
Electronic Theses and Dissertations, 2004-2019

2013

An Examination Of The Impact Of Student Characteristics And Teacher Experience And Preparation Program Attended On Student Achievement In A Small School District

Michael O'Neal
University of Central Florida



Part of the [Special Education and Teaching Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

O'Neal, Michael, "An Examination Of The Impact Of Student Characteristics And Teacher Experience And Preparation Program Attended On Student Achievement In A Small School District" (2013). *Electronic Theses and Dissertations, 2004-2019*. 2672.

<https://stars.library.ucf.edu/etd/2672>



University of
Central
Florida

STARS
Showcase of Text, Archives, Research & Scholarship

AN EXAMINATION OF THE IMPACT OF STUDENT CHARACTERISTICS AND
TEACHER EXPERIENCE AND PREPARATION PROGRAM ATTENDED ON STUDENT
ACHIEVEMENT IN A SMALL SCHOOL DISTRICT

by

MICHAEL O'NEAL
BBA Cameron University, 1987
M.Ed. Texas A&M University – Commerce, 2001

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the College of Education
at the University of Central Florida
Orlando, Florida

Summer Term
2013

Major Professor: Rebecca A. Hines

© 2013 Michael O'Neal

ABSTRACT

Demonstrating a direct link between teacher education programs and student growth is, to say the least, complex. Yet, using value-added systems as a means of holding teacher preparation programs accountable for the effectiveness of their graduates is a growing trend. However, few quantitative studies linking TPPs with the effectiveness of their graduates exist. The availability of student test scores linked to specific teachers in administrative databases makes it possible to use value-added modeling to obtain estimates of teacher effects. Only recently have researchers tapped into this expanding volume of data in an attempt to examine Teacher Preparation Programs as variables of student achievement. This study uses methodologies developed in the early stages of the Value-Added Teacher Preparation Program Assessment Model developed in Louisiana in 2006 as a guide. Using the HLM 7.0 software package, a statistical model was developed to determine if it were feasible to conduct an analysis using data from a single small school district and whether the results of such an analysis showed an impact of student characteristics and teacher experience and preparation program on student outcomes in mathematics.

ACKNOWLEDGMENTS

“Not all those who wander are lost.”

— J.R.R. Tolkien, *The Fellowship of the Ring*

The completion of this project has indeed been a long journey, a journey that included two job changes, two cross-country moves, and the birth of two beautiful children. However, life doesn't stand still for the sake of the dissertation nor does the dissertation wait for the best possible time to be completed. There have been many who questioned not only my commitment to my dissertation, but also whether it would ever be finished at all. I on the other hand, always knew, in spite of the pure joy and utter frustration that life's moments and this project brought, that I would finish. I just had to do it in my own time and in my own way.

While not always evident on the surface, my dissertation has always been a priority to me. However, as most know, there are many priorities in life at any given moment. Unfortunately, due to circumstances being what they were, I could not always keep the dissertation number one. In the end, I have finished, albeit not on my own, and I am truly happy. There are many people to whom I owe a sincere thank you, too many to list individually though I wish I could. Take heart in knowing that while you may not be mentioned by name, I am aware of your contributions and they are appreciated.

There are those, however, that I feel I must recognize. I could never have completed the dissertation without you.

To my chair, Dr. Rebecca Hines, it was only through your willingness to allow me to be who I am and progress in my own way and in my own time that I was able to finish at all. Your

flexibility, general demeanor, and gentle nudges at just the right times kept me going and saw me through to the end. Thank you.

To the remainder of my committee, Dr. Debbie Hahs-Vaughn, Dr. Lee Cross, and Dr. Maria Reyes-Macpherson a wholehearted and sincere thank you. You have helped me in so many ways that I can't count them all. From agreeing to a part of the committee to the guidance, suggestions, and support you've provided along the way, your contributions are appreciated and will never be forgotten.

To Dr. Wilfred Wienke, Dr. Lisa Dieker, and Dr. Cynthia Pearl, you were mentors and friends. Your patience and faith coupled with genuine caring and concern during my entire program touched me in so many ways. You are kind, motivating, encouraging and never judgmental. It is these qualities that allowed me to maintain perspective and never want to disappoint you by not finishing. I will be forever grateful.

To my children, Jamie, Rebecca, and Noah, it was you who provided me the reprieve from work and school that re-energized me and allowed me to keep going. You have been the constant bright spot throughout as well as the providers of so many wonderful memories. The smiles on your faces and happiness in your eyes each and every time I walked through the door are forever imprinted on my heart. I love you all very much.

Last, but most certainly not least, I must acknowledge my wife, Sara Aronin. Her love, support, and belief in me are unsurpassed. She personally endured so much, becoming the proverbial dissertation widow. She has taken care of everything that needed to be taken care of without complaint, just so I could focus on finishing. She has waited patiently, enduring too many long hours alone while I worked. At the same time, she has given me so much and I know I could never have finished without her by my side. She has taken every step with me and there

are no words that can express my gratitude and appreciation for her. I love you with all my heart and am forever indebted to you for giving me your love and support.

TABLE OF CONTENTS

LIST OF FIGURES	xii
LIST OF TABLES	xiii
LIST OF ACRONYMS/ABBREVIATIONS	xiv
CHAPTER ONE: THE PROBLEM AND ITS CLARIFYING COMPONENTS	1
Introduction.....	1
Theoretical Framework.....	2
Teacher characteristics and student outcomes.	3
Statement of the Problem.....	4
Purpose of the Study	6
Significance of the Study	6
Research Question	7
Delimitations.....	7
Limitations	8
Assumptions.....	11
Definition of Terms.....	11
Summary	14
CHAPTER TWO: REVIEW OF LITERATURE.....	15
Teacher Preparation in the University Setting	15
The colonial era.....	16
The early national period.	17
Early teacher preparation.	18
The common school.....	18

The public high school.....	19
The normal school.....	19
The university.	21
Prior Research on the Effectiveness of Teacher Preparation Programs	21
Linking Teacher Preparation Programs and Student Outcomes	22
Teacher effect.....	22
Teacher preparation.	23
Government Involvement in Accountability	24
The Higher Education Opportunity Act of 2008.	25
The No Child Left Behind Act of 2001.	26
The Individuals with Disabilities Act of 2004.....	28
The Education Sciences Reform Act of 2002.....	29
Race to the Top.	29
Research Related to Legislative Effectiveness	30
Development of Value-Added Models	32
Covariates	34
Student gender.	35
Student minority status.	37
Student language program status.	38
Student socioeconomic status	38
Student ESE status.....	39
Student gifted status.....	41
Teacher years of experience.....	41

Legislative Issues Surrounding Value-Added Models	43
Research Addressing the use of Value-Added Models.....	43
Teacher Preparation Program Effectiveness and Value-added Models.....	45
CHAPTER THREE: METHODOLOGY	55
Overview.....	55
Research Design.....	56
Participants.....	56
Students.....	56
Teachers.....	57
Data Source.....	57
Instrumentation	57
Vertical scale score.	61
Reliability.....	61
Validity.	62
Data Analysis	63
Three-level unconditional means model.....	66
Three-level unconditional growth model.....	68
Three-level conditional growth model.....	70
The Level 1 conditional model.	70
The Level 2 conditional model.	71
The Level 3 conditional model.	72
CHAPTER FOUR: RESULTS	74
Participants.....	74

Students.....	74
Teachers.....	75
Multilevel Modeling Assumptions	76
Normal distribution of residuals.	76
Linearity.....	79
Homogeneity of Variance.	80
Two Level Hierarchical Linear Models.....	85
Null model.	86
Unconditional growth model.	88
Conditional growth model.	90
Full model.	94
CHAPTER FIVE: DISCUSSION.....	100
Purpose of the Study	100
Summary of Findings.....	100
Implications of Findings Related to Student Characteristics	103
Gender.....	104
Gifted status.	105
Minority status.	105
Language program status.....	106
Socioeconomic status.....	106
Exceptional students.	107
Summary of Student Level Covariates	108
Implications of Findings Related to Teacher Characteristics	109

Limitations	111
Future Study	112
Conclusion	112
APPENDIX A: INSTITUTIONAL REVIEW BOARD.....	114
APPENDIX B: EXCERPT OF FEDERAL REGISTER REGARDING NCLB REAUTHORIZATION.....	116
APPENDIX C: DATA REQUEST LETTER	120
APPENDIX D: RESPONSE TO DATA REQUEST	122
APPENDIX E: PERMISSION TO REPRINT RESPONSE LETTER	124
APPENDIX F: COPYRIGHT PERMISSION FOR TAKS ITEMS.....	126
APPENDIX G: HLM RESULTS, NULL MODEL.....	128
APPENDIX H: HLM RESULTS, UNCONDITIONAL GROWTH MODEL	132
APPENDIX I: HLM RESULTS, TEST OF SIGNIFICANCE FOR STUDENT DEMOGRAPHIC COVARIATES.....	136
APPENDIX J: HLM RESULTS, CONDITIONAL GROWTH MODEL.....	155
APPENDIX K: HLM RESULTS, FULL CONDITIONAL MODEL.....	159
REFERENCES	164

LIST OF FIGURES

<i>Figure 1.</i> Astin's input-environment-output (I-E-O) model (1991).....	3
<i>Figure 2.</i> Hierarchical Structure of Data.	55
<i>Figure 3.</i> Sample Items 2009 Grade 4 Mathematics TAKS.....	58
<i>Figure 4.</i> Q-Q Plot of Level 1 residuals used to evaluate the normality assumption.	77
<i>Figure 5.</i> Q-Q Plot of Level 2 intercept residuals used to evaluate the normality assumption. ...	78
<i>Figure 6.</i> Q-Q Plot of Level 2 slope residuals used to evaluate the normality assumption.....	78
<i>Figure 7.</i> Scatterplot used to evaluate the linearity assumption at Level 1.	79
<i>Figure 8.</i> Residuals plotted against the male covariate to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.	81
<i>Figure 9.</i> Residuals plotted against the minority covariate to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.....	82
<i>Figure 10.</i> Residuals plotted against langprog to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.....	82
<i>Figure 11.</i> Residuals plotted against ESE to evaluate the assumption of homogeneity of variance at Level 2 intercept and intercept.....	83
<i>Figure 12.</i> Residuals plotted against gifted to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.....	84
<i>Figure 13.</i> Residuals plotted against FRL to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.	84

LIST OF TABLES

Table 1	<i>Summary of Commonly Used Value-Added Models.</i>	33
Table 2	<i>Estimates of Within School Variation in Teacher Effectiveness.</i>	44
Table 3	<i>2010-2011 TAKS Mathematics Assessments.</i>	59
Table 4	<i>2011 TAKS Blueprint for Grades 3-8 Mathematics.</i>	60
Table 5	<i>TAKS Grades 3-8 Mathematics Reliability Estimates.</i>	62
Table 6	<i>Demographic Information of Student Participants.</i>	75
Table 7	<i>Covariates Included at Each Level of Final Model.</i>	86
Table 8	<i>Estimation of Fixed Effects and Variance Components, Null Model (One-Way Random Effects ANOVA).</i>	87
Table 9	<i>Estimation of Fixed Effects and Variance Components, Unconditional Growth Model</i>	90
Table 10	<i>Test of Significance of Student Predictor Variables</i>	91
Table 11	<i>Estimation of Fixed Effects, Conditional Growth Model</i>	93
Table 12	<i>Estimation of Variance Components, Conditional Growth Model</i>	94
Table 13	<i>Description of Teachers' Groups</i>	95
Table 14	<i>Estimation of Fixed Effects, Full Conditional Model</i>	98
Table 15	<i>Estimation of Variance Components, Full Conditional Model</i>	99

LIST OF ACRONYMS/ABBREVIATIONS

AACTE	American Association of Colleges for Teacher Education
America COMPETES	America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science
ARRA	American Recovery and Reinvestment Act
AYP	Adequate Yearly Progress
B.A.	Bachelor of Arts
B.S.	Bachelor of Science
DQC	Data Quality Campaign
EHA	Education for All Handicapped Children Act
ELA	English Language Arts
ESEA	Elementary and Secondary Education Act
ESRA	Education Sciences Reform Act
FRL	Free or Reduced Price Lunch
GI	Government Issue
HEA	Higher Education Opportunity Act
HLM	Hierarchical Linear and Nonlinear Modeling
IDEA	Individuals with Disabilities Educational Act
IES	Institute of Education Sciences
IHE	Institutions of Higher Education
M.A.	Master of Arts
M.Ed.	Master of Education
M.S.	Master of Science

NAEP	National Assessment of Education Progress
NCATE	National Council for Accreditation of Teacher Education
NCLB	No Child Left Behind Act
NCTQ	National Council on Teacher Quality
RTTT	Race to the Top
SFSF	State Fiscal Stabilization Funds
TAKS	Texas Assessment of Knowledge and Skills
TMVSS	TAKS Mathematics Vertical Scale Score
TPP	Teacher Preparation Program
USDOE	United States Department of Education
VAM	Value-added Model(ing)

CHAPTER ONE: THE PROBLEM AND ITS CLARIFYING COMPONENTS

Introduction

While differences between K-12 education and higher education exist, they also share common political and policy agendas (Wellman, 2001). Consequently, trends in K-12 standards-based reforms have stimulated an interest in extending K-12 accountability systems to higher education (Wellman). Policymakers have demanded greater accountability from institutions of higher education (IHE) as well, with teacher preparation programs (TPPs), at the forefront of the discussion (United States Department of Education [USDOE], 2009a). A key element of such discussions is the ideology that TPPs be held accountable for their impact on K-12 student outcomes as is indicated in the Higher Education Act Amendments of 1998 (HEA, 1998). To wit, United States Secretary of Education, Arne Duncan, stated, “It is time to start holding teacher preparation programs more accountable for the impact of their graduates on student learning” (Duncan, 2010).

As previously stated, Congress inserted a provision into the Higher Education Act Amendments of 1998 (HEA, 1998) requiring states to hold teacher preparation programs¹ (TPPs) accountable for their graduates’ effectiveness. Since that time, states have been subjected to mounting pressure from federal government, accreditation agencies, philanthropic organizations, sundry think tanks, etc. to not only develop such systems of accountability, but to publicize the effectiveness of their graduates (Imig, Wiseman, & Imig, 2011). However, 46 states do not currently share teacher performance data with TPPs (Data Quality Campaign, 2012).

¹ Traditional teacher preparation programs generally serve undergraduate students who have no prior teaching or work experience, and lead at least to a bachelor’s degree. Some traditional teacher preparation programs may lead to a teaching credential but not to a degree. (Duncan, 2011, p.1).

In an effort to address the requirements of the HEA (1998), the Louisiana Boards of Regents along with the Board of Elementary and Secondary Education established the Blue Ribbon Commission on Teacher Quality (BRC). The charge to the commission was in essence to make recommendations that would bring about a PK-16 system that held universities (teacher preparation programs) as well as schools accountable for student achievement (Council for a Better Louisiana, 2001).

One response to this charge was the development of the Louisiana Value Added Teacher Preparation Assessment Model, making Louisiana the first state to use value added modeling to address the relationship between teacher preparation programs and student achievement (Gansle, Burns, & Noell, 2011). The Board of Regents contracted Dr. George Noell to conduct pilot studies in 2003 and 2004 to determine the feasibility of using the model statewide. Based upon the results of those pilot studies, a decision was made to pursue the use of the Value Added Teacher Preparation Assessment Model that was fully implemented in 2005 using data from all 66 school districts and the 21 public and private universities with teacher preparation programs in the state of Louisiana.

Theoretical Framework

Astin (1991) concluded that input variables must be included to comprehend the relationships between processes and outcomes. While there have been many models designed and used for this purpose in education, in this study the theoretical framework is based on Astin's input-environment-output model (Figure 1).

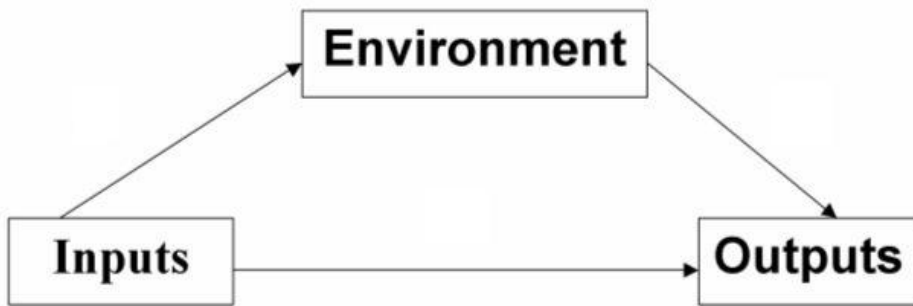


Figure 1. Astin's input-environment-output (I-E-O) model (1991).

Astin's (1991) model was chosen because it permits the assumption that differences among input variables can be controlled, thus providing a more objective estimate of the impact of environment on outcomes. While simplistic in its nature, Astin's model is also practical because opinions formed during program evaluation require comparative analyses. Therefore, making a change to an input or environmental element will result in improved outcomes, whereas choosing to do nothing suggests that the status quo is preferred to any available alternatives.

Using the Astin model, it is theoretically possible to determine whether new teachers who are graduates of a particular teacher preparation program are successful in the context of student outcomes while providing statistical evidence to this end. The model also provides a relevant method for the enhanced evaluation of teacher preparation programs as outlined by current educational policy.

Teacher characteristics and student outcomes. The publication of Equality of Educational Opportunity stirred controversy by weighting the role families and peers play on student outcomes over the role of schools and teachers (Coleman et al., 1966). Since that time

the relationship between teacher influence and student outcomes has become generally accepted (Aaronson, Barrow, & Sander, 2007; Kane & Cantrell, 2010; Rivkin, Hanushek, & Kain, 2005; Rowan, Correnti, & Miller, 2002; Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997). Although the magnitude of the effect has proven difficult to pinpoint, existing research has shown that teachers have greater influences on mathematics outcomes than on reading or Language Arts outcomes (Kane & Cantrell, 2010).

In their seminal study, Sanders and Rivers (1996) found that students who had three effective teachers in a row scored more than 50 percentile points higher on standardized mathematics assessments than students assigned to a series of three ineffective teachers despite beginning with comparable scores. Rowan et al. (2002) estimated that teacher effects explain 8-18% of the variance in student achievement in mathematics, and 52-72% of the variance in student growth in mathematics. Meanwhile, Rivkin et al. (2005) found that an increase of one standard deviation in teacher effectiveness corresponded to a 0.11 standard deviation increase in student mathematics achievement. These studies indicate that student growth is an important part of defining effective teaching. While students in some classrooms have higher scores or show greater improvement than students in other classrooms, it is important to question to what extent the differences are attributable to teacher quality.

Statement of the Problem

A key part of assessing teacher effectiveness based on student outcomes is the assumption that there is a valid way to do so. Even though value-added models (VAMs) are in use in varying degrees and gaining support among policymakers as a means of measuring teacher quality, little research addresses exactly how to tie student growth to teacher performance (Steele, Hamilton, & Stecher, 2010). Consequently, states and school districts are struggling to

find approaches to measuring student growth; this is particularly true of students with disabilities (Holdheide, Browder, Warren, Buzick, & Jones, 2012).

Furthermore, national policy debates on student achievement and teacher quality continue with policymakers seeking indicators that accurately evaluate not only teacher and school performance, but also the performance of teacher preparation programs (Duncan, 2010). The significance of the issue is reflected in the \$4.35 billion Race to the Top (RTTT) grant program, which rewarded States for increasing student achievement and producing effective teachers. The selection criteria in the grant included a provision for improving the effectiveness of TPPs.

Specifically, points were awarded to applicants based on

(t)he extent to which the State has a high-quality plan and ambitious yet achievable annual targets to link student achievement and student growth data to the students' teachers and principals, to link this information to the in-State programs where those teachers and principals were prepared for credentialing, and to publicly report the data for each credentialing program in the State (USDOE, 2009b, p. 10).

However, few quantitative studies linking TPPs with the effectiveness of their graduates exist (National Research Council [NRC], 2010). While the availability of student test scores linked to specific teachers in administrative databases makes it possible to use value-added modeling to obtain estimates of teacher effects, only recently have researchers tapped into this expanding volume of data in an attempt to examine TPPs as variables of student achievement (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Goldhaber & Liddle, 2012; Harris & Sass, 2007; Henry et al., 2011; Henry, Thompson, Fortner, Zuli, & Kershaw, 2010; Koedel, Parsons, Podgursky, & Ehlert, 2012; Mihaley, McCaffery, Sass, & Lockwood, 2012; Noell, Gansle, Patt, & Schafer, 2009; Noell, Porter, & Patt, 2007; Noell, Porter, Patt, & Dahir, 2008).

Purpose of the Study

The purpose of this study is to examine whether a small school district can use multilevel modeling to determine the impact of various student characteristics and teachers' level of experience and teacher preparation program attended on student mathematics achievement. Moreover, does the effectiveness of new teachers from specific teacher preparation programs differ from that of experienced teachers?

Significance of the Study

While this study does not focus on the evaluation of special education teachers *per se*, it is important to note that the Council for Exceptional Children (CEC), in its position paper on special education teacher preparation stated that “the principles of good evaluation apply to all teachers” (CEC, 2012, p. 74).

Adequate yearly progress (AYP) is the means by which all public schools that receive Title I funds are held accountable for student outcomes through the No Child Left Behind Act of 2001 (NCLB). In order to determine if all students are making progress toward meeting academic standards, each state must develop a statewide accountability system. While each state independently defines AYP, they must ensure that all students are proficient in reading and mathematics by 2014. Additionally, each state must establish proficiency targets for each year leading up to 2014. Proficiency targets must be met, not only by all students as an aggregate, but also by the following specific groups of students: economically disadvantaged students, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency (NCLB, 2002).

If a statistically significant, positive relationship between measured student and teacher characteristics and student achievement can be established, it might be possible to use this

information to aide in the development of policies that would help schools and districts meet the AYP proficiency targets required by NCLB. Knowing if and to what degree a measured characteristic is associated with increased student outcomes could provide school districts a means of targeting resources to address areas of specific needs. School districts could ostensibly create specific targets for recruitment and retention efforts, professional development, and establish a framework for staffing decisions.

Research Question

The current study will address the following research questions:

- (1) What is the effect of the student characteristics of gender, minority status, language program status, exceptional education status, gifted education status, and free or reduced price lunch status on the predicted mathematics achievement of students in grades four through eight?
- (2) To what extent is the predicted mathematics achievement of students affected by teachers' level of experience and in the case of new teachers, their teacher preparation program attended?

Delimitations

The span of years and schools that constitute the focus of this study were specifically selected to include the availability of vertically scaled scores as well as other data necessary to complete the study. Because the study encompassed only a single school district and only student mathematics scores are used in analyses, the effect of the curricular model used by that district is also a limiting factor. Random sampling was not used to select schools or assign students and teachers to groups for analysis. Study variables will be chosen based on availability, and their inclusion or exclusion in this study is a reflection of the author's discretion. Consequently, there

may exist other latent, non-modeled factors at the student or teacher levels, which may have an impact on student outcomes. Values, attitudes, and motivation are not examined in the present study. This delimitation should not be construed as minimizing the impact of these variables on student outcomes nor in any way advocate the use of standardized assessments as proxies for the success, or lack thereof, of students, teachers, or teacher preparation programs.

Limitations

There are limitations associated with both value-added modeling in general and this study specifically. Limitations of this study primarily include issues with generalizability and measurement. Because this study involved only one school district in Texas, results might not be generalizable to school districts that do not have similar characteristics nor to school districts outside the state. Additionally, while there are many different assessments of mathematical knowledge and skills, only the results of the Texas Assessment of Knowledge and Skills (TAKS) are examined. Therefore, a true and complete picture of student learning may not be available. Of particular importance is the fact that the TAKS is designed to measure mastery of the Texas Essential Knowledge and Skills (state standards), not the effectiveness of teachers.

Another limitation of this study is that the effect of the teacher preparation program was not contributed to student achievement after a teacher achieved five years experience. Furthermore, teachers were not nested within their respective schools, they were nested within years of experience except in the case of teachers with fewer than five years experience who were nested within TPP.

Since the mid-1990s, recognized experts in the field have expressed many issues with the use of value-added modeling as an evaluative tool in education in numerous studies, papers, and articles (Amrein-Beardsley, 2008; Baker, Bartson, Darling-Hammond, Haertel, Ladd, Linn, et

al., 2010; Darling-Hammond, 2009; DeVore, 2011; Henry et al., 2011; Kukla-Acevedo, Streams, & Toma, 2009; Kupermintz, 2003; Mihaly, McCaffrey, Sass, & Lockwood, 2012; Noell, 2005).

Some of these issues are:

- There are no studies that conclusively prove the causal effect of teachers on student achievement (DeVore, 2011, p. 4).
- [I]nstability can result from differences in the characteristics of students assigned to particular teachers in a particular year, from small samples of students (made even less representative in schools serving disadvantaged students by high rates of student mobility), from other influences on student learning both inside and outside school, and from tests that are poorly lined up with the curriculum teachers are expected to cover, or that do not measure the full range of achievement of students in the class. (Baker et al., 2010, p. 2).
- A number of factors have been found to have strong influences on student learning gains, aside from the teachers to whom their scores would be attached (Baker et al., 2010, p. 3).
- There are concerns with TPP factors, “including selection of teachers into and out of programs, selection of program graduates into teaching positions within the state, and how teacher performance is measured” (Mihaly, McCaffrey, Sass, & Lockwood, 2012, p. 2).
- TPP practices have changed over the range of graduation dates of teachers included in a value-added analysis. Restricting analysis to only recent graduates or graduate cohorts in an effort to ameliorate this issue would, in turn, cause a further reduction in sample size and could introduce selection bias due to the likely non-random distribution of seniority

levels across schools with different achievement levels (Kukla-Acevedo, Streams, & Toma, 2009, p. 15).

- [U]sing student test score outcomes to measure teaching effectiveness include the limited subjects and grades in which testing is conducted (Henry et al., 2011, p. 2).
- [O]ther important outcomes such as graduation, attitudes toward school and learning, or knowledge of one's rights and obligations as a citizen within a democracy are not captured by these standardized tests (Henry et al., 2011, p. 2).
- Although every effort may be taken to use the best available data to remove the effects of variables such as poverty, it cannot be known whether the groups of teachers have truly been equated statistically on all-important factors (Noell, 2005, p. 5).
- There will always remain some potentially important variables (e.g., parental level of education) for which data will not be available (Noell, 2005, p. 5).
- As student data are aggregated from year to year, the number of missing cases is likely to increase which may have a negative impact on results (Noell, 2005, p. 6).
- [U]sing a spring to spring assessment window means that student gains after the standardized assessment actually contribute to the assessment of the following year's teacher, rather than the teacher who taught the student after testing was completed (Noell, 2005, p. 6).

Historically, the debate surrounding value-added modeling has centered on (a) the proper method of obtaining value added scores; (b) the accuracy of those scores; or (c) their appropriate use (Hill, Kapitula, & Umland, 2010). While this debate will likely continue, in light of recent federal, state, and local educational policy decisions, value-added modeling will ostensibly

remain a part of educational evaluative processes. Rather than join this debate, this study acknowledges previously identified limitations to value-added modeling.

Despite known limitations, there is little disagreement that value-added models provide a means for separating teacher effect from school and student effects. Rowan, Chiang, and Miller (1997) state that after controlling for various student characteristics, effects on student achievement can be attributed to the following three variables (a) their teaching ability, (b) their motivation, and (c) their working conditions.

Assumptions

The following assumptions are made for the current study.

1. All students performed their best on each administration of the TAKS.
2. All students who took the TAKS in any given year had an equal opportunity to perform to the best of their abilities.
3. All students did their own work on all assessments.
4. All students were accurately and appropriately administered the TAKS.
5. The TAKS results are a true and accurate depiction of each student's skill and ability.
6. The TAKS provides an accurate measurement of mathematical knowledge and skills.
7. The demographic data for each student is accurately reported.

Definition of Terms

Accountability: "The concept that individuals (e.g., students, teachers, or administrators) or organizations (e.g., schools, school districts, or state departments of education) should be held responsible for improving student achievement and should be either rewarded for

their success or sanctioned for their lack of success in doing so. In education, accountability requires measurable proof that teachers, schools, districts, and states are teaching students efficiently and well. Usually this proof takes the form of student success rates on various tests. In recent years, most accountability programs have been based on state curriculum standards and state tests derived from those standards” (Ravitch, 2007, 7).

Achievement: “A student’s score on the State’s assessments under the ESEA” (U.S. Department of Education, 2009b, 14).

Educational Reform: “Educational reform is defined as changes of one or more of the following aspects of educational system: goals and objectives, policy making and the managerial system or power structure, financing and budget processes, system organization, curriculum, pedagogy. social relations of teaching and learning, selection, evaluation and promotion, designed both to reflect and advance relatively clear and politically salient ideas about the future shape of a given society and the role of education therein” (Zajda, 2010, p. 50).

Higher Education: “Study beyond the level of secondary education. Institutions of higher education include not only colleges and universities but also professional schools in such fields as law, theology, medicine, business, music, and art. They also include teacher-training schools, community colleges, and institutes of technology. At the end of a prescribed course of study, a degree, diploma, or certificate is awarded” (higher education. 2013. In *Merriam-Webster.com*. Retrieved February 10, 2013, from <http://www.merriam-webster.com/dictionary/higher%20education>).

Proficiency Exam: “A test or other structured method that measures the qualifications of prospective teachers, has a pass-fail outcome and is used by the state for teacher certification or licensure” (<http://title2.ed.gov/Title2STRC/Pages/Glossary.aspx>).

Standards-Based Reforms: “Standards-based reform- is defined as a set of standards for what children should know and be able to do at particular grade-levels, align their curricula and teacher training to the standards, create statewide tests to measure student achievement, and based on the results, provide rewards, sanctions, or assistance” (Lake, Hill, O’Toole, & Celio, 1999).

Student Outcomes: “A student’s score on the State’s assessments under the ESEA” (U.S. Department of Education, 2009b, 14). For the purpose of this study, student outcomes will be measured by the Texas Assessment of Knowledge and Skill mathematics vertical score.

Teacher Preparation Programs: “A state-approved course of study, the completion of which signifies that an enrollee has met all the state’s educational and/or training requirements for initial certification or licensure to teach in the state’s elementary, middle or secondary schools. A teacher preparation program may be either a traditional program or an alternative route to certification, as defined by the state. Also, it may be within or outside an institution of higher education” (<http://title2.ed.gov/Title2STRC/Pages/Glossary.aspx>). For this study, teacher preparation program is any state approved course of study leading to initial licensure regardless of degree awarded.

Teacher Quality: “The knowledge, skills, abilities, and dispositions of teachers” which allow them to “engage students in rigorous, meaningful activities that foster academic learning for all students” (National Research Council [NRC], 2001, pp. 19-22).

Value-added Model: “[A] collection of complex statistical techniques that use multiple years of students’ test score data to estimate the effects of individual schools or teachers”

(McCaffrey, Lockwood, Koretz, & Hamilton, 2003, p. xi).

Years of Service: The number of years of service a teacher has been credited with by the Texas Education Agency

Summary

Accountability for the quality of graduates of teacher education programs is generally the responsibility of state governments and accreditation organizations. While neither of these bodies has traditionally required student outcomes to be considered when determining the quality of teacher preparation programs, there is a growing movement to do so.

Using state assessment data and value added modeling to measure growth in student achievement as a measure of teacher effectiveness is a controversial practice brought about by the belief that teachers should be held accountable for student achievement. As an extension of this belief, teacher preparation programs became subject to using the same method as a measure of accountability for producing high quality teachers.

This chapter has included the theoretical framework that guided the study, a statement of the problem, the purpose and significance of the study, research questions, delimitations and limitations of the study, assumptions, and definition of terms associated with the study.

CHAPTER TWO: REVIEW OF LITERATURE

The purpose of this chapter is to provide an overview of relevant literature related to the accountability of teachers and TPPs based on K-12 student outcomes. Presented first is a historical perspective on the path through which teacher preparation came to reside within the university. Following is a discussion of the legislative connection between K-12 education and TPPs, as it pertains to the evolution and convergence of accountability systems, measured in the quantitative context of student outcomes. Next, the development and expansion of value-added models (VAMs) as a measure of teacher effectiveness is examined followed by a discussion of student and teacher covariates. The chapter culminates with an overview of related research discussing the migration of VAMs from teacher accountability to TPP accountability.

Teacher Preparation in the University Setting

Fraser (2007) described teacher preparation throughout the history of the United States as “a haphazard affair” (p. 3). Early on, teaching was viewed as an occupation for which no professional knowledge was needed—a view that resulted in a marked degree of apathy toward formal teacher preparation both within the existing university system and among the general public as well (Sarason, Davidson, & Blatt, 1986). But as teaching came to be seen more as a profession in the same regard as medicine, law, and the clergy, the training of educators moved beyond the vocational training provided in normal schools to a more theoretical approach and thus, formal teacher preparation became a function of the university. This section presents an examination of teacher preparation in the context of historical development, exploring changes in the social, economic, political, and religious frameworks in America that, either directly or indirectly, brought teacher preparation into the modern research university.

The colonial era. Originally considered a joint enterprise of church and home, education rapidly came to be recognized as necessary to sustain prosperity and liberty (Reese, 2005). Education in the colonies was solely at the discretion of parents and provided through a wide range of alternatives. While most children were schooled at home, primary school options included church schools, dame schools, tutorial schools, old-field schools, pauper or charity schools for the poor, and private tutors. Secondary offerings generally consisted of Latin grammar schools, academies, and seminaries. Public primary schools, where they existed, were largely laissez-faire endeavors and funded only in part through taxation. There were no public secondary schools.

