Implications of Interactions Among Society, Education and Technology: A Comparison of Multiple Linear Regression and Multilevel Modeling in Mathematics Achievement Analyses

Pamela Rose Deering
West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
Deering, Pamela Rose, "Implications of Interactions Among Society, Education and Technology: A Comparison of Multiple Linear Regression and Multilevel Modeling in Mathematics Achievement Analyses" (2014). Graduate Theses, Dissertations, and Problem Reports. 123.
https://researchrepository.wvu.edu/etd/123

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.
Implications of Interactions Among Society, Education and Technology: A Comparison of Multiple Linear Regression and Multilevel Modeling in Mathematics Achievement Analyses

Pamela Rose Deering

Dissertation submitted
to the College of Education and Human Resources
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Education in
Technology Education

Sebastián R. Díaz, Ph.D., J.D., Advisor
R. Neal Shambaugh, Ph.D., Chair
Allison Swan Dagen, Ph.D.
Ugur Kale, Ph.D.
Gayle A. Neldon, Ed.D.

Department of Learning Sciences and Human Development
College of Education and Human Services

Morgantown, West Virginia
2014

Keywords: Multilevel Modeling; MLM; HLM; Multiple Linear Regression; OLS; Mathematics Achievement; Predictive Modeling; Benchmark Tests; Accountability; Appalachia; Teacher Experience; Teacher Certification; Student-Centered Funding

Copyright 2014 Pamela Rose Deering
Abstract

Implications of Interactions Among Society, Education and Technology:
A Comparison of Multiple Regression and Multilevel Modeling in Mathematics Achievement Analyses

Pamela Rose Deering

This research compares and contrasts two approaches to predictive analysis of three years’ of school district data to investigate relationships between student and teacher characteristics and math achievement as measured by the state-mandated Maryland School Assessment mathematics exam. The sample for the study consisted of 3,514 students taught by 99 teachers in a small Appalachian school district in western Maryland. The first analytic approach, standard multiple linear regression, produced a model in which each of the predictors is statistically significant: student gender, prior math achievement, student performance on school district mathematics benchmark exams, teacher years of experience, and advanced teacher certification. In the second approach—multilevel modeling with students as the level-1 unit of analysis and teachers as the level-2 unit of analysis—student characteristics are significant predictors of math achievement, and teacher characteristics are insignificant predictors. The study is set within a context of an exploration of relationships among society, education, and technology. Implications of the study’s results for K-12 mathematics education practice and policy are discussed including: the need to define teacher effectiveness and to identify teacher characteristics that contribute to student achievement; the promise of benchmarking exam systems; the necessity of effective math education, minimally from early education through Algebra II; the need to evaluate teacher certification criteria and the efficacy of teacher preparation programs; the importance of using appropriate statistical modeling approaches in education research; and a call to put students back into the education equation through student-centered funding models.
Acknowledgements

I offer my thanks to personnel from the Allegany County Public School system: to William AuMiller, Superintendent of Allegany County Public Schools (retired), for approving my request to conduct this research project; to Don Knotts, Supervisor of Math (retired), Nil Grove, Chief Technology Officer for discussing the project and its feasibility with me, and to Steve Milburn, PC Analyst, for retrieving the data that was used in the research. Without your help, this project would not have been possible.

Thank you, Ann Crabtree, Program Assistant (retired) for the Technology, Learning & Culture Department, for your service to, encouragement of, and watchful eye over me and other doctoral students at West Virginia University. Your presence in the department was heartening. Thank you, Char Allen, Academic Advisor, for your guidance on necessary graduate student processes and procedures.

Thank you, Gayle Neldon; thank you, Allison Swan Dagen; and thank you, Ugur Kale for your unhesitating willingness to join my committee late in the game. Thank you for the bolstering encouragement of your “yes” as I climbed my last mountain in this journey. I appreciate your gifts to me—you time and your feedback; they have contributed to this work.

Thank you, Neal Shambaugh, chair of my committee, for your time, feedback, and administrative contributions. Moreover, thank you for fostering growth in me throughout my doctoral studies: by encouraging me to see the big picture, through thoughtful feedback on my assignments and projects that often provoked deeper thinking and inspired me to keep going, by modeling excellence in teaching and in the scaffolding of projects for your students, by encouraging creativity, and by taking risks in the classroom. Thank you for flinging open new gates of exploration and qualitative research to me.
Sebastian Diaz, thank you for the many hours you have given to me as my dissertation advisor—introducing me to multilevel modeling; planning, discussing and mapping out the research with me; teaching me about data preparation and the ins and outs of SPSS; reviewing my work; and providing input and feedback throughout this long process. Thank you for seeing me as a colleague—listening to and valuing my opinions, collaborating in brain-storming sessions, taking time to discuss topics beyond this study. Thank you for challenging me, in your classroom and through this study. In both contexts, you compelled me to think beyond the immediate issues at hand; you broadened the scope of my thinking. Thank you for caring: your question of “How’s life?”, which you pose before “How’s your dissertation coming?” or “How’s work going?” speaks loudly to me of your regard for and treatment of others as human beings. I am ever so grateful for what you have taught me about life, work, statistics, and research.

Dear One, Gerry Wojnar, thank You for standing by me; thank You for your endless patience as I rambled on about dissertation ideas or issues until my path became clear; thank You for your concern and empathy during the times I felt over-burdened by life/work/dissertation; and thank You for seeing me through to the end of this arduous journey. Many, many thanks to You, Dear One.
Dedication

To my entire family, with special dedication to my husband, Gerry; and to my children: Larry, Belinda, Daniel, Rachel, and Dunel. I am blessed and inspired by the diversity of your individualities, your beauties, your perspectives, your many talents; by your courage in the face of adversities, by our collective wisdom and our individual and collective strength; by your exuberance for life, your passion, your humor, your humanity, your growth, your desire to make a difference; by your complexity, your passions, and your friendship. I love you.
# Table of Contents

Acknowledgements ........................................................................................................................ iii

Dedication ....................................................................................................................................... v

List of Tables .................................................................................................................................. xi

List of Figures .............................................................................................................................. xiii

Chapter 1 – Introduction ................................................................................................................. 1

  Statement of the Problem ........................................................................................................ 2

  Research Questions ................................................................................................................ 3

  Significance of the Study ......................................................................................................... 4

  Definition of Terms and Acronyms .................................................................................... 6

  Variable Notation .................................................................................................................. 9

Chapter 2 – Literature Review ...................................................................................................... 10

  Background: Relationships Among Society, Education and Technology ....................... 10

    Allocation of resources to education .................................................................................. 10

    Teacher and Principal Quality .......................................................................................... 12

    Assessment ......................................................................................................................... 12

    Statistics .............................................................................................................................. 17

    Research .............................................................................................................................. 17

    Legislation ........................................................................................................................... 19

    Education accountability .................................................................................................... 21
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology connections</td>
<td>26</td>
</tr>
<tr>
<td>Education data systems</td>
<td>27</td>
</tr>
<tr>
<td>Data-informed decision making</td>
<td>32</td>
</tr>
<tr>
<td>Accountability systems</td>
<td>35</td>
</tr>
<tr>
<td>Research Context: Society, Mathematics Education, and Technology</td>
<td>38</td>
</tr>
<tr>
<td>Mathematics for the well-being of U.S. society</td>
<td>38</td>
</tr>
<tr>
<td>Mathematics achievement: An indication of success within U.S. society</td>
<td>41</td>
</tr>
<tr>
<td>U.S. students’ mathematics achievement</td>
<td>47</td>
</tr>
<tr>
<td>Trends in International Mathematics and Science Study (TIMSS)</td>
<td>48</td>
</tr>
<tr>
<td>The Programme for International Student Assessment (PISA)</td>
<td>50</td>
</tr>
<tr>
<td>A National Education Goal: Improved mathematics achievement</td>
<td>52</td>
</tr>
<tr>
<td>State of Maryland context</td>
<td>54</td>
</tr>
<tr>
<td>Factors That May Affect Mathematics Achievement</td>
<td>57</td>
</tr>
<tr>
<td>Student characteristics</td>
<td>57</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>57</td>
</tr>
<tr>
<td>Gender differences</td>
<td>59</td>
</tr>
<tr>
<td>Teacher characteristics</td>
<td>62</td>
</tr>
<tr>
<td>Years of experience</td>
<td>64</td>
</tr>
<tr>
<td>Professional certification</td>
<td>66</td>
</tr>
<tr>
<td>Benchmarking systems</td>
<td>70</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>--------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Statistical Models</td>
<td>73</td>
</tr>
<tr>
<td>Multiple linear regression (MLR)</td>
<td>73</td>
</tr>
<tr>
<td>Multilevel modeling (MLM)</td>
<td>76</td>
</tr>
<tr>
<td>Theoretical matters</td>
<td>78</td>
</tr>
<tr>
<td>Comparative studies</td>
<td>81</td>
</tr>
<tr>
<td>Chapter 3 – Methodology</td>
<td>86</td>
</tr>
<tr>
<td>Research Questions</td>
<td>86</td>
</tr>
<tr>
<td>Data Acquisition and Preparation</td>
<td>87</td>
</tr>
<tr>
<td>Population and Sample</td>
<td>90</td>
</tr>
<tr>
<td>Variables</td>
<td>91</td>
</tr>
<tr>
<td>Criterion variable</td>
<td>92</td>
</tr>
<tr>
<td>Predictor variables</td>
<td>93</td>
</tr>
<tr>
<td>Models</td>
<td>94</td>
</tr>
<tr>
<td>Multiple linear regression model (MLR)</td>
<td>94</td>
</tr>
<tr>
<td>MLR equation</td>
<td>94</td>
</tr>
<tr>
<td>MLR Null and Alternate Hypotheses</td>
<td>95</td>
</tr>
<tr>
<td>Multilevel Models (MLMs)</td>
<td>96</td>
</tr>
<tr>
<td>MLM Null and Alternate Hypotheses</td>
<td>97</td>
</tr>
<tr>
<td>Model 0: Null Model</td>
<td>97</td>
</tr>
<tr>
<td>Model 1: Student-Level (L1) random intercept model</td>
<td>98</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td>Research Questions</td>
</tr>
<tr>
<td></td>
<td>Multiple Linear Regression Model</td>
</tr>
<tr>
<td></td>
<td>Null and alternate hypotheses</td>
</tr>
<tr>
<td></td>
<td>Model fit and parameters</td>
</tr>
<tr>
<td></td>
<td>Evaluation of model assumptions</td>
</tr>
<tr>
<td></td>
<td>MLR summary and model equation</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Discussion</td>
</tr>
<tr>
<td></td>
<td>Research Question 1</td>
</tr>
<tr>
<td>Model 2</td>
<td>Teacher-level (L2) random intercept model</td>
</tr>
<tr>
<td>Model 3</td>
<td>Random Slope and Intercept Model</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td>102</td>
</tr>
<tr>
<td>Model 0</td>
<td>Null model</td>
</tr>
<tr>
<td>Model 1</td>
<td>Student-Level (L1) Random Intercept Model</td>
</tr>
<tr>
<td>Model 2</td>
<td>Teacher-level (L2) random intercept model</td>
</tr>
<tr>
<td>Model 3</td>
<td>Random slope and intercept model</td>
</tr>
<tr>
<td>MLM summary</td>
<td>133</td>
</tr>
</tbody>
</table>

