

Goos, M. (2013). Knowledge for teaching secondary school mathematics: What counts? *International Journal of Mathematical Education in Science & Technology, 44*(7), 972-983.

Gordon, R., Kane, T. J., & Staiger, F. O. (2006). Identifying Effective Teachers Using Performance on the Job. *Hamilton Project: Discussion Papers*, (1), 1-35.

Gottlieb, D. (2013). Eisner's evaluation in the age of race to the top. *Curriculum & Teaching Dialogue, 15*(1), 11-25.

Graue, M.E., Delaney, K.K., & Karch, A.S. (2013). Ecologies of education quality. Education Policy Analysis Archives, Vol. 21(8), 1-36.

Green III, P. C., Baker, B. D., & Oluwole, J. (2012). The legal and policy implications of value-added teacher assessment policies. *Brigham Young University Education & Law Journal, 2012*(1), 1-30.

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher*, *43*(6), 293-303

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2009). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education*, *119*(3), 445-470.

Hagenaars, J.A. & McCutcheon, A. (2002). Applied latent class analysis. Cambridge: Cambridge University Press.

Hajdin, G. & PaŽur, K.(2012). Differentiating between student evaluation of teacher and teaching effectiveness. *Journal of Information & Organizational Sciences*, *36*(2), 123-134.

Halpin, P. F., & Kieffer, M. J. (2015). Describing profiles of instructional practice: A new approach to analyzing classroom observation data. *Educational Researcher*, *44*(5), 263-277.

Hamachek, D. (1969). Characteristics of good teachers and implications for teacher education. *Phi Delta Kappa Inc.*, *55*(6), 341-345.

Hamre, B. K., & Pianta, R. C. (2010). Classroom environments and developmental processes: Conceptualization and measurement. In J. Meece & J. Eccles (Eds.), Handbook of research on schools, schooling, and human development (pp. 25- 41). New York, NY: Routledge.

Hanushek, E.A. (1971). “Teacher characteristics and gains in student achievement: Estimation using micro data”. American Economic Review 60 (2), 280–288.

Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2014). The value of smarter teachers: International evidence on teacher cognitive skills and student performance. (Research Report No. 20127). Retrieved from the NBER Working Papers: http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=100008334&site=eds-live&scope=site

Harris, D. (2012, December 4). Creating a valid process for using teacher value-added measures. *Washington Post*. Retrieved from http://www. washingtonpost.com/blogs/answer-sheet/wp/2012/12/04/the-future- of-teacher-evaluation-how-value-added-could-be-used/

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How Teacher Evaluation Methods Matter for Accountability: A Comparative Analysis of Teacher Effectiveness Ratings by Principals and Teacher Value-Added Measures. *American Educational Research Journal*, *51*(1), 73-112.

Hattie, J., & Clinton, J. (2001). The assessment of teachers. *Teaching Education, 12*(3), 279-300.

Henry, G. T., Bastian, K. C., & Fortner, C. K. (2011). Stayers and leavers: Early-career teacher effectiveness and attrition. *Educational Researcher, 40*(6), 271-280.

Henry, G. T., Campbell, S. L., Thompson, C. L., Patriarca, L. A., Luterbach, K. J., Lys, D. B., & Covington, V. M. (2013). The predictive validity of measures of teacher candidate programs and performance: Toward an evidence-based approach to teacher preparation. *Journal of Teacher Education, 64*(5), 439-453.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A Validity Argument Approach to Evaluating Teacher Value-Added Scores. *American Educational Research Journal*, (3), 794-831

Ho, A.D., & Kane, T.J. (2012). *The reliability of classroom observations by school personnel*. MET Project, Bill and Melinda Gates Foundation.

Hubert, T. (2014). Learners of mathematics: High school students' perspectives of culturally relevant mathematics pedagogy. *Journal of African American Studies, 18*(3), 324-336.

ICPSR. (2016) *Membership in ICPSR*. Retrieved from https://www.icpsr.umich.edu/icpsrweb/content/membership/about.html.

Imig, D. G., & Imig, S. R. (2006). The teacher effectiveness movement: How 80 years of essentialist control have shaped the teacher education profession. *Journal of Teacher Education,* (2), 167-180.

Jackson, C. K. (2014). Teacher quality at the high school level: The importance of accounting for tracks. *Journal of Labor Economics, 32*(4), 645-684.

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics, 6*, 801-825.

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *Future of Children, 17*(1), 129-153.

Jacob, B. A., and Lefgren, L., (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics. 26*(1): 101-136.

Jepsen, C. (2005). Teacher characteristics and student achievement: Evidence from teacher surveys. *Journal of Urban Economics, 57*, 302-319.

Johnson, B.L.J. (1997). An organizational analysis of multiple perspectives of effective teaching: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 69-87.

Johnson, C. C., Kahle, J. B., & Fargo, J. D. (2007). Effective teaching results in increased science achievement for all students. *Science Education, 91*(3), 371-383.

Kahle, J. B., & Kronebusch, M. (2003). Science teacher education: From a fractured system to a seamless continuum. *Review of Policy Research*, *20*(4), 585-602.

Kane, T. J., McCaffrey, D. F., Miller, T., Staiger, D. O., & Bill and Melinda Gates, F. (2013). Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Research Paper. MET Project. *Bill & Melinda Gates Foundation*.

Kane, T.J., McCaffrey, D. F, & Staiger, D.O. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching Project.* Seattle, WA: Bill & Melinda Gates Foundation.

Kane, T.J., McCaffrey, D. F, & Staiger, D.O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Bill & Melinda Gates Foundation.

Kane, T.J., and Staiger, D.O. (2012). *Gathering Feedback for Teaching: Combining high quality observations with student surveys and achievement gains.* The Bill and Melinda Gates Foundation.

Klecka, W.R. (1980). *Discriminant Analysis*. Newbury Park: Sage Publications.

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Review: Value-added modeling: A review. *Economics Of Education Review, 47*, 180-195.

Kupermintz, H., Shepard, L., & Linn, R. (2001, April). *Teacher effects as a measure of teacher effectiveness: Construct validity considerations in TVAAS (Tennessee value added assessment system).* Paper presented in: D.Koretz (Chair),New Work on the Evaluation of High-Stakes Testing Programs. Symposium at the National Council on Measurement in Education (NCME) Annual Meeting, Seatle, WA.