In the early colonial educational system, there was relatively no distinction between secular and religious (Marshall, 1962). Education was approached with a spirit of true piety and devotion to vocation for public good. Colonial education systems evolved with distinct regional differences. The colonies developed systems of education to promote their culture, their traditions, and their religions.

Later, in the New England colonies, Massachusetts' Puritan leaders came to believe that simply being a congregant was not enough to defeat evil. A proper level of knowledge of the Scriptures, which could only be attained through reading and writing, was needed to counter the work of *that old deluder Satan*. As a result, the General Court of the colony enacted statutes designed to promote education (Alexander & Alexander, 2001).

The Massachusetts School Law of 1642 made education a state responsibility. While schools were not required, education was, and all children were to learn to read and write. The Massachusetts School Law of 1647 required that all towns of fifty or more households hire a teacher to provide instruction in reading and writing. Towns with 100 or more households were

compelled to open grammar schools to prepare children for university attendance. Most New England colonies had similar laws on the books by 1720 (Alexander & Salmon, 1995).

What began as piety emerged as a realization that education was essential to sustaining government and promoting the general welfare of society (Alexander & Alexander, 2001). Local governing bodies taxed their citizens to provide children an appropriate education (Alexander & Salmon, 1995). The New England legislation laid the foundation for state involvement in education by setting academic standards and providing resources for education.

The early national period. The framers of the constitution chose not to explicitly address public education. During debates at the Constitutional Convention, they recognized education as necessary to sustaining democracy but they could not envision a system allowing federal control of education (Good & Teller, 1973). In fact, many existing state Constitutions already addressed government's role in education. Becoming entangled with discussions on the separation of church and state, the topic of education proved too controversial and was abandoned. This left education a states matter and the existing educational systems remained largely unchanged (Good & Teller). Instead, the federal government turned to implicit tactics, in the form of land grant legislation, to commit to public education (Alexander & Salmon, 1995).

The General Land Ordinance of 1785 set aside land to be ceded to states when they joined the Union (Souder & Fairfax, 1996). The Northwest Ordinance (1787) created a system of territorial governance and dictated the process by which those territories could become states. To qualify for statehood, territories had to provide for, among other things, public education. When admitted to the union, a state would receive their school land as well as additional land to support other public institutions (Tyack, James, & Benavot, 1987).

Under the General Land Ordinance of 1785, land was granted as Congressional townships divided into thirty-six one mile square sections (six miles by six miles) with the sixteenth section designated for promoting education. The land could be used for public education or sold, with the proceeds used for education (Alexander & Alexander).

Early teacher preparation. The formal preparation of teachers during this period was essentially nonexistent. Teachers relied on their natural abilities, personal knowledge, and lessons learned on the job as the means of preparation (Hinsdale, 1900). Teachers in secondary schools were sometimes college students needing financial support for their studies or college graduates who taught temporarily while awaiting their first congregational appointment or other professional apprenticeship (Allmendinger, 1975).

Primary teachers, on the other hand, often lacked any semblance of formal training and rarely was any required. Many were those looking to avoid manual labor or who had failed in other professions (Butts & Cremin, 1953). In most cases, anyone willing to teach could, with the prerequisites more focused on religion than pedagogy (Fraser, 2007). There would not be a concerted effort to address formal teacher preparation until approximately 1820 (Woodring, 1975).

The common school. American education in 1820 was remarkably similar to that of the 1600s. However, education would experience far-reaching reforms during the 19th and 20th centuries. One such reform movement was an effort to create a system of universal public education generally referred to as the common school movement (Fraser, 2007).

Precipitated by social factors such as the spread of capitalism, urbanization, industrialization, population growth (including immigration), and westward expansion, interest in common schools grew rapidly. Reformers saw in the common school a way to provide every

American with, at minimum, enough knowledge to be productive, patriotic, and law-abiding citizens while curing the social ills that prevented them from doing so (Johnson).

The public high school. By facilitating the idea of free tax supported schools, providing an institutional framework upon which to be built, and preparing students for attendance, common schools paved the way for the expansion of public high schools.

By 1820 private academies replaced Latin grammar schools as the primary providers of secondary education and their numbers steadily increased until the mid 1800s. However, their popularity waned as public high schools flourished after court decisions in Illinois, Wisconsin, Kansas, Missouri, and most notably Michigan (see *Stuart v. School District No. 1 of Village of Kalamazoo*, 1874) established solid legal footing for public financing of high schools. The fate of academies as a mainstream provider of secondary education was sealed. By the late 1800s academies had all but disappeared. Though starting slowly, public high schools quickly found widespread acceptance. The number of public high schools in the U.S. grew from 321 in 1860 to some 10,000 by 1910, (Kirschenbaum, Simon, & Napier, 1971).

The normal school. The preparation of teachers for the rapidly expanding common schools was carried out in a multitude of institutions. Normal schools, however, quickly moved to the forefront of teacher preparation and by the end of the nineteenth century, prepared the majority of teachers in the United States. Normal schools were established with a singular purpose, to provide their students the instructional and classroom management skills necessary to teach in public schools (Goudie, 1988). Combining methodological study and classroom experience, normal schools sought to strengthen their students' pedagogical skills.

In the late 19th century, contention between normal schools and colleges regarding who should prepare secondary teachers arose when normal schools expanded to include the

preparation of high school teachers. Partly in response to this fray, normal schools initiated substantial institutional changes. They replicated models found in liberal arts colleges, which until this time had claimed the preparation of secondary teachers as its domain. Normal schools also adopted standards developed by accrediting and professional associations. The curriculum was lengthened to two years of collegiate level work for common school teachers and four years for high school teachers. Professors were recruited from liberal arts colleges and research practices were implemented. In making the transition from offering what was essentially an eighth-grade education to providing college-level courses of study, normal schools turned themselves into de facto liberal arts colleges (Fraser, 2007).

During this transformation, normal school curriculum gained depth and breadth in the arts and sciences. Many adopted the name college and began granting bachelor's degrees in a number of fields, including education (Ogren, 2005; Urban, 1996). Eventually, the word teachers was removed from their title, and their name changed to the more marketable state college (Labaree, 2008). By 1930 the majority of normal schools had become colleges and by the 1950s, ceased to exist (Ogren, 2005). The process of institutional evolution was actualized during the 1950s, 1960s, and 1970s. One by one, former normal schools were conferred the title university (Labaree, 2008).

Normal schools, the entities that had supplied the majority of teachers to the nation's schools, became a casualty of educational reform (Fraser, 2007). Clifford and Guthrie (1988) stated that while normal schools never attained the status their supporters wanted, their departure left two voids in teacher preparation; professional schools dedicated to only teacher preparation and a focus on pedagogy.

The university. As normal schools were expanding and undergoing the transformation to universities, existing universities were also establishing various teacher preparation programs. Schools or colleges of education were created at Iowa and Ohio State (1907), Berkeley (1913), Stanford (1917), Harvard (1920), and Michigan (1921), universities at the top of the higher education hierarchy (Clifford & Guthrie, 1988, pp. 64-65). Soon, advanced degrees in pedagogy were offered (Smith, 1980).

The newly created education schools viewed themselves as having a vastly different mission than that of the normal schools (Powell, 1976). Normal schools concentrated on the needs of a growing educational system through the mass preparation of teachers for common schools, while professors at universities focused on educational research and the preparation of high school teachers and school administrators (Labaree, 2008). These distinctly contrasting objectives are the cornerstone of a continuing dichotomy in missions characterizing the modern universities.

In becoming incorporated within the university, teacher education merely followed the path of other professions. Universities provided the liberal component of education for the high professions (medicine, law, clergy, etc.) as early as the 18th century. Eventually, professional schools in major fields existed solely within the university. Teacher preparation, as with the higher professions, was being professionalized and therefore destined for the university (Labaree, 2008).

Prior Research on the Effectiveness of Teacher Preparation Programs

Relatively little quantitative research linking teacher preparation programs with the quality of their graduates currently exists (National Research Council, 2010). Indeed, the need for and value of educational research was not recognized until after the advent of graduate

education in the 1920s (Reese, 1999). The earliest research focused on effective teaching and attempted to delineate the characteristics of effective teachers from less effective teachers. Researchers, guided by the notion that what teachers do is paramount to their effectiveness, instituted a series of methods experiments aimed at building a reliable knowledge base for teacher education. However, many of the early studies were designed with the student as the unit of analysis rather than the teacher, making generalizations to teachers not actually participating in the investigation all but impossible. Consequently, the mixed results of this early research proved to be inconclusive (Lederman & Niess, 2001).

Lederman and Niess (2001) found that research on teacher preparation conducted between 1920 and 2000 could be categorized into six phases, each focusing on student outcomes as related to: 1. Teacher characteristics; 2. Teaching methods; 3. Teacher behaviors; 4. Mastering competencies; 5. Appropriate use of competencies; and, 6. Subject-specific instructional knowledge and skills. Subsequent literature focused primarily on the effectiveness of alternative versus traditional pathways to certification (Darling-Hammond, 2009; Lasonde, 2010; Reese, 2010; Xu et al., 2012). Most recently researchers have begun using longitudinal data in an attempt to link teacher preparation programs to teacher effectiveness and student outcomes so that conclusions regarding teacher preparation programs effectiveness could be drawn (Boyd, et al., 2009; Goldhaber & Liddle, 2012; Harris & Sass, 2007; Henry et al., 2011; Henry, et al., 2010; Koedel, et al., 2012; Mihaley, et al., 2012; Noell, et al., 2009; Noell, et al., 2007; Noell, et al., 2008).

Linking Teacher Preparation Programs and Student Outcomes

Teacher effect. As previously stated, the publication of *Equality of Educational Opportunity* stirred controversy by emphasizing the role that families and peers play on student

outcomes over the role of schools and teachers (Coleman, 1966). Since that time the significance of teacher influence on student outcomes has become generally accepted (Aaronson, et al., 2007; Cochran-Smith & Zeichner, 2005; Crowe, 2010; Duncan, 2010; Goe, 2007; Heck, 2008; Levine, 2006; Rivkin, et al., 2005; Rowan, et al., 2002; Sanders & Rivers, 1996; Schwerdt & Wuppermann, 2009; Wright, et al., 1997). Frase (2005) went so far as to assert, “[t]he importance of teachers to the educational process has seldom, if ever, been seriously questioned by either academicians or lay people” (p. 437). However, exactly which teacher characteristics are factors in student outcomes (Aaronson, et al.) and the magnitude to which those characteristics affect student outcomes (Rivkin, et al.) continue to be points of debate.

The inability of observable teacher characteristics to explain significant amounts of variability in student outcomes between teachers led to suggestions of identifying effective teachers in terms of student performance through the development and implementation of value-added models (Braun, 2005; Gordon, Kane, & Staiger, 2006).

Teacher preparation. The idea of linking teacher preparation to student outcomes is not new. Throughout the 19th century, higher education controlled almost every aspect of secondary education. Higher education prepared teachers, developed tests, designed or approved curriculum and decided who would be allowed to attend college (Haycock, 1994). Between then and now, an easily identifiable gap developed between K-12 education and teacher preparation (Maeroff, Callan, & Usdan, 2001; Smith, Robb, West, & Tyler, 2010).

Callan (1998) noted that growth in K-12 and higher education enrollments during the 20th century put pressure on both systems to ensure provision of high quality education to all students, particularly at the K-12 level. Postsecondary systems were especially impacted by the passage of federal legislation such as the GI Bill, the Civil Rights Act of 1964, and federal

financial aid initiatives which provided older and more diverse groups of students access to higher education for the first time.

Callan (1998) went on to describe the outcome as a “friendly divorce” (p. 51) between K-12 and higher education beginning in the 1960s. From that time K-12 and higher education, including teacher preparation, evolved as separate and independent systems without any formal mechanism in place to connect them. The result was a level of disconnect between the systems, which in the mid 1980s became increasingly viewed as problematic (Futrell, 2010; Kirst & Venezia, 2001).

An ensuing wave of educational reform created renewed interest in cooperation between the two systems. Increased emphasis on teacher quality not only raised the question of how much teacher preparation contributed to the learning success of K-12 students, but also called for reestablishing a connection by linking TPP accountability to K-12 education student outcomes (American Association of Colleges for Teacher Education [AACTE], 2011; Crowe, 2010; Duncan, 2010; USDOE, 2009b). Since that time, there has been no lack of scrutiny of teacher preparation programs in today’s climate of increased accountability.

Government Involvement in Accountability

Though not mentioned in the Constitution, the federal government has assumed de facto control over education by tying the receipt of federal funds to numerous compliance regulations (Bankston, 2010; Phillips & Hawthorne, 1978). Federal economic and political policies designed to advance the state of education have significantly influenced the evolution of accountability and expanded the federal government’s role in education (Adams & Kirst, 1999). Recently, five pillars of educational reform were identified by President Obama (2009): (a) investing in early childhood initiatives; (b) encouraging better standards and assessments; (c) recruiting, preparing,

and rewarding outstanding teachers; (d) promoting innovation and excellence in schools; and (e) providing every American with a quality higher education. Federal policies supporting these reforms include (a) The Higher Education Opportunity Act of 2008 (HEOA, 2008), (b) the No Child Left Behind Act of 2001 (NCLB, 2002), (c) the Individuals with Disabilities Educational Act of 2004 (IDEA, 2004), (d) the Education Sciences Reform Act of 2002 (ESRA, 2002), and (e) the Race to the Top (RTTT, 2009).

The Higher Education Opportunity Act of 2008. The Higher Education Opportunity Act of 2008 (HEOA) was enacted on August 14, 2008. Although largely focused on expanding college access and preparing minority students for competitive and innovative careers, the act also deals with the issue of accountability (HEOA, 2008).

Partly in an effort to increase the accountability of IHEs for student learning outcomes, some policy makers sought to include accountability measures similar to those of the No Child Left Behind Act of 2001 in the HEOA (Lowry, 2009). The Spellings commission recommended that student achievement be measured using a value added approach with the results being made public. Furthermore, the commission recommended that the results should be presented in such a way as to allow all stakeholders to make comparative judgments about the relative effectiveness of different IHEs (Spellings, 2006) These suggestions were resisted by IHEs (Lowry, 2009) and in the end, they were left to define student success for themselves and not required to apply external standards for judging the success of IHEs (HEOA, 2008, section 496).

However, there was an exception for teacher-training programs. The Higher Education Amendments of 1998, Title II, in part, was designed to improve the quality of teaching. Additionally, this provision authorizes accountability and reporting systems regarding the quality of teacher preparation, including the pass rates of graduates of schools of education on

certification exams. Furthermore, section 208 provides for IHEs to obtain K-12 student data from the states in order to evaluate the effectiveness of both program graduates and the program itself.

The No Child Left Behind Act of 2001. Originally enacted as part of the Johnson Administration's War on Poverty (Kantor, 1991), the Elementary and Secondary Education Act of 1965 (ESEA), currently The No Child Left Behind Act of 2001 (NCLB), retains the mission of improving educational outcomes for disadvantaged students (ESEA, 1965; NCLB, 2002). While the current law's requirements for school accountability in the form of testing and adequate yearly progress (AYP) receive the most attention, one provision of NCLB (2002) requires that all teachers be highly qualified.

NCLB (2002) mandates all public schools in America to ensure that all students are taught by highly qualified teachers. In order to be considered highly qualified under the provisions of NCLB (2002), an educator must be licensed or certified by a state, hold, at minimum, a bachelor's degree, and demonstrate thorough knowledge of the subject matter being taught. Demonstration of thorough knowledge may be in the form of a proficiency exam administered by the state, attainment of a degree in that subject, or by some level of experience as defined by the state (pp. 1959-1960).

Though NCLB expired in 2007, it is automatically extended until officially reauthorized by congress and signed by the President. To date, the reauthorization of NCLB has been addressed by the Obama administration in the ESEA Blueprint for Reform, draft legislation in the U.S. Senate, and finally by additional draft legislation in the U.S. House of Representatives.

The Obama administration's blueprint for the reauthorization of the NCLB states that the current mandates concerning teachers will remain in effect, though with more flexibility, as stakeholders transition from a focus on teacher qualifications to one on teacher effectiveness.

While highly qualified has been supplanted by highly effective in language surrounding teachers, the reauthorization of NCLB still identifies teacher qualifications as a key indicator of performance. The blueprint also promotes the public reporting of teacher preparation programs graduates' impact on student outcomes (USDOE, 2010).

The Senate draft of the proposed legislation is titled the Elementary and Secondary Education Reauthorization Act of 2011. In a letter to the bill's co-sponsors, Senator Lamar Alexander summarized the legislation as keeping the reporting requirements of NCLB while giving educational authorities at the state and local level the discretion to determine whether schools are succeeding yet still rife with burdensome federal mandates (Alexander, 2011, S6572). Some of the key provisions of the bill are:

- It would encourage states and districts to tie teacher evaluations to student achievement and consider student growth as well as minimum grade-level standards.
- It would eliminate adequate yearly progress for most public schools but reinstate it through mandates, definitions and regulations tied to identifying low performing schools and requiring the continuous improvement of all schools.
- It would retain federal control of determining whether teachers are highly qualified.
- It would prevent school districts from having the discretion of how to best spend federal funds.

In the House of Representatives a piecemeal reauthorization of ESEA is taking place. A series of legislative acts designed to reform No Child Left Behind are currently moving through the House. At the time of this writing, the Setting New Priorities in Education Spending Act (H.R. 1891, 2011), the Student Success Act (H.R. 3989, 2012), the Encouraging Innovation and Effective Teachers Act (H.R. 3990, 2012) and the State and Local Funding Flexibility Act (H.R. 2445, 2011) are on the Union Calendar of the House. The Empowering Parents through Quality

Charter Schools Act (H.R. 2218, 2011) has passed the House and is currently in committee in the Senate.

The more significant changes to the current legislation by the Student Success Act, which addresses provisions of Title I of No Child Left Behind, are greater flexibility in the use of Title I funds while eliminating AYP, testing of students in science, the School Improvement Grant program, and the highly qualified teacher requirement are provided. The term effective has generally replaced highly qualified.

As with the Student Success Act, the Encouraging Innovation and Effective Teachers Act focuses on effective. The bill makes teacher evaluation a function of the states and in effect, eliminates award grants to states and school districts to improve student achievement using evidence-based and innovative practices.

The Individuals with Disabilities Act of 2004. The Individuals with Disabilities Improvement Act of 2004 (IDEA) has undergone several reauthorizations since it originated in 1975 as the Education for All Handicapped Children Act (EHA). One purposes of the 2004 reauthorization is to ensure that mechanisms are in place to improve outcomes for students with disabilities one of the specific items mentioned in the law is coordinated research and personnel preparation (IDEA, 2004). In 2004, the Individuals with Disabilities Act was reauthorized and brought into line with NCLB in requiring special education teachers to meet the same highly qualified standards as general education teachers. (National Dissemination Center for Children with Disabilities [NICHCY], 2010).

The application of NCLB criteria to exceptional educators, especially those who teach multiple core academic subjects, placed many of them in limbo. While a teacher may be

considered highly qualified in exceptional education, they might not be considered highly qualified to teach core content (NCLB, 2002).

The Education Sciences Reform Act of 2002. The Education Sciences Reform Act of 2002 (ESRA, 2002) established the Institute of Education Sciences (IES), the research arm of the USDOE. The IES' mission is to expand the general public's knowledge and understanding of early childhood through post-secondary education by providing reliable information regarding the current state of education, educational practices that promote student success, and the effectiveness of educational programs (ESRA, 2002).

The intent of the legislation was to increase the standards of educational research and make education an evidence-based field in which policy makers used data to drive decisions affecting large numbers of students (Feistritzer & Haar, 2006).

The application of data driven decision-making became part of the foundation for evaluating teacher preparation due in large part to complimentary federal legislation (e.g. the Higher Education Act, IDEA 2004, Race to the Top, and No Child Left Behind) (Feistritzer & Haar, 2006). Legislation that not only instituted standards related to teacher preparation but also requirements that states, institutions of higher education, and other entities publicly report their success in meeting those standards (Feistritzer & Haar).

Race to the Top. The Race to the Top Fund (RTT) was a \$4.35 billion grant program funded through the American Recovery and Reinvestment Act of 2009 (ARRA, 2009). The ARRA supported investments in innovative strategies likely to lead to improved student outcomes (ARRA, 2009). While there were many components of RTT related to improving student outcomes, a key focus was on teacher preparation.

As evidenced by accounting for the greatest number of points (158) in the grant selection criteria, improving teacher quality is an essential component of education reform. Furthermore, the RTT executive summary included specific definitions for the terms:

effective teachers: those “whose students achieve acceptable rates (at least one grade level in an academic year) of student growth” (p. 12)

student achievement: “a student’s score on the State’s assessments under the ESEA [Elementary and Secondary Education Act]; and, as appropriate ... other measures of student learning ... provided they are rigorous and comparable across classrooms” (p. 14), and

student growth: “the change in student achievement for an individual student between two or more points in time” (p. 14)

The precision of these definitions lays the foundation of RTT’s teacher quality initiatives and preparation program accountability requirements. Additionally, RTT asks grantees to enact rigorous accountability standards while establishing teacher preparation programs “that are successful at producing effective teachers” (p. 10). Thus, RTT requires grantees to (a) link student outcomes to their teachers; (b) tie student outcome data to teacher preparation programs; and (c) publicly report teacher preparation program effectiveness.

Research Related to Legislative Effectiveness

While each piece of legislation described above plays an important role in the accountability and teacher quality debate, it is NCLB that has become synonymous with accountability (Graue & Johnson, 2011). Thus the impact of NCLB on student achievement has been the question of foremost importance in existing literature and as such will be the focus here.

Citing results from the National Assessment of Education Progress (NAEP), studies showed a trend of increasing mathematics scores since NCLB was enacted (Dee & Jacob, 2011; Lee, 2006; & Lee & Reeves, 2012; Nichols, Glass, & Berliner, 2012). Despite NAEP indicating

a positive impact of NCLB on mathematics scores, there have been no significant changes identified in reading achievement. Between 2001 and 2007 the percent of students at or above proficient in mathematics in 4th grade increased by 12 percentage points, and four percentage points in 8th grade. Long-term trend NAEP revealed similar mathematics achievement growth for 9 and 13 year-olds beginning in 1999. Fourth grade reading achievement declined during the 1990s but increased in both 2002 and 2005. Eighth grade reading scores have been consistent since 1992. Similar results are shown for 9 year-olds who achieved slight increases since 1999 while 13 year olds remained relatively constant. Additionally, a Center on Education Policy (2007) report analyzed state reported data on students scoring proficient and effect sizes. The findings suggested that since 2002, student achievement improved in substantially more states than it declined. On the other hand, Dee and Jacob (2011) point to research (see Ladd, 2007) which suggests that the positive effects of NCLB on student achievement could be overstated and attributable to factors outside the purview of NCLB.

The concept of highly effective teachers being key to improving the student performance is fundamental to current and proposed federal legislation (Heine, 2006). The federal role in teacher quality is a relatively recent development. Beginning in the 1950s with *Brown v. Board of Education* and the National Defense Education Act of 1958, which included teacher preparation components, the federal government began an expanding role into the issue of teacher quality (Superfine, Gottlieb, & Smylie, 2012). Later, teacher preparation would be included in the HEA and the ESEA. Today, the federal role in teacher quality is found throughout federal education legislation (Superfine, Gottlieb, & Smylie).

Development of Value-Added Models

Value-added modeling (VAM) measures students' learning gains while controlling for external variables such as prior knowledge and demographic characteristics (National Council on Teacher Quality [NCTQ], 2008). VAM originated within the field of economics in the 1960s (Miller & Modigliani, 1961). As applied to education, the origin of VAM is credited to Dr. William Sanders who first published his methodology for the analysis of educational data in 1997 (Sanders, Saxton, & Horn, 1997).

Sanders and Horn (1998) initial model premised that because a student's prior achievement is a controlled variable, there was no need to take other available variables (such as race, peer effects, and socioeconomic status) into account. Sanders and Horn further asserted that the longitudinal nature of value-added models allows each student to act as his or her own control, thereby eliminating the necessity of extrinsic co-variables in the estimate of teacher effects because those variables are already included in students' previous test scores, which are used to predict students' future test scores. However, this is not to imply that student characteristics do not influence student achievement. There are in fact many student related factors such as race, ethnicity, SES, disability status, etc. that have been shown to influence student achievement (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Goldhaber & Liddle, 2012; Lomax & Kuenzi, 2012; McCaffrey et al., 2003).

Since 1997, the development of value added models has continued. Rose, Henry, and Lauen (2011) composed a list of eight commonly used models with brief descriptions of each (Table 1).

Table 1

Summary of Commonly Used Value-Added Models.

Model	Description
Two-level hierarchical linear model (HLM2)	A random effects model that accounts for the clustering of students with teachers in each year and grade level.
Three-level hierarchical linear model (HLM3):	A random effects model that accounts for the clustering of students with teachers in each year and grade level, and of these teachers in each school.
Univariate response model (URM):	An Education Value-Added Assessment System (EVAAS) random effects model that accounts for the clustering of students with teachers and incorporates two previous years' end-of-grade performance but not student background characteristics.
Multivariate response model (MRM):	The original EVAAS model. This model is a “multiple membership, multiple classification” random effects model that accounts for multiple years of students clustering with teachers. The MRM accounts for the effects of all other past and future teachers that a student has.
Student fixed effects (SFE) model:	A longitudinal, within-student (fixed effects) model that controls for all between-student variation by using each student as his or her own control over the duration of the panel.
Teacher fixed effects (TFE) model:	A longitudinal, within-teacher (fixed effects) model that captures between-teacher differences by incorporating an indicator variable for each teacher in the model.
Student fixed effects instrumental variable (SFEIV) model:	An instrumental variable model that uses a variable that is putatively unrelated to student performance to adjust students' prior test scores for unobserved effects that may confound measurement of the teacher effect. The fixed effects imply a longitudinal within student model in which each student is used as his or her own control.
Teacher fixed effects instrumental variable (TFEIV) model:	Same as the SFEIV, except that the fixed effects are estimated directly by teacher indicator variables in the model.

Note. Adapted from *Technical Briefing Report: Comparing Value-Added Models for Estimating Teacher Effectiveness – Executive Summary*, by R. A. Rose, G. T. Henry, and D. L. Lauen, 2011, p. 1. Retrieved from <http://www.dpi.state.nc.us/docs/rttt/reports/2012/vam-summary.pdf>

For those who support value-added modeling, it is seemingly more justifiable to assess teacher effectiveness based on student growth rather than a student simply meeting a minimum standard regardless of how far he or she started above or below that standard. Those who support VAM further maintain that they prevent penalizing teachers assigned large numbers of reluctant learners or unduly rewarding teachers having a disproportionate number of above average students in their class (Ballou, 2002). Admittedly, value-added modeling exhibits some weaknesses. Primarily, it is highly unlikely that every variable influencing student achievement can be identified and measured (Rivken, 2007) making it difficult to justify holding teachers individually accountable for student achievement. Rivken (2007) further asserts that non-random assignment of students to teachers and teachers to schools and classrooms, validity and reliability issues with student assessments, and focusing instruction on only what is tested hinder the determination of true estimates of teacher effects.

Despite these weaknesses, the use of value-added modeling continues to grow as does the divide between those who believe in the virtues of value-added models and support their use and those who doubt their validity (Braun, 2005).

Covariates

As previously stated, student demographic characteristics have been shown to influence student achievement. Because one goal of value added modeling is to isolate the contribution of the teacher to student learning from the contribution of other factors, covariates are utilized. In theory, a covariate will prevent any student learning attributable to that factor from being attributed to the teacher and vice versa. However, as McCaffrey et al. (2004) point out, there are difficulties associated with both including and excluding student level covariates. On the one hand, including a

covariate may attribute some of the teacher effect to that covariate but on the other hand excluding the covariate may attribute part its effect to the teacher.

Whether to include student characteristics as statistical controls in multilevel modeling has been a point of debate. Thum and Bryk (1997) addressed the issue of including what they termed fairness variables. They approached the issue from two points of view. One, that if the interest was in holding teachers and schools accountable for student learning, then it was appropriate and necessary to include covariates to level the playing field because some student groups are more difficult to teach than others. But, if the interest was in high academic standards for every student, the covariates were neither necessary nor appropriate. Because multilevel models are relatively new, there is less information available regarding the importance of demographic control variables in the models than is available regarding status models and it cannot be assumed that they have equal influence in both (McCaffrey et al., 2003). Regardless, it has become common statistical practice to include necessary control variables in multilevel models. As a result of the reporting requirements of NCLB, schools, districts, and states now routinely collect data regarding the following student demographic characteristics, which in turn are often included as covariates in studies using multilevel model.

Student gender. The existence of a gender gap in education has been a point of interest for decades. Hyde and Linn (2006) analyzed 46 meta-analyses of gender differences in several cognitive domains, including mathematical ability. These meta-analyses synthesized over 5000 studies with approximately 7 million participants. Their findings were reported on a common scale using the d statistic, which measured the distance between the means of males and females in standard deviation units. Classifying effect sizes of 0.00 to 0.10 as trivial and 0.11 to 0.35 as

small, they found that 78% the effects for gender differences were either small (48%) or trivial (30%).

Kafer (2007) using data from the National Center for Educational Statistics (NCES), found that between in 2005 the National Assessment of Educational Progress (NAEP) average scale math score for males was only two points higher than that of females. Kafer also found that the Long-Term Trend Test given to 9, 13, and 17 year old students, showed that between 1973 and 2004, the mathematics gap between males and females had closed to within three points with the difference in average scale score declining from eight to three points.

Dee (2007), using data from the Early Childhood Longitudinal Study and NAEP, found that there is no gender gap in mathematics or reading upon entering kindergarten. However, in the third grade a slight achievement gap in mathematics, in favor of males, has appeared. By the time students are 13, this gap has increased by approximately two-thirds, though not statistically significant and for 13- to 17-year olds the gap remains stable.

Ellison and Swanson (2009) examined the gender gap at the upper levels of achievement. Examining data from the American Mathematics Competitions (AMC), they found that the gender gap around the mean student score is so small as not to have any practical importance. However, when examining the upper tail, they found a substantial gender gap in favor of males. For example, the ratio of males to females scoring at least 100 on the AMC 12 was 4.2 to 1. Furthermore, when examining students above the 99th percentile, the ratio increases to more than 10 to 1.

Though somewhat smaller, the gender gap in high achievers persists when examining 2012 SAT (<http://media.collegeboard.com/digitalServices/pdf/research/TotalGroup-2012.pdf>). The score distributions show the ratio of males to females scoring 700 – 800 to be 1.7 to 1.

Student minority status. The existence of an achievement gap based on minority status is not in dispute. NAEP data reveal that between 1990 and 2009, the statistically significant difference in average mathematics scale score between white and Hispanic students remained fairly constant. The average scale score was from 19 – 26 points lower for fourth grade Hispanic students and 24 – 36 points lower for eighth grade Hispanic students. The statistically significant difference in average mathematics scale score between white and black students was larger. The average scale score was from 26 – 31 points lower for black fourth graders and 31 – 40 points lower for black eighth graders (Hemphill & Vanneman, 2011; Vanneman, Hamilton, Baldwin, & Rahman, 2009).

Using the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K), a nationally representative sample of over 20,000 children entering kindergarten in 1998, Fryer and Levitt (2006) found that in the fall of their kindergarten year, after controlling for other factors, black students score 0.099 standard deviations lower than white students. Hispanic students and students of other races score 0.197 and 0.158 standard deviations below white students, respectively. Asian students on the other hand scored 0.258 standard deviations above white students. By the spring of third grade, black students scored 0.382 standard deviations lower than white students. Hispanic students and students of other races scored 0.078 and 0.244 standard deviations below white students, respectively and Asian students scored 0.163 standard deviations above white students.

Clotfelter, Ladd, & Vigdor (2009), studied third through eighth grade students in North Carolina and found large gaps in the mean achievement between black students and white students but the gaps did not grow over time. They also found that while other minority groups (Hispanic & American Indian) also had achievement gaps, they were not as large as the black-

white achievement gap and tended to dissipate over time. Additionally, the study, consistent with other research, found that Asian students performed better than white students.

Student language program status. Abedi and Dietel (2004) found that the number of ELL students deemed proficient (each state independently defines proficiency) on state assessments was generally 20 – 30 percentage points lower than the number of non-ELL students. Other studies have consistently shown the existence of an ELL achievement gap as well.

Abedi and Gándara (2006) point out that 79% of ELLs did not meet proficiency standards for the California state assessment in 2005. Fry (2007) found that the 2005 NAEP indicated that 46% of ELL fourth grade students were below basic in mathematics and that the gap widened at the eighth grade. Fry (2008) found that ELLs tend to go to public schools. That those public schools generally have lower overall achievement scores and have higher concentrations of students who traditionally perform poorly on standardized tests. And that for ELLs who are not in those schools, the achievement gap narrows significantly.

Utilizing the NAEP Data Explorer (<http://nces.ed.gov/nationsreportcard/naepdata/>), reports were generated to obtain NAEP ELL data for 1996 through 2011. Between 1996 and 2011, the difference in average mathematics scale score between ELL and non-ELL fourth grade students was from 22 – 26 points lower for ELL students. The difference in average mathematics scale score between ELL and non-ELL eighth grade students was from 35 – 44 points lower for ELL students.

Student socioeconomic status. As with previous covariates, the impact of student socioeconomic status (SES) has been well documented. Sirin (2005) conducted a meta-analysis of journal articles published between 1990 and 2000. The sample included 101,157 students,

6,871 schools, and 128 school districts. The results showed a medium relationship between SES and student achievement at the student level and a strong relationship at the school level. The mean effect size at the student level was approximately 0.28 and the mean effect size at the school level was approximately 0.66.