*Running head: COMPARISON OF MLR & MLM MATH ACHIEVEMENT ANALYSES*
Research Question 2 ................................................................................................... 136

MLM Model 1: Student-level random intercept model.............................................. 137
MLM Model 2: Teacher-level random intercept model............................................. 138
MLM Model 3: Random slope and intercept model.................................................. 139

Research Question 3 .................................................................................................. 140
Importance of the study’s larger context................................................................. 141
Key differences and similarities in modeling results................................................. 144
Practice and policy differences relating to teacher experience.............................. 146
Practice and policy differences relating to teacher certification............................ 148
Gender gap in mathematics performance .............................................................. 149
Early mathematics education.................................................................................. 150
The promise of benchmarking systems..................................................................... 152
Teacher effectiveness............................................................................................... 155
Teacher preparation................................................................................................. 157
Student-centered funding: put the student back in the education equation ....... 159

Summary ..................................................................................................................... 163
Suggestions for Further Research............................................................................. 165
List of Tables

Table 1: Selected U.S. Academic Improvement Education Programs, 2008-2010 ....................... 13
Table 2: U.S. Teacher and Principal Quality Education Programs, 2008-2010 ............................. 15
Table 3: Comparison of U.S. TIMSS Results with Other Jurisdictions, 1995-2007 ....................... 49
Table 4: Comparison of U.S. PISA Results with Other Jurisdictions, 2000-2009 .......................... 51
Table 5: Variables and Their Statistical Functions ........................................................................ 92
Table 6: MLR Model Descriptive Statistics ............................................................................... 105
Table 7: Analysis of Variance ..................................................................................................... 106
Table 8: MLR Model Summary .................................................................................................. 106
Table 9: Pearson Correlation Matrix ........................................................................................... 107
Table 10: MLR Coefficients and Their Confidence Intervals, Correlations and Colinearity Statistics .......................................................................................................................... 111
Table 11: Variance of Predictor Variables .................................................................................. 112
Table 12: Wald Z Statistic and 95% Confidence Intervals for Variance in the MLM Models .. 120
Table 13: MLM Estimates: Intraclass Correlation, Reliability, and Reduction in Variance..... 121
Table 14: MLMs’ Fixed Effects Estimates and Variance-Covariance Estimates for Math Achievement Predictors ........................................................................................................... 122
Table 15: Null Model Estimates of Fixed Effects ........................................................................ 124
Table 16: MLM Null Model Covariance Parameters Estimates ................................................. 124
Table 17: Level 1 Random Intercept Model Estimates of Fixed Effects ...................................... 126
Table 18: Level 1 Random Intercept Model Estimates of Covariance Parameters .................... 126
Table 19: Level 2 Random Intercept Model Estimates of Fixed Effects ...................................... 128
Table 20: Level 2 Random Intercept Model Estimates of Covariance Parameters .................... 129
List of Figures

Figure 1. Predicted percentage increase in earnings resulting from an additional math course in high school. ....................................................................................................... 47

Figure 2. Education units of analysis continuum........................................................................................................... 81

Figure 3. Scatterplot of standardized predicted value against standardized residual. .............. 112

Figure 4. Partial regression plot: teacher years of experience. ................................................... 113

Figure 5. Partial regression plot: prior math MSA z-score ................................................................. 113

Figure 6. Partial regression plot: mean of benchmark exam z-scores. ........................................ 114

Figure 7. MLR standardized residual histogram........................................................................ 116

Figure 8. Normal probability plot of MLR standardized residual. .............................................. 117
Chapter 1 – Introduction

A society’s values and perceived needs drive practices, decisions and policies of that society. As “all technologies embody the explicit and implicit values of their creators” (Cheek, 2000), technologies are created and implemented, based on a society’s values, by that society to address its problems and to meet its needs. As society realizes its increased need for a workforce trained in mathematics, engineering, science, and technology, and as emerging technologies become available to educators, how does society use those technologies to help meet the society’s perceived education needs?

This study explores connections among society, education and technology. Whereas many studies that focus on technology and education investigate outcomes related to the educational use of a particular technology, this study concentrates on how emerging technology used by educational leaders and researchers can ultimately impact K-12 practice and policy. With the United States’ current emphasis on mathematics performance and on accountability in K-12 education, this study takes a practical and pertinent approach to investigating the impact of technology on practice and policy: the study compares and contrasts two approaches to the analysis of mathematics achievement data.

The first approach is multiple linear regression (MLR), a traditional statistical methodology that attempts to determine whether a significant predictive model exists for the given data set. In the context of this study, MLR is used to explain how student characteristics and teacher characteristics affect student performance on a state-mandated mathematics achievement exam. The second approach is multilevel modeling (MLM), a more recent and more sophisticated approach that accounts for naturally occurring hierarchical data structures. In this study, a two-level multilevel model simultaneously analyzes student characteristics in level 1
(L1) and teacher characteristics in level 2 (L2) to investigate the influence of these parameters on students’ mathematics achievement.

Chapter 1 introduces the study. The chapter contains a statement of the problem and the list of research questions, explains the significance of the study, and defines the terms and abbreviations used in the document. In Chapter 2, the literature review, some relationships among society, education and technology are established. Then, honing in on the context of this research, relationships among society, mathematics education and technology are explored. Chapter 3 details the methodology of the study and includes descriptions of data acquisition and preparation, the population and sample for the study, the variables included in the study, and the statistical models used in the study. Chapter 4 presents results of the two statistical modeling approaches used in the study. Chapter 5 compares and contrasts the results of the two predictive modeling approaches and discusses implications for K-12 policy and practice.

**Statement of the Problem**

Complex relationships exist among society, education and technology. Evidence of those relationships is present in the United States’ concern for student achievement trends and the impact of those trends on the nation’s ability to compete in a global economy in the 21st century. Such concerns

- drive research agendas and pursuit of methods to determine what works to improve student achievement;
- drive research-based, data-informed education reform and policy revision to improve student achievement;
- emphasize a need for continued development and implementation of information systems to gather, store, and track data related to student achievement to shape
evolving educational practices, to provide accurate data for deeper research, and to satisfy political policy makers; and

- require accountability systems to promote local educational agency (LEA) effectiveness, teacher effectiveness and student achievement.

As data become more readily available and demands for education accountability increase, the sophistication of quantitative evaluative techniques used to inform these accountability efforts increases. As emerging technologies and more sophisticated analytic methods become available to educators and researchers, what impact does the increased sophistication of such technologies and methodologies have on education? In particular, how do differing statistical approaches to the analyses of data that are available from contemporary information systems impact K-12 education?

**Research Questions**

This study investigates how the analyses of data made available through contemporary information systems impact mathematics education. In particular, the study identifies similarities and differences of a multiple linear regression approach and a multilevel modeling approach to predicting mathematics achievement of Appalachian fourth- through eighth-grade students as measured by the Maryland School Assessment (MSA) mathematics exam. The investigation considers the effects of student and teacher characteristics on math achievement, including the effect of the benchmarking system the school district implemented to improve student achievement. The following research questions (RQs) focused the investigation:

RQ1: How does multiple linear regression (MLR) inform the prediction of mathematics achievement from a linear combination of student characteristics—gender, previous mathematics
achievement, district benchmark exam performance—and teacher characteristics—professional
certification level and years of experience?

RQ2: How does multilevel modeling (MLM), with students and teachers as the respective
level-1 and level-2 sampling units, inform the prediction of mathematics achievement from
linear combinations of the same student characteristics—gender, previous mathematics
achievement, district benchmark exam performance—and the same teacher characteristics—
professional certification level and years of experience?

RQ3: How do MLR and MLM results compare for informing K-12 practice and policy
with respect to student mathematics achievement?

Significance of the Study

The American Recovery and Reinvestment Act of 2009 identified four foundational
education reform areas : (U.S. Department of Education, 2009b):

- adopting standards and assessments that prepare students to succeed in college and
  the workplace and to compete in the global economy;
- building data systems that measure student growth and success, and inform teachers
  and principals about how they can improve instruction;
- recruiting, developing, rewarding, and retaining effective teachers and principals,
  especially where they are needed most; and
- turning around the lowest-achieving schools.

Mathematics achievement in middle and high school correlates with student success in
college and the workplace; and workers who are knowledgeable and skillful in applying
mathematics in a wide variety of fields are needed for the U.S. to compete in the global
economy. The building of data systems to store and make data available to educators is but one
term of the equation for improving mathematics education. To inform educators about how they
can improve instruction and to enable them to recruit, develop, and fairly reward effective
mathematics teachers, the first question to ask is, “What contributes to students’ mathematics
achievement?” This is a difficult question to answer—student, teacher, classroom, and school
attributes, as well as many other variables, effect mathematics achievement.