Lazarev V., Newman D. (2014, March). Can Multifactor Models of Teaching Improve Teacher Effectiveness Measures? In. Corcoran S. (Chair), Policy Considerations in the Implementation of MMTES. Paper session conducted at The Association for Education Finance and Policy, San Antonio, TX. Available at SSRN: http://ssrn.com/abstract=2493544

Leinhardt, G., Wise, A. E., Rand Corporation., & National Institute, o. E. (1984). Modeling and measuring educational treatment in evaluation. *Review of Educational Research, 50*(2), 393-420.

Levine, M., & Levine, A. (2013). Holding accountability accountable: A cost–benefit analysis of achievement test scores. *American Journal of Orthopsychiatry*, *83*(1), 17-26.

Little, O., Goe, L., Bell, C., & National Comprehensive Center for,Teacher Quality. (2009). *A practical guide to evaluating teacher effectiveness* National Comprehensive Center for Teacher Quality.

Little, R.J.A. and Schemker, N. (1995). Missing Data. In Gehard Arminger, Clifford C.Clogg, and Michael E. Sobel (Eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences,* (pp. 39-76). New York: Plenum Press.

Lubke G., Neale M. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. Multivariate Behavioral Research, 43, 592–620.

MacCalla, N. M. (2014) *Evaluating teacher effectiveness in a professional development program: Considering measures for inclusion in a comprehensive teacher evaluation system* (Doctoral Dissertation). Retrieved from eScholarship (UCLA: Education 0249).

Mangiante, E. M. S. (2011). Teachers matter: Measures of teacher effectiveness in low-income minority schools. *Educational Assessment, Evaluation and Accountability, 23*(1), 41-63.

Martin, D. P., & Rimm-Kaufman, S. (2015). Do student self-efficacy and teacher-student interaction quality contribute to emotional and social engagement in fifth grade math? *Journal of School Psychology, 53*, 359-373.

Medley, D. (1971, May 11). *Indicators and Measures of Teacher Effectiveness: A Review of the Research*. Lecture presented at the University of Virginia.

Mihaly, K., McCaffrey, D.F., Staiger, D.O., & Lockwood, J.R. (2013). A composite estimator of effective teaching. Measures of Effective Teaching Longitudinal Database. Retrieved from http://www.icpsr.umich.edu/files/METLDB/MET_Composite_Estimator_of_Effective_Teaching_Research_Paper.pdf

Milanowski, A. T., Heneman, H. I., Kimball, S. M., & Wisconsin Center for Education Research. (2011). Teaching Assessment for Teacher Human Capital Management: Learning from the Current State of the Art. (Working Paper No. 2011-2). Retrieved from *Wisconsin Center For Education Research*: http://files.eric.ed.gov/fulltext/ED517293.pdf

Miller, K. W., & Davison, D. M. (2006). What makes a secondary school science and/or mathematics teacher "highly qualified?". *Science Educator, 15*(1), 56-59.

Moss, P.A., Schutz, A.M., and Collins, K.M. (1998). An integrative approach to portfolio evaluation for teacher licensure. *Journal of Personnel Evaluation in Education, 12*(2), 139-161.

Mplus 7.4©. (2015). How to calculate the power to detect that a parameter is different from zero. Retrieved from http://www.statmodel.com/power.shtml.

Muñoz, M., Scoskie, J., & French, D. (2013). Investigating the 'black box' of effective teaching: the relationship between teachers' perception and student achievement in a large urban district. *Educational Assessment, Evaluation & Accountability*, *25*(3), 205-230.

Murray, S. (2011). Secondary students' descriptions of "good" mathematics teachers. *Australian Mathematics Teacher, the*, *67*(4), 14-21.

Muthén B. O., Muthén L. (2014). Mplus 7.3 [computer software]. Los Angeles, CA: Muthén & Muthén.

Ndum, E., Allen, J., Fisher, T., & ACT, I. (2012). *Examining the relationship of QualityCore[R] implementation and student achievement in algebra I. ACT research report series, 2012 (4)*, 1-60.

Nebesniak, A. L. (2012). Effective instruction: A mathematics coach's perspective. *Mathematics Teacher, 106*(5), 354-358.

Ng, W., Nicholas, H., & Williams, A. (2010). School experience influences on pre-service teachers' evolving beliefs about effective teaching. *Teaching and Teacher Education, 26*(2), 278-289.

O'Donnell, B. (2009). What effective math teachers have in common. *Teaching Children Mathematics, 16*(2), 118-125.

Patrick, J., & Smart, R. M. (1998). An empirical evaluation of teacher effectiveness: The emergence of three critical factors. *Assessment & Evaluation in Higher Education, 23*(2), 165-179.

Peterson, K.D., Wahlquist, C., & Bone, K. (2000). Student surveys for school teacher evaluation *Journal of Personnel Evaluation in Education, 14*(2), 135-53.

Pianta, R.C., Hamre, B.K., Hayes, N., Mintz, S., and LaParo, K.M. (2008). *Classroom Assessment Scoring System-Secondary (CLASS-S)* Charlottesville, VA: University of Virginia.

Polikoff, M. S. (2015). The stability of observational and student survey measures of teaching effectiveness. *American Journal Of Education*, *121*(2), 183-212.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. Marketing Science, 12(1), 103−124.

Rebora, A. (2010). *Teachers beg to differ on effectiveness measures.* Teacher Professional Development Sourcebook, Bethesda: Editorial Projects in Education, Inc.

Rice, J.K. (2010). The impact of teacher experience: Examining the evidence and policy implications. (Issue Brief No. 11). *National Center for Analysis of Longitudinal Data in Education Research.*

Rockoff, J.E., and Speroni, C. (2011). Subjective and objective evaluations of teacher effectiveness: Evidence from New York City. *Labour Economics, 18 (5),* 687-696.

Rosenberg, A. (2008). *Philosophy of Social Science*. (3rd ed.). Boulder, CO: Westview Press.

Rubin, D.B. (1976). Inference in missing data. *Biometrika, 70,* 41-55.

Rowan, B., Harrison, D., & Hayes, A. (2004). Using instructional logs to study mathematics curriculum and teaching in the early grades. Elementary School Journal, 105(1), 103– 127.

Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement. *Journal of Personnel Evaluation in Education 12*(3), 247-256.

SAS Institute, Cary NC, (2012). The [power analysis] for this paper was generated using SAS software. Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Schafer, W. D., Lissitz, R. W., Zhu, X., Zhang, Y., Hou, X., & Li, Y. (2012). Evaluating teachers and schools using student growth models. *Practical Assessment, Research & Evaluation, 17*(16), 1-21.