Graham and Provost (2012) used data from the Early Childhood Longitudinal Study to predict the growth in mathematics achievement between kindergarten and eighth grade for average students. The researchers found that SES had a substantial effect on mathematics achievement growth between kindergarten and eighth grade.

Using cutoff points of the 90th, (high SES) and 10th (low SES) percentile in family income, Graham and Provost (2012) determined that children from low SES families enter kindergarten with lower mathematical achievement and make fewer gains during elementary and middle school than do their affluent peers.

Utilizing the NAEP Data Explorer (<http://nces.ed.gov/nationsreportcard/naepdata/>), reports were generated to obtain NAEP FRL data for 1996 through 2011. Between 1996 and 2011, the difference in average mathematics scale score between FRL and non-FRL fourth grade students was from 22 – 26 points lower for FRL students. The difference in average mathematics scale score between FRL and non-FRL eighth grade students was from 26 – 30 points lower for FRL students.

Student ESE status. There can be little doubt that there is an achievement gap between exceptional education and general education students. Eckes and Swando (2009) found that students with disabilities (SWD) not meeting proficiency standards is the primary cause of schools not making adequate yearly progress (AYP) as defined by NCLB. They state that in

Indiana, 50% of schools did not make AYP for the 2005 – 2006 school year and 80% of those were because the exceptional education subgroup did not make AYP.

In reviewing data covering two school years from California, Texas, and Florida, Eckes and Swando (2009) found, with a single exception, that the exceptional education subgroup had markedly lower percentages of students achieving mathematics proficiency than any other subgroup. The researchers found that while both SWD and their non-disabled peers increase their proficiency over time, SWD simply do not close the gap.

Wei, Lenz, and Blackorby (2012) used data from the Special Education Elementary Longitudinal Study (SEELS) to analyze the mathematics achievement of a nationally representative sample of 7 to 17 year old students. Their results indicated that SWD had lower initial math achievement and grew more slowly than their non-disabled peers at the elementary level. However, at the secondary level, the SWD rate of growth plateaued and became similar to that of non-disabled students. While SWD started school with widely varying achievement scores, they made similar gains in mathematics achievement regardless of their qualifying condition.

Utilizing the NAEP Data Explorer (<http://nces.ed.gov/nationsreportcard/naepdata/>), reports were generated to obtain NAEP SWD data for 1996 through 2011. Between 1996 and 2011, the difference in average mathematics scale score between SWD and non-SWD fourth graders was from 20 – 29 points lower for SWD. The difference in average mathematics scale score between SWD and non-SWD eighth graders was from 37 – 46 points lower for SWD.

In addition to the achievement issues is the issue of disproportionality in exceptional education. Each of the other traditionally low performing subgroups (minority, ELL, and

economically disadvantaged) has historically been overrepresented in exceptional education (Linn & Hemmer, 2012; Skiba et al., 2008; U. S. Commission on Civil Rights [USCCR] 2009).

Student gifted status. A search for recent literature on whether gifted students experience higher mathematics achievement than non-gifted students produced limited results. Delcourt et al. (2007) compared students in gifted programs to high achieving students in districts without gifted programs and found that gifted participants performed better on achievement tests. Bhat (2009) drew a sample of 5,265 students in 530 schools from the National Educational Longitudinal Survey of 1988 (NELS) and found the gifted education had a strong effect on mathematics achievement. Bui, Craig, & Imberman, (2011) used a regression discontinuity design and found that participation in gifted programs had no impact on standardized test scores. Adelson et al. (2012) used a propensity score matching analysis and found that the achievement of gifted students is no higher than that of non-gifted students. In studies on the characteristics of students in gifted programs, males and students with high SES (McBee, 2006) were among those most likely to be recommended for gifted programs.

The paucity of research relating to the achievement of gifted students is not surprising. It is generally assumed that gifted students would naturally be high achieving students. Therefore, much of the research focuses on training teachers of the gifted or on identifying the characteristics of gifted students (Bhat, 2009).

Teacher years of experience. Evidence suggests that the effects of teacher preparation programs on student outcomes decay over time (Goldhaber, Liddle, & Theobald, 2013). However, this decay is accompanied by increased rates in teacher effectiveness during the initial three to five years of experience (Clotfelter et al., 2007; Goldhaber & Liddle, 2012; Henry, Fortner, Bastian, 2012; Koedel, et al., 2012). There is additional evidence that while the rate of

growth slows, teachers' level of experience continues to affect student achievement beyond the first five years (Papay & Kraft, 2010; Teach Plus, 2009; Wiswall, 2010).

The importance of years of experience as a factor in student outcomes is made clear when teacher attrition is examined. Henry et al. (2012) state that the mode value of years of experience for teachers in the United States dropped from 15 in 1988 to 1 in 2008. They further assert that after five years approximately 50% of novice teachers have left the profession. Also of note was, that for some subjects, teachers who persisted beyond five years were more effective in their novice years than teachers who left the profession.

In an examination of the effects of teacher turnover on 850,000 fourth- and fifth- grade student observations over eight years in New York City, Ronfeldt, Loeb, and Wycoff (2013) found that students are negatively impacted by higher rates of teacher turnover in both mathematics and language arts. Furthermore, these effects are stronger in schools with higher concentrations of typically low performing student groups. As large an issue teacher turnover is, it becomes an even larger issue when exceptional education teachers are considered.

According to a fact sheet prepared by the Higher Education Consortium for Special Education ([HECSE] hecse.net/policy_documents/FactSheetSPED%20Shortages.pdf), the attrition rate of exceptional education teachers (13% annually) is double that of general education teachers. Additionally, 60% of alternatively certified teachers and 30% of traditionally certified teachers leave exceptional education within three years of certification. Finally, the cost of replacing exceptional education teachers who leave is estimated to be between \$2.2 and \$2.6 billion dollars a year.

Legislative Issues Surrounding Value-Added Models

In 2005, the United States Department of Education (USDOE, 2005) announced it would accept applications to allow as many as ten states to utilize value-added modeling to measure and report Adequate Yearly Progress (AYP). That same year, the Data Quality Campaign (DQC) identified ten essential elements of a statewide longitudinal data system needed to improve student outcomes. In 2005 zero states implemented all ten; by 2011, there were 36 (<http://www.dataqualitycampaign.org>). In 2007, the America COMPETES Act (2007) codified “Required Elements of a Statewide P-16 Education Data System” (§ 6401(e)(2)(D)), which incorporated the ten DQC elements. In 2009, the American Recovery & Reinvestment Act (ARRA, 2009) made State Fiscal Stabilization Funds (SFSF) available to states committed to establish data systems that contained these elements. Also in 2009, the number of states using value-added models to measure and report AYP had grown to 15 (USDOE, 2009b).

Research Addressing the use of Value-Added Models

The concept of applying the econometric production function to education dates to the release of *Equality of Educational Opportunity* (Coleman, 1966), commonly referred to as the Coleman Report. Because the production function in the context of economics includes human resources, developing an education production function to measure teacher effects became a focal point of much educational research. Such studies typically examined the relationship between the resources put in to the educational process and student outcomes. Generally, these studies were limited to a single dependent variable (e.g. some standardized test score) and dependent variables that were some measure of student, teacher, or school characteristics (Klein, 2007).

Inspired by the seminal research findings of Sanders & Rivers (1996), many researchers modified their research agendas to examine teacher effects as opposed to school effects such as school size, class size, and per student funding in order to determine the effects of teachers on student achievement as measured by standardized tests. Hanushek and Rivkin (2010) summarized research finding significant differences among teacher effects on student achievement (Table 2).

Table 2

Estimates of Within School Variation in Teacher Effectiveness.

Study	Location	Test subject	
		Reading	Math
Rockoff (2004)	New Jersey	0.10	0.11
Nye, Konstantopoulos, and Hedges (2004)	Tennessee	0.26	0.36
Rivkin, Hanushek, and Kain (2005)	Texas	0.15	0.11
Aaronson, Barrow, and Sanders (2007)	Chicago		0.13
Kane et al. (2008)	New York City	0.08	0.11
Jacob and Lefgren (2008)	Undisclosed city	0.12	0.26
Kane and Staiger (2008)	Los Angeles	0.18	0.22
Koedel and Betts (2009)	San Diego		0.23
Rothstein (2010)	North Carolina	0.11	0.15
Hanushek and Rivkin (2010a)	Undisclosed city		0.11
Average		0.13	0.17

Note. All estimates indicate the standard deviation of teacher effectiveness in terms of student achievement standardized to mean zero and variance one. All variances are corrected for test measurement error and except Kane and Staiger (2008) are estimated within school-by-year or within school-by-grade-by-year. Corrected reading estimates are included for Rivkin et al. (2005).

Note. From: Generalizations about using value-added measures of teacher quality, by E. A. Hanushek and S. G. Rivkin, (2010). *American Economic Review*, 100(2), pp. 267-271.

The studies indicate that a one standard deviation difference in teacher effectiveness led to a change in student achievement of 0.11 – 0.36 student-level standard deviations in mathematics. Considering that research in Tennessee produced effect sizes of 0.2 student-level

standard deviations as a result of decreasing class size from 22 to 15 students gives some perspective to the significance of these findings (Krueger, 2003).

Teacher Preparation Program Effectiveness and Value-added Models

For all intents and purposes, teacher preparation programs are generally regulated through accreditation and evaluation. Yet many view accrediting agencies as ineffective instruments of quality control (Cochran-Smith & Zeichner, 2005; Crowe, 2010; Duncan, 2010; Levine, 2006; NCATE, 2010). The perceived lack of quality control in teacher preparation has led to calls for reform including holding programs accountable for K-12 student achievement (Berry, Fuller, Reeves, & Laird, 2007; Duncan, 2010; Teaching Commission, 2006). Additionally, gauging the success of teacher preparation programs by the effectiveness of their graduates was a principle edict of the Race to the Top grant competition.

In spite of the fact that there are more than 1400 teacher education programs in the United States there is little research linking those programs to the performance of their graduates as measured by student achievement (National Research Council, 2010). Only recently have researchers used large databases to extend the link between student achievement and teacher effectiveness to teacher preparation programs (Boyd, et al., 2009; Goldhaber & Liddle, 2012; Harris & Sass, 2007; Henry et al., 2011; Henry, et al., 2010; Koedel, et al., 2012; Mihaley, et al., 2012; Noell, et al., 2009; Noell, et al., 2007; Noell, et al., 2008). The paucity of studies is not due to a lack of interest in tying student outcomes to teacher preparation programs, but rather because the data and the means to process the data did not exist until recently.

Boyd et al. (2009) examined the distribution of the average value-added scores of teachers from different teacher preparation programs supplying many elementary teachers to New York City schools. The analyses employed several hierarchical linear models utilizing fixed

school effects, random school effects, and ordinary least squares (OLS) specifications, in order to test the robustness of the results.

The initial analysis of teachers grouped by program and institution was modeled with student achievement as a function of their prior achievement, time-varying and fixed student characteristics, classroom characteristics, teacher characteristics, teacher preparation program completed (fixed), a fixed effect for school, and a random error term. The sample consisted of multiple cohorts of first and second year teachers and 31 teacher preparation programs (26 traditional and 5 alternative) (Boyd et al., 2009).

Their findings suggest that significant variation exists in the effectiveness of teachers prepared at different programs. The variation between teachers from the average and highest performing programs was similar to the variation between those students eligible for free or reduced price lunch and those who were not. The variation was also similar in both language arts and mathematics. Likewise, programs that produce effective language arts teachers also tend to produce effective mathematics teachers. The findings also indicate that programmatic features of preparation programs can also affect student outcomes (Boyd et al., 2009).

Goldhaber and Liddle (2012) attempted to estimate models that would identify the effectiveness of teachers who received their initial certification from different preparation programs from within and without the state of Washington by regressing student achievement against prior achievement controlling for student, classroom, teacher, credentialing program, school, and district characteristics. The teacher characteristics included variables associated with their credentials. The sample included approximately 8,700 elementary teachers and 294,000 students in grades 3 through 6 and spanned five academic years. The key area of interest of the study was the estimated coefficients for the various credentialing programs.

One of the confounding issues for the study was reported as being able to separate the individual attributes of teachers from the effects of their training program. One way the researchers tried to account for the causal impact of teacher preparation program was by including as control variables the results of tests that potential teachers take before entering a teacher preparation program (Goldhaber & Liddle, 2012).

As with other studies, the results noted that there is a disparity in student achievement between subgroups based on race/ethnicity. Also, student achievement rises with years of teaching experience, plateauing at year 5. As to the key area of interest, what proportion of student achievement is explained by teacher preparation program, the results indicated that less than one percent (.65% in mathematics) of the total variation student achievement. However, the amount of variance in mathematics achievement explained by preparation program was greater than that explained by teacher and credentialing characteristics such as race, gender, degree level, and experience (Goldhaber & Liddle, 2012).

Harris and Sass (2007) investigated the relationship between teacher productivity and teacher training in Florida. Restricting their analysis to only students who received instruction in the relevant subject area in only one classroom, they parse the teacher-school effect into three parts, (a) teacher effect resulting from undergraduate education, (b) teacher effect attributable to pre-college ability, and (c) school effect. The impact of pre-service education on student achievement is estimated by “regressing the estimated teacher-school effects on a vector of pre-service education variables for teacher, their entrance exam scores, a set of school indicators, and a random error” (pp. 14-15).

Findings indicated that there is no relationship between teachers’ undergraduate preparation and student achievement regardless of the type of undergraduate degree the teacher

held. Of note, however, is the fact that while not associated with student achievement, controlling for pre-college ability in the form of SAT or equivalent entrance exam scores rendered all other college major effects insignificant. While the study found no relationship between teacher preparation programs and student achievement, it does lend weight to the premise that such analyses are viable (Harris & Sass 2007).

Henry et al. (2011) sought to estimate the effect of different methods of teacher preparation on student achievement in mathematics and other subjects across elementary and secondary grade levels by estimating the effectiveness of teachers who entered teaching via one of 11 different portals (six traditional and five alternative). This was accomplished by comparing the test score gains of students taught by teachers who entered teaching through the various portals to the gains of those taught by traditionally prepared teachers, controlling for an array of student, classroom, and school characteristics. The analyses were performed using year to year value added models because the researchers believed these models would adjust for differences between students or schools related to test score gains when estimating teacher preparation program effects.

The researchers began by standardizing all student test scores. Then covariates at the student, classroom, and school levels were chosen in an effort to allow for the statistical adjustment of potential plausible threats to imbalance. Because value added models with student fixed effects use students as their own control, they were only viable for students in grades 3-8. Estimates of preparation program effects were made by comparing each portal for each grade/subject to a reference group comprised of in-state public undergraduate prepared teachers in their fifth year of teaching resulting in 97 comparisons (Henry et al., 2011).

Of the 97 comparisons, students taught by in- state public undergraduate prepared teachers performed better in 14% of the comparisons, worse in 9%, and without significant differences 76% of the comparisons. Interestingly, findings of the study indicate that a mismatch may exist between where some teachers are placed and where they might be more effective (Henry et al., 2011).

Henry et al. (2010) conducted a study to isolate the effects of University of North Carolina system teacher preparation programs on student outcomes in North Carolina. To accomplish this task, the authors linked individual student test scores to the teacher who taught the class in a tested subject (i.e. mathematics). Based on the principle that the effects a teacher preparation program will diminish over time, the researchers limited their sample to teachers with less than ten years experience. To isolate teacher preparation program effects, a year-to-year multilevel, value added model with a large number of controls consisting of student, classroom, teacher, and school data was employed.

These controls included prior year test scores in reading and mathematics. The stated purpose of including prior year scores, was to ensure that “neither individual teachers nor teacher preparation programs get credit or blame for factors that are beyond their control” (p. 4). The authors pointed out that in controlling for the chosen variables, they were not assuming that said variables in any way impacted student achievement. Instead, the inclusion of controls allowed the models to detect effects of the controls if any existed and allowed those effects to be separated from the effects of the teacher preparation programs (Henry et al., 2010).

The findings of this study show that it is possible to estimate the effects of teacher preparation programs on student achievement of students who their graduates teach. If the

teacher preparation programs in the University of North Carolina system were judged solely on this criteria, traditional undergraduate programs could be considered as doing a slightly better job of preparing teachers than all other programs and the Master of Arts in teaching programs neither better nor worse than other pathways (Henry et al., 2010).

Koedel et al. (2012) examined the extent to which teachers prepared at different teacher preparation programs differ in effectiveness using value-added models. The study sample consisted of 1,309 teachers and their students (61,150 mathematics) at 656 elementary schools (389 of which employ teachers from multiple programs) certified through one of 24 major (having prepared in excess of 15 teachers) preparation programs in Missouri with programs producing 50 or more teachers evaluated separately. To be included in the study, teachers must have had begun their teaching career no earlier than 2004, been recommended for certification within three years of the date of initial employment, and teach in grades 4, 5, or 6. The study spanned the 2008 through 2011 academic years.

As with other studies, there are confounding issues addressed. First, there is the issue of selection. For this study, ACT scores were used to investigate the impact of selection. Results indicated that the variance of average ACT scores is largely within institutions. For example, graduates from one university who enter public school teaching have lower ACT scores than other students from the same university while at other universities, future public school teachers have ACT scores similar to those of students who do not enter teaching. These findings indicate that teachers from institutions with more stringent entrance requirements did not outperform other teachers (Koedel et al., 2012) .

Second, the study included only traditional teacher preparation programs. The authors acknowledge that had alternative certification programs been included in the study, additional

heterogeneity might have existed across programs leading to increased variances (Koedel et al., 2012).

The models were constructed with some containing student and school level characteristics as controls. The student characteristics included race, gender, free/reduced-price lunch status, language-learner status, and mobility status. The school level characteristics included aggregates for each of the student characteristics. Models were specified both with and without fixed school effects. Of note for this study is the fact that the “standard errors are clustered at the individual-teacher level throughout our analysis to properly reflect the data structure” (Koedel et al., 2012, p. 10). The authors reported on three specific models, (a) Model A which includes the lagged student test score, student-level controls, controls for teacher experience, and the preparation program indicators; (b) Model B which includes everything in Model A, plus school-level aggregates analogous to the student-level controls; and Model C which includes everything in Model A plus school fixed effects (p. 14).

The main findings of the study are that the variance in student achievement attributable to teacher preparation program is very small and that differences in the variance lie within programs and not necessarily between programs. Of additional note is the authors’ discussion regarding the effects of at which level of a model clustering of the standard errors occurs can influence the interpretation of results. The authors state that previous studies have employed incorrect clustering leading to reported standard errors that are too small. The authors further suggest that the individual teacher level is the appropriate level of clustering (Koedel et al., 2012).

Mihaly et al. (2012) investigated the issues involved with using school fixed effects when using multilevel models to estimate the effectiveness of teacher preparation programs. School

fixed effects are often included in models to estimate teacher preparation program effects as they control for the school level characteristics. Because school fixed effects rely on differences among student outcomes within the same school to identify teacher preparation program effects, teachers from different preparation programs must teach in that school. However, if the training programs are not connected to one another the model estimates may not be feasible.

Of primary concern was that the clustering of graduates of a specific teacher preparation program in a particular school district or geographic location would inflate the variances of the estimates of teacher preparation program effects. Using data from the 2000 through 2004 academic years strong indications of regional clustering among program graduates was detected. However, there were enough graduates working a great enough distance from their preparation program and enough programs located in close geographic proximity to one another so that the entire network of programs was fully connected provided at least three years of data were combined (Mihaly et al., 2012).

The authors also addressed selection bias as have other studies relating to the effectiveness of teacher preparation programs. But added that selection bias was not a part of this particular study. The findings related to the issue at hand for this study found that while graduate clustering confounded the results of estimating teacher preparation program effectiveness when employing school fixed effects in models, this issue could be overcome with the combination of data across a large enough time span (three years) (Mihaly et al, 2012).

Noell et al. (2007, 2008, 2009), utilizing the state of Louisiana's educational administrative database, conducted value-added analyses to determine the effectiveness of new teachers as compared to that of more experienced teachers from state's teacher preparation

programs. These studies yielded three reports that describe the development of the Value Added Teacher Preparation Assessment Model as well as the results of the analyses.

Effect estimates were generated using a multilevel model (MLM) for all teacher preparation pathways. The results showed the mean expected effect of a teacher preparation program compared to that of experienced certified teachers. For example, an effect estimate of 5.0 indicated that a student of the average completer of a specific university's teacher preparation program would score 5.0 points higher on the state exam than students taught by experienced certified teachers. An effect estimate of -5.0 would indicate that said student would score 5.0 points lower (Noell et al., 2007; 2008; 2009).

Rather than ranking the state's teacher preparation programs, the researchers chose to place them in one of five levels

1. programs for which there is evidence that new teachers are more effective than experienced teachers, but this is not necessarily a statistically significant difference,
2. programs whose effect is more similar to experienced teachers than new teachers,
3. programs whose effect is typical of new teachers,
4. programs for which there is evidence that new teachers are less effective than average new teachers, but the difference is not statistically significant, and
5. programs whose effect estimate is statistically significantly below the mean for new teachers (Noell et al., 2007; 2008; 2009).

The 2007 study was complicated by the fact that the state's teacher preparation programs were going through a statewide redesign immediately prior to the analysis and thus only three programs qualified for the study post re-design and all three were alternative certification

programs. For these three programs, one was at level 1 (see levels in preceding paragraph), one was at level 2 and the third was at level 3 in mathematics. Of the 12 programs for whom pre-redesign data were analyzed, one was at level 5, two were at level 4, and the remaining eight were at level 3 in mathematics (Noell et al., 2007).

In 2008, the study contained only six programs (five alternative certification and one traditional) who met study inclusion qualifications in mathematics. Of these six programs, one was designated level 1, two level 2, and three level 3 (Noell et al., 2008). In 2009, the study contained eight programs (two traditional and six alternative certification) who met study inclusion qualifications in mathematics. Of these eight programs, one was designated level 1 and seven level 3 (Noell et al., 2009).

Their findings were generally consistent over time showing variability in teacher effectiveness across teacher preparation programs. Additionally, most programs remained at the same effectiveness level across years moving only one level if they moved at all. The results also suggest that given sufficient data, producing value added estimates of teacher preparation program effectiveness that are reasonably stable is possible (Noell et al., 2007; 2008; 2009).

Though previous research addresses teacher preparation, it is important to note the improbability, if not impossibility, of completely disentangling the effects attributable to program selection criteria from those of the actual training participants received while in a program. Provided the limiting factors of the data, it is highly improbable that the effects of candidate selection criteria and the effectiveness of the teacher preparation program attended can be separate. Therefore, in accord with previous research, program estimates produced by this study will in all likelihood reflect the combined effects of selection criteria and training.

CHAPTER THREE: METHODOLOGY

Overview

This study examines the relationship between student characteristics, teacher experience and preparation program attended, and student mathematics achievement as measured by the Texas Assessment of Knowledge and Skills (TAKS). The overall methodology for the study including study design, sample data, data collection procedures, instrumentation, and statistical procedures are discussed. Data were collected from a small school district in north-central Texas over a span of four academic years. The data were analyzed using HLM 7 for Windows (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011). The program also produced a residual file for each level of the model. These files were analyzed using SPSS 17 to determine whether statistical assumptions were satisfied.

A three-level multilevel model (MLM) was employed with repeated measures of mathematics scores at Level 1, students at Level 2, and teachers at Level 3. The purpose of this design was to analyze an organizational structure where individual student scores are nested within each student and students are nested within teachers as depicted in figure 2.

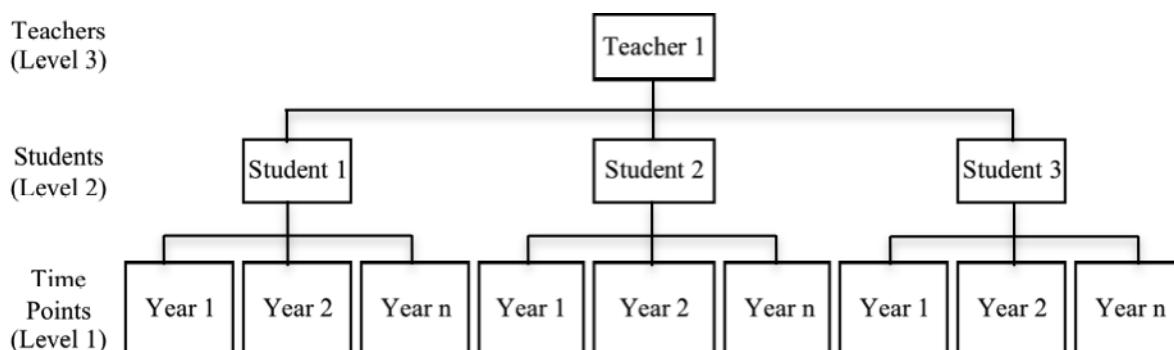


Figure 2. Hierarchical Structure of Data.

Research Design

The research design was non-experimental ex post facto using a census of an intact group of fourth through eighth grade students enrolled in one Texas school district. Setting

The setting for the current study was a Texas school district containing four schools: two elementary schools (K-5), one middle school (6-8), and one high school (9-12). During the 2010 – 2011 academic year, there were approximately 2,000 students and 170 teachers in the district according to the Texas Education Agency (TEA) LONESTAR reporting system (<http://loving1.tea.state.tx.us/lonestar/Home.aspx>),

Participants

Students. The students included in this study were delimited to the following:

1. Student must have been enrolled in grade 4, 5, 6, 7, or 8 during the 2011 TAKS administration.
2. Student must have participated in the 2011 administration of the TAKS Mathematics assessment.
3. Student must be a member of the 2011 campus-level accountability subset².
4. Student must have at least three vertical scale scores during the study period (2008-2011).

In 2011, the TAKS was administered to students in grades 3 through 10 and at an exit level. Vertical scaling (which will be discussed in a following section) of the TAKS began in 2008, but only for grades 3 through 8. Therefore students in grades 9 and beyond during the

² If a student was reported in membership at one campus on October 29, 2010, but moves to another campus before the test, that student's performance was removed from the accountability results for both campuses, whether the campuses were in the same district or different districts. Campuses were held accountable only for those students reported to be enrolled in the campus in the fall and tested in the same campus in the second semester.

2011 administration were excluded. Additionally, any student not having adequate scores to allow computation of a growth trajectory after the 2011 administration were excluded. Similarly, because all students in the state are given the same assessment and student level data is retained by the state, scores from 2008, 2009, and 2010 need not have come from the school district in the current study.

Teachers. Teachers were selected for participation based on the sole criteria of being the teacher of selected students.

Data Source

Data for the current study were obtained from two sources. After written permission was requested from (see Appendix C) and granted by the district's administration (see Appendix D), student data were collected from a data management system which included all student level covariates (i.e., gender, race/ethnicity, exceptional education, gifted, language program , and free/reduced price lunch. The teacher level covariates (teacher preparation program attended and years of service) were obtained from the district personnel office.

Instrumentation

The Texas Assessment of Knowledge and Skills-Mathematics (TAKS) was designed to measure “the extent to which a student has learned and is able to apply the defined knowledge and skills at each tested grade level” (TEA, 2011a, p. 69). In grades 3 – 8 mathematics, the TAKS covers six objectives including (a) numbers, operations, and quantitative reasoning, (b) patterns, relationships, and algebraic reasoning, (c) geometry and spatial reasoning, (d) measurement, (e) probability and statistics, and (f) mathematical processes and tools. Every TAKS test is directly aligned to the Texas Essential Knowledge and Skills (TEKS) (p. 69). The

TEKS were adopted by the Texas State Board of Education in 1997 and implemented as the statewide curriculum in the 1998-1999 academic year.

The TAKS is a criterion-referenced assessment. The items on the grades 3-8 TAKS are primarily multiple choice with some student generated response items (Figure 3).

<p>DIRECTIONS</p> <p>Read each question. Then fill in the correct answer on your answer document. If a correct answer is not here, mark the letter for “Not here.”</p> <p>SAMPLE A</p> <p>Which digit is in the thousands place in the number 4,861,392?</p> <p>A 6</p> <p>B 4</p> <p>C 1</p> <p>D Not here</p>	<p>SAMPLE B</p> <p>Joey has 8 books. Roberto has twice as many books as Joey has. How many books does Roberto have?</p> <p>Record your answer and fill in the bubbles on your answer document. Be sure to use the correct place value.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td>1</td> <td>6</td> <td>.</td> </tr> <tr> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> <tr> <td><input type="radio"/></td> <td><input checked="" type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> <tr> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> <tr> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> <tr> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> <tr> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> <tr> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> <tr> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> <tr> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> <tr> <td><input type="radio"/></td> <td><input type="radio"/></td> <td><input type="radio"/></td> <td></td> </tr> </table>		1	6	.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
	1	6	.																																										
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>																																											
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>																																											
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>																																											
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>																																											
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>																																											
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>																																											
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>																																											
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>																																											
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>																																											
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>																																											

Figure 3. Sample Items 2009 Grade 4 Mathematics TAKS. Adapted from “Grade 4 (English and Spanish) Test Administration Directions 2009 Writing, Mathematics, Reading”. Copyright 2009 by the Texas Education Agency. Reprinted with permission.

There are six versions of the TAKS which a student might be administered, commonly referred to as forms. The test form used by students who receive no testing accommodations is simply referred to as the TAKS, of which there is also a Spanish version for some grade levels. The TAKS forms are described in Table 3. For the purposes of this study, only scores from the TAKS English form were used in the data analysis. This was due to the fact that it is the most widely used assessment and that the scores from the other assessments are not reported on the same scale as the TAKS English version.

Table 3

2010-2011 TAKS Mathematics Assessments.

Assessment	Description
TAKS English: Available for grades 3-10 and exit level	The English language version of the criterion-referenced assessment used to evaluate the academic skills of students who receive academic instruction in English and do not meet eligibility requirements for other forms.
TAKS Spanish: Available for grades 3-5	Spanish-version assessments are designed to evaluate the academic skills of English language learners (ELLs) who receive academic instruction in Spanish while they learn English (TEA, 2011a, p. 70).
TAKS Accommodated: Available for all English-and Spanish-forms TAKS	[F]or students receiving special education services who meet the eligibility requirements for specific accommodations. This is a general assessment based on the same grade-level academic achievement standards as TAKS. The TAKS (Accommodated) form includes format changes (larger font, fewer items per page) and contains no embedded field-test items (TEA, 2011a, p. 70).
Linguistically Accommodated Testing: Available for grades 3-8 and 10	LAT is an assessment process for eligible immigrant ELLs who are granted a limited English proficiency (LEP) exemption under state law but are required to be assessed in certain grades and subjects under federal law. The LAT process enables eligible immigrant ELLs to be assessed with linguistic accommodations that help them better understand the language used on the tests (TEA, 2011a, p. 70).
TAKS Modified: Available in English for the same grades and subjects as TAKS	An alternate assessment based on modified academic achievement standards designed for students receiving special education services who meet participation requirements. It covers the same grade-level content as the TAKS, but the format (larger font, fewer items per page, etc.) and test design (fewer answer choices, simpler vocabulary and sentence structure, etc.) have been changed (TEA, 2011b, p. 117).
TAKS Alternate: Available for the same grades and subjects as TAKS	An alternate assessment based on alternate academic achievement standards designed for students with significant cognitive disabilities receiving special education services who meet the participation requirements. It is not a traditional paper or multiple-choice test. Instead, the assessment involves teachers observing students as they complete standardized state-developed assessment tasks that link to the grade-level Texas Essential Knowledge and Skills (TEKS) (TEA, 2011c, p. 135).

Note. Compiled from information contained within the Texas Educational Agency Technical Digest 2010-2011.

The TAKS assesses six objectives with various numbers of questions for each objective at each grade level (Table 4).

Table 4

2011 TAKS Blueprint for Grades 3-8 Mathematics.

Objectives	Number of Items Measuring Objective					
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Objective 1 - Numbers, operations, and quantitative reasoning	10	11	11	10	10	10
Objective 2 - Patterns, relationships, and algebraic reasoning	6	7	7	9	10	10
Objective 3 - Geometry and spatial reasoning	6	6	7	7	7	7
Objective 4 - Measurement	6	6	7	5	5	5
Objective 5 - Probability and statistics	4	4	4	6	7	8
Objective 6 - Mathematical processes and tools	8	8	8	9	9	10
Total number of items	40	42	44	46	48	50

Note: Adapted from "Texas Assessment of Knowledge and Skills (TAKS) Blueprint for Grades 3-8 Mathematics" Retrieved from <http://www.tea.state.tx.us/student.assessment/taks/blueprints/>

The number of questions a student answers correctly is the raw score. The raw score is useful if multiple forms have identical levels of difficulty from administration to administration. However, this is generally not the case and the number or percentage of items correct on two different forms won't result in comparable assessments of students' knowledge or skills across forms (i.e. different grade levels). To make assessment results meaningful, scores from different forms must be comparable (Livingston, 2004).

Until 2008, Texas reported scores only on a horizontal scale. The horizontal scale allowed comparison across test administrations but not across grade levels. In other words, the

TAKS could be used to determine whether a student achieved proficiency on a specific assessment, to compare one student to another if both were assessed at the same grade level in the same subject, or to compare one group of students to another group at the same grade level in the same subject in different years. In 2008, the TEA developed and began reporting TAKS scores on a vertical scale for grades 3 – 8 to comply with Section 39.036 in S.B. No.1031 (TEA, 2009a).