Consequent to recent education reform and legislation, education accountability systems
that inform high-stakes decision and policy making are being established across the United
States. The decisions and policies may have direct, significant consequences for teachers.
Therefore, it is vitally important that educators, leaders and researchers take great care in
answering questions such as, “What factors contribute to students’ mathematics achievement,
how do these factors affect students’ mathematics achievement, and to what extent do these
factors influence students’ mathematics achievement?” From a design perspective, comparing
the two analyses of mathematics achievement results via (a) a traditional MLR methodology and
(b) the more recently developed MLM methodology (specifically designed for analyses of nested
data) may improve researchers’ ability to more effectively identify and assess factors that
significantly contribute to students’ mathematics achievement. Therefore, this study holds
promise in contributing to more effective K-12 decision and policy making.
### Definition of Terms and Acronyms

The following terms/abbreviations and their respective definitions are used throughout this document:

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AASCU</td>
<td>American Association of State Colleges and Universities</td>
</tr>
<tr>
<td>Accountability</td>
<td>The state of districts, schools, educators, and/or students being held responsible by policies or processes for performance or outcomes (American Association of School Administrators, 2002; National School Boards Foundation, 2001)</td>
</tr>
<tr>
<td>Accountability System</td>
<td>A “system that assigns incentives to progress in meeting objectives” (Baker, 2005)</td>
</tr>
<tr>
<td>API</td>
<td>Academic Performance Index. A single number that indicates a local educational agency’s, a school’s or a subgroup’s performance level (based on the results of statewide testing) which is used to measure academic performance and growth (Association of California School Administrators Accountability and Assessment Task Force, 2010)</td>
</tr>
<tr>
<td>AYP</td>
<td>Adequate Yearly Progress. State-defined yearly measurement of school achievement required by the <em>No Child Left Behind Act of 2001</em> “Assessments which are administered intermittently throughout the school year and at specified times during a curriculum sequence to evaluate students’ knowledge/skills relative to a particular set of longer-term learning goals” (Paige, 2002)</td>
</tr>
<tr>
<td>CVR</td>
<td>Covariance ratio</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Data Warehouses</td>
<td>“Electronic data collection and storage systems that provide access to current and historical data” (Means, Padilla, &amp; Gallagher, 2010)</td>
</tr>
<tr>
<td>EDS</td>
<td>Education data system</td>
</tr>
<tr>
<td>ERIC</td>
<td>Education Resources Information Center</td>
</tr>
<tr>
<td>GNP</td>
<td>Gross National Product</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>“The opposite of homoscedasticity. This occurs when the residuals at each level of the predictor variable(s) have unequal variances” (Field, 2005, p. 732)</td>
</tr>
<tr>
<td>Homoscedasticity</td>
<td>“An assumption in regression analysis that the residuals at each level of the predictor variable(s) have similar variances” (Field, 2005, p. 733)</td>
</tr>
<tr>
<td>ID</td>
<td>Identification code</td>
</tr>
<tr>
<td>IEA</td>
<td>International Association for the Evaluation of Educational Achievement</td>
</tr>
<tr>
<td>L1</td>
<td>Level 1 in a multilevel model</td>
</tr>
<tr>
<td>L2</td>
<td>Level 2 in a multilevel model</td>
</tr>
<tr>
<td>LEA</td>
<td>Local educational agency</td>
</tr>
<tr>
<td>MLM</td>
<td>Multilevel model/ing</td>
</tr>
<tr>
<td>MLR</td>
<td>Multiple linear regression</td>
</tr>
<tr>
<td>MRC</td>
<td>Multiple regression/correlation</td>
</tr>
<tr>
<td>MSA</td>
<td>Maryland School Assessment. Maryland’s reading, mathematics and science achievement test that meets testing requirements of the federal No Child Left Behind Act of 2001</td>
</tr>
</tbody>
</table>
Multicollinearity

“A situation in which two or more variables are very closely linearly related” (Field, 2005, p. 738)

OLS
Ordinary least squares

PASW
Predictive Analytics SoftWare, a version of SPSS Statistics

RQ
Research question

SIS
Student information system. Education data system that provides real-time access to student data; e.g., demographics, schedules, attendance, test scores, and grades (Means et al., 2010)

STEM education
Science, technology, engineering and mathematics education

VIF
Variance inflation factor

WLS
Weighted least squares
Variable Notation

Variables used in this study are denoted in equations and tables throughout the document by the following notation:

- $BM_{zMean}$: The mean of a student’s math benchmark exams $z$-scores for a given year
- $BM_{zMean\_mean}$: The aggregate mean, by teacher, of the mean of his/her students’ benchmark exam $z$-scores
- $maxtCert$: The maximum level of certification acquired by a teacher during 2007-2009
- $mMSA$: Mathematics MSA score
- $PmMSAz$: A student’s prior math MSA $z$-score
- $PmMSAz\_mean$: The aggregate mean, by teacher, of his/her students’ prior math MSA $z$-scores
- $stuGend$: Student gender
- $tCert$: A teacher’s level of professional certification
- $tExp$: A teacher’s years of teaching experience
- $tExpMean$: The mean of a teacher’s years of experience over the 2007-2009 time frame
Chapter 2 – Literature Review

The literature review explores and identifies relationships among society, education and technology and describes the context for the study. The review begins with a background for the study which explores relationships between society, education and technology by looking at selected United States education programs and at legislation that authorized and funded educational research and development. Continuing, accountability, data systems, data-informed decision making and accountability systems are considered. Then the context of the study is discussed, including the importance of mathematics achievement in society, factors included in the study that may impact mathematics achievement, and the statistical models used in the study.

Background: Relationships Among Society, Education and Technology

A society’s values and perceived needs drive the decisions, practices, and policies of that society. Allocation of resources to education and the types of programs that are funded show U.S. society’s educational needs. The types of programs for which funds are appropriated to education indicate national education needs and/or goals. Legislation enacted by the government for education research and development further illustrates the importance of education in U.S. society. Finally, elements of education accountability in the United States demonstrate U.S. society’s concern for the quality of its educational systems.

Allocation of resources to education. The United States’ historical increase in allocation of financial resources to education is an indicator of the importance the nation has placed on education. From 1890 to 1990, public expenditures on primary and secondary in the United States grew almost one hundredfold from $2 billion to $187 billion, outpacing the triple growth rate of GNP during that period (Hanushek, 1989). In 1890, educational expenditures were less than 1% of GNP; in 1990 they were 3.4% of GNP (Hanushek, 1989). On a per student basis,
expenditures quintupled in each fifty-year period between 1890 and 1980. Per student expenditures equaled $164 in 1980, $772 in 1940 and $4,622 in 1990 (Hanushek, 1989). Recently, the U.S. Department of Education’s discretionary budget was the third largest of Cabinet agencies; only the Department of Defense and the Department of Health and Humans Services had larger discretionary budgets (U.S. Department of Education, n.d.b).

The types of programs for which funds are appropriated to education indicate national education needs and/or goals. Among the U.S. education programs authorized and funded by federal law are programs that emphasize or promote

- academic and/or school improvement;
- measurement of academic achievement through development of academic standards and assessment instruments, long-term trend analyses, and longitudinal data systems;
- collection, analysis and reporting of data related to academic/school improvement,
- teacher and/or principal quality; and
- rigorous education research upon which practice and policy can be established.

For example, the Guide to U.S. Department of Education Programs: Fiscal Year 2010 gives an overview of recent education programs authorized and funded by federal law. Academic achievement is a key goal of many of these programs funded by the federal government. During fiscal years 2008 - 2010, more than $57 billion was allocated to programs aimed at academic improvement. Table 1 highlights some of the U.S. Academic Improvement Education Programs, many of which aim for improvement of student achievement and/or attainment of state performance standards or school goals that demonstrate adequate yearly progress (AYP) required by the No Child Left Behind Act of 2001 (NCLBA). In addition, the Academic Improvement Programs encourage identification of models that are known to improve schools and that are
based on scientific research. The federal government authorized 25 School Improvement Programs from 2008 through 2010 and funded those programs with nearly $4 billion dollars in appropriations (Education Sciences Reform Act of 2002).

**Teacher and Principal Quality.** The number of and types of federally funded programs intended to promote teacher and principal quality are further indication of society’s concern for education. In fiscal years 2008-2012, appropriations for Teacher and Principal Quality programs totaled more than $16.2 billion (U. S. Department of Education, Office of Communications and Outreach, 2010). One of the overall goals of Teacher and Principal Quality programs is to improve student achievement; the improvement was to come in part through implementing teacher incentives, holding local education agencies and schools accountable for student achievement, developing performance-based compensation systems, and making teacher-preparation programs accountable for preparing effective teachers. Table 2 provides further details of some of the federal appropriations for Teacher and Principal Quality programs from 2008-2012.

**Assessment.** Because academic achievement must be measured, assessment is another category of Education Programs authorized and funded by the federal government. More than $1.75 billion was appropriated to four major Assessment Education Programs from 2008-2010 (U. S. Department of Education, Office of Communications and Outreach, 2010, p. 20-24):

1. The Grants for Enhanced Assessment Instruments program had an allocation of $30,196,480 to improve the quality, validity and reliability of state academic assessments; to measure student achievement using multiple measures; to chart student progress over time; and to evaluate achievement through the development of
Table 1

**Selected U.S. Academic Improvement Education Programs, 2008-2012**

<table>
<thead>
<tr>
<th>Program</th>
<th>Goal</th>
<th>Year</th>
<th>Appropriations</th>
</tr>
</thead>
<tbody>
<tr>
<td>21st-Century Community Learning Centers</td>
<td>Students attainment of performance standards in core subjects, such as reading and math</td>
<td>2008</td>
<td>$1,081,166,187</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$1,131,166,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$1,166,166,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$1,153,853,668</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$1,151,673,216</td>
</tr>
<tr>
<td>Comprehensive School Reform Program</td>
<td>Student and school attainment of state standards through development and implementation of comprehensive school reforms grounded in scientifically based research and effective practices</td>
<td>2008</td>
<td>$1,605,454</td>
</tr>
<tr>
<td>High School Graduation Initiative</td>
<td>To improve retention of high school students through dropout prevention and reentry programs</td>
<td>2010</td>
<td>$50,000,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$48,902,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$48,809,575</td>
</tr>
<tr>
<td>Improving Literacy Through School Libraries</td>
<td>To improve reading achievement in schools where at least 20% of students are living in poverty</td>
<td>2008</td>
<td>$19,144,597</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$19,145,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$19,145,000</td>
</tr>
<tr>
<td>Investing in Innovation</td>
<td>Development, identification and use of replicable models for driving dramatic gains in student achievement</td>
<td>2009</td>
<td>$19,145,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$646,071,398</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$149,700,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$149,417,067</td>
</tr>
<tr>
<td>Jacob K. Javits Gifted and Talented Student Education</td>
<td>To build and enhance the ability of elementary and secondary schools to meet the special education needs of gifted and talented students</td>
<td>2008</td>
<td>$7,463,298</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$7,463,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$7,463,000</td>
</tr>
</tbody>
</table>
Table 1, (continued)