Scherrer, J. (2013). *Investigating the relationship between "effective" teachers and theoretical notions of effective teaching: An analysis of whole-group discussions.* (Doctoral dissertation, University of Pittsburgh). Retreived from http://d-scholarship.pitt.edu/19570/1/Scherrer,_Jimmy,_investigating_the_relationship_between_effective_teachers_and_theoretical_notions_of_effective_teaching_EDT_Pitt2013.pdf

Schumacker, R.E. & Lomax, R.G. (2016). A Beginner's Guide to Structural Equation Modeling. 4rd Ed. New York, NY: Routledge Taylor & Francis Group.

Schumacker, R. E. and Pugh, J. (2013). Identifying reading and math performance in school systems with latent class longitudinal growth modeling. *Journal of Educational Research and Policy*, 13 (3), 51- 62.

Searles, W. E., & Kudeki, N. (1987). A comparison of teacher and principal perception of an outstanding science teacher. *Journal of Research in Science Teaching*, *24*(1), 1-13.

Strickland Jr., J. F., & Page, J. A. (1991). Qualities exhibited by effective mathematics teachers: Perceptions of preservice and inservice. *Journal of Instructional Psychology, 18*(4), 250-255.

Stronge, J.H., Ward, T.J., and Grant, L.W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievements. *Journal of Teacher Education*, *62*(4), 339-355.

Tairab, H. H., & Wilkinson, W. J. (1991). Evaluation of biology trainee teachers' effectiveness using their students' perceptions of teaching behavior. *Research in Science & Technological Education*, *9*(1), 51-62.

The Danielson Group. (2015). *The Framework.* Retrieved from https://www.danielsongroup.org/framework/

The New Teacher Project. (2010). *Teacher Evaluation 2.0.* Retrieved from http://tntp.org/assets/documents/Teacher-Evaluation-Oct10F.pdf

Teacher quality [electronic resource]: sustained coordination among key federal education programs could enhance state efforts to improve teacher quality : report to the Chairman of the Subcommittee on Higher Education, Lifelong Learning, and Competitiveness, Committee on Education and Labor, House of Representatives. (2009). [Washington, D.C.]: U.S. Govt. Accountability Office, [2009].

Torff, B., & Sessions, D. (2009). Principals' perceptions of the causes of teacher ineffectiveness in different secondary subjects. *Teacher Education Quarterly, 36*(3), 127-148.

Van Tassel-Baska, J., Quek, C. and Feng, A.X. (2006). The development and use of a structured teacher observation scale to assess differentiates best practice. *Roeper Review*, *29*(2), 84-92.

Watson, S. (2006). Novice science teachers: Expectations and experiences. *Journal of Science Teacher Education*, *17*(3), 279-290.

Webster, W. J., Mendro, R. L., Orsak, T. H., & Weerasinghe, D. (1998, April). An application of hierarchical linear modeling to the estimation of school and teacher effects. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Weisberg, D, Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect. The New Teacher Project. Available: http://widgeteffect.org/downloads/TheWidgetEffect.pdf

Wright, S. P, Horn, S.P., and Sanders, W.L. (1997). Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation. Journal of Personnel Evaluation in Education, 11(1), 57-67.

Appendix A: IRB Certification

Office for Research

Institutional Review Board for the Protection of Human Subjects

THE UNIVERSITY OF ALABAMA

R E S E A R C H

July 14, 2015

Lauren Holmes
Dept. of ESPRMC
College of Education
Box 870231

Re: IRB#: 15-OR-222 "Holmes Dissertation"

Dear Ms. Holmes:

The University of Alabama Institutional Review Board has granted approval for your proposed research.

Your application has been given expedited approval according to 45 CFR part 46. Approval has been given under expedited review category 5 as outlined below:

(5) Research involving materials (data, documents, records or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis)

Your application will expire on July 13, 2016. If your research will continue beyond this date, complete the relevant portions of the IRB Renewal Application. If you wish to modify the application, complete the Modification of an Approved Protocol Form. Changes in this study cannot be initiated without IRB approval, except when necessary to eliminate apparent immediate hazards to participants. When the study closes, complete the appropriate portions of the IRB Request for Study Closure Form.

Should you need to submit any further correspondence regarding this proposal, please include the above application number.

Good luck with your research.

Sincerely,

358 Rose Administration Building
Box 870127
Tuscaloosa, Alabama 35487-0127
(205) 348-8461
FAX (205) 348-7189
TOLL FREE (877) 820-3066

Appendix B: Agreement for Use of Confidential Data Form

Agreement for the Use of Confidential Data
from the
Measures of Effective Teaching Longitudinal Database
at the
Inter-university Consortium for Political and Social Research (ICPSR)

I. DEFINITIONS

A. "Investigator" is the person primarily responsible for analysis and other use of Confidential Data obtained through this Agreement.

B. "Research Staff" are persons authorized by the Investigator's institution, excluding the Investigator, who will have access to Confidential Data obtained through this Agreement. Research Staff include project staff or students conducting dissertation or thesis research.

C. "Participants" are persons, other than Investigator and Research Staff, who will be provided access to Confidential Data by the Investigator. For example, research subjects who will view videos included in the Confidential Data as part of an IRB approved research protocol are Participants for this agreement. Institution is responsible for ensuring Participant compliance with all aspects of this agreement.

D. "Institution" is the university or research institution at which the Investigator will conduct research using Confidential Data obtained through this Agreement.

E. "Representative of the Institution" is a person authorized to enter into contractual agreements on behalf of Investigator's Institution.

F. "Confidential Data" consist of data, images, videos and any objects derived from them with information that is linkable to a specific individual either directly or indirectly, and for which the individual (whether a person or organization) has the expectation that the information will not be released in a manner allowing public identification of the individual or causing some harm to the individual.

G. "Private Person" means any individual (including an individual acting in his official capacity) and any private (i.e., non-government) partnership, corporation, association, organization, or entity (or any combination thereof), including family, household, school, neighborhood, health service, or institution.

H. "ICPSR" is the Inter-university Consortium of Political and Social Research.

I. "Restricted Data Contracting System" ("RDCS") is the web-based system for data contracts at ICPSR.

J. "Data Security Plan" is a component of this Agreement, found as Attachment A, which specifies permissible computer configurations for use of Confidential Data through Investigator responses to a series of questions, and records what the Investigator commits to do in order to keep Confidential Data secure.