Vertical scale score. Vertical scaling is the result of a process whereby two or more assessments that have similar constructs measured, but at different levels of difficulty and content are statistically linked. Thus allowing their scores to be expressed on a common scale. This process is known as calibration (Kolen, 2004). The specific actions taken by TEA to develop the TAKS English mathematics vertical scales are reported in the *2008 TAKS English Vertical Scaling Study Report* (TEA, 2009a).

Reliability. Reliability is a measure of how consistently a score is achieved when an assessment is scored on different occasions (Worthen, Borg, & White, 1993). Because reliability estimates can change with each administration of an assessment, it is important to report them with each administration. It is important to understand that reliability estimates are a function of an assessment's score and not of the assessment itself (Thompson, 1999).

There are many methods for estimating reliability. Those that are the result of a single administration are referred to as internal consistency measures. The coefficient alpha is the most commonly used method for obtaining internal consistency reliability estimates. There are three different measures of the coefficient alpha, (a) Cronbach's alpha; (b) the Kruder Richardson 20 (KR20); and (c) Hoyt's method (Crocker & Algina, 1986).

The TEA estimates TAKS reliability using the KR20 for assessments with only multiple choice items and stratified coefficient alpha for tests with a mixture of multiple choice and partial credit items. As a general rule, reliability coefficients from 0.70 to 0.79 are considered adequate, 0.80 to 0.89 are considered good, and above 0.90 are considered excellent (TEA, 2009c). The TAKS mathematics reliability estimates across grades 3-8 for the 2008, 2009, 2010, and 2011 administrations ranged from 0.876 - 0.908 (TEA, 2008: 2009b; 2010; 2011d) (Table 5).

Table 5

TAKS Grades 3-8 Mathematics Reliability Estimates.

Year	KR20 Reliability Estimates					
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
2008	0.8778	0.889	0.893	0.915	0.919	0.912
2009	0.890	0.902	0.902	0.908	0.908	0.905
2010	0.878	0.888	0.902	0.909	0.904	0.907
2011	0.876	0.887	0.902	0.908	0.904	0.906

Note. Compiled from TAKS 2008, 2009, 2010, 2011 mean *p*-values and internal consistency values by objective and subject area, (TEA, 2008, 2009b; 2010; 2011d)

Validity. Validity refers to the extent a score reflects a test takers true knowledge and skills (Worthen et al., 1993). That is, is what is supposed to be measured being measured? There are essentially three types of validity; (a) content validity, (b) criterion validity, and (c) construct validity. TEA explains that because their assessment program “is concerned with the general question of to what extent test scores help educators make appropriate judgments about student performance” (TEA, 2009c, p. 71), they are seeking evidence of content validity. Such evidence would support the assumption that the TAKS assesses students’ knowledge and understanding of the TEKS.

Establishing evidence of validity for scores from the TAKS is an ongoing process and the following steps are taken annually:

- writing items based on test objectives and item guidelines
- reviewing items on more than one occasion for appropriateness of item content and identification of item bias
- field-testing of items
- reviewing field-test data with educators
- building tests to pre-defined criteria
- reviewing high-school tests for accuracy of the advanced content by university-level experts (TEA, 2009c, p. 72)

The Texas Education Agency working in conjunction with Pearson has conducted extensive analyses to determine that the TAKS scores are valid (TEA, 2008b). The methods and results are explained in detail in the TAKS Technical Digest (TEA, 2008b) For example, TEA reports the following:

Results of the study indicated that the TAKS scale scores at the Met Standard performance level predicted ACT scale scores of approximately 20 for mathematics. Based on a national study of high school graduates from 2002 to 2004, 50% of students scored at or above this ACT score. The TAKS scale scores at the Met Standard performance level predicted ACT scale scores of approximately 18 for English. Of the high school students in the ACT data, 67% scored at least this high on the ACT English test (TEA, 2008b, p. 165).

Data Analysis

While several procedures for analyzing hierarchical data exist, a multilevel model (MLM) was deemed most appropriate for this study . MLM is able to simultaneously identify the relationships within each level and between levels and requires meeting fewer assumptions than

other methods (Hoffman, 1997; Raudenbush & Bryk, 2002). Woltman, Feldstain, MacKay, and Rocchi (2012) describe this phenomenon:

HLM can accommodate non-independence of observations, a lack of sphericity, missing data, small and/or discrepant group sample sizes, and heterogeneity of variance across repeated measures. Effect size estimates and standard errors remain undistorted and the potentially meaningful variance overlooked using disaggregation or aggregation is retained (p. 56).

There are several other advantages to using MLM. One advantage of MLM in examining repeated-measures is that at Level 1 it allows the inclusion of all students, including those with missing observations. Assuming that observations were missing at random, each student test observation is treated as a separate case so that only missing data points and not the students having missing data are excluded from analyses. Furthermore, the assumption that every subject must be measured at identical points in time and for an identical number of occasions is unnecessary. This may also be considered a disadvantage because missing data points are allowed only at Level 1. If there are missing data points at levels 2 or 3, those groups will be excluded. Another advantage of MLM is that both continuous and categorical predictors can be used to examine relationships between growth rates and correlates. Finally, MLM allows growth parameters to be estimated even when including relatively small numbers of students (Bryk & Raudenbush, 1992).

Another issue surrounding MLM is that of power and sample size. Generally, MLM requires a large sample size to achieve adequate power. Hoffman (1997) discussed this issue and found that the power of Level 1 largely depended on total sample size (the number of observations) whereas the power at higher levels was dependent upon the number of groups. Hoffman pointed out that a sample of thirty groups with thirty observations ($n=900$) has the same power as one hundred fifty groups with 5 observations ($n=750$). Hoffman went on to state that

the a preference should be placed on collecting data from many groups rather than more individuals per group.

Based on the work of Raudenbush and Bryk (2002), a three-level multilevel growth model was constructed to examine the relationship between student characteristics, teacher experience and preparation program attended, and student TAKS mathematics scores for the 2010-2011 academic year. More specifically, whether teachers' experience and preparation program attended are related to student TAKS achievement when controlling for student level variables will be examined.

Building a three-level growth model with students who were measured on at least three occasions allows for the estimation of an individual growth trajectory and thus an examination of the relationship between teachers' experience, preparation program attended and estimates of students' achievement. Furthermore, incorporating additional covariates such as gender, race, and participation in exceptional education, gifted, language, or free and reduced price lunch programs at Level 2 of the model provides more information regarding whether and to what degree student characteristics contribute to the relationship between teachers' experience, preparation program attended and TAKS Mathematics Vertical Scale Score (TMVSS). Raudenbush and Bryk (2002) highlight the importance of including statistical adjustments for individuals' demographic information as people are not randomly assigned to groups such as gender and race. Failing to control for these variables could bias the effect of experience and teacher preparation program. Additionally, if a predictor variable is strongly related to the outcome variable, controlling for it will reduce the amount of unexplained variance.

The current model was constructed in phases following generally accepted model-building protocols. First, a fully unconditional (Null) model was built, followed by an unconditional growth model, and finally a conditional growth model.

Included in Level 1 are TMVSS and time of TAKS data collection across 4 academic years (i.e., spring 2008, spring 2009, spring 2010, spring 2011). Level 2 variables include various student demographic information including gender, minority status, exceptional education status, gifted status, language program status, and free or reduced price lunch status. Due to the small numbers of student being members of the associated subgroups, minority, exceptional education, language program and free or reduced price lunch status were treated as dichotomous with students either belonging or not belonging to the group as opposed to including each subgroup separately. For example, rather than group exceptional education students according to one of the 13 qualifying categories, all exceptional education students were put into a single group. Finally, the Level 3 variables will be teachers' experience level (experience level refers to the number of years of service a teacher has been credited with by the Texas Education Agency) and preparation program attended (any state approved course of study leading to initial licensure regardless of degree awarded).

Three-level unconditional means model. The first step was to build a model void of predictor variables (Equations 2 – 4). Because the model contains no predictor variables, it is called an unconditional means model or null model. The purpose of the null model is to partition the variance in TMVSS into three components: (1) σ^2 (within students), (2) τ_π (between students within teachers), and (3) τ_β (between teachers) so that:

$$\text{Level 1 (within students):} \quad Y_{ij} = \pi_{0ij} + e_{ij} \quad (1)$$

$$\text{Level 2 (between students):} \quad \pi_{0ij} = \beta_{00j} + \Upsilon_{0ij} \quad (2)$$

$$\text{Level 3 (between teachers/TPPs):} \quad \beta_{00j} = \gamma_{000} + \mu_{00j} \quad (3)$$

where:

- Y_{tij} is the TMVSS at time t (grade level), for student i , of teacher j ;
- π_{0ij} is the initial TMVSS for student i , of teacher j , when the centered grade level equals 0 (third grade);
- e_{tij} is a random effect representing the deviation of the TMVSS of student i , of teacher j , from the predicted score based on the Level 2 model. It is assumed that e_{tij} is normally distributed with a mean of zero and a variance of σ^2 ;
- β_{00j} is the mean grade 3 TMVSS within teacher j ;
- Υ_{0ij} is the random error of the intercept at the student. It is assumed that Υ_{0ij} is normally distributed with a mean of zero and a variance of τ_π .
- γ_{000} is the overall mean grade 3 TMVSS across students and teachers; and
- μ_{00j} is a random error of the intercept at the teacher level. It is assumed that μ_{00j} is normally distributed with a mean of zero and a variance of τ_β .

Furthermore, equations 2 – 4 lay the groundwork for the calculation of the intraclass correlation coefficient (ICC) which shows the proportion of variance at each level (equations 5 – 7). In the three-level model, the ICC at each level is calculated as follows:

$$\text{Proportion of variance at Level 1} = \frac{\sigma^2}{(\sigma^2 + \tau_\pi + \tau_\beta)} \quad (4)$$

$$\text{Proportion of variance at Level 2} = \frac{\tau_\pi}{(\sigma^2 + \tau_\pi + \tau_\beta)} \quad (5)$$

$$\text{Proportion of variance at Level 3} = \frac{\tau_\beta}{(\sigma^2 + \tau_\pi + \tau_\beta)} \quad (6)$$

Three-level unconditional growth model. In the next step, the extent to which each student's TMVSS increases, beginning with their 2008 or 2009 measurement and continuing to the 2011 time point is examined using a random effects three-level linear model. The choice of the random effects model is based on the likelihood that the linear growth slopes of the TMVSS are not fixed, but in fact vary across time. Also, rather than reporting the amount of variance at each level, as in the unconditional model, the unconditional growth model shows the variance attributable to time effects and whether patterns of change vary significantly between students over time (Holcomb, Combs, Sirmon, & Sexton, 2010). This is accomplished by adding the covariate $gradecod_{tij}$ and its corresponding slope coefficient π_{1ij} to the unconditional model Level 1 equation. Grade is coded as "0" for third, "1" for fourth, "2" for fifth, "3" for sixth, "4" for seventh, and "5" for eighth. This sets the initial score at year zero.

The Level 2 model shows the individual student intercepts and slopes as a function of their mean intercepts and slopes. Thus, equation 9a defines the mean initial status of student i of teacher j as a function of the mean initial score within teacher j (β_{00j}), plus a student deviation (Υ_{0ij}) from this mean initial score. Equation 9b defines student ij 's growth as a function of the mean growth within teacher j (β_{10j}). Additionally, the residual, Υ_{1ij} allows the linear trend for the slope coefficient to vary randomly between students within teachers.

At Level 3, the mean initial mean status within teacher j , β_{00j} , is modeled as function of the overall initial mean status of all students (γ_{000}) and a random variance (μ_{00j}). β_{10j} is the mean growth within teacher j , while γ_{100} , is the overall mean growth in TMVSS. As at Level 2, a residual, μ_{10j} , is added to allow the slope coefficient to vary between teachers. Thus, the three level unconditional growth model is:

$$\text{Level 1: } Y_{ij} = \pi_{0ij} + \pi_{1ij}(\text{gradecod}) + e_{ij} \quad (7)$$

$$\text{Level 2: } \pi_{0ij} = \beta_{00j} + \Upsilon_{0ij} \quad (8a)$$

$$\pi_{1ij} = \beta_{10j} + \Upsilon_{1ij} \quad (8b)$$

$$\text{Level 3: } \beta_{00j} = \gamma_{000} + \mu_{00j} \quad (9a)$$

$$\beta_{10j} = \gamma_{100} + \mu_{10j} \quad (9b)$$

where:

- Y_{ij} is the TMVSS at time t (grade level), for student i , of teacher j ;
- π_{0ij} is the initial TMVSS for student i , of teacher j , when the centered grade level equals 0 (third grade);
- π_{1ij} is the slope or growth rate of student i , of teacher j over the academic year;
- e_{ij} is a random effect representing the deviation of the TMVSS of student i , of teacher j , from the predicted score based on the Level 2 model. It is assumed that e_{ij} is normally distributed with a mean of zero and a variance of σ^2 ;
- β_{00j} is the mean grade 3 TMVSS within teacher j ;
- Υ_{0ij} is the random error of the intercept at the student. It is assumed that Υ_{0ij} is normally distributed with a mean of zero and a variance of τ_π .
- β_{10j} is the slope or mean growth rate within teacher j ;
- Υ_{1ij} is the random error at the student level;
- γ_{000} is the overall mean grade 3 TMVSS across students and teachers; and
- μ_{00j} is a random error of the intercept at the teacher level. It is assumed that μ_{00j} is normally distributed with a mean of zero and a variance of τ_β .

- γ_{100} is the overall mean learning rate over the academic year across students and teachers; and
- μ_{10j} is the random error at the teacher level.

The time variable at Level 1 and the demographic controls at Level 2 will be uncentered and teacher experience and preparation program attended at Level 3 will be grand mean centered.

The MLM models will be analyzed using Full Maximum Likelihood estimation (FEMML).

Three-level conditional growth model. The final step was to construct a conditional growth model that estimated the effects of teacher preparation program on student learning. The method chosen to accomplish this was a three-level multi-level model (HLM3) of TMVSS. In answering the research questions, consideration of including or excluding various Level 2 and 3 predictor variables was required. Evaluation of demographic variables is unnecessary at Level 1 as it includes individual student TMVSS over four years modeled as a function of time. For experience, teachers are assigned to one of five groups. Group 1 contains teachers from teacher preparation program “A” who have less than 5 years teaching experience. Group 2 contains teachers from teacher preparation program “B” who have less than 5 years teaching experience. Group 3 contains teachers from teacher preparation program “C” who have less than 5 years teaching experience. Group 4 contains teachers with 5 – 19 years of experience regardless of their teacher preparation program. Group 5 was used as a reference group and contained teachers with 20+ years of experience.

The Level 1 conditional model. The model for Level 1 (student growth over time) is identical to the Level 1 model of the three-level unconditional growth model

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{gradecod}_{tij}) + e_{tij} \quad (10)$$

where

- Y_{tij} is the TMVSS at time t (grade level), for student i , of teacher j ;
- π_{0ij} is the initial TMVSS for student i , of teacher j , when the centered grade level equals 0 (third grade);
- π_{1ij} is the slope or growth rate of student i , of teacher j over the academic year;
- e_{tij} is a random effect representing the deviation of the TMVSS of student i , of teacher j , from the predicted score based on the Level 2 model. It is assumed that e_{tij} is normally distributed with a mean of zero and a variance of σ^2 .

The Level 2 conditional model. In the development of the Level 2 conditional model, a linear regression will be performed to determine the proportion of variability (R^2) attributable to each of the potential predictor variables that might affect Y_{tij} . This initial model will be

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{male})_{ij} + \beta_{02j}(\text{minority})_{ij} + \beta_{03j}(\text{ESE})_{ij} + \beta_{04j}(\text{langprog})_{ij} + \beta_{05j}(\text{gifted})_{ij} + \beta_{06j}(\text{FRL})_{ij} + \Upsilon_{0ij} \quad (11a)$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}(\text{male})_{ij} + \beta_{12j}(\text{minority})_{ij} + \beta_{13j}(\text{ESE})_{ij} + \beta_{14j}(\text{langprog})_{ij} + \beta_{15j}(\text{gifted})_{ij} + \beta_{16j}(\text{FRL})_{ij} + \Upsilon_{1ij} \quad (11b)$$

where

- π_{0ij} is the student specific TMVSS parameter;
- π_{1ij} is the teacher specific TMVSS parameter;
- β_{00j} is the mean initial TMVSS for the covariates coded 0 within teacher j ;
- β_{10j} is the mean expected linear change of TMVSS for the covariates coded as 0 within teacher j ; and
- Υ_{0ij} and Υ_{1ij} are residuals.

Each additional factor is a regression coefficient that expresses the relationship between current achievement and each of the demographic controls of students of teacher j ;

- male, minority, ESE (students receiving exceptional education services), langprog (students participating in a language program), gifted, and FRL (Free/Reduced Price Lunch) are dummy variables constructed dichotomously with 1 indicating membership in a group and 0 not a member.

Should the Level 3 sample size prove inadequate for the inclusion of a third level in the model, the teacher preparation program (TPP) and experience differences will be accounted for at Level 2 by entering TPP and experience as a set of dummy variables. Additionally, teacher variables at Level 2 will be grand mean centered should a two level model prove necessary

Should this occur, Level 2 will be fully modeled as follows:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{male})_{ij} + \beta_{02j}(\text{minority})_{ij} + \beta_{03j}(\text{ESE})_{ij} + \beta_{04j}(\text{langprog})_{ij} + \beta_{05j}(\text{gifted})_{ij} + \beta_{06j}(\text{FRL})_{ij} + \beta_{07j}(\text{TPP}_{1\dots n})_{ij} + \beta_{08j}(\text{Group}_{1\dots n})_{ij} + \Upsilon_{0ij} \quad (11c)$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}(\text{male})_{ij} + \beta_{12j}(\text{minority})_{ij} + \beta_{13j}(\text{ESE})_{ij} + \beta_{14j}(\text{langprog})_{ij} + \beta_{15j}(\text{gifted})_{ij} + \beta_{16j}(\text{FRL})_{ij} + \beta_{17j}(\text{TPP}_{1\dots n})_{ij} + \beta_{18j}(\text{Group}_{1\dots n})_{ij} + \Upsilon_{1ij} \quad (11d)$$

The Level 3 conditional model. Level 3 of this model will show how the estimates for the growth curves (intercept and time slopes) vary based upon teacher preparation program attended and experience. As stated earlier, teachers are assigned to one of five groups. Group 1 contains teachers from teacher preparation program “A” who have less than 5 years teaching experience. Group 2 contains teachers from teacher preparation program “B” who have less than 5 years teaching experience. Group 3 contains teachers from teacher preparation program “C” who have less than 5 years teaching experience. Group 4 contains teachers with 5 – 19 years of

experience regardless of their teacher preparation program. Group 5 was used as a reference group and contained teachers with 20+ years of experience.

Level 3 will estimate variability among teachers in two β coefficients. The model will predict teacher mean initial status and growth rates for teacher j (Equation 13a). The Level 3 analyses will further predict whether any gaps in growth rates vary across teachers as a function of group membership (Equation 13b).

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(\text{TPP})_j + (\text{Group}_{n-1})_j + \mu_{00j} \quad (12a)$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}(\text{TPP})_j + (\text{Group}_{n-1})_j + \mu_{10j} \quad (12b)$$

where

- β_{00j} is the mean initial TMVSS within teacher j ;
- β_{10j} is the mean academic year TMVSS growth rate within teacher j ;
- γ_{000} is the overall mean initial TMVSS;
- γ_{100} is the overall mean TMVSS growth rate; and
- μ_{00j} and μ_{10j} are residuals that represent the deviation of teacher j 's coefficient from its predicted value based on this model.

Because the research questions are concerned only with the relationship between teacher preparation program attended and TMVSS and not the relationship between teacher preparation program attended and the level 2 covariates, only the equations for β_{00j} and β_{10j} were modeled. For example, the current study is not concerned with whether teacher preparation program attended is a good predictor of Race/Ethnicity slope differences.

CHAPTER FOUR: RESULTS

Because the 15 Level 3 participants were placed into five groups, a two level model was developed with teacher characteristic variables accounted for in Level 2.

In this chapter the results of the analyses are presented. The chapter begins with a description of the study participants followed by a discussion of the tests of multilevel modeling assumptions. Finally, the results of the multilevel models are presented and discussed.

Participants

The final sample consisted of 2,172 TAKS observations nested within 687 students nested within 15 teachers.

Students. Demographic information of the final student sample is as follows. Of the 687 student participants, 1% were in fourth grade (only fourth graders who had repeated either third or fourth grade had three TMVSS), 27.2% in fifth grade, 21.1% in sixth grade, 28.0% in seventh grade, and 22.6% in eighth grade. Males made up 53.1% and females 46.9%. No third grade students were represented in the sample as none had the requisite number of observations to create a growth trajectory. Thirty-six percent of participants were minorities, 41.3% were eligible for free or reduced price lunch, 3.6% were identified as LEP, 3.6% were identified as eligible for exceptional education, and 4.6% were identified as gifted (Table 6).

Table 6

Demographic Information of Student Participants.

Demographic Information	Frequency (Percentage)
Grade	
4	7 (1.0%)
5	187 (27.2%)
6	145 (21.1%)
7	192 (28.0%)
8	155 (22.6%)
Gender	
Male	365 (53.1%)
Female	322 (46.8%)
Minority	
Yes	248 (36.1%)
No	439 (63.9%)
Free/Reduced Price Lunch	
Eligible	284 (41.3%)
Not Eligible	403 (58.7%)
Language Program	
Yes	25 (3.6%)
No	662 (96.4%)
Special Education	
Yes	25 (3.6%)
No	662 (96.4%)
Gifted	
Yes	32 (4.6%)
No	655 (95.4%)

Teachers. Fifteen teachers were eligible to participate in this study. Three had fewer than 5 years experience, eight had between 5 and 19 years, and four had 20 or more years. Each of the three teachers with fewer than 5 years completed a different teacher preparation program. Because level three contained only 15 participants distributed among five groups which is below the minimum needed to obtain reliable statistical estimates (Raudenbush & Bryk, 2002), a decision was made to model the teacher effects in level 2.

Multilevel Modeling Assumptions

There are many assumptions associated with hierarchical linear modeling. Among them are proper model specification, adequate sample size, normal distribution of residuals, linearity, multicollinearity, normal distribution of variables, and homogeneity of variance of residuals.

- Normal distribution of residuals is the assumption that the error terms at each level are normally distributed.
- Linearity is the assumption that TMVSS scores will generally increase with the passage of time. In other words, it is assumed that students will score higher in eighth grade than in seventh and higher in seventh grade than in sixth, etc.
- Multicollinearity is a situation where two or more predictor variables are correlated thereby providing redundant information. This can cause inflation in the standard error of estimates and generate misleading results.
- Normal distribution of the variable is the assumption that independent variables have a normal distribution.
- Homogeneity of variance of residuals is the assumption that the standard deviation and variance of the error terms are constant for all response variables and that the error terms are drawn from the same population.

Normal distribution of residuals. The HLM program generated two residual files, one at each level, containing the Empirical Bayes (EB) residuals, fitted values, OLS residuals, and EB coefficients. While there are several methods for assessing the normality of residuals, for this study they were evaluated by examining a Q-Q Plots. The more points lie on a straight line at approximately a 45 degree angle the more they are normally distributed. The Q-Q Plot for the

Level-1 residuals (the difference between the expected and observed TMVSS) approximated a normal distribution (see Figure 4).

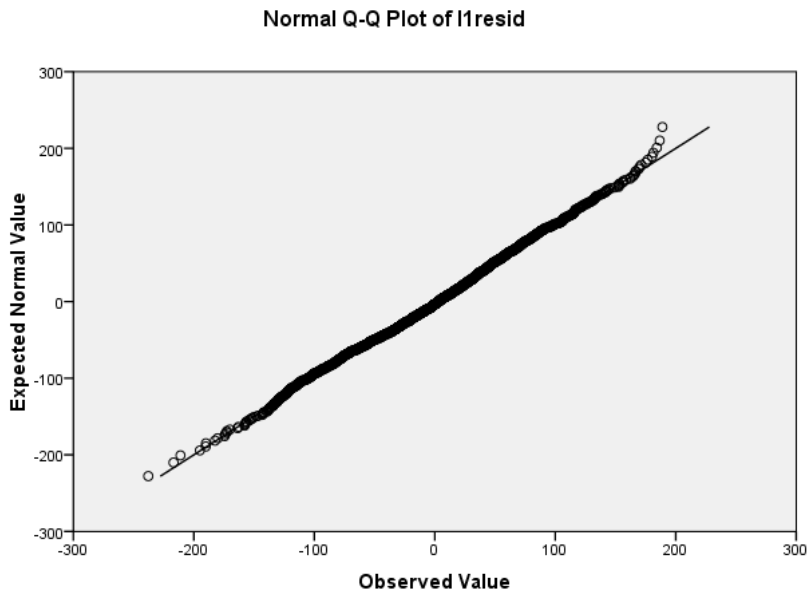


Figure 4. Q-Q Plot of Level 1 residuals used to evaluate the normality assumption.

Normality of residuals at Level 2 was also assessed with TMVSS as the outcome. The raw residuals for intercept and slope in each model were examined, using the following two Q-Q Plots (see Figures 5 & 6).

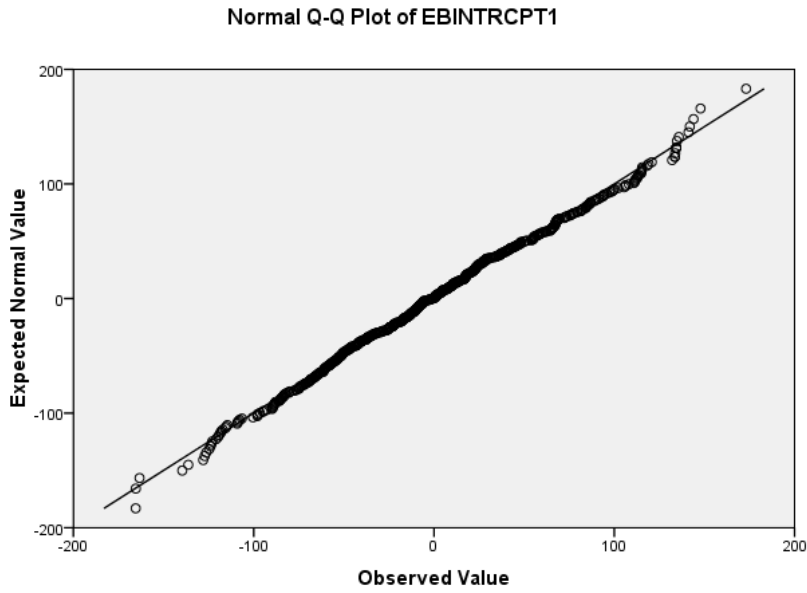


Figure 5. Q-Q Plot of Level 2 intercept residuals used to evaluate the normality assumption.

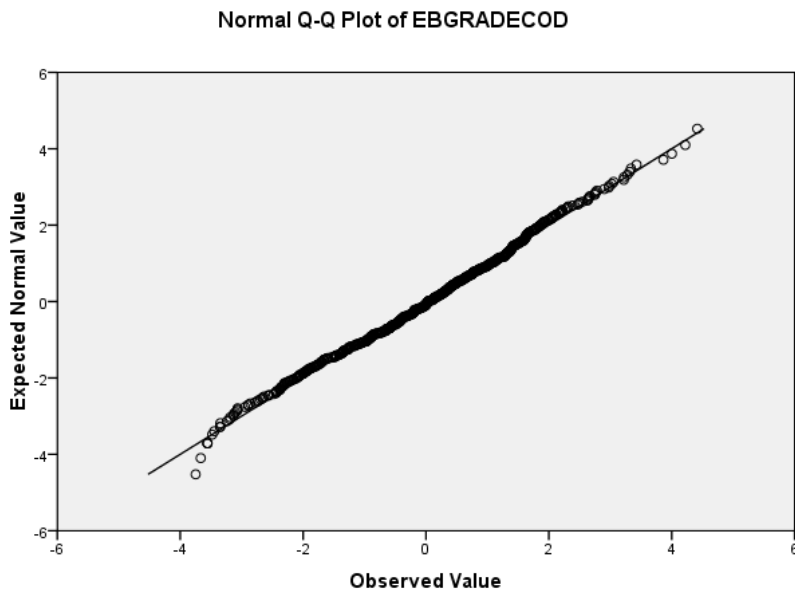


Figure 6. Q-Q Plot of Level 2 slope residuals used to evaluate the normality assumption.

As with the Level 1 normality analysis, a visual inspection of the Q-Q Plots revealed that each appeared to approximate a normal distribution.

Linearity. To examine linearity at Level 1, the researcher examined growth plots. Scatterplots were generated at Level 1 for TMVSS outcomes. While all students were examined (i.e., $N = 687$), only the mean TMVSS score as the outcome was included in the graph (see Figure 7). The figure shows the scatterplot of Grade on the X-axis, across the 2008-2011 test administrations and the overall mean across students for TMVSS on the Y-axis. A visual inspection indicates a general linear trend, thereby satisfying the assumption of linearity.

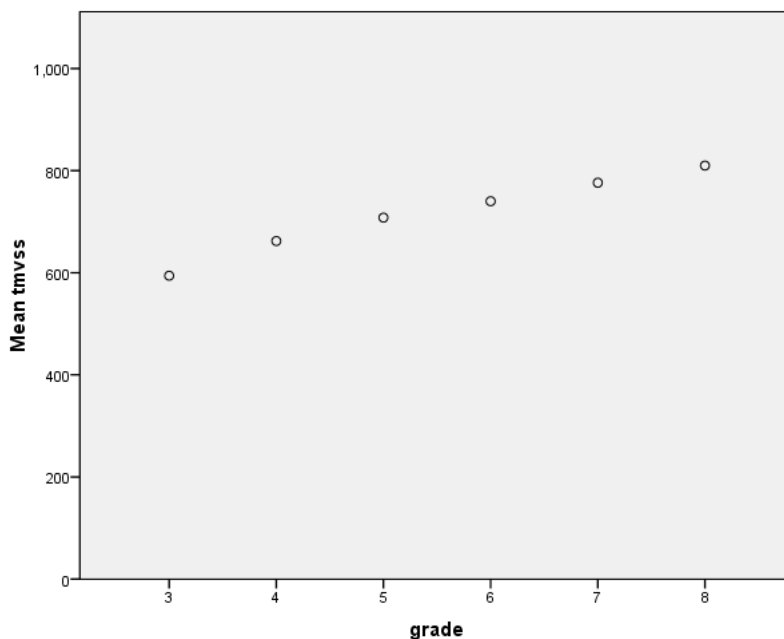


Figure 7. Scatterplot used to evaluate the linearity assumption at Level 1.

The growth plots for individual student's TMVSS suggested that most students experienced a linear change across grades. For a small number of students, growth trajectories appeared curvilinear or having no linear relationship across grades. This phenomenon could be due, in part, to the small number of data points (three or four observations across the 2008-2011

academic years) making an accurate assessment of growth difficult. Because all level 2 predictors are dichotomous, linearity at Level 2 was not evaluated.

Homogeneity of Variance. Homogeneity of variance at Level 1 was evaluated using the Chi-Square generated by the HLM software in the fully conditional model (which is discussed in a following section). The results suggest that the assumption of homogeneity of variance for Level 1 was violated (TMVSS: $\chi^2 = 1039.92$, $df = 486$, $p = 0.000$). This could be caused, in part, by unobserved violations of normality.

Homogeneity at Level 2 was examined by plotting the Empirical Bayes intercept and slope against each of the covariates (Male, Minority, ESE, LangProg, Gifted, and FRL). In order to satisfy the assumption, residual variability needs to be approximately equal for every predictor value. All six covariates were examined and are presented in Figures 8, 9, 10, 11, 12, and 13. The graphs for the remaining scatterplots are included in the appendices.

In Figure 8, 0 represents female and 1 represents male. The residual scatterplot for those values should be somewhat congruous. However, as can be seen, that while the Level 2 residual scatterplot for males and females are somewhat congruous for intercept, they are not congruous for slope. It appears that the slope scatterplot for females show greater variability (is more spread out) and has a greater number of outliers than the male scatterplot. Thus, for the gender covariate, the homoscedasticity assumption does not appear to be satisfied.

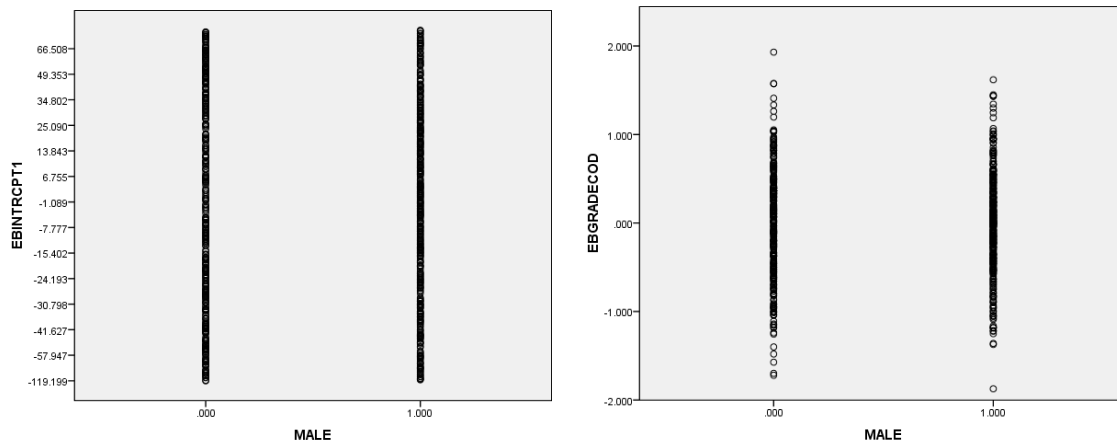


Figure 8. Residuals plotted against the male covariate to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.