*Selected Recent U.S. Academic Improvement Education Programs, 2008 - 2012*

<table>
<thead>
<tr>
<th>Program</th>
<th>Goal</th>
<th>Year</th>
<th>Appropriations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race to the Top - District</td>
<td>To support bold, locally directed learning/teaching improvements that will improve educator effectiveness and student achievement</td>
<td>2012</td>
<td>$383,000,000</td>
</tr>
<tr>
<td>Race to the Top Incentive Grants</td>
<td>To close the achievement gap and drive substantial gains in student achievement, graduation rates, and college success through systemic reform and continuous improvement</td>
<td>2009</td>
<td>$4,350,000,000</td>
</tr>
<tr>
<td>Race to the Top – Early Learning Challenge</td>
<td>To improve the quality of early learning and close the achievement gap for children with high needs; based on successful state systems; high-quality, accountable programs; promoting early learning and development outcomes; a great early childhood education workforce; measuring outcomes and progress</td>
<td>2011</td>
<td>$497,294,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$132,934,875</td>
</tr>
<tr>
<td>Rural and Low-Income School Program</td>
<td>To help rural schools meet their states’ definition of Adequate Yearly Progress</td>
<td>2008</td>
<td>$85,927,161</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$86,691,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$87,441,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$87,266,118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$89,596,342</td>
</tr>
<tr>
<td>Small Rural School Achievement</td>
<td>To help rural school districts meet their states’ definition of Adequate Yearly Progress</td>
<td>2008</td>
<td>$85,927,161</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$86,691,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$87,441,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$87,266,118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$89,596,342</td>
</tr>
<tr>
<td>Smaller Learning Communities</td>
<td>To prepare all students to succeed in postsecondary education and careers without need for remediation</td>
<td>2008</td>
<td>$80,107,636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$80,107,636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$80,107,636</td>
</tr>
</tbody>
</table>

Table 2

**U.S. Teacher and Principal Quality Education Programs, 2008-2012**

<table>
<thead>
<tr>
<th>Program</th>
<th>Goal</th>
<th>Year</th>
<th>Appropriations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Certification or Advanced</td>
<td>To encourage and support teachers seeking advanced certification or advanced credentialing through high-quality professional teacher enhancement programs designed to improve teaching and learning; to support the development of a school leadership credential</td>
<td>2008</td>
<td>$9,649,427</td>
</tr>
<tr>
<td>Credentialing</td>
<td></td>
<td>2009</td>
<td>$10,649,427</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$10,649,000</td>
</tr>
<tr>
<td>Full-Service Community Schools</td>
<td>Through collaboration between public elementary and secondary schools or between community-based organizations and public-private ventures to improve the quality of elementary and secondary education at the state and local levels and help all children meet challenging academic content and achievement standards</td>
<td>2008</td>
<td>$4,912,650</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$5,000,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$10,000,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$10,000,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$10,093,886</td>
</tr>
<tr>
<td>Improving Teacher Quality State Grants</td>
<td>To increase academic achievement by improving teacher and principal quality by increasing the number of highly qualified teachers, the number of highly qualified principals and assistant principals, and the effectiveness of teachers and principals by holding local education agencies and schools accountable for improvements in student academic achievement</td>
<td>2008</td>
<td>$2,935,248,441</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$2,947,749,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$2,947,749,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$2,464,876,077</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$2,466,567,300</td>
</tr>
<tr>
<td>Mathematics and Science Partnerships</td>
<td>To improve the content knowledge of teachers and the performance of students in the areas of mathematics and science</td>
<td>2008</td>
<td>$178,977,665</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$178,978,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$180,478,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$175,127,044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$149,716,500</td>
</tr>
</tbody>
</table>
## U.S. Teacher and Principal Quality Education Programs, 2008-2012

<table>
<thead>
<tr>
<th>Program</th>
<th>Goal</th>
<th>Year</th>
<th>Appropriations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Incentive Fund</td>
<td>To develop and implement performance-based compensation systems for teachers, principals, and other personnel in high-need schools; to improve student achievement by increasing teacher and principal effectiveness; to reform compensation systems so that teachers, principals, and other personnel in high-need schools are rewarded for increases in student achievement; to increase the number of effective teachers teaching poor, minority, and disadvantaged students in hard-to-staff subjects; and to create sustainable performance-based compensation systems</td>
<td>2008</td>
<td>$97,270,470</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$97,270,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$400,000,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$399,200,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$299,433,000</td>
</tr>
<tr>
<td>Teacher Quality Partnership Grants</td>
<td>To improve student achievement and the quality of teachers working in high-need schools and Early Childhood programs by improving the preparation of teachers and enhancing professional development activities for teachers; holding teacher preparation programs accountable for preparing effective teachers; and recruiting highly qualified individuals, including minorities and individuals from other occupations, into the teaching force; to help create effective pathways into teaching and support the nation’s teaching force in effectively improving student outcomes</td>
<td>2008</td>
<td>$33,662,460</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009</td>
<td>$50,000,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010</td>
<td>$43,000,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>$42,914,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>$42,832,893</td>
</tr>
</tbody>
</table>

comprehensive assessment instruments, such as performance- and technology-based academic assessments (p. 20).

2. The National Assessment of Educational Progress program had $358,363,340 in appropriations to support national, state, trial urban district, and long-term trend assessments (p. 21).

3. The National Assessment of Educational Progress program had $358,363,340 in appropriations to support national, state, trial urban district, and long-term trend assessments (p. 22-23).

4. The Statewide Longitudinal Data Systems program had allocations of $171,543,000 to design, develop and implement statewide longitudinal data systems to efficiently and accurately manage, analyze, disaggregate, and use individual student data consistent with the *Elementary and Secondary Education Act of 1965* (pp. 23-24).

*Statistics.* From 2008 to 2010, the federal government increased National Center for Education Statistics (NCES) support by about $10 million a year for a three-year total of nearly $300 million. The NCES works with public, private, for-profit and nonprofit organizations, agencies, institutions and consortia to collect, analyze, and report the significance of statistics as well as to assist in improving statistical systems of local educational agencies and postsecondary institutions (U. S. Department of Education, Office of Communications and Outreach, 2010, pp. 23-24).

*Research.* More than $990 million of federal funds were appropriated to six Department of Education research programs from 2008 through 2010 (U. S. Department of Education, Office of Communications and Outreach, 2010, pp. 23-24):
1. The Education Research program had $527,088,000 in appropriations for research in education research training; statistical and research methodology in education; cognition and student learning; teacher quality in reading, writing, mathematics and science education; education leadership; education policy; educational technology; and more (p. 189).

2. The Education Resources Information Center was allocated $25,967,388 for the Education Resources Information Center (ERIC) information system (p. 189).

3. The Jacob K. Javits Gifted and Talented Students Education Program National Research and Development Center had $5,223,000 in appropriations for research on methods and techniques for identifying and teaching gifted and talented students (p.189).

4. The Regional Education Laboratories were allocated $203,788,000 to support laboratories that conduct applied research, develop multimedia educational materials, provide technical support, and disseminate information to help others improve education (p. 194).

5. The Research in Special Education program had $212,255,000 in appropriations to research early intervention and early childhood special education; mathematics and science education; behavioral outcomes to support learning; reading, writing and language development; teacher quality in special education; autism spectrum disorders; transition outcomes for special education; and more (pp. 194-195).

6. The Small Business Innovation Research program was allotted $19,066,614 for research and development projects that propose a sound approach to investigating
important assistive technology, education, and science and/or engineering questions (pp. 194-195).

Education research is not only an important part of the education programs instituted by the government; legislation has supported education research and development to contribute to quality education in the United States.


1. In 1972, Congress established the National Institute of Education (NIE) to provide leadership in conducting and supporting scientific inquiry into the educational process (U.S. National Archives and Records Administration, n.d.).


3. In 1978, Congress authorized the National Assessment of Education Progress (NAEP) to provide a fair and accurate assessment of academic achievement.

4. In 1979, Congress established the Department of Education to:
   - strengthen commitment to equal educational opportunity for every individual;
   - improve the quality of education by supplementing and complementing educational efforts of states, local school systems, the private sector, public and private educational institutions, community-based organizations, parents and students;
increase parents’, students’, and the public’s involvement in Federal education programs;

- improve the quality and usefulness of education through federally supported research and evaluation as well as sharing of information;

- increase accountability of Federal education programs to the President, the Congress and the public (U.S. Department of Education, 2009a).

- improve coordination of Federal education programs;

- improve management and efficiency of Federal education activities; and

- increase accountability of Federal education programs to the President, the Congress and the public (S. Rep. No. 107-337, 2002).

5. In 1985, Congress established the Office of Educational Research and Improvement (OERI), and terminated the NIE. The OERI was “charged with extending knowledge about education through its research activities and to provide leadership in the conduct and support of scientific inquiry into the educational process” (S. Rep. No. 107-337, 2002).

6. In 1988, Congress reauthorized the NCES and the NAEP.

7. In 1994, Congress reauthorized the OERI, the NCES and the NAEP.

8. In 2002, Congress reauthorized the NAEP, requiring state participation in the NCLBA.

The United States’ enacted legislation and federally funded education programs show the focus on education and student achievement in U.S. society. Further evidence of U.S. society’s concern for education and student achievement can be seen in the education accountability movement.
**Education accountability.** Accountability can be defined as the state of districts, schools, educators, and/or students being held responsible by policies or processes for performance or outcomes (American Association of School Administrators, 2002; National School Boards Foundation, 2001). Accountability in education serves three main purposes: “to inform, to reorient action, and to justify what is done” (Benveniste, 1985). Broadly speaking, those currently accountable for education in the United States include federal and state legislators, state boards and departments of education, teacher education providers, state commissioners, local school boards, individual school districts and schools, principals/administrators, teachers, students, parents, and community/business leaders (Education Commission of the States, 1998; Gaskie & Gaskie, 1975).