K. "Deductive Disclosure" is the discerning of an individual's identity or confidential information through the use of known characteristics of that individual. Disclosure risk is present if an unacceptably narrow estimation of an individual's confidential information is possible or if determining the exact attributes of the individual is possible with a high level of confidence.

L. "Derivative" is a file, video, image, or statistic derived from the Confidential Data that poses disclosure risk to any Private Person in the Confidential Data obtained through this Agreement. Derivatives include copies of the Confidential Data received from ICPSR, subsets of the Confidential Data, and analysis results that do not conform to the guidelines in Section VI.G.

II. DESCRIPTION OF DISCLOSURE

Deductive disclosure of an individual's identity from research material is a major concern of federal agencies, researchers, and Institutional Review Boards. If a person is known to have participated in ANY study or if information is known to be included in files or a database from which the Confidential Data were obtained, then a combination of his or her personal characteristics may allow someone to determine which record corresponds to that individual. Investigators and Institutions who receive any portion of Confidential Data are obligated to protect the individual's confidential information from deductive disclosure risk by strictly adhering to the obligations set forth in this Agreement and otherwise taking precautions to protect the Confidential Data from non-authorized use.

III. REQUIREMENTS OF INVESTIGATORS

A. Investigators must meet the following criteria:

1. Have a PhD or other terminal degree; and

2. Hold a faculty appointment or research position at Institution.

B. The Investigator assumes the responsibility of completing the RDCS online application and required documents, reports, and amendments. The Investigator agrees to responsibly manage and use Confidential Data and implement all Confidentiality Data security procedures per the Data Security Plan.

C. The Investigator will provide ICPSR any publications or public presentations derived from the Confidential Data.

IV. REQUIREMENTS OF INSTITUTION

The Institution must meet the following criteria:

A. Be an institution of higher education, a research organization, a research arm of a government agency, or a nongovernmental, not for profit, agency.

B. Have a demonstrated record of using Confidential Data according to commonly accepted standards of research ethics and applicable statutory requirements.

V. OBLIGATIONS OF ICPSR

In consideration of the promises made in Section VI of this Agreement, ICPSR agrees to:

A. Provide access to the Confidential Data requested by the Investigator in the Confidential Data Order within a reasonable time of execution of this Agreement by appropriate ICPSR officials. Quantitative Confidential Data will be made available via the Virtual Data Enclave, a secure remote-access work space. Video files and accompanying metadata will be made available via the MET LDB online secure streaming system. Access to both requires proper authentication. ICPSR will provide instructions on establishing user accounts within a reasonable amount of time after the execution of the agreement.

B. Provide electronic documentation of the origins, form, and general content of the Confidential Data, in the same time period and manner as the Confidential Data.

ICPSR MAKES NO REPRESENTATIONS NOR EXTENDS ANY WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED. THERE ARE NO EXPRESS OR IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, OR THAT THE USE OF THE CONFIDENTIAL DATA WILL NOT INFRINGE ANY PATENT, COPYRIGHT, TRADEMARK, OR OTHER PROPRIETARY RIGHTS. Unless prohibited by law, Investigator and Institution assume all liability for claims for damages against them by third parties that may arise from the use or disclosure of the Confidential Data.

VI. OBLIGATIONS OF INVESTIGATOR, RESEARCH STAFF, AND INSTITUTION

Confidential Data provided under this Agreement shall be accessed by the Investigator, Research Staff, Participants, and Institution in strictest confidence and can be disclosed only in compliance with the terms of this Agreement. In consideration of the promises in Section V of this Agreement, and for use of Confidential Data from ICPSR, the Investigator, Research Staff, Participants, and Institution agree:

A. That the Confidential Data will be used solely for research or statistical purposes relative to the research project identified on the Application for Obtaining Confidential Data accompanying this Agreement, and for no other purpose whatsoever without the prior consent of ICPSR. Further, no attempt will be made to identify private persons, no Confidential Data of private person(s) will be published or otherwise distributed, and Confidential Data will be protected against deductive disclosure risk by strictly adhering to the obligations set forth in this Agreement and otherwise taking precautions to protect the Confidential Data from non-authorized use.

B. To supply ICPSR with a completed RDCS online Application for Obtaining Confidential Data that will include the following:

1. A signed Agreement

2. A Research Plan describing inquiry and publications consistent with the objectives of the Measuring Effective Teaching Project to advance knowledge about effective teachers and teaching.

3. Confidential Data Order Summary specifying which files and documentation are requested

4. A copy of a document signed by the Institution's Institutional Review Board (IRB) approving or exempting the research project

from key variables. For example, in no table should all cases in any row or column be found in a single cell.

4. No release of the statistic if the total, mean, or average is based on fewer cases than the appropriate threshold as determined from the sample characteristics.

5. No release of the statistic if the contribution of a few observations dominates the estimate of a particular cell. For example, in no instance should the quantity figures be released if one case contributes more than 60 percent of the quantity amount.

6. No release of data that permits disclosure when used in combination with other known data. For example, unique values or counts below the appropriate threshold for key variables in the Confidential Data that are continuous and link to other data from ICPSR or elsewhere.

7. No release of minimum and maximum values of identifiable characteristics (e.g., income, age, household size, etc.) or reporting of values in the "tails," e.g., the 5^{th} or 95^{th} percentile, from a variable(s) representing highly skewed populations.

8. Release only weighted results if specified in the data documentation.

9. No release of ANOVAs and regression equations when the analytic model that includes categorical covariates is saturated or nearly saturated. In general, variables in analytic models should conform to disclosure rules for descriptive statistics (e.g., see #7 above) and appropriate weights should be applied.

10. In no instance should data on an identifiable case, or any of the kinds of data listed in preceding items 1-9, be derivable through subtraction or other calculation from the combination of tables released.

11. No release of sample population information or characteristics in greater detail than released or published by the researchers who collected the Confidential Data. This includes but is not limited to publication of maps.

12. No release of anecdotal information about a specific private person(s) or case study without prior approval.

13. The above guidelines also apply to charts as they are graphical representations of cross-tabulations. In addition, graphical outputs (e.g., scatterplots, box plots, plots of residuals) should adhere to the above guidelines.