In Figure 9, 1 represents those students reported as minority and 0 represents those students reported as not being minorities. The residual scatterplots for those values should again be somewhat congruous if the assumption of homoscedasticity is to be satisfied. However, as can be seen, the Level 2 residual scatterplots for minorities and non-minorities are not congruous for either slope or intercept. While the intercept scatterplot show a more equal spread than the slope scatterplot, neither shows enough similarity to consider the homoscedasticity assumption satisfied.

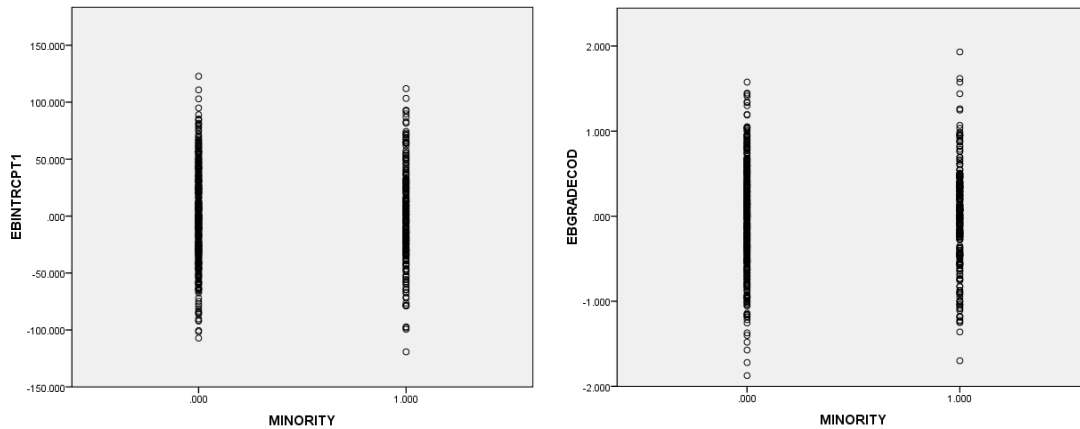


Figure 9. Residuals plotted against the minority covariate to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.

In Figure 10, 1 represents those students reported as participating in a language program (LEP/Bilingual) and 0 represents those students reported as not participating in a language program. The residual scatterplots for those values should again be somewhat congruous if the assumption of homoscedasticity is to be satisfied. However, as can be seen, the Level 2 residual scatterplots for language program participants and non-language program participants are not congruous for either slope or intercept and neither shows enough similarity to consider the homoscedasticity assumption satisfied.

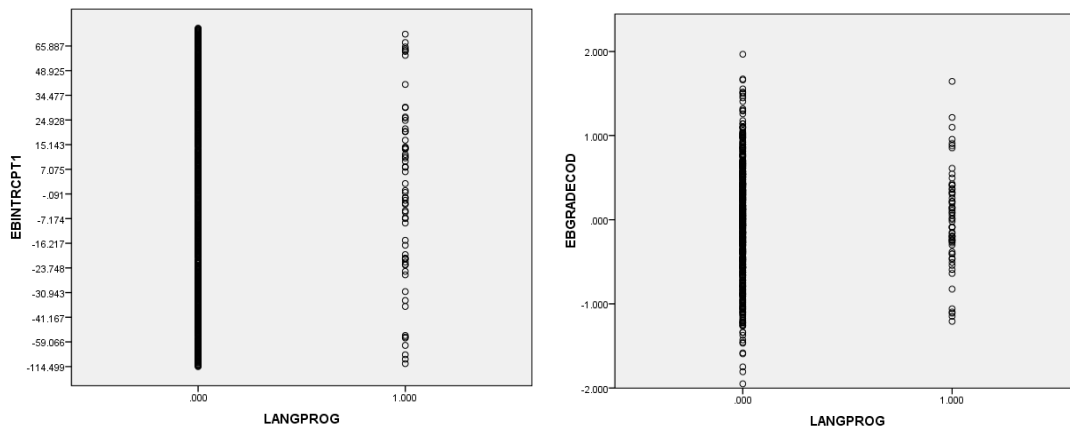


Figure 10. Residuals plotted against langprog to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.

In Figure 11, 1 represents those students reported as participating in exceptional education and 0 represents those students reported as not participating in exceptional education. The residual scatterplots for those values should once more be somewhat congruous if the assumption of homoscedasticity is to be satisfied. However, as can be seen, the Level 2 residual scatterplots for exceptional education participants and non-exceptional education participants are not congruous for either slope or intercept and neither shows enough similarity to consider the homoscedasticity assumption satisfied.

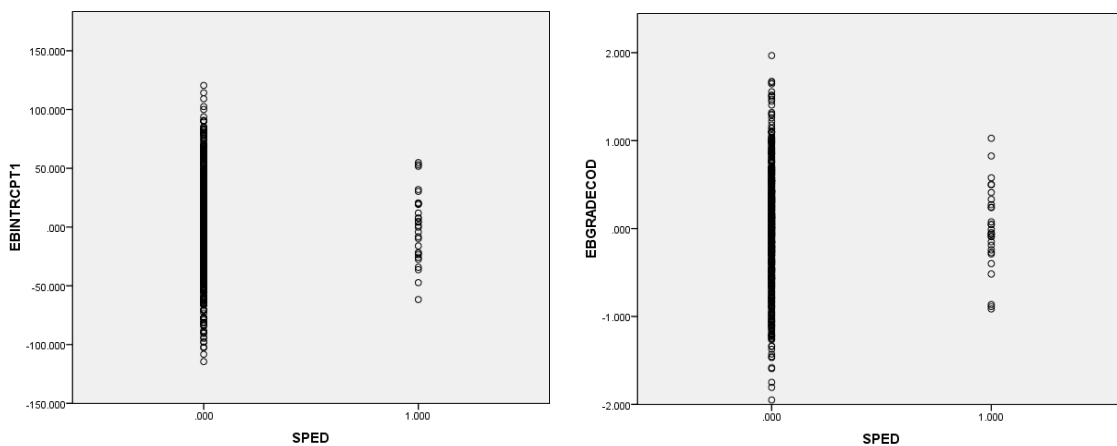


Figure 11. Residuals plotted against ESE to evaluate the assumption of homogeneity of variance at Level 2 intercept and intercept.

In Figure 12, 1 represents those students reported as participating in gifted education and 0 represents those students reported as not participating in gifted education. The residual scatterplots for those values should also be somewhat congruous if the assumption of homoscedasticity is to be satisfied. However, as can be seen, the Level 2 residual scatterplots for exceptional education participants and non-exceptional education participants are not congruous for either slope or intercept and neither shows enough similarity to consider the homoscedasticity assumption satisfied.

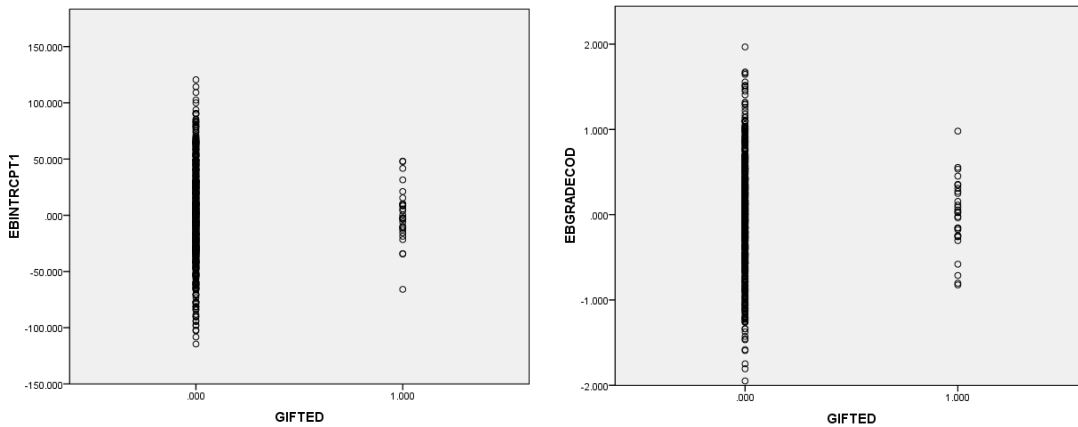


Figure 12. Residuals plotted against gifted to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.

In Figure 13, 1 represents those students reported as participating in the free and reduced price lunch program and 0 represents those students reported as not participating in the free and reduced price lunch program. The residual scatterplots for those values should once again be somewhat congruous if the assumption of homoscedasticity is to be satisfied. However, as can be seen, the Level 2 residual scatterplots for exceptional education participants and non-exceptional education participants are not congruous for either slope or intercept and neither shows enough similarity to consider the homoscedasticity assumption satisfied.

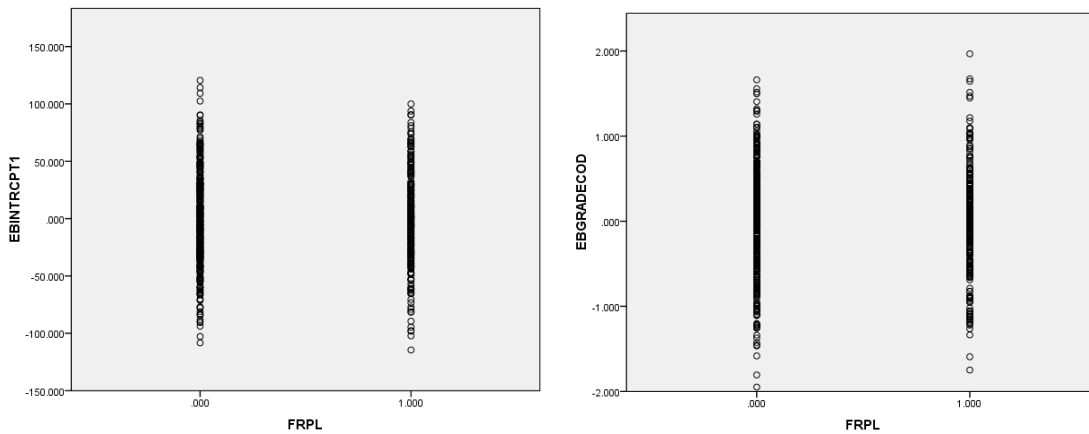


Figure 13. Residuals plotted against FRL to evaluate the assumption of homogeneity of variance at Level 2 intercept and slope.

As previously noted, some assumptions have been violated, which could increase the likelihood of Type I or Type II Errors. Based on these findings, the results herein should be interpreted with caution. As is advised when all assumptions have not been satisfied, the results are reported for the robust standard errors. However, when the coefficients in the HLM output with and without robust standard errors do not differ significantly, the interpretation and validity of the results are generally accepted.

Two Level Hierarchical Linear Models

A two level hierarchical linear growth model was constructed to examine the relationship between teacher preparation program attended and growth in student achievement in grades 4-8 across four academic years beginning in 2008 as measured by vertical scores on the TAKS mathematics assessment. More specifically, the aim was to investigate whether there is a relationship between student demographics, teacher level of experience and teacher preparation program, and growth in student mathematics achievement. The model was constructed using generally accepted model-building practices. First a null model was constructed, followed by an unconditional growth model, followed by various contextual models, and culminating with a full model. The covariates at each level are depicted in Table 7.

Table 7

Covariates Included at Each Level of Final Model.

Level	Covariates Included at Each Level
Level 1	Grade TAKS Mathematics Vertical Scale Score
Level 2	Male Minority Status Language Program Status Exceptional Education Status Gifted Education Status Free/Reduced Price Lunch Status New Teacher University A Status New Teacher University B Status New Teacher University C Status Experienced Teacher Status

Null model. The primary purpose of the null model, also known as a one-way ANOVA with random effects, is to examine the variances within-students (Level 1) and between-students (Level 2). The null model also allows for the computation of the proportional variance at each level, known as the intraclass correlation coefficient (ICC). The ICC is used to determine whether MLM is appropriate for further analysis of data. Additionally, the results serve as a point of reference when analyzing other, more complex models. The model specified was:

$$\text{Level-1 Model: } \text{TMVSS}_{ti} = \pi_{0i} + e_{ti} \quad (13)$$

$$\text{Level-2 Model: } \pi_{0i} = \beta_{00} + r_{0i} \quad (14)$$

The results of the null model are found in Table 8. The intercept is the mean of the mean of all TMVSS of each student regardless of grade. This value is also known as the expected TMVSS (γ_{00}) and was significantly different from zero ($\gamma_{00} = 694.73$, $t = 216.45$, $df = 686$, $p <$

0.001). Table 8 also shows the estimated variance components (random effects) of the model which are statistically significant, meaning substantial variation exists in student TMVSS means between-students ($\gamma_0 = 5301.17$, $df = 686$, $X^2 = 2892.41$, $p < 0.001$). This suggests that the variance is too large to simply assume it is attributable to only sampling error and further indicates that analysis should continue to examine other factors that might account for the between-student within-teacher and between-teacher variation in intercepts.

The proportion of variance at Level 1 (within students) was 49.92% and 50.08% at Level 2 (between students). The proportions were calculated using equations 5-6 (repeated).

$$\text{Proportion of variance at Level 1} = \frac{\sigma^2}{(\sigma^2 + \tau_\pi)} \quad (4 \text{ restated})$$

$$\text{Proportion of variance at Level 2} = \frac{\tau_\pi}{(\sigma^2 + \tau_\pi)} \quad (5 \text{ restated})$$

where $\sigma^2 = 5284.99$ and $\tau_\pi = 5301.17$. Because there was significant unexplained variability, multilevel modeling is used for further analyses and a time factor (*gradecod*) was added as a Level 1 predictor to explain the effect of time in TMVSS.

Table 8

Estimation of Fixed Effects and Variance Components, Null Model (One-Way Random Effects ANOVA).

Fixed Effects	Coefficient (SE)	t (df)	p
Model for Initial TMVSS Status (π_{0i})			
Expected Mean TMVSS (γ_{00})	694.73 (3.21)	216.45 (686)	< 0.001
Random Effects	Variance Component	X ²	p
Within Students 1 (e_{ii})	5284.99		
Between Students (γ_{0i})	5301.17	2892.41	< 0.001

Note. $df = 686$; Deviance = 25743.47; Number of estimated parameters = 2.

Unconditional growth model. For the unconditional growth model the Level 1 model was extended to include the predictor *gradecod* (time) and the model becomes:

$$\text{Level-1 Model: } \text{TMVSS}_{ti} = \pi_{0i} + \pi_{1i}(\text{gradecod}_{ti}) + e_{ti} \quad (15)$$

$$\text{Level-2 Model: } \pi_{0i} = \beta_{00} + r_{0i} \quad (16a)$$

$$\pi_{1i} = \beta_{10} + r_{1i} \quad (16b)$$

By including a time factor (*gradecod*) as a predictor at Level 1 uncentered but adding no predictors at Level 2, γ_{00} represents the expected average TMVSS when *gradecod* is 0 (third grade). The results (Table 9) showed that variances had changed relative to the null model. Using the equation suggested by Snijders and Bosker (1999),

$$1 - \left(\frac{\sigma_{\text{current model}}^2 + t_{\rho(\text{current model})}}{\sigma_{\text{null model}}^2 + t_{\rho(\text{null model})}} \right) = \frac{2635.96 + 4267.09}{5284.94 + 5301.17} = 0.3479 \quad (17)$$

the model explains 34.79% of total variance. The effect of *gradecod* was calculated to be .5010. That is, when *gradecod* is zero, approximately 50.12% of the explainable variance in expected TMVSS is explained by *gradecod* at Level 1 and 19.51% at Level 2. This is calculated by subtracting the total within student variance of the current model from the total within student variance that can be explained by any level 1 model and dividing the result by the total within student variance that can be explained by any level 1 model

$$\left(\frac{\sigma_{\text{null}}^2 - \sigma_{\text{current model}}^2}{\sigma_{\text{null}}^2} \right) = \left(\frac{5284.94 - 2635.96}{5284.94} \right) = 0.5012 \quad (18)$$

The proportional variances were calculated to be 38.19% within students and 61.81% between students using equations 5 and 6, respectively.

Mean TMVSS across all students was significantly different from zero ($\gamma_{00} = 612.04$, $t = 183.89$, $df = 686$, $p < 0.001$). Additionally, there was a significant difference in the *gradecod* slope (the

mean rate of change across all students) ($\gamma_{10} = 43.73$, $t = 43.86$, $df = 686$, $p < 0.001$). On average, there was a 43.73 point annual increase in student TMVSS. Initial status and linear growth were negatively correlated with a correlation of -0.56 ($p < 0.001$). This means that students who had a low initial TMVSS grew at a faster rate than those with a high initial TMVSS. It should be noted however, that this phenomenon could have been due to a plateau effect. That is students with lower scores had the opportunity to increase their scores by a larger percentage than higher scoring students who were closer to the maximum score on a particular assessment.

Statistically significant variability still exists in TMVSS means after controlling for time ($\tau_{\pi 0} = 4267.09$, $X^2 = 1647.28$, $df = 686$, $p < 0.001$), however, Level 2 slope variance (between-student individual difference in growth rates) ($\tau_{\pi 1} = 10.18$, $X^2 = 712.94$, $df = 686$, $p = 0.231$) is not statistically significant. This finding indicates that all students have comparable growth rates. While the growth rate is not statistically significant without covariates, their inclusion may show that they have significant influence on the growth rate and the error term for slope will continue to vary randomly.

Table 9

Estimation of Fixed Effects and Variance Components, Unconditional Growth Model

Fixed Effects	Coefficient (SE)	<i>t</i> (df)	<i>p</i>
Model for Initial TMVSS Status (π_{0i})			
Intercept (γ_{00})	612.04 (3.33)	183.89 (686)	< 0.001
Model for TMVSS Growth Rate (π_{1i})			
Intercept (γ_{10})	43.73 (1.00)	43.86 (686)	< 0.001
Random Effects	Variance Component	X^2	<i>p</i>
Level 1 Intercept (e_{ti})	2635.96		
Level 2 Intercept (Υ_{0i})	4267.09	1647.28	< 0.001
Slope (growth rate)	10.18	712.94	0.231

Note. *df* = 686; Deviance = 24443.41; Number of estimated parameters = 4;

Conditional growth model. The first step in the construction of the conditional model was to determine which predictors should be retained. Because the inclusion of multiple predictor variables tends to complicate multilevel models, predictors should be entered when there is evidence of an association between the predictor and the dependent variable. A predictor variable can also be included in the model in the absence of said association if there is a good theoretical reason for keeping it. .

To determine whether a predictor variable should be included in the model, a conditional model was created for each of the variables individually as follows:

$$\text{Level-1 Model: } \text{TMVSS}_{0i} = \pi_{0i} + e_{ti} \quad (19)$$

$$\text{Level-2 Model: } \pi_{0i} = \beta_{00} + \beta_{01} (\text{predictor variable}_i) + r_{0i} \quad (20)$$

Of the six variables tested (see Table 10), 5 (minority status, language program status, exceptional education status, gifted status, and free/reduced price lunch status) proved to be

statistically significant and one (male) was not. This means that students do not make comparable gains in TMVSS based on minority status, language program status, exceptional education status, gifted status, and free/reduced price lunch status. Raudenbush & Bryk (2002) suggest excluding predictors when preliminary t-ratios for their effects are less than 1.0. Thus, only variables with t-ratios with an absolute value greater than 1.0 were entered into the model.

Table 10

Test of Significance of Student Predictor Variables

Predictor Variable	Coefficient	SE	t-ratio	p-value
MALE	1.51	6.44	0.234	0.815
MINORITY	-20.91	6.52	-3.20	0.001
LANGPROG	-57.54	9.57	-6.01	< 0.001
ESE	-31.12	13.97	-2.23	0.026
GIFTED	79.93	12.37	6.46	< 0.001
FRL	-20.62	6.44	-3.20	0.001

Note: $df = 685$

Development of the conditional growth model continued with the Level 1 model including the predictor *gradecod* and the Level 2 error term representing between-student individual difference in growth rates fixed. The model becomes:

$$\text{Level-1 Model: } \text{TMVSS}_{0i} = \pi_{0i} + \pi_{1i} * (\text{gradecod}_{i}) + e_{ti} \quad (21)$$

$$\begin{aligned} \text{Level-2 Model: } \pi_{0i} = & \beta_{00} + \beta_{01} * (\text{MINORITY}_i) + \beta_{02} * (\text{LANGPROG}_i) + \beta_{03} * (\text{ESE}_i) \\ & + \beta_{04} * (\text{GIFTED}_i) + \beta_{05} * (\text{FRL}_i) + r_{0i} \end{aligned} \quad (22a)$$

$$\begin{aligned} \pi_{1i} = & \beta_{10} + \beta_{11} * (\text{MINORITY}_i) + \beta_{12} * (\text{LANGPROG}_i) + \beta_{13} * (\text{ESE}_i) \\ & + \beta_{14} * (\text{GIFTED}_i) + \beta_{15} * (\text{FRL}_i) \end{aligned} \quad (22b)$$

The results shown in Table 11 and Table 12 indicate that variances had changed relative to the null model with the current model explaining 39.94% of the total variance. The proportional variances were 41.21% at Level 1 and 58.79% at Level 2. The inclusion of the five statistically significant variables accounted for 50.43% of the explainable within student variance and 29.49% of the explainable between student variance. The result was statistically significant ($\tau_{\pi 0} = 3737.89$, $X^2 = 1508.26$, $df = 681$, $p < 0.001$) and indicates that differences between students that might be accounted for by other Level 2 predictors remain.

Overall mean TMVSS across students is statistically significant ($\gamma_{00} = 622.76$, $t = 132.28$, $df = 681$, $p < 0.001$). That is the initial mean TMVSS is 622.76 when all Level 2 predictor variables equal zero. Additionally, there was a significant difference in the *gradecod* slope (the mean rate of change across all students) ($\gamma_{10} = 42.37$, $t = 29.26$, $df = 681$, $p < 0.001$). On average, there was a 42.38 point annual increase in student TMVSS. Initial status and linear growth were negatively correlated with a correlation of -0.52 ($p < 0.001$). This means that students who had a low initial TMVSS grew at a faster rate than those with a high initial TMVSS. As previously noted, this phenomenon could have been due to a plateau effect. That is students with lower scores had the opportunity to increase their scores by a larger percentage than higher scoring students who were closer to the maximum score on a particular assessment.

The effect of gifted status was positive and statistically significant ($\gamma_{04} = 87.49$, $t = 6.83$, $df = 681$, $p < 0.001$). The positive coefficient represents the increase, on average, in a student's mean TMVSS of a student identified as a member of the gifted group. Free or reduced price lunch was negative and statistically significant ($\beta_{05} = -22.14$, $t = -3.18$, $df = 681$, $p = 0.001$). This means there were no statistically significant differences in scores for exceptional education students, minorities, or students in language programs.

The effect of the demographic predictors on TMVSS growth over time was not statistically significant for any student level predictors (minority: $\beta_{11} = 0.56$, $t = 0.26$, $df = 681$, $p = 0.794$; langprog.: $\beta_{12} = 6.66$, $t = 1.23$, $df = 681$, $p = 0.220$); ESE: ($\beta_{13} = -6.01$, $t = -1.26$, $df = 681$, $p = 0.208$; gifted: $\beta_{14} = -4.03$, $t = -0.82$, $df = 681$, $p = 0.412$; or free/reduced price lunch: $\beta_{15} = 3.44$, $t = 1.66$, $df = 681$, $p = 0.098$).

Table 11

Estimation of Fixed Effects, Conditional Growth Model

Fixed Effect	Coefficient	Standard error	t-ratio	p-value
For INTRCPT1, π_0				
INTRCPT2, β_{00}	622.76	4.71	132.28	<0.001
MINORITY, β_{01}	-9.99	7.19	-1.39	0.165
LANGPROG, β_{02}	-8.10	10.54	-0.77	0.443
ESE, β_{03}	-31.08	16.90	-1.84	0.066
GIFTED, β_{04}	87.49	12.81	6.83	<0.001
FRL, β_{05}	-22.14	6.96	-3.18	0.002
For GRADECOD slope, π_1				
INTRCPT2, β_{10}	42.37	1.45	29.26	<0.001
MINORITY, β_{11}	0.56	2.15	0.26	0.794
LANGPROG, β_{12}	6.66	5.42	1.23	0.220
ESE, β_{13}	-6.01	4.77	-1.26	0.208
GIFTED, β_{14}	-4.03	4.90	-0.82	0.412
FRL, β_{15}	3.44	2.08	1.66	0.098

Note: $df = 681$

Table 12

Estimation of Variance Components, Conditional Growth Model

Random Effect	Standard Deviation	Variance Component	χ^2	<i>p</i> -value
INTRCPT1, r_0	61.14	3737.89	1508.26	<0.001
GRADECOD slope, slope, r_1	2.75	7.57	704.32	0.260
level-1, e	51.18	2619.81		

Note. *df* = 681; Deviance = 24372.71; Number of estimated parameters = 16.

Which predictor variables to include in the full model was given careful consideration. A comparison of the results from conditional models, one including and one excluding the “male” predictor variable revealed that excluding it at Level 2 produced results similar to including it. The difference in the explained variance between the two models was 0.02% and there was no change in the coefficient values of the other predictor variables.

Other factors were considered as well. Literature referenced in previous sections, suggest that the included variables have been historically stable and statistically significant predictors of student achievement. Furthermore, while not statistically significant in this study, the included variables often have practical significance. Based on this rationale, all variables that indicated statistical significance in the independent analyses of each variable will be included in the full model as varying randomly.

Full model. The full model was constructed with time in Level 1, the predictors of minority status (*minority*), language program status (*langprog*), exceptional education status (*ESE*), gifted status (*gifted*), free/reduced price lunch status (*FRL*) and teacher group membership in Level 2. Following the work of Noell (2006), five dummy variables were created to represent various levels of teacher experience. Teachers with 20 or more years experience were coded “1”

for group 5 and all other teachers were coded “0”. This group was used as a reference group. Teachers with 5 – 19 years experience were coded “1” for group 4 and all other teachers were coded “0”. Teachers with fewer than five years experience were grouped by teacher preparation program, with three programs represented (Table 13).

Table 13

Description of Teachers' Groups

Group	Description
Group 1	Teachers with fewer than 5 years experience from University “A”.
Group 2	Teachers with fewer than 5 years experience from University “B”.
Group 3	Teachers with fewer than 5 years experience from University “C”.
Group 4	All teachers with 5 – 19 years experience.
Group 5	All teachers with 20 or more years experience (reference group).

Thus, the full model is:

Level-1 Model

$$TMVSS_{ti} = \pi_{0i} + \pi_{1i}*(GRADECOD_{ti}) + e_{ti} \quad (23)$$

Level-2 Model

$$\begin{aligned} \pi_{0i} = & \beta_{00} + \beta_{01}*(MINORITY_i) + \beta_{02}*(LANGPROG_i) + \beta_{03}*(ESE_i) + \beta_{04}*(GIFTED_i) \\ & + \beta_{05}*(FRL_i) + \beta_{06}*(GROUP1_i) + \beta_{07}*(GROUP2_i) + \beta_{08}*(GROUP3_i) \\ & + \beta_{09}*(GROUP4_i) + r_{0i} \end{aligned} \quad (24a)$$

$$\begin{aligned} \pi_{1i} = & \beta_{10} + \beta_{11}*(MINORITY_i) + \beta_{12}*(LANGPROG_i) + \beta_{13}*(ESE_i) + \beta_{14}*(GIFTED_i) \\ & + \beta_{15}*(FRL_i) + \beta_{16}*(GROUP1_i) + \beta_{17}*(GROUP2_i) + \beta_{18}*(GROUP3_i) \\ & + \beta_{19}*(GROUP4_i) + r_{1i} \end{aligned} \quad (24b)$$

The results, shown in Table 14 and Table 15, show that variances had changed relative to the null model with the current model explaining 40.51% of the total variance. The proportional variances were 41.40% at Level 1 and 58.60% at Level 2. The model accounted for 50.66% of the explainable within student variance and 30.38% of the explainable between student variance. The result was statistically significant ($\tau_{\pi 0} = 3690.51$, $X^2 = 1487.51$, $df = 677$, $p < 0.001$) and suggests that substantial differences exist between students that might be accounted for by additional Level 2 predictors. The growth rate variance remains statistically non-significant ($\tau_{\pi 1} = 3.52$, $X^2 = 704.02$, $df = 677$, $p = 0.229$).

Overall mean TMVSS across students is statistically significant ($\gamma_{00} = 619.42$, $t = 130.28$, $df = 681$, $p < 0.001$). The effect of gifted status ($\beta_{04} = 88.30$, $t = 6.56$, $df = 679$, $p < 0.001$) was positive and statistically significant. A positive coefficient represents the increase, on average, in a student's mean TMVSS of a student identified as a member of that particular group. There was a negative statistically significant effect for free/reduced price lunch status ($\beta_{05} = -21.15$, $t = -3.05$, $df = 679$, $p = 0.002$) on mean TMVSS. A negative coefficient represents the decrease, on average, in a student's mean TMVSS of a student identified as a member of that particular group.

There was no statistically significant effect of minority status ($\beta_{01} = -8.67$, $t = -1.22$, $df = 679$, $p = 0.223$), language program status ($\beta_{02} = -10.67$, $t = -0.99$, $df = 679$, $p = 0.320$) or exceptional student education status ($\beta_{03} = -22.39$, $t = -1.38$, $df = 679$, $p = 0.167$) on mean TMVSS. This means there were no statistically significant differences in scores for any student based on being a member of one of those groups. The effect of time on TMVSS was positive and statistically significant ($\beta_{10} = 45.35$, $t = 29.63$, $df = 679$, $p < 0.001$) when all predictor variables are zero. This indicates that each year there was an average increase of 45.35 points in TMVSS when all predictors were zero.

Teachers with 5 – 19 years experience showed a positive, but not statistically significant ($\beta_{09} = 6.75, t = 0.95, df = 677, p = 0.344$) effect on expected TMVSS. New teachers from University “A” showed a negative but not statistically significant effect on expected TMVSS ($\beta_{06} = -19.72, t = -1.49, df = 677, p = 0.137$). New teachers from University “B” showed a negative but not statistically significant effect on expected TMVSS ($\beta_{07} = -82.67, t = -1.71, df = 677, p = 0.087$). New teachers from University “C” showed a negative but not statistically significant effect on expected TMVSS ($\beta_{08} = -18.53, t = -1.77, df = 677, p = 0.077$).

In examining the effect of the teacher groups on TMVSS growth over time, new teachers from Universities “A” and “B” had negative, but not statistically significant results ($\beta_{16} = -1.62, t = -0.51, df = 677, p = 0.612; \beta_{17} = -4.07, t = -0.35, df = 677, p = 0.724$) while new teachers from University “C” showed large, statistically significant results in growth ($\beta_{18} = 14.88, t = 3.50, df = 677, p = <0.001$) and teachers with 5- -19 years experience showed positive, statistically significant growth over time ($\beta_{19} = 7.40, t = 2.84, df = 677, p = 0.005$).

Table 14

Estimation of Fixed Effects, Full Conditional Model

Fixed Effect	Coefficient	Standard error	<i>t</i> -ratio	<i>p</i> -value
For INTRCPT1, π_0				
INTRCPT2, β_{00}	619.42	4.75	130.28	<0.001
MINORITY, β_{01}	-8.67	7.11	-1.22	0.223
LANGPROG, β_{02}	-10.67	10.72	-1.00	0.320
ESE, β_{03}	-22.39	16.17	-1.38	0.167
GIFTED, β_{04}	88.30	13.41	6.58	<0.001
FRL, β_{05}	-21.15	6.94	-3.05	0.002
GROUP1, β_{06}	-19.72	13.23	-1.49	0.137
GROUP2, β_{07}	-82.67	48.25	-1.71	0.087
GROUP3, β_{08}	-18.53	10.46	-1.77	0.077
GROUP4, β_{09}	6.75	7.13	0.95	0.344
For GRADECOD slope, π_1				
INTRCPT2, β_{10}	45.35	1.53	29.63	<0.001
MINORITY, β_{11}	0.73	2.11	0.35	0.730
LANGPROG, β_{12}	4.19	5.37	0.78	0.435
ESE, β_{13}	0.90	5.16	0.17	0.862
GIFTED, β_{14}	-7.71	4.42	-1.74	0.082
FRL, β_{15}	3.86	2.06	1.87	0.062
GROUP1, β_{16}	-1.62	3.19	-0.51	0.612
GROUP2, β_{17}	-4.07	11.53	-0.35	0.724
GROUP3, β_{18}	14.89	4.25	3.50	<0.001
GROUP4, β_{19}	7.44	2.60	2.86	0.004

Note: *df* = 677

Table 15

Estimation of Variance Components, Full Conditional Model

Random Effect	Standard Deviation	Variance Component	χ^2	<i>p</i> -value
INTRCPT1, r_0	60.75	3690.51	1487.51	<0.001
GRADECOD				
slope, r_1	3.52	12.38	704.02	0.229
level-1, e	51.06	2607.44		

Note: $df = 677$

CHAPTER FIVE: DISCUSSION

This chapter discusses the results of the study. It begins with a brief summarization of the purpose of the study followed by a summary of pertinent results. The chapter concludes with a discussion of the implications of the research, limitations, and future research.