Early in education’s history, teachers who received room and board in exchange for teaching a family’s children were accountable to the parents of that family (Gaskie & Gaskie, 1975). In the 18th- and early 19th-century, a visiting committee of town elders held school teachers accountable (Dorn, 1998). However, oftentimes teachers were just required to pass a knowledge test and interview with the local school board to be accepted for service. Once a teacher was hired, the teachers’ students were held accountable for their learning and held back if they failed to perform history, geography and arithmetic test at satisfactory levels (Ravitch, 2002). Over time, accountability practices changed. For example, an 1817 Georgia state law, which applied to schools that served poor families, prohibited commissioners from paying any salary to a teacher whose students’ examination results indicated insufficient progress (Lessinger, 1970). During the 1860s, a system of “accountability by results” that produced a “payment by results” system was implemented in Great Britain (Kuchapski, n.d.).
Education accountability is often associated with “scientific management” and the efficiency movement of the early 20th century (Kuchapski, n.d.). The focus at that time on preparing students for vocations resulted in competency-based education, functional literacy and accountability (Wise, as cited in Kuchapski, n.d.). A more modern push for accountability in education is associated with Russia’s successful launch of the Sputnik in 1957 which resulted in a criticism of math and science education in American schools (Ahearn, 2000). The Sputnik launch added fuel to accountability concerns because it created a link between education and national security (Wise, as cited in Kuchapski, n.d.).

The landmark report, *Equality of Educational Opportunity* (Coleman et al., 1966), not only sparked interest in accountability, but also impacted education research in that the authors of the report decided to investigate how school resources affected achievement. This shifted research focus from inputs to results (Ravitch, 2002). The researchers in that study explored the relation of achievement (measured by achievement tests) to school characteristics, academic practices and academic characteristics of the teachers and students in the schools across the United States (Coleman et al., 1966). Before Coleman’s report, education reform emphasized resources; after the report, reformers sought changes in performance (Ravitch, 2002; The Abell Foundation, 2001).

The education accountability movement gained momentum in the early 1970s (Gaskie & Gaskie, 1975). Ever-rising school costs coupled with stagnated or lackluster student performance resulted in demands by the public for evidence of the effects of educational expenditures (Gaskie & Gaskie, 1975; Hanushek & Raymond, 2003). Rather than focusing on the traditional regulation of process and inputs, reformers emphasized regulation of outcomes (Hanushek &
Raymond, 2003). The emphasis in the 1970s on students graduating with at least a set of basic skills led to minimal competency assessments (Education Commission of the States, 2003).

As high stakes related to student achievement became more apparent, parents, taxpayers, the public and business leaders demanded improved student achievement, holding schools accountable for results (National School Boards Foundation, 2001). By 1975, accountability of some sort had been written into the statutes of more than 30 states (Gaskie & Gaskie, 1975). By 1997, 95% of states in the U.S. had or were developing content standards and accompanying assessments (Education Commission of the States, 1997).

In the early 1980s, the National Commission Excellence in Education—formed by the Secretary of Education—examined the quality of education in the United States. The Commission’s report, *A Nation at Risk: The Imperative for Educational Reform* ignited a new wave of education reform. New demands for education accountability from the mid-1980s through the 1990s characterized the nation’s commitment to quality education (Yang, 2004). Reforms in the 1980s focused on better use of time, higher teacher quality and increased graduation requirements (Education Commission of the States, 2003) and included enactment of new student testing requirements in 40 states (Kuchapski, n.d.). According to the Education Commission of the States (2003), during the 1990s challenging academic standards that all students were expected to attain were defined and adopted, and tests were designed to assess students’ progress toward attainment of those standards. By 1999, only three states had no state-mandated accountability system components in place (in statutes, regulations or by executive order), 18 states had one or two components in place, and 29 states had at least three components in place (Education Commission of the States, 1999). By 2001, some legislatures
were mandating the dismantling of low-performing schools or state takeovers of failing districts (Education Commission of the States, 1999).

“Education has become a very high profile issue for politicians mainly in response to public demands for improvements in learning” (Ahearn, 2000). One can see a shift in U.S. legislation toward accountability practices—from an emphasis on increasing the quantity and quality of American education and bringing better education to disadvantaged youth in the 1960s to closing the achievement gap with accountability, flexibility and choice in the 21st century. For example, the *Elementary and Secondary Education Act of 1965* was enacted to bring better education to millions of disadvantaged youth, to provide the best educational equipment and ideas and innovations to all students, and to advance the technology of teaching and the training of teachers (Office of Education, U.S. Department of Health, Education & Welfare, 1965).

The 2001 reauthorization of the *Elementary and Secondary Education Act of 1965*, known as the *No Child Left Behind Act of 2001* (NCLBA), professed the following goals: to “close the achievement gap with accountability, flexibility and choice, so that no child is left behind” (*No Child Left Behind Act of 2001*, 2002, p. 1). Connections to previous education reforms and the accountability movement are apparent in Title I of the Act. The stated goal of Title I of the Act was to improve the academic achievement of the disadvantaged: “The purpose of this title is to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging state academic achievement standards and state academic assessments” (*No Child Left Behind Act of 2001*, 2002, SEC. 1001). This was to be accomplished in part by
aligning academic assessments and accountability systems as well as curriculum, instructional materials, and teacher training with challenging state academic standards so common academic achievement expectations could be measured;

- holding states, LEAs and schools accountable for improving the academic achievement of all students;

- identifying and improving low-performing schools that had failed to provide a high-quality education to their students;

- using state assessment systems to improve and strengthen accountability, teaching and learning;

- providing more decision-making authority and flexibility to schools and teachers on the condition that they assume greater responsibility for student achievement; and

- promoting school-wide reform and the access of children to effective scientifically based instructional strategies.

Grants to support accountability are listed in Subpart 1 of Title VI of the NCLBA; grants were included to

- develop additional state standards and assessments,

- administer state assessments,

- develop multiple measures to increase validity and reliability of state assessments,

- promote validity and reliability of state assessments, and

- develop information and reporting systems to identify research-based best educational practices (No child left behind act of 2001, 2002).
Although the intentions of the accountability movement are noble—to improve education, to increase the level of student achievement, and to make better decisions for future improvement—the movement has had problems and cause for concern, including

- potential negative consequences for schools, school districts, and/or teachers due to low test results (Gaskie & Gaskie, 1975);
- concern that state funding would become contingent upon assessment results (Gaskie & Gaskie, 1975);
- concern that teachers will be deeply involved in implementing accountability systems for which they had little input into developing (Gaskie & Gaskie, 1975);
- demoralization of teachers (Benveniste, 1985);
- unfair consequences for students due to implementation of high-stakes student accountability policies before effective school accountability systems are in place; e.g., promotion policies and exit examinations (Goertz, 2001);
- potential subtle curriculum changes geared to high-stakes tests (Gaskie & Gaskie, 1975);
- administrators’ loss of local autonomy under state-imposed systems (Gaskie & Gaskie, 1975);
- increased bureaucratization in schools resulting in less opportunity for teacher discretion, innovation, risk-taking or adaptaption for varying student needs (Benveniste, 1985); and
- potential negative effects of increased testing for students (Gaskie & Gaskie).

**Technology connections.** “Technology is in part a social process…supported to serve the society that generates and controls it through society’s private and public institutions and people”
In spite of concerns, accountability has become a critical component in the United States’ efforts to address the quality of education and to promote workforce competency and global competitiveness in the 21st century. The U.S. Department of Education stated in its Strategic Plan for Fiscal Years 2011-2014: “We also will promote rigorous and fair accountability for states, districts, and schools based on the collection and use of comprehensive performance data, including student academic growth, that inform educational improvements to drive higher student achievement and improved educational outcomes for all students” (U.S. Department of Education, n.d.c, p. 2). Technology enables society to collect and use data for promoting education accountability. “The connection between accountability and technology can be best summarized in one term: feedback. Technological supports provide ways of designing, collecting and sharing information in order to provide the basis for the improvement of systems and outcomes” (Baker, 2005, p. 1). Recently, education data systems (EDSs) have been widely implemented in the United States to collect and share information for the purpose of satisfying state and/or federal demands.

*Education data systems.* As the competitiveness of the United States in the global economy of the 21st century is dependent upon effectively educating its students, education can be considered an environment in which threats and opportunities arise. Management guru Peter F. Drucker emphasized the increasing importance of data in such critical environments: “the need for information on the environment where the major threats and opportunities are likely to arise will become increasingly urgent” (Drucker, 2006, p. 150). Accountability conditions and the push for data-informed decision making have created a greater need for relevant data that is organized in a format that is accessible and useful for decision makers (Supovitz, 2008). Increasingly, tools are being developed to collect and store education data in the United States.
Multiple types of data systems have been and are being implemented in school districts across the United States. Wayman (2005) listed some of the common types of data systems used in education:

1. Student information systems (SIS) provide real-time access to information such as enrollment and attendance but are not designed to analyze data or provide historical data. Most school districts had SISs in place; as part of the 2006 national *Study of Education Data Systems and Decision Making*, a nation-wide survey of school districts in 2007 revealed that 70% of respondents had had SISs for six or more years (Means et al., 2010).

2. Assessment systems organize and analyze data for benchmark assessments given within a school year. Seventy-nine percent of respondents in the 2006 national *Study of Education Data Systems and Decision Making* reported that their districts had assessment systems that organize and analyze benchmark assessment data (Means et al., 2010).

3. Data warehouses give access to all types of historical data but usually do not provide immediate turnaround of new data. Seventy-seven percent of respondents in the 2006 national *Study of Education Data Systems and Decision Making* reported that their districts have data warehouses that provide current and historical student data as well as other data regarding district functioning (Means et al., 2010).