I. To mitigate the risk of disclosing identities or private information derived from the Confidential Information by following practices that include, but are not limited to the following:

1. No streaming video from the Confidential Data may be captured on any computer or other medium.

2. No excerpts, images or other derivatives from the Confidential Data may be published or disseminated in any way.

3. No descriptions of individuals, activities, environments, or other aspects of the Confidential Data may be released in a way that would lead to identification of individuals. Information about

objects in the Confidential Data (such as school, grade, subject) may not be included in presentations or publications if they may increase the risk of disclosure. Special care should be used in describing attributes of individuals that in combination might uniquely identify an individual, such as school, grade, age, race, gender, "gifted," "special education," "English language learner," or physical attributes (height, weight, hair color, etc.).

4. No anecdotal descriptions or verbatim transcripts may be released if they can be linked to information that increases the risk of identification of individuals.

5. No information from quantitative and video objects in the Confidential Data may be linked for the purpose of identifying individuals.

6. No identifying information revealed by individuals depicted in the Confidential Data may be recorded in any way. For example, names of persons, places, or events written on blackboards or spoken by an individual may not be written on paper or typed into a computer document. This type of information may never be released in public presentations or publications. If there is any doubt about whether a research note may pose a disclosure risk, it should be created within the Virtual Data Enclave.

J. That if the identity of any private person should be discovered, then:

1. No use will be made of this knowledge;

2. ICPSR will be advised of the incident within five (5) business days of discovery of the incident;

3. The information that would identify the private person will be safeguarded or destroyed as requested by ICPSR; and

4. No one else will be informed of the discovered identity..

K. Unless other provisions have been made with ICPSR, all access to the Confidential Data will be terminated on or before completion of this Agreement or within five (5) days of written notice from ICPSR. Investigators requiring access to the Confidential Data beyond completion of this Agreement should submit a request for continuation three months prior to the end date of the Agreement.

L. To ensure that the Confidential Data are managed and used in compliance with the terms and conditions of this Agreement and with all applicable statutes and regulations. Noncompliance with this Agreement by any Research Staff or Participant hereto shall be deemed noncompliance and a breach by Investigator and Institution for purposes of section VII below.

M. To notify ICPSR of a change in institutional affiliation of the Investigator. Notification must be in writing and must be received by ICPSR at least six (6) weeks prior to Investigator's last day of employment with Institution. Investigator's separation from Institution terminates this Agreement. Investigator may reapply for access to Confidential Data as an employee of the new institution. Re-application requires:

1. Execution of a new Agreement for the Use of Confidential Data by both the Investigator and the proposed new institution;

2. Execution of any Supplemental Agreement(s) with Research Staff and Pledges of Confidentiality by Research Staff and Participants at the proposed new institution;

3. Preparation and approval of a new Data Security Plan; and

4. Evidence of approval or exemption by the proposed new institution's IRB.

These materials must be approved by ICPSR before Confidential Data or any derivatives or analyses may be accessed at the new institution.

N. That if the Investigator who is changing institutions does not have the new agreement executed by the time they leave their institution, ICPSR will temporarily deactivate the Investigator's account but will maintain the Investigator's profile to save their work during the transition. Upon approval of the new RDCS online application, ICPSR will reactivate the Investigator's account. If a new agreement is not executed within three (3) month, the Investigator's account will be deleted.

O. That any books, articles, conference papers, theses, dissertations, reports, or other publications that employed the Confidential Data or other resources provided by ICPSR reference the bibliographic citation provided by ICPSR in the study description.

P. That use of the Confidential Data will be consistent with the Institution's policies regarding scientific integrity and human subjects research.

Q. To respond fully and in writing within ten (10) working days after receipt of any written inquiry from ICPSR regarding compliance with this Agreement.

VII. VIOLATIONS OF THIS AGREEMENT

A. The Institution will treat allegations by ICPSR or other parties of violations of this Agreement as allegations of violations of its policies and procedures on scientific integrity and misconduct. If the allegations are confirmed, the Institution will treat the violations as it would violations of the explicit terms of its policies on scientific integrity and misconduct.

B. In the event Investigator or Institution breaches any provision of this Agreement, they shall be jointly and severally responsible to promptly cure the breach and mitigate any damages. Investigator and Institution hereby acknowledge that any breach of the confidentiality provisions herein may result in irreparable harm to ICPSR not adequately compensable by money damages. Investigator and Institution hereby acknowledge the possibility of injunctive relief in the event of breach, in addition to money damages. In addition, ICPSR may:

1. Terminate this Agreement upon notice and terminate access to the Confidential Data and any derivatives thereof;

2. Deny Investigator future access to Confidential Data; and/or

3. Report the inappropriate use or disclosure to the appropriate federal and private agencies or foundations that fund scientific and public policy research.

C. Institution agrees, to the extent permitted under the law, to indemnify, defend, and hold harmless The University of Michigan, ICPSR, RAND Corporation, Bill & Melinda Gates Foundation, and the sources of Confidential Data from any or all claims and losses accruing to any person, organization, or other legal entity as a result of Investigator's, Research Staff's, Participant's, and/or Institution's acts, omissions, or breaches of this Agreement.

VIII. CONFIDENTIALITY

The Institution is considered to be a contractor or cooperating agency of ICPSR;, as such, the Institution, the Investigator, and Research Staff are authorized to protect the privacy of the individuals who are the subjects of the Confidential Data by withholding their identifying characteristics from all persons not connected with the conduct of the Investigator's research project. Identifying characteristics are considered to include those data defined as confidential under the terms of this Agreement.

IX. INCORPORATION BY REFERENCE

All parties agree that the following documents are incorporated into this Agreement by reference:

A. The Application for Obtaining Confidential Data

B. A copy of the Institution's IRB approval or exemption of the Research Project

C. The Data Security Plan proposed by the Investigator and approved by ICPSR

X. MISCELLANEOUS

A. All notices, contractual correspondence, and return of data under this Agreement on behalf of the Investigator shall be made in writing and delivered to the address below:

MET Longitudinal Database Restricted Data Manager
ICPSR
P.O. Box 1248
Ann Arbor, MI 48106-1248

B. This agreement shall be effective for 24 months from execution.

C. The respective rights and obligations of ICPSR and Investigator, Research Staff, and Institution pursuant to this Agreement shall survive termination of the Agreement.