Purpose of the Study

The purpose of this study is to examine whether a small school district can use multilevel modeling to determine the impact of various student characteristics and teachers' level of experience and teacher preparation program attended on student mathematics achievement. Moreover, does the effectiveness of new teachers from specific teacher preparation programs differ from that of experienced teachers. The research questions addressed were:

- (1) What is the effect of the student characteristics of gender, minority status, language program status, exceptional education status, gifted education status, and free or reduced price lunch status on the predicted mathematics achievement of students in grades four through eight?
- (2) To what extent is the predicted mathematics achievement of students affected by teachers' level of experience and in the case of new teachers, their teacher preparation program attended?

In order to answer these questions, data were obtained from a small school district in Texas and analyzed by building a two-level multilevel model using HLM 7 and SPSS.

Summary of Findings

A series of two level hierarchical linear models were constructed to determine whether they were a viable means of examining the relationship between student test scores, teacher level

of experience, and teacher preparation program attended and if so, the nature of the relationship. Following standard protocols for hierarchical linear modeling (MLM), a null model devoid of any predictor variables was constructed first. The null model suggested that about 50% of the variation in expected TMVSS between students could be explained by the level 2 predictor variables.

Because there was a significant amount of unexplained variability, a second model, the unconditional growth model, was constructed which extended the null model by adding a time factor, students' grade level, at Level 1 to explain the effect of time on TMVSS. The unconditional growth model explained about 35% of the total variance. The mean TMVSS across all students was 612 and there was, on average, a 44 point gain in student TMVSS annually. Additionally, initial TMVSS and linear growth were negatively correlated which means that students who had a low initial TMVSS grew at a faster rate than those with a high initial TMVSS. While statistically significant variability still existed in mean TMVSS after controlling for time, the individual difference in growth rates was not statistically significant indicating that all students have comparable growth rates.

The next step was the construction of a conditional growth model. To begin the process, a determination of which student demographic predictor variables should be retained in the model was made. This was done by creating a conditional model for each of the variables individually to determine the effect of that variable on TMVSS. Five of the six variables were statistically significant and one was not. The five significant variables were entered into the model as varying randomly. The model now explained 39% of the total variance in TMVSS. The inclusion of the five predictor variables accounted for 50% of the explainable within-student variance and 29%

of the explainable between-student variance. The result indicated that differences between students that might be accounted for by other Level 2 predictors remain.

Overall mean initial TMVSS across students was 622 when all Level 2 predictor variables were zero (i.e. white students who are not LEP, not identified as a student with a disability, gifted, and not receiving free or reduced price lunch). On average, there was a 42 point annual increase in student TMVSS. Initial status and linear growth continued to be negatively correlated.

In examining the effects of the predictor variables on TMVSS, a positive coefficient represents an increase, on average, in a student's mean TMVSS and a negative coefficient represents a decrease. The effect of gifted was positive and statistically significant while the effect of free or reduced price lunch was negative and statistically significant. The effects of minority, language program, exceptional education status, were all negative and not statistically significant but were included in the final model because they have a historical basis of impacting achievement.

The effect of the demographic predictors on the rate of TMVSS growth was not statistically significant for any student demographic predictors and statistically significant variability still existed in TMVSS means after adding the student demographic, however, Level 2 between-student individual difference in growth rates was still not statistically significant.

The final step was the construction of the full model with the effect of time modeled in Level 1, the student demographic predictor variables and teacher group membership modeled in Level 2. Group 1 contained teachers from teacher preparation program "A" who have less than 5 years teaching experience. Group 2 contained teachers from teacher preparation program "B" who have less than 5 years teaching experience. Group 3 contained teachers from teacher

preparation program “C” who have less than 5 years teaching experience. Group 4 contained teachers with 5 – 19 years of experience regardless of their teacher preparation program. Group 5 was used as a reference group and contained teachers with 20+ years of experience. The model explained 40% of the total variance. Overall mean TMVSS across students is 619 and increased 45 points a year. As with the previous model, the effect of gifted was positive and statistically significant while the effect of free or reduced price lunch was negative and statistically significant. The effects of minority, language program, exceptional education status, were all negative and not statistically significant

In regard to the teacher group variables, teachers with 5 – 19 years of experience showed a positive effect on expected TMVSS and all three groups of teachers with fewer than 5 years experience showed negative effects. However, none of these effects were statistically significant. In examining the effect of the teacher groups on TMVSS growth over time, new teachers from universities “A” and “C” showed negative effects that were not statistically significant while new teachers from university “B” and teachers with 5 – 19 years of experience showed positive effects that were statistically significant.

Implications of Findings Related to Student Characteristics

The first question posed in this study dealt with the impact of student characteristics (gender, minority status, language program status, exceptional education status, gifted education status, and free or reduced price lunch status) on the predicted mathematics achievement of students. A discussion on the impact of the student level variables should include their convoluted nature and attempts to disentangle them. The student level predictors commonly associated with low achievement (minority, low SES, ESE, and ELL) are often intertwined and in many instances, low performing students have membership in more than one if not all of these

groups (Linn & Hemmer, 2012; Skiba et al., 2008; U. S. Commission on Civil Rights, 2009). If that is the case, then which gap should be the focus of schools' efforts to close first? Will closing the SES achievement gap ameliorate the racial and ELL gaps? These questions are not the focus of this research but the findings herein, while not conclusive, suggest that the magnitude of these gaps is not as strong in the district of this study as in a nationally representative sample (NAEP). In the discussion that follows, a comparison will be made between NAEP average scale scores and TAKS Mathematics Vertical Scale Scores. The NCES reported that 74% of the NAEP grade 4 and 81% of the NAEP grade 8 assessment standards are either fully or partially addressed by the TAKS assessment standards (http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/techbrief/tr_00708.pdf). It should be noted that the scores are not reported on the same scale. NAEP scores are reported on a scale of 0 – 500 and TMVSS are reported on a scale of 0 – 1000.

Also, in the current study, only two of the five student level predictor variables were statistically significant. There are many causes for findings not being statistically significant. These include non-random assignment of teachers to schools and students to teachers, sampling error, lack of power, random chance and of course, that the covariates are simply not related to the outcome. While not all covariates showed statistical significance, the model did account for almost 41% of the explainable variance in TMVSS and may therefore exhibit some practical significance.

Gender. The issue of whether there truly is a “gender gap” has been a topic of much research and debate (Kafer, 2007). A seemingly common public perception is that males generally outperform females in mathematics. However, research has shown that when an

achievement difference between males and females does exist, it is usually small and not statistically significant (Dee, 2007; Hyde & Linn, 2006).

Whether to include gender as a covariate in this study was given thorough consideration. Models were constructed both with and without gender as a predictor variable and there was a very small difference (.02%) in the amount of variance explained when the gender covariate was included and the difference in the mean TMVSS across grades for males was only 5 points higher than that of females. Furthermore, when analyzed independently, gender was not a statistically significant predictor of student achievement and exhibited a low *t*-ratio. Based on these factors, the final decision was to exclude gender as a predictor variable.

Gifted status. Conventional wisdom suggests that students identified as gifted are high achievers or they probably wouldn't be identified as gifted. That said, there is not an abundance of literature that examines whether gifted students achieve at higher levels than non-gifted students and the literature that is available shows mixed results (Bui, Craig, & Imberman, 2011). The results of the current study suggest that students identified as gifted have an 88 point higher initial TMVSS than non-gifted students. However, gifted students rate of growth is lower than that of non-gifted students. This is most likely due to a plateau effect. Students who have higher initial scores have less room to grow.

Minority status. Unlike the gender gap, the existence of a minority gap is not in question. In reviewing NAEP data, it is quickly evident that the achievement gap between white, black and Hispanic students is substantial and has persisted over time (Hemphill & Vanneman, 2011; Vanneman, Hamilton, Baldwin, & Rahman, 2009). According to data retrieved using the NAEP Data Explorer, on average, black and Hispanic students' average scale scores were about

23 points lower than those of white students in fourth grade and 28 points in eighth grade on the 2011 assessment (<http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx>).

In the school district studied, the mean difference between minority and non-minority students on the TMVSS across grades four through eight was 16 points in favor of non-minority students. The results of the current study also suggest that, on average, having minority membership results in an eight point lower initial TMVSS and less than a one point difference in expected annual growth. Neither of the results was statistically significant.

The NAEP and the TAKS do not measure the same knowledge and skills. Nor is the NAEP given across all grades 3 – 8 as is the TAKS. However, it would appear that in the school district of interest, minority students achieve at a level more commensurate with non-minority students than students in the nationally representative sample.

Language program status. As with minority status, the impact of ELL status is well documented with as much as 20 – 30 percentage points below that of non-ELLs (Abedi & Dietel, 2004; Abedi & Gándara, 2006; Fry, 2007, 2008). The 2011 NAEP data show gaps in favor of non-ELLs of 24 average scale score points for fourth graders and 41 points for eighth graders. The current study suggests that having ELL membership results in a ten point lower initial TMVSS and a four point decrease in expected annual growth, on average. However, the district maintains the 20 percentage point difference in the number of ELL and non-ELL students meeting proficiency requirements on state assessments.

Socioeconomic status. Socioeconomic status may be the most studied and historically significant predictors of student achievement. Beginning with the Coleman report in 1966 and continuing to the present, there is no shortage of literature documenting the impact of SES on student success (Coleman et al, 1966; Sirin, 2005).

The 2011 NAEP data show gaps in favor of non-FRL students of 23 average scale score points for fourth graders and 26 points for eighth graders. The current study suggests that having FRL membership results in a statistically significant 21 point decrease in initial TMVSS and a four point decrease in expected annual growth, on average. These results seem similar to the nationally representative sample of NAEP and support the findings of previous research that as a general rule, poor students do not achieve as highly as more affluent students in mathematics. However, even though they do not achieve at levels similar to non-economically disadvantaged students, 95% of economically disadvantaged meet state proficiency requirements.

Exceptional students. The factors that create and sustain disproportionality in exceptional education are complex and even to this day not fully understood. While a number of contributing factors have been identified, none of them has proven definitive as the single cause of disproportionality. The best conclusion that can be drawn is that disproportionality is the product of many factors both within and without the education system. It is likely that disproportionality is caused by the interaction of these many factors including SES, race, ethnicity, and ELLs (Linn & Herman, 2012; Skiba et al., 2008).

In the school district in this study, disproportionality is not present. Thirty-seven percent of the sample is minority and 9% ELLs. Within the exceptional education students represented in the sample, 34% are minority and 7% are ELLs. This suggests that any differences in achievement attributable to exceptional education group membership are not the result of disproportionality.

The 2011 NAEP data show gaps in favor of non-exceptional education students of 25 average scale score points for fourth graders and 38 points for eighth graders. The current study suggests that ESE students have an initial TMVSS 21 that is points lower than non-ESE students

and a one point positive differential in expected annual growth, on average. These results seem similar to the nationally representative sample of NAEP and support existing literature that exceptional education students do not achieve at the same levels as non-exceptional education students in mathematics (Eckes & Swando, 2009; Wei, Lenz, & Blackorby, 2012). However, even though they do not achieve at levels similar to non-exceptional education students, approximately 85% of exceptional education students meet state proficiency requirements.

Summary of Student Level Covariates

As previously stated, student level predictors traditionally associated with low achievement are often intertwined. Minority students are more likely than non-minority students to be poor, ELLs, and identified for exceptional education or any combination thereof. However, in the district in this study, some of these factors appear to have been mitigated to some degree.

The school district in this study is not nationally representative. It should also be kept in mind that there are probably very few school districts in the country that are nationally representative. This makes it difficult to draw inferences comparing the students in a particular school or district to nationally representative samples. Within this district however, 95% or more of all students met state proficiency requirements. Furthermore, with the exception of exceptional education and ELL students, at least 95% of every subgroup met state proficiency requirements. Possible factors that could account for this phenomenon are discussed below.

Because the district contains only two elementary schools and one middle school (grades of interest), members of traditionally low performing student groups may not have been subject to the clustering that is typically seen in large urban school districts. Another factor that is related to the clustering is that because this is a small school district, there is likely to not be a large disparity between schools in per student spending.

Another advantage of a small community in overcoming achievement gaps is a greater sense of community. Perhaps, because the district in this study is in a small community with only one high school, the school tends to be the focal point of the community, leading to increased sense of belonging among all students and increased parental involvement as well. To be sure, further study is warranted to attempt to discover what occurs to counteract the traditional effects of student level covariates on student achievement.

Implications of Findings Related to Teacher Characteristics

Question 2 asked whether when using a multilevel model, is the mathematics achievement of students taught by teachers with fewer than five years experience (new teachers) from specific teacher preparation programs comparable to the mathematics achievement of students taught by teachers with five to nineteen years experience (experienced teachers).

As previously discussed, several studies (Boyd, et al., 2009; Goldhaber & Liddle, 2012; Harris & Sass, 2007; Henry et al., 2011; Henry, et al., 2010; Koedel, et al., 2012; Mihaley, et al., 2012; Noell, et al., 2009; Noell, et al., 2007; Noell, et al., 2008) have employed value-added models (VAMs) to examine issues surrounding teacher preparation. These studies suggest that VAMs can provide data supporting the fact that some teacher preparation programs are more effective than others at preparing teachers who have an immediately positive impact on student achievement.

The results suggest that students of teachers with fewer than five years experience from all three TPPs in this study had lower expected initial TMVSS than students of teachers with five to nineteen years experience. Though these results had no statistical significance, they may have practical significance. Additionally, the experienced teacher group (5 – 19 years) showed a statistically significant positive influence on the average annual expected growth in TMVSS.

New teachers (less than five years experience) from university “C” showed a statistically significant positive influence on the average annual expected growth in TMVSS as well.

In the context of the current study, the experience covariate possibly warrants the most attention. Given the effects of teacher preparation programs on student outcomes decay over time, usually four or five years (Clotfelter et al., 2007; Goldhaber & Liddle, 2012; Goldhaber, Liddle, & Theobald, 2013; Henry, Fortner, Bastian, 2012; Koedel, et al., 2012), school districts may wish to focus their resources on attracting and keeping effective teachers while developing professional development models and induction programs that increase the rate at which new teachers become more effective. Schools could also work in conjunction with universities to improve preparation programs. For example, most new teachers feel their preparation programs did not adequately prepare them for the classroom (Levine, 2006). Perhaps pre-service teachers could accumulate more contact hours with students earlier in their programs (roughly 1,000 hours equals one year of service).

The importance of experience is magnified when the discussion turns to exceptional education. As indicated earlier, exceptional education students are a traditionally low achieving student group (Eckes & Swando, 2009; Linn & Hemmer, 2012; Skiba et al., 2008; Wei, Lenz, and Blackorby, 2012). This situation is compounded when exceptional education students experience higher rates of teacher turnover than non-exceptional education students (Ronfeldt, Loeb, & Wycoff, 2013). Given that the attrition rate in exceptional education is twice that of general education and the majority of exceptional education teachers who leave the profession leave within the first three years HECSE (hecse.net/policy_documents/FactSheetSPED%20Shortages.pdf), it seems as if many exceptional education students are taught by a never-ending string of new teachers.

The district in this study does not seem to experience as high a turnover rate as is suggested by HECSE. Of the 15 teachers who participated in this study, one was in their second year, two were in their fourth year, eight had between five and nineteen years, and four had twenty or more years. The mean experience was 15 years.

There are again, any number of reasons teacher turnover is not as high here as is reported in the literature. Perhaps the sense of community of a small town comes into play. Perhaps, the teachers have strong ties to the community, or salaries could be a factor. Whatever the reason, as with the student characteristics, further investigation is warranted.

Limitations

In addition to the limitations associated with hierarchical linear modeling outlined in Chapter One, other limitations were revealed as the study progressed. First, the research questions focus on a single school district in Texas. This affects the generalizability of the study as well as restricting the sample size. The sample size was further restricted by including (a) only students in 4th – 8th grade, (b) including only students with an adequate number of observations to compute a growth curve, (c) utilizing only mathematics scores, and (d) using only vertical scales scores which limited the span of years to four.

Second, as with any study attempting to measure the effects of some variable on outcomes, it is impossible to identify and include every factor that influences student achievement. While rigorous and complex statistical methods were employed to isolate the effect of various student and teacher characteristics on student achievement, the credibility of the effects reported herein are limited by the intrinsic nature of the dependent variable. As is the case with most educational research, this consideration should be a part of any effort to interpret or apply the findings.

Future Study

As the debate over the use of student outcomes to evaluate teachers continues, researchers will also continue in their efforts to develop and refine value added models. As models improve, the method may become more widely accepted and utilized to its full potential.

Future research should investigate the following areas:

1. Replicate this study in other states and districts of similar size and compare the results to the results of this study.
2. Analyze the scores on each objective of the assessment and use the findings to determine which objectives are best taught by which teachers and why.
3. Investigate methods for analyzing data to disentangle non-tested subjects effects. A study of this nature should be helpful in identifying characteristics that are responsible for improving student achievement.
4. Each of the above suggestions should add a qualitative component to explain larger portions of the variance between students, schools, districts, university programs for teacher preparation, etc. and to further disentangle effects attributable to specific variables.

Conclusion

The purpose of this study was to determine if the methods applied to large scale studies (statewide) could be used on a much smaller scale (a single district) to determine the impact of student and teacher characteristics on student mathematics achievement. While not conclusive, the results obtained in this study are encouraging. However, the results seem to have created more questions than they answered.

While multilevel modeling remains controversial, its use continues to grow.

Demonstrating a direct link between measured student and teacher characteristics and student growth is, to say the least, complex (Cochran-Smith & Zeichner, 2005; Schalock, Schalock, & Ayres, 2006). Yet, using multilevel modeling as a means to do so is a growing trend (NCTQ, 2007). By tying teachers' effectiveness back to student and teacher characteristics, districts could ostensibly focus their resources on recruiting and retaining teachers that are most effective with their student demographic.

**APPENDIX A:
INSTITUTIONAL REVIEW BOARD**



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901, 407-882-2012 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

NOT HUMAN RESEARCH DETERMINATION

From : **UCF Institutional Review Board #1
FWA00000351, IRB00001138**

To : **Michael O'Neal**

Date : **March 28, 2013**

Dear Researcher:

On 3/28/2013 the IRB determined that the following proposed activity is not human research as defined by DHHS regulations at 45 CFR 46 or FDA regulations at 21 CFR 50/56:

Type of Review: Not Human Research Determination
Project Title: An Investigation Of The Effectiveness Of Teacher
Preparation Programs As Measured By Student
Achievement
Investigator: Michael O'Neal
IRB ID: SBE-13-09148
Funding Agency:
Grant Title:
Research ID: N/A

University of Central Florida IRB review and approval is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are to be made and there are questions about whether these activities are research involving human subjects, please contact the IRB office to discuss the proposed changes.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Patria Davis on 03/28/2013 09:43:50 AM EST

A handwritten signature in black ink, appearing to read 'Patria Davis', written over a circular stamp or seal.

IRB Coordinator

**APPENDIX B:
EXCERPT OF FEDERAL REGISTER REGARDING NCLB REAUTHORIZATION**

been done. So I hope my colleagues here in the Senate on both sides of the aisle will come together and recognize that and repeal once and for all this very bad piece of legislation.

It was good news when the administration recognized they couldn't implement it, it was not workable. It would be better news for the American taxpayers and for future generations of Americans if the Congress would repeal this legislation and do it soon.

I yield the floor.

Mr. INHOFE. Mr. President, I ask unanimous consent that at the conclusion of the remarks of the Senator from Tennessee, I be recognized in morning business. What I am going to do is try to clear up some of the misunderstanding about the troops who have gone into Uganda and other areas on the LRA, Lord's Resistance Army.

The ACTING PRESIDENT pro tempore. Without objection, it is so ordered.

The Senator from Tennessee.

EDUCATION

Mr. ALEXANDER. Mr. President, last month several Republican Senators came to the floor and offered legislation to fix No Child Left Behind, the legislation that was passed nearly 10 years ago to try to address our Nation's 100,000 public schools. In that legislation, we sought to fix problems with the legislation, not just to create another big reauthorization bill. The ideas we had were not all our ideas. They included many ideas from President Obama and his excellent Education Secretary, Secretary Duncan, as well as Democratic and Republican Members of Congress. They included having more realistic goals for No Child Left Behind. The original goal set in 2001 would, according to Secretary Duncan, create an unworkable situation where 80,000 of the 100,000 schools might be identified as failing in the next few years.

A second goal of our legislation was to move decisions about deciding whether schools and teachers were succeeding or failing out of Washington, DC, and back to State and local governments. A lot has happened in the last 10 years in the States—really the last 20 or 25 but especially in the last 10 years. We have better reporting requirements from No Child Left Behind. We have new State common standards, higher academic standards. We have new State tests that have been created—not here but by the States to do that. And now States are working together to create accountability systems. So there is a much better chance that States and local school districts can create an environment where students learn what they need to know and be able to do.

Our legislation encourages States to create what I think is the holy grail of public education; that is, principal-teacher evaluation related to student achievement. I know from experience

that is hard to do. In 1983 and 1984, when I was Governor of Tennessee, we became the first State to pay teachers more for teaching well. It took us a year and a half and a huge battle with the National Education Association in order to put it in place, but 10,000 teachers became master teachers. It was a good first step. Tennessee is already doing it again.

Here is my local newspaper: Evaluation of teachers contentious. There is nothing more contentious, and the last thing we need is Washington sticking its nose into that, other than to create an environment where State and local governments can use Federal money to pay for their own State and local programs. We propose consolidating programs, making it easier for school districts to transfer Federal money and expand choices and expand charter schools.

Now, today, the chairman and ranking member of the Senate education committee—the HELP Committee, as we call it—have introduced another draft piece of legislation to fix No Child Left Behind. I intend to vote to move this bill out of committee, although it is not yet the kind of legislation that I would be willing to vote to send to the President, but it is a good place to start.

There is a good deal of agreement in terms of what we want to do in our legislation from a few weeks ago and the Harkin-Enzi bill. Among the agreements is moving decisions about whether schools are succeeding or failing out of Washington. Another is to encourage principal-teacher evaluation without mandating, defining, and regulating it from Washington, DC. Another good provision is to encourage but not define and mandate and regulate using measures of growth of students—not just whether they achieved something but whether they are making rapid progress toward a goal. The idea is to make that in terms of whether schools and students are succeeding.

There are many provisions in the Harkin-Enzi bill that have been suggested by both Democrats and Republicans, but there are a number of provisions—not in our legislation—that I don't support, and I am going to seek to amend them. I have indicated to Senators that I intend to offer seven amendments which, in my view, would take out of the legislation provisions that tend to create a national school board. One is the so-called achievement gap. One is the so-called highly qualified teachers provision. These are all provisions that substitute the judgment of people in Washington for that of mayors, local school boards, governors, and legislators. So I don't think we need a national school board, and neither do most Americans.

Some will say: Well, then, why would you support a bill that you don't entirely agree with? The reason is we have a process in Congress. This isn't like the health care bill a few years ago when we had 40 Republican Senators

and Speaker PELOSI was in charge of the House of Representatives. We now have 47 Republican Senators, we have a Republican House of Representatives, and we need to get started fixing this problem. We need to do something a little different around here. Instead of just beating our chests, we need to find a way to put our heads together, head toward a reasonable result, come up with a solution, and offer it to the President and to the American people.

There is no reason in the world why we can't, with the amount of agreement we already have, send to the President by Christmas legislation fixing No Child Left Behind. We should do it because if we don't, Congress's inaction will mean we will transform the U.S. Education Secretary into a waiver-granting czar for 80,000 schools in this country which, according to this law, will be identified as failing.

Well, if we were to have an education czar, or if we were to have a chairman of a national education school board, Secretary Arne Duncan would be a good one. But I don't think we want one in the United States of America. So I think we should act before Christmas in order to avoid creating a waiver education czar, and we should act before Christmas in a way that does not create a national school board.

There is one other suggestion I would make to the authors of this bill. In our earlier meetings with the President, Congressman GEORGE MILLER of California, who was a key leader in developing No Child Left Behind, said this bill to fix No Child Left Behind ought to be a lean bill. I agree with Congressman MILLER. The legislation Republicans introduced a few weeks ago totaled 221 pages in its five bills. The comparable section of the Harkin-Enzi draft is 517 pages. I urge us to follow Congressman MILLER's advice in the final result and be much more succinct than that.

So despite these concerns, I will vote on Wednesday or Thursday, whenever we finish, in favor of bringing this base bill out of the HELP Committee and on to the Senate floor where we can have full amendments. I am going to do my best to improve it in committee and on the Senate floor to make it more like the legislation we introduced a month ago. I am going to continue to do that in the conference we have with the House of Representatives. I think it is time we recognize the American people expect us to step up to major issues, to put our best ideas together, and come up with a result. We are part way there. There is a good place to start.

I thank Senator HARKIN and Senator ENZI for the work they have done, as well as Representative KLINE and Representative MILLER, and I thank the President and Secretary Duncan for their attitude. I look forward to working with them to come to a conclusion.

One last thing: We talk a lot about jobs around here. Every American knows better schools mean better jobs, and they all know schools are a lot like

jobs. We can't create them from Washington, but we can create an environment in which people in their own communities, and families and States can create better schools and better jobs. This is a good place to start.

Mr. President, I ask unanimous consent to have printed in the RECORD a letter of support which also outlines my objections to the legislation that was introduced today, and a copy of an article from the Maryville Alcoa Daily Times today which reminds us of how difficult it is to evaluate teachers fairly and how wise we would be if we satisfied ourselves with creating an environment in which that could happen but did not mandate it, define it, and regulate it from Washington, DC.

There being no objection, the material was ordered to be printed in the RECORD, as follows:

OCTOBER 16, 2011.

Hon. TOM HARKIN,
*Chairman, Committee on Health, Education,
Labor & Pensions, U.S. Senate, Wash-
ington, DC.*

Hon. MIKE ENZI,
*Ranking Member, Committee on Health, Edu-
cation, Labor & Pensions, U.S. Senate,
Washington, DC.*

DEAR TOM AND MIKE: Thank you for the opportunity to participate in discussions about fixing the problems with No Child Left Behind.

I support your base bill (the Elementary and Secondary Education Reauthorization Act of 2011) as a first step in the right direction that will enable our Health, Education, Labor and Pensions (HELP) Committee to start working now to fix the problems with No Child Left Behind. I will vote to move it out of committee, although it is not yet legislation that I could vote in favor of sending to the President.

I have attached a summary of 7 amendments I will offer. Most of these are intended to stop the legislation from creating a national school board that would substitute its judgment for that of governors, state legislatures, mayors, local school board members, parents, principals and teachers. Hopefully, substitute language including these amendments will be the final product of our legislative work.

Despite these misgivings, I believe the HELP Committee should start now with this base bill and try to move an improved bill to the Senate floor where there needs to be a full and complete amendment process to further improve it and send it to a conference with the House of Representatives.

There is no reason why Congress should not be able to send legislation fixing No Child Left Behind to the President by Christmas. If Congress does not act now, our inaction will transform the U.S. Secretary of Education into a waiver-granting czar over an unworkable law that has identified what he says may be as many as 80,000 "failing" public schools, a development even worse than provisions in this draft that would make him a chairman of a national school board. If we were to have such a czar or chairman, Arne Duncan would be a good one, but I do not believe that we should have one in our country.

The strengths of the base bill are that it moves most decisions about whether schools are succeeding or failing out of Washington and back to states and communities. It keeps the valuable reporting requirements of No Child Left Behind. It should help to produce an environment in which states and school districts are more likely to create

principal teacher evaluation systems related to student achievement. It will encourage schools to recognize growth in student academic achievement as well as grade-level performance. The base bill further includes many good provisions suggested by Secretary Duncan and congressional Republicans, as well as Democrats.

The base bill's main weakness is that it contains provisions that would transform the U.S. Secretary of Education into chairman of a national school board. Chief among these problems are federal mandates, definitions and regulations for identifying "achievement gap" schools and the "continuous improvement" of all 100,000 public schools. Although the draft eliminates the concept of "Adequate Yearly Progress" for 95% of schools, these provisions attempt to reinstate it through the back door. In addition, the bill retains in Washington, DC decisions about whether our 3.2 million teachers are "highly qualified" or not. It does not sufficiently consolidate programs and actually creates several new ones that have no real chance of ever being funded. And it does little to make it easier for local school districts to transfer and use federal funds more efficiently or to simplify the burdensome Peer Review process for state plans that must be submitted to the U.S. Department of Education.

There is one other important flaw: the bill is wordy. It is at least 860 pages. When several of us met with President Obama to discuss fixing No Child Left Behind, we agreed to take Congressman George Miller's advice to produce "a lean bill." The five bills offered last month by Senators Isakson, Burr, Kirk and I, along with several other Republican Senators, totaled 221 pages. The comparable sections of your draft total 517 pages. We can be more succinct than that.

Despite these concerns, I will vote in favor of this base bill being reported out of the HELP Committee and look forward to working with you and our colleagues in the Senate and House to improve the bill so that the President can sign it into law this year.

Sincerely,

LAMAR ALEXANDER.

[From the Daily Times (Maryville, TN), Oct. 17, 2011]

GROWING PAINS: BLOUNT SCHOOLS STRUGGLE WITH TEACHER EVALUATION
(By Matthew Stewart)

Blount County Schools have experienced some difficulties in implementing the state's teacher evaluation model, and educators want state lawmakers to give them a voice in the process.

"We don't mind accountability, but it has to be fair," said Grady Caskey, who serves as the Blount County Education Association's president. "The system has to be based on achievable expectations and goals."

Blount County Schools is using the Tennessee Educator Acceleration Model (TEAM), which was developed by the state Department of Education. Alcoa City Schools and Maryville City Schools are using the Teacher Instructional Growth for Effectiveness and Results (TIGER) model, which was developed by the Association of Independent and Municipal Schools.

Both Alcoa and Maryville field-tested evaluation models. However, Blount County didn't field-test a model.

Many county educators have become frustrated with TEAM's implementation, Caskey said. "People are throwing up their hands and saying, 'I'm done.'" Teachers are asking more and more about early retirement requirements. We have two seasoned teachers who are retiring mid-year. Several more are considering it. We're losing our best, most experienced teachers."

BCEA has learned about many implementation problems, he said.

Blount County's principals haven't set uniform requirements, Caskey said. "Some are requiring lesson plans for the entire school year. Others are only requiring observation plans, which is what the law actually requires. I recently received an email from a teacher who puts his kids to bed at 8 p.m. then writes lesson plans until midnight or 1 a.m."

Educators also don't have a template for their lesson plans, he said. "They've got several different versions floating around. It's causing a lot of busy work. I thought the governor said this was going to be less paperwork. We're drowning in it."

Educators need to start talking with lawmakers about the evaluation process, Caskey said. "TEAM is counterproductive. I know we can identify better ways to improve teachers. Legislators are going to have to change it. Politics got us into this mess, and politics will get us out. Education isn't a business. We're not an assembly line. We're not turning out widgets but humans."

STUDENTS IN LIMBO

Many educators are also worried about the evaluation model.

"TEAM has some good points," said Rebecca Dickenson, who is Bagleton Elementary School's librarian. "However, it was implemented in a huge hurry without enough explanation for teachers and principals."

"It's left teachers in limbo with their kids," said Mark Williamson, who teaches social studies at William Blount High School. "Principals are trying their best, but things are constantly changing."

Williamson, a former BCEA president who currently serves on the executive board, thinks the evaluation model has affected his students academically. "I spent 15 hours working on a lesson plan for my first evaluation. At the end of the day, it took 15 hours away from my kids. I couldn't plan ahead, find updated information or seek out current events such as the Arab Spring. I was trying to do what I needed to do according to the lesson plan."

Teacher morale has been impacted as well, he said.

"I haven't seen my principal as much," said Dickenson, who also serves as BCEA's vice president. "I'm used to her walking through the library and getting the opportunity to see what I'm doing in class. However, she's been inundated with evaluations this year."

Lawmakers need to lessen the workload for observers, she said.

RESOLVING PROBLEMS

School officials are working to address teacher concerns, said Director of Schools Rob Britt. "It hasn't been implemented consistently across the state. So, you're going to see these things in every system. We're personally experiencing a lot of growing pains."

Britt and Dr. Jane Morton, supervisor of instruction for grades 6-12, organized two forums with teachers before fall break. They gathered input and created a list of nearly 35 concerns.

School officials are seeking answers from the state Department of Education, Britt said. "I know teachers are concerned about TEAM, and I am as well. We're making efforts to try to get answers for teachers and get more direction for principals. We're very sensitive to teacher concerns. It's high stakes, and we're performing our due diligence for them."

School officials are also working to create supports for teachers, he said. "We want to keep our teachers. We want to support them and help them grow. We're committed as administrators to making it as palatable as possible."

The school district's observers will require more training, Britt said. "Most are implementing the way that they were trained. The state didn't provide exhaustive training. It was more surface-level, which was a good beginning. However, it wasn't thorough. We need more follow-up in a timely manner."

FUTURE PLANS

The state Department of Education is currently evaluating TEAM.

State officials are committed to gathering feedback that will help determine where the evaluation model needs revision, and stakeholders are providing input through several channels.

The Tennessee Consortium on Research, Evaluation and Development (TN CREDE) is launching a statewide survey in spring 2012 and conducting focus groups throughout the year. State officials are also traveling across the state to meet with stakeholders.

The state Department of Education's Advisory Group will bring revision recommendations to Education Commissioner Kevin Huffman. Based on the proposed revisions, the recommendations might need to be brought before the State Board of Education.

I thank the President, and I yield the floor.

The ACTING PRESIDENT pro tempore. The Senator from Oklahoma.

Mr. INHOFE. Mr. President, I asked for unanimous consent to be recognized following the remarks by the Senator from Tennessee. It has been called to my attention that the Senator from Virginia would like to have the floor at this time, so I renew my unanimous consent request that I be recognized at the conclusion of the remarks of the Senator from Virginia.