The strongest driver for developing academic data warehouses is the need for continuous decision making and for quality enhancement (Wierschem, McMillen, & McBroom, 2003). Properly structured and administered data warehouses can provide stable data and security control that is needed to address accountability issues (Wierschem et al., 2003). EDSs can
facilitate reporting requirements of the NCLB Act of 2001, provide to teachers timely information about transfer students, and provide information to educators, parents and policymakers for improving schools (Dougherty, 2002). EDSs provide or contribute to accountability functions identified by Baker (2005):

- storing and organizing information,
- generating reports,
- performing computations and analyses to summarize and compare data, (d)
- providing access to information and methods to cross-check its accuracy,
- encouraging interaction and communication for problem solving, and
- monitoring costs. Education data systems must be capable of providing a longitudinal record of each student’s educational performance and experiences (Means, Padilla, DeBarger, & Bakia, 2009).

Statewide longitudinal data systems enable a state to

- monitor students’ academic progress from grade to grade,
- determine specific programs’ and schools’ educational efficiencies and added value,
- identify consistently high-performing schools to explore best practices,
- evaluate the effects of teacher preparation/training on student achievement, and
- focus the institution on preparation of a higher percentage of students who are college and career ready (Data Quality Campaign, n.d.a).

In November 2005, ten founding organizations launched The Data Quality Campaign (DQC) to improve not only the collection of high quality education data, but also the availability and use of data to improve student achievement (Data Quality Campaign, 2009). The DQC identified ten essential elements of longitudinal EDSs as well as ten state actions for effective data use to promote the goal of having policies and practices in place that make sure that
stakeholders are able to access, to understand, and to use data for continuous improvement (Data Quality Campaign, n.d.b). DQC’s ten essential elements of a longitudinal data system are:

- a unique identifier for each student that links student data across vital databases and across years;
- student-level information on demographics, enrollment, and program participation;
- the capability to measure individual students’ growth from year to year by matching students’ test records from year to year;
- information on untested students, including the reasons they were not tested;
- a system to identify teachers that matches teachers to students;
- student-level transcript data that includes data on course completion and grades earned;
- student-level college readiness test scores;
- student-level dropout and graduation data;
- matching capability between students’ P-12 records and postsecondary systems; and
- a state data audit system assessing quality, reliability and validity.

DQC’s ten state actions for effective data use promote attainment of specific goals as follows (Data Quality Campaign, n.d.b, p. 5):

1. The goal of expanding the ability of state data systems to link across P-20/workforce pipelines is promoted through the state actions of

   - linking data systems;
   - creating stable, sustained support;
   - developing governance structures; and
   - building state data repositories.
2. The goal of ensuring that data can be accessed, analyzed and used is promoted through the state actions of (a) implementing systems to provide timely access to information, (b) creating progress reports with individual student data to improve student performance, and (c) creating reports with longitudinal statistics to guide systemwide improvement efforts.

3. The goal of building the capacity of all stakeholders to use longitudinal data is promoted through the state actions of (a) developing a research agenda, (b) promoting educator professional development and credentialing, and (c) promoting strategies to raise awareness of available data.

Florida Department of Education’s data warehouse is a good example of a longitudinal EDS. The data warehouse contains student data from preschool, through high school and college, and into the workforce; this data system is one of eleven statewide longitudinal data systems that meets the DQC’s ten longitudinal data system essential elements (Partnership for Learning, 2010). Florida’s data warehouse was the first data warehouse of its kind to be developed; the warehouse is nationally recognized for its student data which ranges from pre-school through college and into the workforce (Partnership for Learning, 2010). From data housed in this data warehouse, the state produces reports that help high schools gauge how well prepared their students are for college and beyond. The data is linked to community colleges, colleges and universities in the state; therefore, high schools can determine how many of their students passed freshman math and English, how many took remedial courses and how many maintained at least a 2.0 GPA in college (Partnership for Learning, 2010). The data warehouse improved Florida’s ability to conduct studies of student performance over time by (a) providing a process that is more efficient and consistent for compiling longitudinal student data; (b) establishing enhanced
student privacy-protection procedures; and (c) serving a range of users, including department employees, managers, external researchers, and policymakers (Del Monte et al., 2009).

Florida’s data warehouse illustrates Drucker’s claim of a transition to a “new and radically different view of the meaning and purpose of information: as a measurement on which to base future action rather than as a postmortem and a record of what has already happened” (Drucker, 2006, p. 150). Drucker highlighted the importance of not just the technical tool, but the interdependence and interaction among concepts and tools. In education, this interdependence and interactivity between EDSs and the concepts of improving education and accountability takes form in the move toward data-informed decision making and policy.

Data-informed decision making. “The use of student data systems to improve education and help students succeed is a national priority” (Means et al., 2010, p. ix); use of data in decision making is expected to cover all levels of the education system—from the classroom level to the school, district, state and federal levels classroom levels (Means et al., 2009). For example, the National School Boards Foundation proposed that a vital step in attaining their mission of fostering “excellence and equity in public education through innovation in school board leadership and community involvement” is to “educate board members about the necessity of using good data to inform the decisions they make that affect districtwide [sic] goals for improving student achievement” (National School Boards Foundation, 2001, p.1). Assessment data is an important component of decision making: states use assessment data to evaluate the effectiveness of their educational systems, districts use it to keep account of their instructional programs, and teachers use it to identify students’ curricular strengths and weaknesses (Stecker, Lembke, & Foegen, 2008). The National School Boards Foundation (2001) emphasized the need to know how students are currently achieving—on state-mandated tests, on standards-based tests,
and over time—in order to improve student learning. And to help all students achieve, the
Foundation promoted analysis of data from multiple perspectives—race, culture, socio-economic

Data-informed decision making can contribute to:

- identification of teachers’ professional development needs (American Association of
  School Administrators, 2002; Bray, 2003; Goertz, 2001);
- increased educators’ understanding of assessment or attainment of student outcomes
  (Bray, 2003; Goertz, 2001);
- evaluation of the effectiveness of instruction, education initiatives and programs
  (American Association of School Administrators, 2002; Dougherty, 2002);
- curriculum development guidance (American Association of School Administrators,
  2002; Goertz, 2000);
- wise resources allocation (American Association of School Administrators, 2002;
  National School Boards Foundation, 2001);
- accountability practices (American Association of School Administrators, 2002);
- identification of highly successful schools and their practices; and
- development of ways to address the achievement gaps between students with
  differing educational needs and those of different racial, socioeconomic, ethnic, and
  language backgrounds (Goertz, 2001).

Some educators also have concerns about using data for decision-making and policy.
Some are skeptical and resistant of the practice, believing that data analysis based on student test
results might
- be used against them (American Association of School Administrators, 2002),
Evidence exists in the literature to support the notion that data-informed decision making in education is related to positive outcomes. Williams and Kirst (2006) surveyed 257 principals and more than 5,500 teachers across the state of California to identify school practices that contribute to high-performing elementary schools that serve largely low-income students. The use of data to improve student achievement and instruction strongly correlated with a higher school academic performance index (API).

At the Evelyn S. Thompson Elementary School in Houston, Texas, skill specialists (experienced teachers who support classroom teachers in specific subjects), regularly analyzed student performance data and worked with teachers to identify skill/knowledge gaps and plan actions to address the gaps. They also met with administrators to plan support strategies for teachers and students who needed help. In 1994—before implementing data-informed instruction—65% of third and fourth grade students met state standards in mathematics; in 2005, 94% met the state standards. Acquisition of state standards for reading improved as well, from 74% of students attaining the standards in 1994 to 92% in 2005 (Petersen, 2007).

Carlson, Borman and Robinson (2011) estimated the one-year impacts of a data-driven reform initiative conducted across 500 schools within 59 districts and seven states. The reform initiative included implementation of quarterly benchmark assessments along with consulting services that included data reviews with district and school leaders and training in the interpretation and use of data for district, school and teacher leaders. Relative to operation as
usual (the control condition), the data-driven initiative showed statistically significant
district-wide improvements in student mathematics achievement.

On the other hand, some Boston public schools implemented an intervention in which
teachers were coached to use formative assessment feedback to inform instructional strategies. Although the teachers reported that the professional development they received contributed to their understanding of data and their ability to work with students, there were no significant differences in student achievement at the schools that implemented the intervention (Quint, Sepanik, Smith, & MDRC, 2008).

Data-informed decision making in education is one part of a web of connections among society, education and technology: For U.S. society to remain secure and competitive globally in the 21st century, the government has made it a priority to improve education for all students; data systems are used to gather, store and provide access to data; the data are used to track student achievement, to identify achievement gaps among different groups of students, to analyze trends, to research what works to improve education, to contribute to decision making and to promote accountability; and accountability systems “embody prevailing societal values and aspirations” (Anderson, 2006, p. 2), helping to ensure student achievement.

**Accountability systems.** To realize the benefits of accountability, individuals or organizations must know what the goals are, must measure progress toward the goals, and must reward success and address failure (Barber, 2004). Baker (2005, p. 2) listed a four-part scheme for accountability systems:

1. Valued and attainable goals are identified;
2. Responsible parties (RPs) are identified, and feedback on progress toward the goals is given;
3. Different levels of performance or rates of progress receive differentiated rewards and sanctions; and

4. Improvement occurs because RPs modify their actions in order to receive rewards and avoid sanctions.

The combination of components which comprise education accountability systems contribute to an institution’s ability to realize accountability benefits:

- established standards/goals (Baker, 2005; Education Commission of the States, 1998; Education Commission of the States, 1999),
- standards/goals progress indicators (Baker, 2005; Education Commission of the States, 1997; Education Commission of the States, 1998; Education Commission of the States, 1999; Goertz, 2001), and

It is important to include multiple types of achievement progress indicators in an accountability system. Ananda and Rabinowitz (2001) advised inclusion of assessments as well as non-assessment indicators that can influence achievement such as attendance and retention. Baker, Linn, Herman and Koretz (2002) encouraged use of student and teacher characteristics data for interpreting student achievement. The Association of California School Administrators Accountability and Assessment Task Force (2010) recommended the use of formative assessments in addition to summative assessments.