D. This Agreement may be amended or modified only by the mutual written consent of the authorized representatives of ICPSR and Investigator and Institution. Investigator's research project, Data Security Plan, Research Staff, or Participants may be amended or modified only by submitting such amendments or modifications to the RDCS and receiving approval from the authorized representatives of ICPSR. This Agreement may be extended only by submitting an extension request to the RDCS and receiving approval from the authorized representatives of ICPSR. Investigator and Institution agree to amend this Agreement to the extent necessary for ICPSR to comply with the requirements of any applicable regulatory authority.

E. The persons signing this Agreement have the right and authority to execute this Agreement, and no further approvals are necessary to create a binding agreement.

F. The obligations of Investigator, Research Staff, Participants, and Institution set forth within this Agreement may not be assigned or otherwise transferred without the express written consent of ICPSR.

Appendix C: MET Data Files

MET Data File	File #	Data Type	Variable	N
ICPSR Study # 344414 Core Files	34414-0001	District file	Principal Survey	152
	34414-0002	Teacher file	Teacher Study Status, Teacher personal/professional background, Principal Ratings of Teacher Effectiveness	152
	34414-0003	Class Section file	Student Survey Aggregates	603
	34414-0004	Student file	Student Test Scores, Student Survey	152
	34414-0005	CLASS file	Observation Scores	152
	34414-0006	FFT file	Observation Scores	152

ICPSR # 3868 Item-Level Supplemental Test Files	34868-0002	ACT QualityCore ® -Year2	Student Algebra I Scores	480

ICPSR Study # 34346 Item-Level Supplemental Observational Scores	34346-0002	CLASS, Year 2	Item-Level Observational Scores, Phase 2, 9th	1816 videos
	34346-0004	FFT, Year 2	Item-Level Observational Scores, Phase 2, 9th	908 videos

Appendix D: ACT QualityCore® Algebra I

ACT Course Standards
Algebra I

A set of empirically derived course standards is the heart of each QualityCore® mathematics course. The ACT Course Standards represent a solid evidence-based foundation in mathematics. They were developed from an intensive study of high-performing high schools with significant minority and low-income enrollments that produced many graduates who met or exceeded ACT College Readiness Benchmark Scores (See http://www.act.org/path/policy/reports/success.html).

This document contains a list of ACT Course Standards for a rigorous Algebra I course—what students should know and be able to do in the course—and a worksheet teachers can use to compare their course content to these standards. The ACT standards encompass the following overarching themes and/or foundational concepts:

A. Prerequisites

B. Exploring the Skills and Strategies Underlying Mathematics

C. Establishing Number Sense and Operations Skills

D. Exploring Expressions, Equations, and Functions in the First Degree

E. Exploring Quadratic Equations and Functions

F. Exploring Advanced Functions

G. Organizing and Analyzing Data and Applying Probability

ACT Course Standards—Algebra I

A. Prerequisites
1. Skills Acquired by Students in a Previous Course and Refined in This Course
a. Set up and solve problems following the correct order of operations (including proportions, percent, and absolute value) with rational numbers (integers, fractions, decimals)
b. Find the greatest common factor and least common multiple of a set of whole numbers
c. Use rational numbers to demonstrate knowledge of additive and multiplicative inverses
d. Simplify ratios
e. Use scientific notation when working with very large or very small quantities
f. Add, subtract, multiply, and divide rational numbers, including integers, fractions, and decimals, without calculators
B. Exploring the Skills and Strategies Underlying Mathematics
1. Mathematical Processes Learned in the Context of Increasingly Complex Mathematical and Real-World Problems (Note: These mathematical processes are the same for Algebra I, Geometry, Algebra II, and Precalculus.)
a. Apply problem-solving skills (e.g., identifying irrelevant or missing information, making conjectures, extracting mathematical meaning, recognizing and performing multiple steps when needed, verifying results in the context of the problem) to the solution of real-world problems

1

b.	Use a variety of strategies to set up and solve increasingly complex problems
c.	Represent data, real-world situations, and solutions in increasingly complex contexts (e.g., expressions, formulas, tables, charts, graphs, relations, functions) and understand the relationships
d.	Use the language of mathematics to communicate increasingly complex ideas orally and in writing, using symbols and notations correctly
e.	Make appropriate use of estimation and mental mathematics in computations and to determine the reasonableness of solutions to increasingly complex problems
f.	Make mathematical connections among concepts, across disciplines, and in everyday experiences
g.	Demonstrate the appropriate role of technology (e.g., calculators, software programs) in mathematics (e.g., organize data, develop concepts, explore relationships, decrease time spent on computations after a skill has been established)
h.	Apply previously learned mathematical concepts in more advanced contexts
C. Establishing Number Sense and Operation Skills	
1. Foundations	
a.	Evaluate and simplify expressions requiring addition, subtraction, multiplication, and division with and without grouping symbols
b.	Translate real-world problems into expressions using variables to represent values
c.	Apply algebraic properties (e.g., commutative, associative, distributive, identity, inverse, substitution) to simplify algebraic expressions
d.	Add and subtract polynomials
e.	Factor a monomial from a polynomial
f.	Multiply monomials, binomials, trinomials, and polynomials
D. Exploring Expressions, Equations, and Functions in the First Degree	
1. Expressions, Equations, and Inequalities	
a.	Solve single-step and multistep equations and inequalities in one variable
b.	Solve equations that contain absolute value
c.	Solve formulas for a specified variable
d.	Write and graph linear equations and inequalities from real-world situations (e.g., a constant-rate distance/time problem)
e.	Write linear equations in standard form and slope-intercept form when given two points, a point and the slope, or the graph of the equation
f.	Identify, formulate, and obtain solutions to problems involving direct and inverse variation
g.	Solve systems of two equations using various methods, including elimination, substitution, and graphing with and without technology
2. Graphs, Relations, and Functions	
a.	Graph linear inequalities in one variable on the real number line to solve problems
b.	Give the domain and range of relations and functions
c.	Evaluate functions at given values