The ACTING PRESIDENT pro tempore. Without objection, it is so ordered.

The Senator from Virginia.

NATIONAL CRIMINAL JUSTICE COMMISSION ACT

Mr. WEBB. Mr. President, I wish to thank my colleague from Oklahoma for giving me the courtesy of speaking, and I thank him again for the work he has done on the Foreign Relations Committee, Subcommittee on East Asian Affairs, where he is the ranking Republican, and the other work he has done on the Armed Services Committee.

Today I rise to speak about the National Criminal Justice Commission legislation which I introduced more than 2 years ago and which the leader and the managers of this bill are now going to offer as an amendment to the pending legislation. First of all, I thank the leader and the managers of the bill for calling up this legislation. I also thank my principal Republican cosponsor, Senator LINDSEY GRAHAM, for all the work he has done.

There are good national commissions and bad national commissions and redundant national commissions and sometimes there are national commissions which are not only needed but vital to the resolution of issues we face.

I am thinking, as I speak, of the first Commission on Wartime Contracting

which Senator CLAIRE MCCASKILL and I introduced 4 years ago and which resulted in a finding of approximately \$30 billion in fraud, waste, and abuse in contracts that had gone to Iraq and Afghanistan and which provided a model for the way we should be approaching such contracts in the future. I would put this particular national commission in that category. It was put together after much thought and many hearings. It is paid for, it is sunsetted at 18 months, and it is dedicated to helping us resolve an issue of very serious national purpose.

I began on this issue before I came to the Senate—the issue of the imbalance in our criminal justice system and the need to bring a comprehensive resolution in terms of how we handle crime and reentry in this country. We have had more than 2½ years of hearings since I came to the Senate. After I introduced this legislation, we met—at staff levels, since I am not on the Judiciary Committee—with representatives from more than 100 different organizations across the country and across the philosophical spectrum.

This chart is an indication of the type of support we have received for this commission. I will not read the names, and I don't expect anyone viewing the TV screen to be able to read all the names, but this is an unusual circumstance. We have organizations as philosophically diverse as the ACLU, the NAACP, the Sentencing Project, the National Organization for Victim Assistance, the ABA Criminal Justice Section, the National Center for Victims of Crime, along with the Fraternal Order of Police, the National Sheriffs Association, and the International Association of Chiefs of Police, which all agree we need to step forward and examine our criminal justice system in a comprehensive way, from point of apprehension to point of return, so that we make better use of our assets and make better use of our own people, quite frankly.

Today we incarcerate more people than any other country in the Western world or in any known country in the world. We have 2.3 million people in our prisons and jails and another 5 million people on probation or in some way under postcorrectional management. Hundreds of thousands of people are being released from jails and prisons every year and reentering society, and at this point we are without a comprehensive structure that will allow those who wish to become productive citizens again the opportunity to have the right kind of transition.

At the same time, we have 7 million people under some form of correctional supervision or in prisons and we don't feel any safer. This is the other beam our analysis has ridden as we looked at this. Even today, if we ask Americans, two-thirds of the people in this country believe crime is more prevalent today than it was a year ago.

So we were tasked—we tasked ourselves—with looking at this problem to

try to figure out how we can do a better job of addressing the issue of criminal justice, spending less money. We are now in a situation where State and local budgets have been stretched to the breaking point. Professor Western of Harvard estimates that annual correctional spending right now is about \$70 billion, with State spending on corrections increasing 40 percent over the past 20 years.

We are witnessing a war on our border with respect to gang warfare. Since President Calderon launched an offensive against drug gangs and cartels in 2006, tens of thousands of people have died in drug trafficking violence along the border. It is estimated that these cartels are now operating in more than 230 cities and towns in the United States. These entities need to be examined in the context of transnational gang activity as they relate to our criminal justice system.

We are also largely housing our Nation's mentally ill in our prison system. The number of mentally ill in prison right now is nearly five times the number of mentally ill in inpatient mental hospitals. Noted experts have cited jails and prisons as the No. 1 holding facility for the mentally ill.

So the conclusion we reached, after listening to dozens of representatives from different organizations across the philosophical spectrum, was that we need to have a long-overdue, top-to-bottom, beginning-to-end examination of how the criminal justice system works in the United States from point of apprehension to the decision of whether to arrest. And, if arrested, what sort of port does a person go into? How long should that person be in prison? What should prison administration look like, and how could that be better adapted? What models do we have out there that can be applied? What should reentry programs look like, and how do we deal with the ever-increasing problems of transnational gangs? We need to examine all of those pieces together.

The last review of this nature that was undertaken was done in 1965 by President Lyndon Johnson. So I introduced the National Criminal Justice Act, the goal of which is to create a blue ribbon national commission, time sunsetted—18 months—to get the finest minds in the country together to examine these different pieces and to come back to the Congress with specific recommendations for reforming our national criminal justice system.

Just last week, in a meeting of the Senate Law Enforcement Caucus, Philadelphia Police Chief Charles Ramsey noted the tremendous influence of this last commission's report, which was reported in 1967—44 years ago—and voiced strong support for the creation of a new commission. We are long overdue to look at what works and what doesn't in our criminal justice system.

This bill has, quite frankly, struck a nerve across the country. I have heard from citizens across all 50 States in support of this initiative. I mentioned

**APPENDIX C:
DATA REQUEST LETTER**

Michael O'Neal

Independent School District

, TX

June 1, 2012

Dear Mr.

I am a doctoral student in the Department of Special Education at the University of Central Florida in Orlando. I am presently involved in research for my Doctoral Dissertation entitled, An Investigation of the Effectiveness of Teacher Preparation Programs as Measured by Student Achievement.

This study will examine the relationship, if any, between teacher preparation program attended and student mathematics achievement growth as measured by the Texas Assessment of Knowledge and Skills (TAKS).

I will be more than willing to share my findings with you and your staff at the conclusion of my research. I would like to request your permission to collect data pertaining to all students who participated in the TAKS during the 2009, 2010, and 2011 administrations and their teachers of record for mathematics in the 2009-2010 academic year. The data will be kept under strict confidence and all participants will remain anonymous in the reporting and analysis of the research data. The data will need to consist of demographic information for both, teachers and students and TAKS results for 2008-2009, 2009-2010, and 2010-2011 academic years.

This research will involve no contact with students or teachers and no time will be taken away from instruction. I would request the assistance of administrative personnel familiar with the data needed for my research and most effective method of retrieving it.

If you consent to the use of this information in my dissertation, please reply in writing so that I might add the permission letter to my dissertation. If you have any questions I may be reached at or by email at .. Thank you for your attention and support.

Respectfully,



Michael O'Neal

**APPENDIX D:
RESPONSE TO DATA REQUEST**

Michael O'Neal

June 12, 2012

Superintendent

Dear Mr. O'Neal

The _____ considered your request and hereby grants permission for your study with the following stipulation:

Assistant Superintendent
Administrative Services

- The data should be gathered using eduphoria! SchoolObjects:aware, the district's current data management system. I will provide any assistance you might need with the program.

Executive Director
Curriculum and
Instructional Services

- All demographic data for participants, whether students or teachers, must be kept confidential and anonymous in analysis and reporting.

- The collection of any data not available through Aware, will be coordinated through my office.

Director of Transportation

If you have any questions or require clarification on any of the above points, please let me know. We look forward to working with you and seeing the results of your study.

Sincerely,

Assistant Superintendent

**APPENDIX E:
PERMISSION TO REPRINT RESPONSE LETTER**

Independent School District

Avenue
Texas

Superintendent

Michael O'Neal

June 12, 2012

Dear Mr. O'Neal

Permission is hereby granted to reprint the original letter from me granting permission for your study.

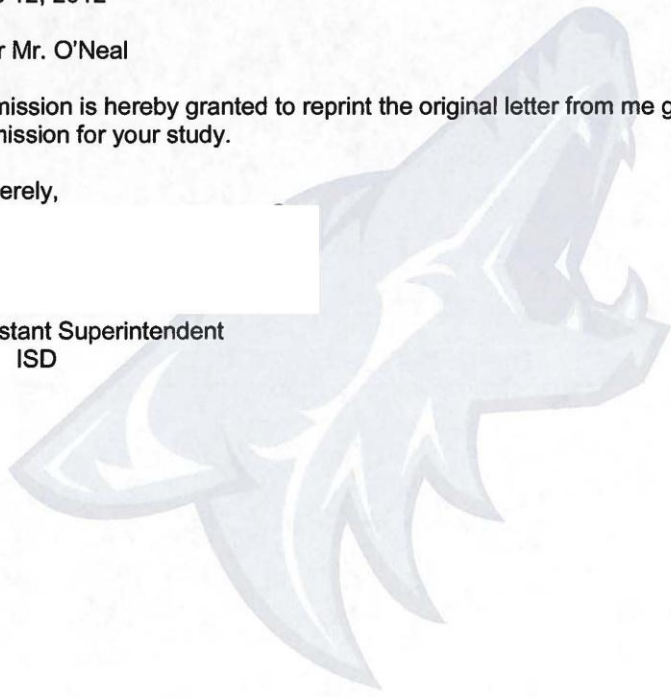
Assistant Superintendent
Administrative Services

Sincerely,

Executive Director
Curriculum and
Instructional Services

Assistant Superintendent
ISD

Director of Transportation



**APPENDIX F:
COPYRIGHT PERMISSION FOR TAKS ITEMS**



Michael Oneal <onealm65@gmail.com>

Released TAKS

Copyrights <Copyrights@tea.state.tx.us> Wed, Dec 19, 2012 at 6:24 PM
To: Michael O'Neal <onealm65@gmail.com>, Copyrights <Copyrights@tea.state.tx.us>

Dear Mr. O'Neal,

The TEA Materials (hereafter called "Materials") you requested are referenced and described in your email attached below are incorporated into as part of this document.

Copyright restrictions are stated below in the TEA Copyright Notice.

Service to Be Performed Description: Exactly as described below in the reprinted "original message" and is made part of this Permission Granted. Your use as described immediately in the email attached below is covered by TEA's Copyright and Terms of Service Policy discussed above and reprinted below; and, you are granted the right to utilize the Materials in accordance with the further terms and conditions stated within this document. You may not market or sell TEA's copyrighted Materials, or TEA-owned Materials that might be contained within a Derivative Work without an additional License Agreement from TEA. You may not give TEA-owned materials to any other persons. The permission has been granted for the term of this project. You must submit a new request once this project has terminated or substantially changed from its original inception.

You are granted permission to use pictures of two released mathematics questions (1 multiple choice and 1 griddable) available from the Agency's website as examples for your paper.

If you perform your activity as described in the below "original message" and attached hereto, then you have agreed to the terms and conditions listed within this communication. Any future exact replication of this activity described in the Service to Be Performed Description section above does require further communication with TEA. Any future changes that modify the terms and conditions or Service to Be Performed Description above and below stated will require you to contact TEA at Copyrights@tea.state.tx.us.

Royalty Payment: None - Educational Purposes

Period: As stated in the request.

Limitation on Number of Copies Produced: As stated in the request.

Acknowledgement of Receipt of Permission:

**APPENDIX G:
HLM RESULTS, NULL MODEL**

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 30 March 2013, Saturday
Time: 22:44:10

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{oi} + e_{ti}$$

Level-2 Model

$$\pi_{oi} = \beta_{00} + r_{oi}$$

Mixed Model

$$TMVSS_{ti} = \beta_{00} + r_{oi} + e_{ti}$$

Final Results – Iteration 5

Iterations stopped due to small change in likelihood function

$\sigma^2 = 5284.94084$

Standard error of $\sigma^2 = 193.70264$

τ

INTRCPT1, π_0 5301.16717

Standard error of τ

INTRCPT1, π_0 386.11101

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.749

The value of the log-likelihood function at iteration 5 = -1.287474E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	694.732733	3.209135	216.486	686	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	694.732733	3.209747	216.445	686	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	72.80911	5301.16717	686	2892.40629	<0.001
level-1, e	72.69760	5284.94084			

Statistics for the current model

Deviance = 25749.479319

Number of estimated parameters = 3

Test of homogeneity of level-1 variance

χ^2 statistic = 1247.08050
degrees of freedom = 686
 p -value = 0.000

**APPENDIX H:
HLM RESULTS, UNCONDITIONAL GROWTH MODEL**

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 30 March 2013, Saturday
Time: 22:57:17

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{0i} + \pi_{1i} * (GRADECOD_{ti}) + e_{ti}$$

Level-2 Model

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

Mixed Model

$$TMVSS_{ti} = \beta_{00} + \beta_{10} * GRADECOD_{ti} + r_{0i} + r_{1i} * GRADECOD_{ti} + e_{ti}$$

Final Results – Iteration 1296

Iterations stopped due to small change in likelihood function

$$\sigma^2 = 2635.95704$$

$$\text{Standard error of } \sigma^2 = 111.49127$$

τ

INTRCPT1, π_0	4267.08891	-117.40283
GRADECOD, π_1	-117.40283	10.18575

Standard errors of τ

INTRCPT1, π_0	415.79710	116.28618
GRADECOD, π_1	116.28618	41.50133

τ (as correlations)

INTRCPT1, π_0	1.000	-0.563
GRADECOD, π_1	-0.563	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.528
GRADECOD, π_1	0.011

The value of the log-likelihood function at iteration 1296 = -1.222453E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	612.038783	3.359040	182.206	686	<0.001
For GRADECOD slope, π_1					
INTRCPT2, β_{10}	43.730854	1.026063	42.620	686	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	612.038783	3.328186	183.896	686	<0.001
For GRADECOD slope, π_1					
INTRCPT2, β_{10}	43.730854	0.997136	43.856	686	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	<i>d.f.</i>	χ^2	<i>p</i> -value
INTRCPT1, r_0	65.32296	4267.08891	686	1647.28007	<0.001
GRADECOD slope, r_1	3.19151	10.18575	686	712.93549	0.231
level-1, e	51.34157	2635.95704			

Statistics for the current model

Deviance = 24449.064693

Number of estimated parameters = 6

Test of homogeneity of level-1 variance

χ^2 statistic = 1064.54089

degrees of freedom = 487

p-value = 0.000

**APPENDIX I:
HLM RESULTS, TEST OF SIGNIFICANCE FOR STUDENT DEMOGRAPHIC
COVARIATES**

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 31 March 2013, Sunday
Time: 1:21:31

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{0i} + e_{ti}$$

Level-2 Model

$$\pi_{0i} = \beta_{00} + \beta_{01}*(MALE_i) + r_{0i}$$

Mixed Model

$$TMVSS_{ti} = \beta_{00} + \beta_{01}*MALE_i + r_{0i} + e_{ti}$$

Final Results – Iteration 5

Iterations stopped due to small change in likelihood function

$\sigma^2 = 5285.00606$

Standard error of $\sigma^2 = 193.70497$

τ

INTRCPT1, π_0 5300.39032

Standard error of τ

INTRCPT1, π_0 386.07053

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.749

The value of the log-likelihood function at iteration 5 = -1.287471E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	693.927904	4.699363	147.664	685	<0.001
MALE, β_{01}	1.508412	6.432583	0.234	685	0.815

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	693.927904	4.728498	146.754	685	<0.001
MALE, β_{01}	1.508412	6.438689	0.234	685	0.815

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	72.80378	5300.39032	685	2891.90978	<0.001
level-1, e	72.69805	5285.00606			

Statistics for the current model

Deviance = 25749.424361

Number of estimated parameters = 4

Test of homogeneity of level-1 variance

χ^2 statistic = 1247.08050
degrees of freedom = 686
 p -value = 0.000

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 31 March 2013, Sunday
Time: 1:22:13

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{oi} + e_{ti}$$

Level-2 Model

$$\pi_{oi} = \beta_{00} + \beta_{01}*(MINORITY)_i + r_{oi}$$

Mixed Model

$$TMVSS_{ti} = \beta_{00} + \beta_{01}*MINORITY_i + r_{oi} + e_{ti}$$

Final Results – Iteration 5

Iterations stopped due to small change in likelihood function

$\sigma^2 = 5287.10635$

Standard error of $\sigma^2 = 193.77448$

τ

INTRCPT1, π_0 5192.41328

Standard error of τ

INTRCPT1, π_0 380.31144

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.745

The value of the log-likelihood function at iteration 5 = -1.286976E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	702.492898	4.010949	175.144	685	<0.001
MINORITY, β_{01}	-20.909288	6.597474	-3.169	685	0.002

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	702.492898	4.076629	172.322	685	<0.001
MINORITY, β_{01}	-20.909288	6.521685	-3.206	685	0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	72.05840	5192.41328	685	2841.14888	<0.001
level-1, e	72.71249	5287.10635			

Statistics for the current model

Deviance = 25739.520822

Number of estimated parameters = 4

Test of homogeneity of level-1 variance

χ^2 statistic = 1247.08050
degrees of freedom = 686
 p -value = 0.000

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 31 March 2013, Sunday
Time: 1:22:54

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{0i} + e_{ti}$$

Level-2 Model

$$\pi_{0i} = \beta_{00} + \beta_{01}*(LANGPROG_i) + r_{0i}$$

Mixed Model

$$TMVSS_{ti} = \beta_{00} + \beta_{01}*LANGPROG_i + r_{0i} + e_{ti}$$

Final Results – Iteration 5

Iterations stopped due to small change in likelihood function

$\sigma^2 = 5287.20851$

Standard error of $\sigma^2 = 193.76729$

τ

INTRCPT1, π_0 5039.95021

Standard error of τ

INTRCPT1, π_0 372.12595

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.739

The value of the log-likelihood function at iteration 5 = -1.286218E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	699.631636	3.290038	212.652	685	<0.001
LANGPROG, β_{01}	-57.547760	11.367236	-5.063	685	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	699.631636	3.338504	209.564	685	<0.001
LANGPROG, β_{01}	-57.547760	9.571235	-6.013	685	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	70.99261	5039.95021	685	2776.44164	<0.001
level-1, e	72.71319	5287.20851			

Statistics for the current model

Deviance = 25724.355652

Number of estimated parameters = 4

Test of homogeneity of level-1 variance

χ^2 statistic = 1247.08050
degrees of freedom = 686
 p -value = 0.000

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 31 March 2013, Sunday
Time: 1:24: 0

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{0i} + e_{ti}$$

Level-2 Model

$$\pi_{0i} = \beta_{00} + \beta_{01}*(SPED_i) + r_{0i}$$

Mixed Model

$$TMVSS_{ti} = \beta_{00} + \beta_{01}*SPED_i + r_{0i} + e_{ti}$$

Final Results – Iteration 5

Iterations stopped due to small change in likelihood function

$\sigma^2 = 5286.51213$

Standard error of $\sigma^2 = 193.75723$

τ

INTRCPT1, π_0 5257.98778

Standard error of τ

INTRCPT1, π_0 383.82192

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.747

The value of the log-likelihood function at iteration 5 = -1.287288E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	696.022793	3.267277	213.028	685	<0.001
SPED, β_{01}	-31.130935	16.129581	-1.930	685	0.054

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	696.022793	3.285956	211.817	685	<0.001
SPED, β_{01}	-31.130935	13.970360	-2.228	685	0.026

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	72.51198	5257.98778	685	2870.24856	<0.001
level-1, e	72.70840	5286.51213			

Statistics for the current model

Deviance = 25745.768223

Number of estimated parameters = 4

Test of homogeneity of level-1 variance

χ^2 statistic = 1247.08050
degrees of freedom = 686
 p -value = 0.000

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 31 March 2013, Sunday
Time: 1:24:31

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{oi} + e_{ti}$$

Level-2 Model

$$\pi_{oi} = \beta_{00} + \beta_{01}*(GIFTED_i) + r_{oi}$$

Mixed Model

$$TMVSS_{ti} = \beta_{00} + \beta_{01}*GIFTED_i + r_{oi} + e_{ti}$$

Final Results – Iteration 5

Iterations stopped due to small change in likelihood function

$\sigma^2 = 5287.21747$

Standard error of $\sigma^2 = 193.76757$

τ

INTRCPT1, π_0 5039.31085

Standard error of τ

INTRCPT1, π_0 372.09179

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.739

The value of the log-likelihood function at iteration 5 = -1.286215E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	691.484556	3.216815	214.959	685	<0.001
GIFTED, β_{01}	79.926566	15.768420	5.069	685	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	691.484556	3.245306	213.072	685	<0.001
GIFTED, β_{01}	79.926566	12.364806	6.464	685	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	70.98810	5039.31085	685	2776.93946	<0.001
level-1, e	72.71326	5287.21747			

Statistics for the current model

Deviance = 25724.294015

Number of estimated parameters = 4

Test of homogeneity of level-1 variance

χ^2 statistic = 1247.08050
degrees of freedom = 686
 p -value = 0.000

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 31 March 2013, Sunday
Time: 1:24:58

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{0i} + e_{ti}$$

Level-2 Model

$$\pi_{0i} = \beta_{00} + \beta_{01} * (FRPL_i) + r_{0i}$$

Mixed Model

$$TMVSS_{ti} = \beta_{00} + \beta_{01} * FRPL_i + r_{0i} + e_{ti}$$

Final Results – Iteration 5

Iterations stopped due to small change in likelihood function

$\sigma^2 = 5285.53242$

Standard error of $\sigma^2 = 193.71727$

τ

INTRCPT1, π_0 5197.72544

Standard error of τ

INTRCPT1, π_0 380.56654

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.745

The value of the log-likelihood function at iteration 5 = -1.286978E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	702.890849	4.092507	171.751	685	<0.001
FRPL, β_{01}	-20.623535	6.518973	-3.164	685	0.002

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	702.890849	4.180108	168.151	685	<0.001
FRPL, β_{01}	-20.623535	6.444524	-3.200	685	0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	72.09525	5197.72544	685	2847.07661	<0.001
level-1, e	72.70167	5285.53242			

Statistics for the current model

Deviance = 25739.550054

Number of estimated parameters = 4

Test of homogeneity of level-1 variance

χ^2 statistic = 1247.08050
degrees of freedom = 686
 p -value = 0.000

**APPENDIX J:
HLM RESULTS, CONDITIONAL GROWTH MODEL**

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 31 March 2013, Sunday
Time: 1:27:40

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{0i} + \pi_{1i}^*(GRADECOD_{ti}) + e_{ti}$$

Level-2 Model

$$\begin{aligned}\pi_{0i} &= \beta_{00} + \beta_{01}^*(MINORITY_i) + \beta_{02}^*(LANGPROG_i) + \beta_{03}^*(SPED_i) + \beta_{04}^*(GIFTED_i) \\ &\quad + \beta_{05}^*(FRPL_i) + r_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11}^*(MINORITY_i) + \beta_{12}^*(LANGPROG_i) + \beta_{13}^*(SPED_i) + \beta_{14}^*(GIFTED_i) \\ &\quad + \beta_{15}^*(FRPL_i)\end{aligned}$$

Mixed Model

$$\begin{aligned}TMVSS_{ti} &= \beta_{00} + \beta_{01}^*MINORITY_i + \beta_{02}^*LANGPROG_i + \beta_{03}^*SPED_i \\ &\quad + \beta_{04}^*GIFTED_i + \beta_{05}^*FRPL_i\end{aligned}$$

file:///Volumes/NO%20NAME/hlm2%20conditional%20growth%20model.html[4/1/13 2:57:32 AM]

$$\begin{aligned}
& + \beta_{10}^{04} \text{GRADECOD}_{ti} + \beta_{11}^{05} \text{MINORITY}_i \text{GRADECOD}_{ti} + \beta_{12} \text{LANGPROG}_i \text{GRADECOD}_{ti} + \\
& \beta_{13} \text{SPED}_i \text{GRADECOD}_{ti} \\
& + \beta_{14} \text{GIFTED}_i \text{GRADECOD}_{ti} + \beta_{15} \text{FRPL}_i \text{GRADECOD}_{ti} \\
& + r_{0i} + e_{ti}
\end{aligned}$$

Final Results – Iteration 7

Iterations stopped due to small change in likelihood function

$$\sigma^2 = 2633.71046$$

Standard error of $\sigma^2 = 96.56933$

τ

INTRCPT1, π_0 3420.51668

Standard error of τ

INTRCPT1, π_0 234.27296

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.794

The value of the log-likelihood function at iteration 7 = -1.218687E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	622.616184	4.546581	136.942	681	<0.001
MINORITY, β_{01}	-10.014057	7.201667	-1.391	681	0.165
LANGPROG, β_{02}	-7.889998	10.995160	-0.718	681	0.473
SPED, β_{03}	-31.407714	17.145938	-1.832	681	0.067
GIFTED, β_{04}	87.495613	15.311646	5.714	681	<0.001
FRPL, β_{05}	-22.176516	6.847483	-3.239	681	0.001
For GRADECOD slope, π_1					
INTRCPT2, β_{10}	42.383433	1.459368	29.042	1479	<0.001
MINORITY, β_{11}	0.615652	2.287479	0.269	1479	0.788
LANGPROG, β_{12}	6.986372	5.388772	1.296	1479	0.195
SPED, β_{13}	-5.777994	5.426090	-1.065	1479	0.287
GIFTED, β_{14}	-4.039636	4.888452	-0.826	1479	0.409
FRPL, β_{15}	3.489532	2.196018	1.589	1479	0.112

Final estimation of fixed effects

(with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	622.616184	4.707081	132.272	681	<0.001
MINORITY, β_{01}	-10.014057	7.184920	-1.394	681	0.164
LANGPROG, β_{02}	-7.889998	10.513374	-0.750	681	0.453
SPED, β_{03}	-31.407714	16.930757	-1.855	681	0.064
GIFTED, β_{04}	87.495613	12.803726	6.834	681	<0.001
FRPL, β_{05}	-22.176516	6.952153	-3.190	681	0.001
For GRADECOD slope, π_1					
INTRCPT2, β_{10}	42.383433	1.449876	29.232	1479	<0.001
MINORITY, β_{11}	0.615652	2.151435	0.286	1479	0.775
LANGPROG, β_{12}	6.986372	5.383381	1.298	1479	0.195
SPED, β_{13}	-5.777994	4.760868	-1.214	1479	0.225
GIFTED, β_{14}	-4.039636	4.905059	-0.824	1479	0.410
FRPL, β_{15}	3.489532	2.077847	1.679	1479	0.093

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	58.48518	3420.51668	681	3559.08003	<0.001
level-1, e	51.31969	2633.71046			

Statistics for the current model

Deviance = 24373.736147

Number of estimated parameters = 14

Test of homogeneity of level-1 variance

χ^2 statistic = 1327.72559

degrees of freedom = 686

p-value = 0.000

**APPENDIX K:
HLM RESULTS, FULL CONDITIONAL MODEL**

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.00.21106.30001)
Date: 31 March 2013, Sunday
Time: 2: 2:12

Specifications for this HLM2 run

Problem Title: no title

The maximum number of level-1 units = 2172
The maximum number of level-2 units = 687
The maximum number of iterations = 100

Method of estimation: full maximum likelihood

The outcome variable is TMVSS

Summary of the model specified

Level-1 Model

$$TMVSS_{ti} = \pi_{0i} + \pi_{1i}*(GRADECOD_{ti}) + e_{ti}$$

Level-2 Model

$$\begin{aligned} \pi_{0i} = & \beta_{00} + \beta_{01}*(MINORITY_i) + \beta_{02}*(LANGPROG_i) + \beta_{03}*(SPED_i) + \beta_{04}*(GIFTED_i) \\ & + \beta_{05}*(FRPL_i) + \beta_{06}*(GROUP1_i) + \beta_{07}*(GROUP2_i) + \beta_{08}*(GROUP3_i) \\ & + \beta_{09}*(GROUP4_i) + r_{0i} \\ \pi_{1i} = & \beta_{10} + \beta_{11}*(MINORITY_i) + \beta_{12}*(LANGPROG_i) + \beta_{13}*(SPED_i) + \beta_{14}*(GIFTED_i) \\ & + \beta_{15}*(FRPL_i) + \beta_{16}*(GROUP1_i) + \beta_{17}*(GROUP2_i) + \beta_{18}*(GROUP3_i) \\ & + \beta_{19}*(GROUP4_i) + r_{1i} \end{aligned}$$

GROUP1 GROUP2 GROUP3 GROUP4 have been centered around the grand mean.

Mixed Model

$$\begin{aligned}
 TMVSS_{ti} = & \beta_{00} + \beta_{01} * MINORITY_i + \beta_{02} * LANGPROG_i + \beta_{03} * SPED_i \\
 & + \beta_{04} * GIFTED_i + \beta_{05} * FRPL_i + \beta_{06} * GROUP1_i + \beta_{07} * GROUP2_i \\
 & + \beta_{08} * GROUP3_i + \beta_{09} * GROUP4_i \\
 & + \beta_{10} * GRADECOD_{ti} + \beta_{11} * MINORITY_i * GRADECOD_{ti} + \beta_{12} * LANGPROG_i * GRADECOD_{ti} + \\
 & \beta_{13} * SPED_i * GRADECOD_{ti} \\
 & + \beta_{14} * GIFTED_i * GRADECOD_{ti} + \beta_{15} * FRPL_i * GRADECOD_{ti} + \beta_{16} * GROUP1_i * GRADECOD_{ti} + \\
 & \beta_{17} * GROUP2_i * GRADECOD_{ti} \\
 & + \beta_{18} * GROUP3_i * GRADECOD_{ti} + \beta_{19} * GROUP4_i * GRADECOD_{ti} \\
 & + r_{0i} + r_{1i} * GRADECOD_{ti} + e_{ti}
 \end{aligned}$$

Final Results – Iteration 1359

Iterations stopped due to small change in likelihood function

$\sigma^2 = 2607.44019$

Standard error of $\sigma^2 = 109.37605$

τ

INTRCPT1, π_0	3690.51153	-162.67421
GRADECOD, π_1	-162.67421	12.38498

Standard errors of τ

INTRCPT1, π_0	379.84658	110.49084
GRADECOD, π_1	110.49084	39.67109

τ (as correlations)

INTRCPT1, π_0	1.000	-0.761
GRADECOD, π_1	-0.761	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.499
GRADECOD, π_1	0.014

The value of the log-likelihood function at iteration 1359 = -1.215696E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	619.415320	4.742380	130.613	677	<0.001
MINORITY, β_{01}	-8.669626	7.394034	-1.173	677	0.241
LANGPROG, β_{02}	-10.668168	11.283137	-0.945	677	0.345

SPED, β_{03}	-22.385215	19.150366	-1.169	677	0.243
GIFTED, β_{04}	88.304324	15.733066	5.613	677	<0.001
FRPL, β_{05}	-21.149821	6.994139	-3.024	677	0.003
GROUP1, β_{06}	-19.724865	14.214799	-1.388	677	0.166
GROUP2, β_{07}	-82.671826	45.991621	-1.798	677	0.073
GROUP3, β_{08}	-18.533364	10.656239	-1.739	677	0.082
GROUP4, β_{09}	6.745858	7.227797	0.933	677	0.351
For GRADECOD slope, π_1					
INTRCPT2, β_{10}	45.354387	1.507630	30.083	677	<0.001
MINORITY, β_{11}	0.726119	2.265187	0.321	677	0.749
LANGPROG, β_{12}	4.194586	5.236975	0.801	677	0.423
SPED, β_{13}	0.897732	6.484895	0.138	677	0.890
GIFTED, β_{14}	-7.713233	4.859725	-1.587	677	0.113
FRPL, β_{15}	3.855607	2.171036	1.776	677	0.076
GROUP1, β_{16}	-1.619708	3.311195	-0.489	677	0.625
GROUP2, β_{17}	-4.068970	12.162507	-0.335	677	0.738
GROUP3, β_{18}	14.888333	4.697826	3.169	677	0.002
GROUP4, β_{19}	7.440778	2.583168	2.880	677	0.004

**Final estimation of fixed effects
(with robust standard errors)**

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2, β_{00}	619.415320	4.754483	130.280	677	<0.001
MINORITY, β_{01}	-8.669626	7.108466	-1.220	677	0.223
LANGPROG, β_{02}	-10.668168	10.723587	-0.995	677	0.320
SPED, β_{03}	-22.385215	16.171836	-1.384	677	0.167
GIFTED, β_{04}	88.304324	13.410464	6.585	677	<0.001
FRPL, β_{05}	-21.149821	6.936932	-3.049	677	0.002
GROUP1, β_{06}	-19.724865	13.234481	-1.490	677	0.137
GROUP2, β_{07}	-82.671826	48.248856	-1.713	677	0.087
GROUP3, β_{08}	-18.533364	10.458280	-1.772	677	0.077
GROUP4, β_{09}	6.745858	7.127352	0.946	677	0.344
For GRADECOD slope, π_1					
INTRCPT2, β_{10}	45.354387	1.530814	29.628	677	<0.001
MINORITY, β_{11}	0.726119	2.106580	0.345	677	0.730
LANGPROG, β_{12}	4.194586	5.365112	0.782	677	0.435
SPED, β_{13}	0.897732	5.158621	0.174	677	0.862
GIFTED, β_{14}	-7.713233	4.421773	-1.744	677	0.082
FRPL, β_{15}	3.855607	2.062318	1.870	677	0.062
GROUP1, β_{16}	-1.619708	3.190434	-0.508	677	0.612

GROUP2, β_{17}	-4.068970	11.533054	-0.353	677	0.724
GROUP3, β_{18}	14.888333	4.254674	3.499	677	<0.001
GROUP4, β_{19}	7.440778	2.602864	2.859	677	0.004

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	<i>d.f.</i>	χ^2	<i>p</i> -value
INTRCPT1, r_0	60.74958	3690.51153	677	1487.51268	<0.001
GRADECOD slope, r_1	3.51923	12.38498	677	704.01682	0.229
level-1, <i>e</i>	51.06310	2607.44019			

Statistics for the current model

Deviance = 24313.918458
Number of estimated parameters = 24

Test of homogeneity of level-1 variance

χ^2 statistic = 1039.92069
degrees of freedom = 486
p-value = 0.000

REFERENCES

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95-135.
- Abedi, J. & Dietel, R. (2004). Challenges in the No Child Left Behind Act for English language learners. *Phi Delta Kappan*, 85(10), 782-785.
- Abedi, J. & Gandara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurements: Issues and Practices*, 26(5), 36-46.
- Adams, J. E., & Kirst, M. (1999). New demands for educational accountability: Striving for results in an era of excellence. In J. Murphey & K. S. Louis (Eds.), *Handbook of research in educational administration* (2nd ed., pp. 463–489). San Francisco: Jossey-Bass.
- Adelson, J. L., McCoach, D. B., & Gavin, M. K. (2012). Examining the effects of gifted programming in mathematics and reading using the ECLS-K. *Gifted Child Quarterly*, 56(1), 25-39.
- Alexander, K., & Alexander, M. D. (2001). *American public school law* (5th ed.). Belmont, CA: Wadsworth.
- Alexander, K., & Salmon R. G. (1995). *Public school finance*. Boston: Allyn and Bacon.
- Alexander, Sen. [TN]. (2011, October 17). Letter to Sens. Harkin [IA] & Enzi [WY]. In *Congressional Record* 144, S6572. Retrieved from Government Printing Office website.
- Allmendinger, D. F. (1975). *Paupers and scholars: The transformation of student life in nineteenth-century New England*. New York: St. Martin's Press.
- America COMPETES Act, Pub. L. No. 110-69, 121 Stat. 572 (2007). Retrieved from the Government Printing Office website.