An accountability system’s consequences come in the forms of rewards and sanctions that “offer the potential for focusing teachers’ work, motivating school improvement efforts and improving teaching and learning” (Education Commission of the States, 1998, p. vi). Rewards go
to schools and teachers based on attainment of stated performance goals; sanctions are applied to schools, teachers or students when student achievement does not meet set standards (Ananda & Rabinowitz, 2001).

Education accountability systems create “transparency in reporting student achievement outcomes for all students and specific subgroups, while simultaneously making visible both systemic accountability innovations and limitations” (Association of California School Administrators Accountability and Assessment Task Force, 2010, p. 1). The systems hold and provide educator access to information that can be used to

- guide improvement and/or meet established standards (Education Commission of the States, 1998; Goertz, 2000; Supovitz, 2009);
- inform policy and decision making (Education Commission of the States, 1998);
- guide resource allocation (Education Commission of the States, 1998), and
- plan staff development.

Accountability systems enable institutions to

- monitor and evaluate student progress toward established standards (Education Commission of the States, 1998),
- motivate improvement (Goertz, Olah, Riggan, & Consortium for Policy Research in Education, 2009; Supovitz, 2009),
- spur alignment of the major educational system components (Supovitz, 2009),
- enable performance evaluation in relation to other systems (Education Commission of the States, 1998), and
- indicate value to stakeholders (Education Commission of the States, 1998; Supovitz, 2009).
As Cheek (2000) put it, accountability systems “are policy formulations whose technical functions of collecting, analyzing, and publishing data coupled with a continuum of state intervention procedures are intended to bring about desired changes in schools and populations of students” (p. 9).

The implementation of accountability systems to ensure student achievement is an indicator of the value of education to society. U.S. society’s concern for quality education also promotes research agendas to determine how to improve education, contributes to creation of education programs to improve education, and drives the data-informed decision making movement in education. Technology enables the gathering, storing, and analysis of data to improve education. An area of vital importance in U.S. society’s education is mathematics.

**Research Context: Society, Mathematics Education, and Technology**

To hone in on the context for this study, in the next section of the literature review the value of mathematics to U.S. society and the individuals within that society is explored; and U.S. students’ mathematics performance is compared with the performance of students in other countries. Then some factors that may affect mathematics performance are discussed. Finally, the background on the study’s statistical models is presented.

**Mathematics for the well-being of U.S. society.** The launch of the Sputnik in 1957 linked mathematics and science education with national security. The fact that many other disciplines—e.g., medicine, computer science, networking, software development, physics, chemistry, engineering, business, commerce, biology, pharmacology, actuarial science, psychology, economics, and geography—depend on a foundation of mathematics and/or statistics shows the importance of mathematics to the well-being of society. The National
Mathematics Advisory Panel (2008, p. xii) explained why mathematics is critical to the well-being of U.S. society:

The eminence, safety, and well-being of nations have been entwined for centuries with the ability of their people to deal with sophisticated quantitative ideas. Leading societies have commanded mathematical skills that have brought them advantages in medicine and health, in technology and commerce, in navigation and exploration, in defense and finance, and in the ability to understand past failures and to forecast future developments. History is full of examples.

During most of the 20th century, the United States possessed peerless mathematical prowess—not just as measured by the depth and number of the mathematical specialists who practiced here but also by the scale and quality of its engineering, science, and financial leadership, and even by the extent of mathematical education in its broad population. But without substantial and sustained changes to its educational system, the United States will relinquish its leadership in the 21st century. The report goes on to explain that although much of the discussion about mathematics and science in the United States emphasizes national economic well-being and competitiveness, there is additional reason for concern: the safety of the U.S. and quality of life in the nation is at stake. The United States may face a potential inability to sustain a quality workforce, due to a combination of looming retirements of a large portion of the current science and engineering workforce and the potential failure of the U.S.’s recent strategy of importing technical talent from abroad. Not only the future workforce, but also citizens and policy leaders, will need to be more skillful at handling quantitative concepts (National Mathematics Advisory Panel, 2008). The United States’ ability to sustain a quality workforce became evident in the summer of 2010
when, with almost ten percent unemployment, businesses complained that workers with needed skills could not be found (Hanushek, Peterson, & Woessmann, 2010).

Other “corporate, government, and national scientific and technical leaders have expressed concern that pressures on the science and technology enterprise could seriously erode this past success and jeopardize future US prosperity” (National Research Council, 2007, p. ix). Participants in a joint meeting of the councils of the National Academy of Sciences and the National Academy of Engineering “expressed concern that a weakening of science and technology in the United States would inevitably degrade its social and economic conditions and in particular erode the ability of its citizens to compete for high-quality jobs” (National Research Council, 2007, p. ix). In response to such concerns, the National Academies funded a study; the Committee on Science, Engineering and Public Policy established a committee of CEOs of Fortune 100 corporations, former presidential appointees, Nobel laureates, and presidents of major universities to investigate actions that federal policy makers could take to improve the United States’ science technology enterprise and facilitate the United States’ competition, prosperity and security in the global community of the 21st century (National Research Council, 2007).

After reviewing trends abroad and in the United States, the committee concluded in its report, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*, that “that the scientific and technological building blocks critical to our economic leadership are eroding at a time when many other nations are gathering strength” (National Research Council, 2007, p. 3). The committee (a) voiced concern about the United States’ K-12 mathematics and science education system, calling it “the foundation of…human capital in today’s global economy” (National Research Council, 2007, p. 30); (b) concluded that
“the science and technology research community and the industries that rely on that research are critical to the quality of life in the United States” (National Research Council, 2007, p. 67); and (c) recommended actions to improve and provide thorough education in mathematics, science and technology for the purpose of developing an innovative science and engineering workforce.

Five years after the release of Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future, the presidents of the National Academy of Sciences, the National Academy of Engineering and the Institute of Medicine, noting that the United States was facing even greater challenges in remaining competitive in the world, requested an update on the report. The update, Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5, indicated that the United States’ outlook has worsened due to (a) economic turmoil; (b) worldwide advancement of knowledge, education, investment, innovation and industrial infrastructure; (c) little improvement, overall, in the United States’ 14,000 public school systems’ mathematics and science education while many other nations were making noticeable progress (National Research Council, 2010). The report ends with a 2001 quote from the United States Commission on National Security for the 21st Century: “Second only to a weapon of mass destruction detonating in an American city, we can think of nothing more dangerous than a failure to manage properly science, technology and education for the common good…” (National Research Council, 2010, p. 67).

**Mathematics achievement: An indication of success within U.S. society.** Because science and technology support many aspects of U.S. society, it is vital that all students acquire science and math skills (Afterschool Alliance, 2010). Achieve, Inc.—a bipartisan, non-profit organization that contributes to raising state academic standards, improving assessment, and strengthening accountability—identified mathematics as a basic building block for success
because mathematics is a foundation for higher-order thinking; it is “an intellectual gateway to abstract reasoning” (Achieve, 2011, p. 4). The organization promotes higher-level mathematics for all high school students because taking higher-level mathematics

- is key to access to higher education;
- is critical for college success and completion;
- better prepares students for the workplace; and
- students for a career in many fields, including aerospace, construction, health, information technology, and manufacturing (Achieve, 2011).

“Mathematics and science education is gaining increasing recognition as key for the well-being of individuals and society” (Culpepper, Basile, Ferguson, Lanning, & Perkins, 2010, p. 157). Mathematics achievement is important for students because it impacts school, college and career readiness. In a study of six longitudinal data sets aimed at finding links between school readiness factors and later success in reading and mathematics, early mathematics skill level (at age five or six) had greater power in predicting later student success than did school-entry reading level and attentions skills of young children (Duncan et al., 2007). Not surprisingly, Duncan et al. found that early mathematics skill was a powerful predictor of later mathematics achievement. In addition, they found that early mathematics skill was a more effective predictor of later reading achievement than early reading skill was in predicting later mathematics achievement.

Researchers have recently focused on the importance of students’ completion of algebra, in part, because “the study of algebra enhances cognitive skills while serving as a gatekeeper to more advanced academic courses and enhanced opportunities after high school” (Spielhagen, 2006, p. 39). There is evidence that students benefit from taking algebra in the eighth grade
rather than in the ninth grade. Spielhagen compared students \( n = 2,643 \) who completed algebra in the eighth grade with those who studied algebra in the ninth grade. She found that students who completed algebra in the eighth grade stayed in mathematics longer and attended college at greater rates than students who did not take algebra in the eighth grade. Edge (2009) found a statistically significant relationship not only between eighth-grade algebra completion and high school math course completion, but also between algebra completion and college readiness. Achieve’s (2009) research indicated that students need four years of rigorous math courses in high school (Algebra II at a minimum) to be prepared for college and career success. Achieve (2004) also recommended that students take geometry, data analysis and statistics in high school so that students are prepared for post-secondary school challenges. Mathematics achievement through the eighth grade seems to be a key to student success. According to Baum and Pavea (2005), high math test scores in eighth grade significantly contribute to high school graduation, enrollment in college, and earning a bachelor’s degree within all socioeconomic groups.

The *Foundations for Success* report (National Mathematics Advisory Panel, 2008) emphasized that success in high school mathematics—in particular, through Algebra II—correlates highly with access to college and graduation from college. Students who complete Algebra II in high school are more than twice as likely to graduate from college as students who do not complete Algebra II (Adelman, 1999; National Mathematics Advisory Panel, 2008). Adelman (1999) declared that “the highest level of mathematics reached in high school continues to be a key marker in precollegiate momentum, with the tipping point of momentum toward a bachelor’s degree now firmly above Algebra 2” (p. xix). Furthermore, “of all pre-college curricula, the highest level of mathematics one studies in secondary school has the strongest
continuing influence on bachelor’s degree completion” (Adelman, 1999, p. vii). Zelkowski (2011) found that taking four years of mathematics in high school is a strong predictor of college readiness and of earning a bachelor’s degree.