d.	Identify graphs of relations and functions and analyze them to determine whether a relation is a function (e.g., vertical line test)
e.	Graph linear inequalities with two variables on the standard (x,y) coordinate plane
f.	Use the terminology associated with the Cartesian plane in describing points and lines
g.	Recognize the concept of slope as a rate of change and determine the slope when given the equation of a line in standard form or slope-intercept form, the graph of a line, two points, or a verbal description
h.	Graph a linear equation using a table of values, x- and y-intercepts, slope-intercept form, and technology
i.	Translate between different representations of relations and functions: graphs, equations, sets of ordered pairs, verbal descriptions, and tables
E. Exploring Quadratic Equations and Functions	
1. Equations and Inequalities	
a.	Factor perfect square trinomials and the difference of two squares
b.	Factor trinomials in the form $ax^2 + bx + c$
c.	Solve quadratic equations using multiple methods, including graphing, factoring, and the square root principle
2. Graphs, Relations, and Functions	
a.	Identify graphs of quadratic functions
b.	Relate factors, solutions (roots), zeros of related functions, and x-intercepts in equations that arise from quadratic functions
F. Exploring Advanced Functions	
1. Rational and Radical Expressions, Equations, and Functions	
a.	Use properties of exponents (including zero and negative exponents) to evaluate and simplify expressions
b.	Evaluate and simplify rational expressions
c.	Add, subtract, multiply, and divide rational expressions
d.	Find rational number square roots (without calculators) and approximate irrational square roots (with and without calculators)
e.	Evaluate and simplify radical expressions
f.	Multiply radical expressions
g.	Simplify an algebraic quotient by rationalizing an irrational monomial denominator
G. Organizing and Analyzing Data and Applying Probability	
1. Data Relations, Probability, and Statistics	
a.	Identify the effect on mean, median, mode, and range when a set of data is changed
b.	Interpret data from line, bar, and circle graphs, histograms, scatterplots, box-and-whisker plots, stem-and-leaf plots, and frequency tables to draw inferences and make predictions
c.	Identify arithmetic sequences and patterns in a set of data
d.	Identify patterns of growth (e.g., patterns of exponential growth) in a set of data
e.	Find the probability of a simple event

f.	Distinguish between independent and dependent events
g.	Identify an approximate line of best fit to model data and make predictions
h.	Identify the most efficient way to display data

Appendix E: MET Student TRIPOD Survey

Class Number

Place student barcode label here.

Student Perception Survey for Secondary Students

Dear Student,

Thank you for participating in this survey. While answering the questions, it is important that you think about your experiences in a specific classroom. The proctor of the survey will tell you the classroom/teacher you should think about. If they have not done so already, please ask.

Your teacher and your principal will not look at your answers. Later, someone from outside of your school will tell your teacher and your principal how the students in your school responded, but not how you or any one individual student answered. Please answer what you really think and feel. You do not have to answer any question that you do not want to answer.

	Totally Untrue	Mostly Untrue	Some-what	Mostly True	Totally True
1. I like the ways we learn in this class.	○	○	○	○	○
2. Students get to decide how activities are done in this class.	○	○	○	○	○
3. My teacher wants us to share our thoughts.	○	○	○	○	○
4. In this class we have to think hard about the writing we do.	○	○	○	○	○
5. Because of my teacher, I think more about going to college.	○	○	○	○	○
6. My teacher asks questions to be sure we are following along when s/he is teaching.	○	○	○	○	○
7. My teacher asks students to explain more about answers they give.	○	○	○	○	○
8. I think we get more homework in this class than kids in other classes.	○	○	○	○	○
9. Everybody knows what they should be doing and learning in this class.	○	○	○	○	○
10. My teacher in this class makes me feel that s/he really cares about me.	○	○	○	○	○
11. Our class stays busy and doesn't waste time.	○	○	○	○	○
12. I get nervous in this class.	○	○	○	○	○
13. My teacher makes us think first, before s/he answers our questions.	○	○	○	○	○
14. My teacher knows when the class understands, and when we do not.	○	○	○	○	○
15. For a new student, this class would be a good one to join.	○	○	○	○	○
16. In this class, I take it easy and do not try very hard to do my best.	○	○	○	○	○
17. If I need help with homework, there is someone at home who can help me.	○	○	○	○	○
18. If you don't understand something, my teacher explains it another way.	○	○	○	○	○
19. My teacher makes learning enjoyable.	○	○	○	○	○
20. Being in this class makes me feel angry.	○	○	○	○	○

	Totally Untrue	Mostly Untrue	Some-what	Mostly True	Totally True
21. Students in this class treat the teacher with respect.	○	○	○	○	○
22. My teacher doesn't let people give up when the work gets hard.	○	○	○	○	○
23. I have learned a lot this year about the state test.	○	○	○	○	○
24. I have pushed myself hard to completely understand my lessons in this class.	○	○	○	○	○
25. My teacher wants me to explain my answers -- why I think what I think.	○	○	○	○	○
26. My teacher has several good ways to explain each topic that we cover in this class.	○	○	○	○	○
27. The comments that I get on my work in this class help me understand how to improve.	○	○	○	○	○
28. Student behavior in this class is a problem.	○	○	○	○	○
29. The teacher in this class encourages me to do my best.	○	○	○	○	○
30. At home, I don't have a quiet place where I can do homework for this class.	○	○	○	○	○
31. When s/he is teaching us, my teacher thinks we understand even when we don't.	○	○	○	○	○
32. In this class, I get to test or try out my ideas to see if they work.	○	○	○	○	○
33. We get helpful comments to let us know what we did wrong on assignments.	○	○	○	○	○
34. Lessons in this class are often hard for the teacher to make clear.	○	○	○	○	○
35. My teacher makes me want to go to college.	○	○	○	○	○
36. I feel stressed out in this class.	○	○	○	○	○
37. My teacher wants us to use our thinking skills, not just memorize things.	○	○	○	○	○
38. Student behavior in this class is under control.	○	○	○	○	○
39. My after-school activities don't leave time to finish my homework for this class.	○	○	○	○	○
40. My teacher really tries to understand how students feel about things.	○	○	○	○	○
41. I hate the way that students behave in this class.	○	○	○	○	○
42. My teacher in this class does not know me very well yet.	○	○	○	○	○
43. My teacher makes lessons interesting.	○	○	○	○	○
44. What I am learning in this class will help me in my life.	○	○	○	○	○
45. My classmates behave the way my teacher wants them to.	○	○	○	○	○
46. My teacher seems to believe in my ability.	○	○	○	○	○
47. This class does not keep my attention--I get bored.	○	○	○	○	○
48. In this class, doing your homework is not very important.	○	○	○	○	○
49. My teacher tells us what we are learning and why.	○	○	○	○	○