- American Association of Colleges for Teacher Education. (2011, June). *Transformations in educator preparation: Effectiveness and accountability*. Retrieved from <http://www.aacte.org>
- American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 15 (2009). Retrieved from the Government Printing Office website.
- Amrein-Beardsley A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65–75.
- Astin, A. W., & American Council on Education. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. New York: American Council on Education.
- Baker, E. L., Bartson, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., et al. (2010). *Problems with the use of student test scores to evaluate teachers*. (EPI Briefing Paper No. 278). Washington, DC: Economic Policy Institute. Retrieved from <http://www.epi.org/publication/bp278/>
- Ballou, D. (2002). Sizing up test scores. *Education Next*, 2(2), 10-15. Retrieved from http://educationnext.org/files/ednext20022_10.pdf
- Bankston, I. C. (2010). Federal control of public schools and the decline of community. *Modern Age*, 52(3), 184-197.
- Berry, B., Fuller, E., Reeves, C., & Laird, E. (2007). Linking teacher and student data to improve teacher and teaching quality. In *National Center for Educational Accountability, Data Quality Campaign*. [http://www. dataqualitycampaign. org/files/Meetings-DQC_Quarterly_Issue_Brief_031207. pdf](http://www.dataqualitycampaign.org/files/Meetings-DQC_Quarterly_Issue_Brief_031207.pdf)

- Bhatt, R. (2009). The impacts of gifted and talented education. *Andrew Young School of Policy Studies Research Paper Series*, (09-11).
- Blue Ribbon Commission on Teacher Quality. (2000). Blue ribbon commission on teacher quality recommendations: Year one report. Retrieved from <http://regents.louisiana.gov/assets/docs/TeacherPreparation/BRCYear1Report.pdf>
- Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, 31(4), 416-440. doi:10.3102/0162373709353129
- Braun, H. I. (2005, September). Using student progress to evaluate teachers: A primer on value-added models. Retrieved from the Educational Testing Service website: <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>
- Brubacher, J. S., & Rudy, W. (1997). *Higher education in transition: A history of American colleges and universities*. New Brunswick, NJ: Transaction Publishers.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis techniques*. Newbury Park, CA: Sage.
- Bui, S. A., Craig, S. G., & Imberman, S. A. (2011). *Is gifted education a bright idea? Assessing the impact of gifted and talented programs on achievement* (No. w17089). National Bureau of Economic Research.
- Butts, R. F., & Cremin, L. A. (1953). *A history of education in American culture*. New York: Holt, Rinehart & Winston.
- Callan, P. (1998). The role of state policy systems in fostering separation or collaboration. In M. Timpane & L. White (Eds.) *Higher education and school reform* (pp. 41-56). San Francisco: Jossey-Bass.

- Center on Education Policy. (2008, June). *Has student achievement increased since 2002? State test score trends through 2006–07*. Retrieved August 1, 2012, from <http://www.cep-dc.org>
- Clifford, G., and Guthrie, J. (1988). *Ed school: A brief for professional education*. Chicago: University of Chicago Press.
- Clotfelter, C. T., Ladd, H. F. & Vigdor, J. L. (2009). The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics*, 91(2), pp. 398-419.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26, 673-682.
- Cochran-Smith, M., & Zeichner, K. (2005). *Studying teacher education: The report of the AERA panel on research and teacher education*. Mahwah, NJ: Lawrence Erlbaum.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Council for a Better Louisiana (2001). *The essential profession: Improving teacher quality in Louisiana*. Baton Rouge, LA: Author.
- The Council for Exceptional Children's Position on Special Education Teacher Evaluation. (2013). *Teaching Exceptional Children*, 45(3), 73-76.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Philadelphia: Harcourt Brace Jovanovich College Publishers.

- Crowe, E. (2010). Measuring what matters: A stronger accountability model for teacher education. *Center for American Progress*. Retrieved from http://www.americanprogress.org/issues/2010/07/pdf/teacher_accountability.pdf
- Darling-Hammond, L. (2009, March). *Educational opportunity and alternative certification: New evidence and new questions*, Policy Brief No. 1. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation: Popular modes of evaluating teachers are fraught with inaccuracies and inconsistencies, but the field has identified better approaches. *Phi Delta Kappan*, 93(6), 8-15.
- Data Quality Campaign. (2012). *Executive summary: All states could empower stakeholders to make education decisions with data—but they aren't yet*. Washington, DC: Author. Retrieved from http://www.dataqualitycampaign.org/stateanalysis/executive_summary/
- Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, 42(3), 528-554. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=25852724&site=ehost-live>
- Dee, T. S. & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.
- Delcourt, M. A., Cornell, D. G., & Goldberg, M. D. (2007). Cognitive and affective learning outcomes of gifted elementary school students. *Gifted Child Quarterly*, 51(4), 359-381.
- DeVore, D. P. (2011). *Evaluation of Public School Teachers and The Role of Value-Added Measures*. Working Paper. Educational Policy Institute of California. La Verne, CA.

Duncan, A. (2010, November 16). Remarks to national council for accreditation of teacher education [Transcript]. Speech at the National Press Club, Washington, D.C. Retrieved from <http://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-national-council-accreditation-teacher-education>

Duncan, A. (2011). *Preparing and credentialing the nation's teachers: The secretary's eighth report on teacher quality based on data provided for 2008, 2009, 2010*. Washington, DC: US Department of Education.

Eckes, S., & Swando, J. (2009). Special education subgroups under NCLB: Issues to consider. *The Teachers College Record*, 111(11), 2479-2504.

Education Sciences Reform Act of 2002, Pub. L. No. 107-279, 116 Stat. 1940 (2002).

Retrieved from the Government Printing Office website.

Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10, 79 Stat. 27 (1965).

Retrieved from <http://www.nctic1p.org/files/40646763.pdf>.

Ellison, G. & Swanson, A. (2012). *The gender gap in secondary school mathematics at high achievement levels: Evidence from the American mathematics competitions* (No. 15238).

NBER Working Paper. Retrieved from <http://www.nber.org/papers/w15238>

Empowering Parents through Quality Charter Schools Act, H.R. 2218, 112th Cong. (2011).

Retrieved from the Library of Congress U.S. Legislative website (THOMAS).

Encouraging Innovation and Effective Teachers Act, H.R. 3990, 112th Cong. (2012).

Retrieved from the Library of Congress U.S. Legislative website (THOMAS).

Feistritzer, C. E., & Haar, C. K. (2006). *Research on alternate routes education research*.

Washington, DC: National Center for Alternative Certification.

- Flexner, A. (1910). *Medical education in the United States and Canada: A report to the Carnegie Foundation for the advancement of teaching*. Boston: Updyke.
- Frase, L. E. (2005). Refocusing the purposes of teacher supervision. In F. W. English (Ed.), *The Sage Handbook of Educational Leadership: Advanced Theory, Research and Practice*. Thousand Oaks, CA: Sage Publications.
- Fraser, J. W. (2007). *Preparing America's teachers: A history*. New York: Teachers College Press.
- Fry, R. (2007). *How far behind in math and reading are English language learners?* Washington, DC: Pew Hispanic Center.
- Fry, R. (2008). *The role of schools in the English language learner achievement gap*. Washington, DC: Pew Hispanic Center.
- Fryer, R. G., & Levitt, S. D. (2006). The black-white test score gap through third grade. *American Law and Economics Review*, 8(2), 249-281.
- Futrell, M.H. (2010). Transforming teacher education to reform America's P-20 education system. *Journal of Teacher Education*, 61(5) 432-440. doi:10.1177/0022487110375803
- Gansle, K. A., Noell, G. H., Knox, R. M., & Schafer, M. J. (2010). Value added assessment of teacher preparation in Louisiana: 2005-2006 to 2008-2009. Retrieved from <http://regents.louisiana.gov/assets/docs/TeacherPreparation/2010VATechnical082610.pdf>
- Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from <http://www.tqsource.org/publications/LinkBetweenTQandStudentOutcomes.pdf>

- Goldhaber, D. & Liddle, S. (2012). *The gateway to the profession: Assessing teacher preparation programs based on student achievement* (No. 65). CALDER Working Paper. Retrieved from <http://www.caldercenter.org/upload/Goldhaber-et-al.pdf>
- Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review, 34*, 29-44.
- Good, H. G., & Teller, J. D. *A History of American Education* (3rd ed.). New York: The MacMillan Co., 1973.
- Gordon, R., Kane, T. J., and Staiger, D. O. (2006). *The Hamilton project: Identifying effective teachers using performance on the job*. (Policy Brief No. 2006-01) Retrieved from The Brookings Institution website: http://www.brookings.edu/views/papers/200604hamilton_1_pb.pdf
- Goudie, J. W. (1988). *The rise and demise of laboratory schools using Western Michigan University's campus school as a case study*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. 303575268.
- Graham, S. E., & Provost, L. E. (2012). *Mathematics Achievement Gaps between Suburban Students and Their Rural and Urban Peers Increase over Time* (No. 52). Carsey Institute Issue Brief.
- Graue, E. & Johnson, E. (2011). Reclaiming assessment through accountability that is “just right”. *Teachers College Record, 113*(8), 1827-1862.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review, 100*(2), 267-271. doi:10.1257/aer.100.2.267

- Harris, D., & Sass, T. (2007). *Teacher training, teacher quality, and student achievement* (No. 3). CALDER Working Paper. Retrieved from <http://www.caldercenter.org/publications/calder-working-paper-3.cfm>.
- Haycock, K. (1994). Higher education and the schools: A call to action and strategy for change. *Metropolitan Universities: An International Forum*, 5, 12-23.
- Heck, R. H. (2008). Teacher effectiveness and student achievement: Investigating a multilevel cross-classified model. *Journal of Educational Administration*, 47(2), 227-249.
doi:10.1108/09578230910941066
- Heine, H. (2006). *Teacher certification systems*. Retrieved from the Pacific Resources For Education And Learning website:
http://www.prel.org/media/140135/98_certification_systems.pdf
- Hemphill, F.C., and Vanneman, A. (2011). *Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress* (NCES 2011-459). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Henry, G. T., Fortner, C. K., & Bastian, K. C. (2012). The Effects of Experience and Attrition for Novice High-School Science and Mathematics Teachers. *Science*, 335(6072), 1118-1121.
- Henry, G. T., Thompson, C. L., Bastian, K. C., Kershaw, D. C., Purtell, K. M., & Zulli, R. A. (2011). *Does teacher preparation affect student achievement?* Working Paper. Carolina Institute for Public Policy. Chapel Hill, NC.

- Henry, G. T., Thompson, C.L., Fortner, C.K., Zulli, R.A., & Kershaw, D.C. (2010). *The Impact of teacher preparation on student learning in North Carolina public schools*. Working Paper. Carolina Institute for Public Policy. Chapel Hill, NC.
- Higher Education Amendments of 1998, Pub. L. No. 105-244, 112 Stat. 1581 (1998). Retrieved from U.S. Government Printing Office website.
- Higher Education Opportunity Act, Pub. L. No. 110-315, 122 Stat. 3078 (2008). Retrieved from U.S. Government Printing Office website.
- Higher Education. (2013). In *Miriam-Webster Concise Encyclopedia*. Retrieved from [http://www.merriam-webster.com/dictionary/higher education](http://www.merriam-webster.com/dictionary/higher%20education)
- Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831. doi:10.3102/0002831210387916
- Hinsdale, B. A. (1898). *Horace Mann and the common school revival in the United States*. New York: C. Scribner's sons.
- Hoffman, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, 23(6), 723-744. doi: 10.1177/014920639702300602
- Holcomb, T. R., Combs, J. G., Sirmon, D. G., & Sexton, J. 2010. Modeling levels and time in entrepreneurship research. *Organizational Research Methods*, 13(2), 348-389.
- Holdheide, L., Browder, D., Warren, S., Buzick, H., & Jones, N. (2012). Using student growth to evaluate educators of students with disabilities: Issues, challenges, and next steps (Forum summary). Washington, DC: National Comprehensive Center for Teacher Quality.
- Hyde, J. S. & Linn, M. C. (2006). Gender Similarities in Mathematics and Science. *Science* 314, 599-600.

Imig, D., Wiseman, D., & Imig, S. (2011). Teacher education in the United States, 2011.

Journal of Education for Teaching, 37(4), 399-408.

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446, 118

Stat. 2647 (2004). Retrieved from U.S. Government Printing Office website.

Johnson, W. R. (1830). *Remarks on the duty of the several states in regard to the public*

education. Philadelphia: W. Sharpness.

Kafer, K. (2007). *Taking the boy crisis in education seriously: How school choice can boost*

achievement among boys and girls. (No. 604). Independent Women's Forum Position

Paper. Washington, DC:.,

Kane, T. J., & Cantrell, S. (2010). *Learning about teaching: Initial findings from the measures of*

effective teaching project. Seattle, WA: Bill & Melinda Gates Foundation.

Kantor, H. (1991). Education, social reform, and the state: ESEA and federal education policy in

the 1960s. *American Journal of Education*, 100(1), 47-83.

Kerr, C. (2001). *The uses of the university*. Harvard University Press.

Kirschenbaum, H., Simon, S. B., & Napier, R. W. (1971). *Wad-ja-get? The grading game in*

American education. New York: Hart.

Kirst, M. W., & Venezia, A. (2001). Bridging the great divide between secondary schools and

postsecondary education. *Phi Delta Kappan*, 83(1), 92-97.

Klein C.C. (2007). *Efficiency versus effectiveness: interpreting education production studies*.

Working paper. Department of Economics and Finance, Middle Tennessee State

University, Murfreesboro, TN.

- Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2012). *Teacher preparation programs and teacher quality: Are there real differences across programs?* (No. 79). CALDER Working Paper. Retrieved from <http://www.caldercenter.org/publications/upload/WP-79.pdf>
- Kolen, M. J. (2004). Linking assessments: concepts and history. *Applied Psychological Measurement*, 28(4), 219-226.
- Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, 113, F34-F63.
- Kukla-Acevedo, S., Streams, M., & Toma, E. F. (2009). *Evaluation of teacher preparation programs: A reality show in Kentucky* (No. 2009-09). IFIR Working Paper. Lexington, KY: Institute for Federalism and International Relations. Retrieved from http://www.ifigr.org/publication/ifir_working_papers/IFIR-WP-2009-09.pdf
- Kupermintz H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational Evaluation and Policy Analysis*, 25, 287–298.
- Labaree, D. F. (2008). An uneasy relationship: The history of teacher education in the university. In M. Cochran-Smith, S. Feiman-Nemser, & J. D. McIntyre (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed., pp. 290-306). Mahwah, NJ: Lawrence Erlbaum.

- Ladd, H. F. (2007, November). *Holding schools accountable revisited*. Spencer Foundation Lecture in Education Policy and Management, Association for Public Policy Analysis and Management Retrieved August 1, 2012, from https://netforum.avectra.com/Public/DocumentGenerate.aspx?wbn_key=291d86fb-80d4-484d-8988-fb8f6a96e90f.doc&Site=APPAM
- Lassonde, C. (2010). The effectiveness of varied pathways of teacher education in the United States: What research says about alternative and traditional routes and providers [Special Issue]. *Excelsior: Leadership in Teaching and Learning*, Summer 2010, 58-75.
- Lederman, N. G., & Niess, M. L. (2001). An attempt to anchor our moving targets. *School Science and Mathematics*, 101(2), 57-60.
- Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps*. Cambridge, MA: Harvard Civil Rights Project.
- Lee, J. & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation And Policy Analysis*, 34(2), 209-231.
doi:10.3102/0162373711431604
- Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist*, 35(2), 125-141.
- Levine, A. (2006). *Educating school teachers*. New York: The Education Schools Project.
Retrieved from: <http://web.indstate.edu/coe/tec/EducatingTeachersReport.pdf>
- Linn, D., & Hemmer, L. (2012). English Language Learner Disproportionality in Special Education: Implications for the Scholar-Practitioner. *Journal of Educational Research and Practice*, 1(1), 10.

- Livingston, S. A. (2004). *Equating test scores without IRT* (ETS Report). Princeton, NJ: ETS.
- Lomax, E. D. & Kuenzi, J. J. (2012, December). *Value-added modeling for teacher effectiveness* (CRS Report No. R41051). Retrieved from www.fas.org/sgp/crs/misc/R41051.pdf
- Lowry, R. C. (2009). Reauthorization of the federal higher education act and accountability for student learning: The dog that didn't bark. *Publius* (39)3, 506-526.
doi:10.1093/publius/pjp010
- Maeroff, G. I., Callan, P. M., & Usdan, M. D. (2001). A relationship waiting to happen. In G. I. Maeroff, P. M. Callan & M. D. Usdan (Eds.), *The learning connection: New partnerships between schools and colleges* (pp. 1-10). New York: Teachers College Press.
- Marshall, R. A. (1962). *The story of our schools: A short history of public education in the United States*. Washington, DC: National Council for the Social Studies.
- McBee, M. T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *Prufrock Journal*, 17(2), 103-111.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004, Spring). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67-101.
- McCaffrey, D. F., Lockwood, J. R., Korte, D. M., & Hamilton, L. S. (2003). Evaluating value-added models for teacher accountability. Santa Monica, CA: RAND Corporation.
- Mihaly, K., McCaffrey, D., Sass, T. R. & Lockwood, J. R. (2012). *Where you come from or where you go? Distinguishing between school quality and the effectiveness of teacher preparation program graduates* (No. 63). CALDER Working Paper. Retrieved from the National Center for Analysis of Longitudinal Data in Education Research website:
<http://www.caldercenter.org/publications/calder-working-paper-63.cfm>

- Miller, M. H. & Modigliani, F. (1961). Dividend policy, growth, and the valuation of shares. *The Journal of Business*, 34(4), 411-433.
- National Council for Accreditation of Teacher Education. (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers*. Washington, DC: Author. Retrieved from <http://www.ncate.org/LinkClick.aspx?fileticket=zzeiB1OoqPk%3d&tab id=715>
- National Council on Teacher Quality (2007). *State teacher policy yearbook: Progress on teacher quality*. Washington, DC: Author. Retrieved from http://www.nctq.org/stpy/reports/stpy_national.pdf
- National Council on Teacher Quality (2008). *State teacher policy yearbook: What states can do to retain effective new teachers*. Washington, DC: Author. Retrieved from http://www.nctq.org/stpy08/reports/stpy_national.pdf
- National Dissemination Center for Children with Disabilities. (2010). IDEA's definition of "Highly Qualified". Retrieved from <http://nichcy.org/schools-administrators/hqt/idea>
- National Research Council. (2001). *America becoming: Racial trends and their consequences, Volume 1*. N.J.Smelser, W.J.Wilson, and F.Mitchell (Eds.), Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Committee on the Study of Teacher Preparation Programs in the United States, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

- Nichols, S. L., Glass, G. V, Berliner, D. C. (2012). High-stakes testing and student achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives*, 20(20), 1-35.
Retrieved August 1, 2012, from <http://epaa.asu.edu/ojs/article/view/1048>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002). Retrieved from the Government Printing Office website.
- Noell, G. H. (2005). *Technical report of: Assessing teacher preparation program effectiveness: A pilot examination of value added approaches*. Retrieved from www.laregentsarchive.com/Academic/TE/technical_report.pdf
- Noell, G. H. (2006). *Annual report of: Value added assessment of teacher preparation*. Baton Rouge, LA: Louisiana Board of Regents.
- Noell, G. H., Gansle, K. A., Patt, R. M., & Schafer, M. J. (2009). Value added assessment of teacher preparation in Louisiana: 2005-2006 to 2007-2008. Baton Rouge, LA: Louisiana State University Department of Psychology. Retrieved from [http://www.laregentsarchive.com/Academic/TE/2009/2008-09VATechnical\(8.24.09\).pdf](http://www.laregentsarchive.com/Academic/TE/2009/2008-09VATechnical(8.24.09).pdf)
- Noell, G. H., Porter, B. A., & Patt, R. M. (2007). *Value added assessment of teacher preparation in Louisiana: 2004–2006*. Retrieved from <http://regents.louisiana.gov/assets/docs/TeacherPreparation/VAATPPTechnicalReport10-24-2007.pdf>
- Noell, G. H., Porter, B. A., Patt, R. M. & Dahir, A. (2008). *Value-added assessment of teacher preparation in Louisiana: 2004-2005 to 2006-2007*. Retrieved from: [http://www.regents.state.la.us/Academic/TE/2008/Final%20ValueAdded%20Report%20\(12.02.08\).pdf](http://www.regents.state.la.us/Academic/TE/2008/Final%20ValueAdded%20Report%20(12.02.08).pdf)

- Obama, B. (2009). *Remarks by the president to the Hispanic Chamber of Commerce*. Retrieved from http://www.whitehouse.gov/the_press_office/remarks-of-the-President-to-the-United-States-Hispanic-Chamber-of-Commerce
- Ogren, Christine (2005) *The American state normal school: An instrument of great good*. New York: Palgrave Macmillan.
- Papay, J. P., & Kraft, M. A. (2011). *Productivity returns to experience in the teacher labor market: methodological challenges and new evidence on long-term career growth*. Working Paper.
- Phillips, J. A. & Hawthorne, R. (1978). Political dimensions of curriculum decision making. *Educational Leadership*, 2, 362-366.
- Powell, A. (1976). University schools of education in the twentieth century. *Peabody Journal of Education*, 54(1), 3-20.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, 2nd edition. Newbury Park, CA: Sage.
- Raudenbush, S., Bryk, S., Cheong, Y. F., Congdon, R., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.
- Reese, S. (2010). Traditional or alternative: Finding new teachers along different pathways. *Techniques (Association For Career And Technical Education)*, 85(1), 16-21.
- Reese, W. J. (1999). What history teaches about the impact of educational research on practice. *Review of Research in Education*, 24, 1-19.
- Reese, W. J. (2005). *America's public schools from the common school to "No Child Left Behind."* Baltimore: Johns Hopkins University Press.

- Rivkin, S. G. (2007). *Value-added analysis and education policy*. (Brief 1) Retrieved from http://urbaninstitute.org/UploadedPDF/411577_value-added_analysis.pdf
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, 50(1), 4-36.
- Rose, R. A., Henry, G. T. & Lauen, D. L. (2011). *Technical Briefing Report: Comparing Value-Added Models for Estimating Teacher Effectiveness – Executive Summary*. Retrieved from <http://www.dpi.state.nc.us/docs/rttt/reports/2012/vam-summary.pdf>
- Rowan, B. B., Correnti, R. R., & Miller, R. J. (2002). What Large-Scale, Survey Research Tells Us About Teacher Effects on Student Achievement: Insights From the Prospects Study of Elementary Schools. *Teachers College Record*, 104(8), 1525-1567.
- Rowan, B. B., Fang-Shen, C., & Miller, R. J. (1997). Using Research on Employees' Performance to Study the Effects of Teachers on Students' Achievement. *Sociology Of Education*, 70(4), 256-284.
- Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256
- Sanders, W. L., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future student academic achievement: Research progress report*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.

- Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value-added assessment system, a quantitative, outcomes-based approach to educational measurement. In J. Millman (Ed.). *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp.137-162). Thousand Oaks, CA: Corwin Press.
- Sarason, S.B., Davidson, K.S., & Blatt, B. (1986). *The preparation of teachers: An unstudied problem in education*. Cambridge, MA: Brookline Books.
- Schallock, D., Schallock, M., & Ayres, R. (2006). Scaling up research in teacher education: New demands on theory, measurement, and design. *Journal of Teacher Education*, 57 (2), 102-119.
- Schwerdt, G. & Wupperman, A.C. (2009). *Is traditional teaching really all that bad? A within-student between-subject approach* (No. 2634). CESIFO Working Paper. Retrieved from http://www.ifo.de/pls/guestci/download/CESifo%20Working%20Papers%202009/CESifo%20Working%20Papers%20April%202009/cesifo1_wp2634.pdf
- Setting New Priorities in Education Spending Act, H.R. 1891, 112th Cong. (2011). Retrieved from the Library of Congress U.S. Legislative website (THOMAS).
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, 75(3), 417-453.
- Skiba, R. J., Simmons, A. B., Ritter, S., Gibb, A. C., Rausch, M. K., Cuadrado, J., & Chung, C. G. (2008). Achieving equity in special education: History, status, and current challenges. *Exceptional Children*, 74(3), 264-288.
- Smith, B. O. (1980). *A design for a school of pedagogy*. Washington, DC: US Government Printing Office.

- Smith, D. D., Robb, S. M., West, J., & Tyler, N. C. (2010). The changing education landscape: How special education leadership preparation can make a difference for teachers and their students with disabilities. *Teacher Education and Special Education*, 33(1), 25-43. doi:10.1177/0888406409358425
- Snijders, T., & Bosker, R. (1999). *Multilevel Analysis: an introduction to basic and advanced multilevel modeling*. London: Sage Publications.
- Souder, J. A., & Fairfax, S. K. (1996). *State trust lands: History, management, and sustainable use*. Lawrence, KS: University Press of Kansas.
- Spellings, M. (2006). *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: U.S. Department of Education. Retrieved October 25, 2011, from <http://www.ed.gov/about/bdscomm/list/hiedfuture/reports/pre-pub-report.pdf>
- State and Local Funding Flexibility Act, H.R. 2445, 112th Cong. (2011). Retrieved from the Library of Congress U.S. Legislative website (THOMAS).
- Steele, J. L., Hamilton, L. S., & Stecher, B. M. (2010). Incorporating Student Performance Measures into Teacher Evaluation Systems. Technical Report. *RAND Corporation*.
- Student Success Act, H.R. 3989, 112th Cong. (2012). Retrieved from the Library of Congress U.S. Legislative website (THOMAS).
- Superfine, B. M., Gottlieb, J. J., Smylie, M. A. (2012). The expanding federal role in teacher workforce policy. *Education Policy*, 26(1), 58-78. doi:10.1177/0895904811435722
- Teach Plus. (2009). *Ready for the next challenge: Improving the retention and distribution of excellent teachers in urban schools*. Cambridge, MA: Author.

- Texas Education Agency. (2008). TAKS 2008 mean p-values and internal consistency values by objective and subject area. In *Technical Digest 2007-2008*. Retrieved from <http://www.tea.state.tx.us/student.assessment/techdigest/yr0708/>
- Texas Education Agency. (2008b). Chapter 17: Validity. In *Technical Digest 2007-2008*. Retrieved from <http://www.tea.state.tx.us/student.assessment/techdigest/yr0708/>
- Texas Education Agency. (2009a). *2008 TAKS English vertical scaling study report*. Retrieved from: <http://www.tea.state.tx.us/student.assessment/techdigest/#scale>
- Texas Education Agency. (2009b). TAKS 2009 mean p-values and internal consistency values by objective and subject area. In *Technical Digest 2008-2009*. Retrieved from <http://www.tea.state.tx.us/student.assessment/techdigest/yr0809/>
- Texas Education Agency. (2009c). Chapter 3: Standard technical processes. In *Technical Digest 2008-2009*. Retrieved from <http://www.tea.state.tx.us/student.assessment/techdigest/yr0809/>
- Texas Education Agency. (2010). TAKS 2010 mean p-values and internal consistency values by objective and subject area. In *Technical Digest 2009-2010*. Retrieved from <http://www.tea.state.tx.us/student.assessment/techdigest/yr0910/>
- Texas Education Agency. (2011a). Chapter 4: Texas assessment of knowledge and skills (TAKS), TAKS (accommodated), and linguistically accommodated testing (LAT). In *Technical Digest 2010-2011*. Retrieved from <http://www.tea.state.tx.us/student.assessment/techdigest/yr1011.aspx>
- Texas Education Agency. (2011b). Chapter 5: TAKS–modified (TAKS–M). In *Technical Digest 2010-2011*. Retrieved from <http://www.tea.state.tx.us/student.assessment/techdigest/yr1011.aspx>

- Texas Education Agency. (2011c). Chapter 6: TAKS–alternate (TAKS–Alt). In *Technical Digest 2010-2011*. Retrieved from <http://www.tea.state.tx.us/student.assessment/techdigest/yr1011.aspx>
- Texas Education Agency. (2011d). TAKS 2011 mean p-values and internal consistency values by objective and subject area. In *Technical Digest 2010-2011*. Retrieved from <http://www.tea.state.tx.us/student.assessment/techdigest/yr1011.aspx>
- Texas Education Agency. (2011e). Chapter 2: Building a high-quality assessment system. In *Technical Digest 2010-2011*. Retrieved from <http://www.tea.state.tx.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147506068&libID=2147506061>
- The Teaching Commission. (2006). *Teaching at Risk: Progress & Potholes*. New York, NY.
- Thompson, B. (1999). Understanding coefficient alpha, really. Paper presented at the annual meeting of the Education Research Exchange, College Station, Texas, February 5, 1999.
- Thum, Y. M., & Bryk, A. S. (1997). Value-added productivity indicators, the Dallas system. In J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp. 100-109). Thousand Oaks: Corwin Press, Inc.
- Tyack, D., James, T., & Benavot, A. (1987). *Law and the shaping of public education, 1785-1954*. Madison, WI: The University of Wisconsin Press.
- U. S. Commission on Civil Rights. (2009). *Minorities in special education: A briefing before the United States Commission on Civil Rights held in Washington, DC, December 3, 2007*. Washington, DC: U.S. Commission on Civil Rights.

U.S. Department of Education (2009b). *Race to the Top: Executive Summary*. Washington, DC:

US Department of Education. Retrieved from

<http://ed.gov/programs/racetothetop/executive-summary.pdf>

U.S. Department of Education (2010). *A blueprint for reform. The reauthorization of the*

Elementary and Secondary School Act. Washington, DC: US Department of Education.

Retrieved from <http://www2.ed.gov/policy/elsec/leg/blueprint/>.

U.S. Department of Education. (2005, November 18). *Secretary Spellings Announces Growth*

Model Pilot, Addresses Chief State School Officers' Annual Policy Forum in Richmond

[Press release]. Retrieved from

<http://http://www2.ed.gov/news/pressreleases/2005/11/11182005.html>

U.S. Department of Education. (2009a, January 8). *Secretary Spellings Approves Additional*

Growth Model Pilots for 2008-2009 School Year [Press release]. Retrieved from

<http://www2.ed.gov/news/pressreleases/2009/01/01082009a.html>

Urban, W.J. (1996). Historical studies of teacher education. In W.R. Houston (Ed.), *Handbook of*

research on teacher education (pp. 59-71). New York: MacMillan Publishing Company.

Vanneman, A., Hamilton, L., Baldwin Anderson, J., and Rahman, T. (2009). *Achievement Gaps:*

How Black and White Students in Public Schools Perform in Mathematics and Reading

on the National Assessment of Educational Progress, (NCES 2009-455). National Center

for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Washington, DC.

Veysey, L. R. (1965). *The Emergence of the American University*. Chicago: The University of

Chicago Press.

- Wei, X., Lenz, K. B., & Blackorby, J. (2012). Math Growth Trajectories of Students With Disabilities: Disability Category, Gender, Racial, and Socioeconomic Status Differences From Ages 7 to 17. *Remedial and Special Education*.
- Wellman, J. V. (2001, March/April). Assessing state accountability systems. *Change*, 33(2), 47-52.
- Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, 100, 61-78.
- Woltman, H., Feldstain, A., Mackay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1), 52-69.
- Woodring, P. (1975) The development of teacher education. In K. Ryan (Ed.), *Teacher education* (74th yearbook of the National Society for the Study of Education, Part 2, pp. 1-24). Chicago: University of Chicago Press.
- Worthen, B. R., Borg, W. R., & White, K. (1993). *Measurement and assessment in schools*. New York: Longman.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(1), 57-67.
- Xu, Z., Hannaway, J. & Taylor, C. (2011). Making a difference? The effects of teach for America in high school. *Journal of Policy Analysis and Management*, 30(3), 447-469.
doi:10.1002/pam.20585
- Zajda, J. I. (2010). *Globalisation, ideology and education policy reforms*. Dordrecht: Springer.