3465

	Totally Untrue	Mostly Untrue	Some-what	Mostly True	Totally True
50. In this class, I stop trying when the work gets hard.	○	○	○	○	○
51. If I am sad or angry, my teacher helps me feel better.	○	○	○	○	○
52. My teacher pushes me to become a better thinker.	○	○	○	○	○
53. My teacher takes the time to summarize what we learn each day.	○	○	○	○	○
54. I often feel like this class has nothing to do with real life outside school.	○	○	○	○	○
55. In this class, we learn a lot almost every day.	○	○	○	○	○
56. My teacher seems to know if something is bothering me.	○	○	○	○	○
57. We spend a lot of time in this class practicing for the state test.	○	○	○	○	○
58. I have done my best quality work in this class all year long.	○	○	○	○	○
59. Students in this class don't get to say much about what we think, we mostly listen.	○	○	○	○	○
60. I feel smart when I am in this class.	○	○	○	○	○
61. My teacher checks to make sure we understand what s/he is teaching us.	○	○	○	○	○
62. Instead of giving us answers, my teacher would rather give us questions to discuss.	○	○	○	○	○
63. My teacher respects my ideas and suggestions.	○	○	○	○	○
64. This class makes me a better thinker.	○	○	○	○	○
65. I am happy with how well I have done in this class.	○	○	○	○	○
66. Student behavior in this class makes the teacher angry.	○	○	○	○	○
67. My teacher explains difficult things clearly.	○	○	○	○	○
68. In this class we have to write every day.	○	○	○	○	○
69. This class is a happy place for me to be.	○	○	○	○	○
70. In this class, my teacher accepts nothing less than our full effort.	○	○	○	○	○
71. Getting ready for the state test takes a lot of time in our class.	○	○	○	○	○
72. We get to do a lot in this class, not just listen to the teacher.	○	○	○	○	○
73. Often, I am not sure what I am supposed to be learning in this class.	○	○	○	○	○
74. My teacher gives us time to explain our ideas.	○	○	○	○	○
75. In this class, we learn to correct our mistakes.	○	○	○	○	○
76. When doing schoolwork for this class, I try to learn as much as I can and I don't worry about how long it takes.	○	○	○	○	○
77. Students speak up and share their ideas about class work.	○	○	○	○	○
78. I read at home almost every day.	○	○	○	○	○

79. When homework is assigned for this class, how much of it do you usually complete? (Select one choice.)

◯ Never assigned ◯ None of it ◯ Some of it ◯ Most of it ◯ All ◯ All plus some extra

80. Outside of class, about how much time in a week do you usually spend doing homework for this class? (Select one choice.)

◯ No time ◯ Half an hour ◯ 1 hour ◯ 2 hours ◯ 3-4 hours ◯ 5-7 hours ◯ 8+ hours

81. During most weeks, how many days a week is there homework to do for this class? (Select one choice.)

◯ 1 day ◯ 2 days ◯ 3 days ◯ 4 days ◯ 5 days ◯ Never assigned

82. Is there a computer at your home? If yes, how many? (Select one choice.)

◯ No ◯ Yes, one ◯ Yes, two ◯ Yes, three or more

83. Are you female or male?

◯ Male ◯ Female

84. How many adults live with you?

◯ One ◯ Two ◯ More than two

85. How many books are there in your home?

◯ 0 to 10 ◯ 11 - 24 ◯ 25 - 100 ◯ 100 - 250 ◯ More than 250

86. What is you race/ethnicity? (mark all that apply)

◯ White ◯ Black or African American ◯ Hispanic/Latino ◯ Asian ◯ Pacific Islander

◯ Arabic/Middle Eastern ◯ West Indian ◯ Native American (Indian) ◯ South Asian or East Indian ◯ Other

87. Counting yourself and all others, how many children live with you?

◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 or more

88. Does anyone in your family speak a language other than English at home?

◯ No ◯ Seldom ◯ Half the time ◯ Mostly ◯ Always

89. Think of the adult at your house who went to school for the most years. This person:

Did not finish high school
◯ No ◯ Yes

Finished high school
◯ No ◯ Yes

Attended some college or earned a 2-year degree
◯ No ◯ Yes

Finished a 4-year college degree
◯ No ◯ Yes

Finished a professional or graduate degree after college
◯ No ◯ Yes

◯ I don't know

90. When you were younger, what kind of marks (or grades) did you usually get in school?

◯ Very high ◯ High ◯ Good ◯ Some good, some not ◯ Usually not very good

91. How long have you been in this class?

◯ For less than two weeks ◯ For about three to four weeks ◯ Between one and two months ◯ More than 2 months

92. What adult was with you while you took this survey?

◯ My teacher for this class

◯ Someone who is my teacher this school year, but not my teacher for this class

◯ Someone who is not teaching me this school year

3465

Appendix F: Program Syntax

1) *Mplus* 7.4© Program

```
Title: Algebra I Teacher Effectiveness LCA;
     Data: File is VARCOMPLETE.dat;
     Variable:
     names = TEACHER_ICPSR_ID T_ENGAGE C_ENGAGE C_UNDER F_UNDER
     T_MANAGE C_MANAGE F_MANAGE;
     usevariables = T_ENGAGE C_ENGAGE C_UNDER F_UNDER T_MANAGE
     C_MANAGE F_MANAGE;
     classes = c(2);
     IDVARIABLE = TEACHER_ICPSR_ID;
     Analysis:
     Type = mixture;
     Model:
     %overall%;
     Plot:
     Type is plot3;
     series is T_ENGAGE (1) C_ENGAGE (2) C_UNDER (3) F_UNDER (4)
     T_MANAGE (5) C_MANAGE (6) F_MANAGE (7);
     Savedata:
     file is variable2.txt;
     SAVE = CPROBABILITIES;
     format is free;
     output:
     tech1 tech3 tech7 tech11;
```

2) IBM SPSS Statistics 22

MANOVA for mean measurements of the observed variables

```
GLM T_ENGAGE C_ENGAGE C_UNDER F_UNDER T_MANAGE C_MANAGE F_MANAGE
BY LCA_2classes
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/EMMEANS=TABLES(OVERALL)
/EMMEANS=TABLES(LCA_2classes)
/PRINT=DESCRIPTIVE ETASQ OPOWER HOMOGENEITY
/CRITERIA=ALPHA (.05)
/DESIGN= LCA_2classes.
```

Discriminant Analysis for ACT scores versus 2 latent class categorization

```
DISCRIMINANT
/GROUPS=LCA_2classses(1 2)
/VARIABLES=ACT_alg_score
/ANALYSIS ALL
/SAVE=CLASS SCORES PROBS
/PRIORS EQUAL
/STATISITICS=MEAN STDDEV BOXM
/CLASSIFY=NONMISSING POOL
```