The authors of a study of 2,108 students found that math ACT scores related significantly to performance of students in college biology and chemistry classes. A five-point difference in ACT math scores corresponded to more than a point difference in general biology course grade and corresponded to a 1.65 difference in chemistry course grade (Culpepper et al., 2010). Davis (2010) investigated early warning indicators of college readiness of 134 early-college high school students in Texas and found that those who maintained an A or B grade in Algebra I were ten times more likely to be college ready and experience success at college—they had a 78% chance of maintaining a 2.5 grade point average their first year in college.

Culpepper (2010) found that students who completed one year of calculus in high school tended to perform a grade level higher in general biology than students who did not take calculus in high school. Sadler and Tai (2007) conducted a study across 77 four-year higher education institutions to determine predictors of student success in college sciences (biology, n = 2,650; chemistry, n = 3,561; and physics, n = 2,263). They found that the number of years of mathematics instruction a student had in high school carried significant benefit across college biology, chemistry and physics courses.

There is evidence that mathematics achievement contributes to retention of some college students. Calculus readiness was a significant predictor of first-year retention of 129 freshman engineering majors (Moses et al., 2011). Min, Zhang, Long, Anderson and Ohland (2011) used a large longitudinal database of more than 100,000 engineering students from nine universities over nine years to analyze the loss rate of undergraduate engineering students. They found that
whereas students in general most often leave engineering programs during the third semester of study, students with SAT math scores less than 550 tend to leave during the second semester. SAT math scores correlated with retention over three years of students majoring in applied mathematics, biology, chemistry, mathematics, microbiology, molecular/cellular biology, physics and zoology at a research one university in the South (Scott, Tolson, & Huang, 2009). Rose and Betts (2001) controlled for student demographic traits, measures of ability, family characteristics and school characteristics in their investigation on the impact of different high-school mathematics courses on students’ predicted probability of graduating from college. Their results showed that students whose highest level math course in high school was vocational math had only a 4.9% probability of graduating from college. Pre-algebra students had an 8.1% probability, algebra/geometry students had a 14.5% probability; intermediate algebra students had a 26.8% probability; advanced algebra students had a 38% probability; and calculus students had a 42.8% probability of graduating from college.

Adequate basic mathematics skills are needed for productive adult life; advanced mathematics skills are vital in many professions. An individual’s past participation in mathematics courses and higher education continue to be correlated with benefits after graduation from college (National Mathematics Advisory Panel, 2008). The *Foundations for Success* report (National Mathematics Advisory Panel, 2008) stated that success in high school mathematics through Algebra II correlates highly with earnings in the top quartile of employment income. Carneval and Desrochers (2002) reported that 84% of the sample in their study who held highly paid professional jobs had completed Algebra II or higher level math courses in high school and that 67% of those who held well-paid, white-collar, skilled jobs had taken Algebra II or higher level math courses. Forbes.com reported that for people with less than
five years’ work experience, engineering-based jobs were the highest paying; but that economics, finance and math majors had better earnings growth potential than engineering majors (Badenhausen, 2008). Majors in which students learn to integrate mathematics and science with the everyday world—mathematics, economics, engineering and finance—had an excellent benefit in terms of earnings potential (Badenhausen, 2008).

Rose and Betts (2001) found that taking additional mathematics courses in high school had direct and indirect impacts on earnings, and that the impact varied by level of mathematics courses taken (see Figure 1). For example, considering just the direct impact of taking additional math courses in high school, if one student takes a pre-algebra class in high school and another student takes two additional math classes (algebra/geometry and intermediate algebra), the pre-algebra student is predicted to earn 6.4% less than the student who took two additional math classes (3.2% less for not having taken algebra/geometry, plus 3.2% less for not having taken intermediate algebra). When the indirect effect that taking additional math courses has on obtaining postsecondary education is considered as well, the student who takes the additional two classes is predicted to earn 10.4% more than the student who took only pre-algebra (the sum of the percentages of direct and indirect impact of the algebra/geometry class and the intermediate algebra class). Note that taking one more vocational math class than the average student does has a negative impact on earnings.

Mathematics achievement will continue to be an indicator of success in U.S. society as “the growth of jobs in the mathematics-intensive science and engineering workforce is outpacing overall job growth by 3:1” (National Mathematics Advisory Panel, 2008, p. xii). Carnegie President Anthony S. Bryk summed up the importance of mathematics to the individual: “Rather
than a gateway to a college education and a better life, mathematics has become an unyielding gatekeeper” (Carnegie Foundation for the Advancement of Teaching, 2010, para. 3).

![Figure 1](image)

Figure 1. Predicted percentage increase in earnings resulting from an additional math course in high school. Adapted from “Math Matters: The Links Between High School Curriculum, College Graduation, and Earnings” by H. Rose and J. R. Betts (2001, p. xv). Copyright 2001 by the Public Policy Institute of California (PPIC). Used with permission of the PPCI.

**U.S. students’ mathematics achievement.** Mathematics achievement is important for the success of the individual in U.S. society and for the well-being of the nation as a society within a global framework. Just how well do U.S. students perform in mathematics? Results of internationally standardized mathematics achievement tests show that U.S. students have consistently underperformed in mathematics compared to some of their peers worldwide.
Trends in International Mathematics and Science Study (TIMSS). The International Association for the Evaluation of Educational Achievement (IEA) has undertaken a series of international studies, Trends in International Mathematics and Science Study (Williams et al., 2000). The tests measure science and mathematics achievement of students in early, middle and final years of their schooling. TIMSS data have been gathered four times—in 1995, 1999, 2003 and 2007. Table 3 details U.S. students’ TIMSS mathematics performance with respect to other jurisdictions. Inspection of TIMSS reports reveals that Belgium-Flemish, Hong Kong (SAR), Japan, Republic of Korea, Netherlands, and Singapore and significantly outperformed U.S. students in TIMSS mathematics performance in at least half of the exam instances.
Table 3

Comparison of U.S. TIMSS Results with Other Jurisdictions, 1995-2007

<table>
<thead>
<tr>
<th>Year</th>
<th>4th-grade TIMSS mathematics averages</th>
<th>U.S. ranking and outperformers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>1995</td>
<td>400 (Kuwait)</td>
<td>625 (Singapore)</td>
</tr>
<tr>
<td>2003</td>
<td>339 (Tunisia)</td>
<td>594 (Singapore)</td>
</tr>
<tr>
<td>2007</td>
<td>224 (Yemen)</td>
<td>607 (Hong Kong SAR)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>8th-grade TIMSS mathematics averages</th>
<th>U.S. ranking and outperformers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>1995</td>
<td>354 (S. Africa)</td>
<td>643 (Singapore)</td>
</tr>
<tr>
<td>1999</td>
<td>275 (S. Africa)</td>
<td>604 (Singapore)</td>
</tr>
<tr>
<td>2003</td>
<td>264 (S. Africa)</td>
<td>605 (Singapore)</td>
</tr>
<tr>
<td>2007</td>
<td>307 (Qatar)</td>
<td>598 (Chinese Taipei)</td>
</tr>
</tbody>
</table>

Note. Fourth grade was not included in the 1999 assessment. Data sources: Beaton et al., 1996; P. Gonzales et al., 2001; P. Gonzales et al., 2004; P. Gonzales et al., 2008; Mullis et al. 1997.
The Programme for International Student Assessment (PISA). Another internationally standardized assessment, the Programme for International Student Assessment, was jointly developed by participating countries. It is administered every three years to 15-year-old school children in countries that partner with the Organisation for Economic Co-operation and Development. PISA assesses students’ acquisition—near the end of compulsory education—of knowledge and skills that are vital for full participation in society. The PISA assessment goes beyond measuring mastery of school curriculum to assess students’ abilities to adapt knowledge acquired at school to real-life situations (Organisation for Economic Cooperation and Development, n.d.).

Table 4 compares U.S. students’ PISA mathematics performance to students from other jurisdictions. Inspection of PISA reports reveals that Australia, Belgium, Canada, Chinese Taipei, Denmark, Estonia, Finland, France, Germany, Hong Kong-China, Iceland, Japan, Republic of Korea, Liechtenstein, Macao-China, Netherlands, New Zealand, Norway, Slovak Republic, Slovenia and Switzerland significantly outperformed U.S. students in at least half the instances of the PISA exam that measured mathematical literacy.
Table 4

Comparison of U.S. PISA Results with Other Jurisdictions, 2000-2009

<table>
<thead>
<tr>
<th>Year</th>
<th>Minimum</th>
<th>Maximum</th>
<th>U.S.</th>
<th>International</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>292 (Peru)</td>
<td>560 (Hong Kong-China)</td>
<td>493</td>
<td>469</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>359 (Tunisia)</td>
<td>550 (Hong Kong-China)</td>
<td>483</td>
<td>488</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>311 (Kyrgyz Republic)</td>
<td>549 (Chinese Taipei)</td>
<td>474</td>
<td>469</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>321 (Kyrgyz Republic)</td>
<td>600 (Shanghai-China)</td>
<td>487</td>
<td>464</td>
</tr>
</tbody>
</table>

Hanushek, Peterson and Woessmann (2010) examined the United States’—and each of the 50 states’—production of high-achieving mathematics students by comparing PISA scores of U.S. high achievers with those of 56 other countries. They found that although the percentage of U.S. students that scored at advanced levels varied considerably among the states, no state did well in international comparison and states in the U.S. are below the levels of many industrialized nations; for instance:

1. Fourteen countries outperformed U.S. first-ranked Massachusetts, which had more than eleven percent of student with advanced scores.

2. Second-ranked Minnesota’s students compared with those of Denmark, Estonia, France, Iceland, Slovenia, and Sweden.

3. Many states’ results were similar to those of developing countries.

4. The lowest performing states—Mississippi, New Mexico and West Virginia—had a smaller percentage of high-performing students than Serbia and Uruguay.

In light of U.S. students’ mathematics performance on the TIMSS exam and the PISA exam, it is understandable that the National Mathematics Panel (2008, p. xii) deemed that “American students have not been succeeding in the mathematical part of their education at anything like a level expected of an international leader.” Furthermore, it is not surprising that improved mathematics achievement has become a national education goal.

A National Education Goal: Improved mathematics achievement. The value the U.S. places on the mathematics achievement of its students is evidenced by recommendations or actions to improve mathematics education initiated by presidents of the United States. An example is one of the six goals adopted in 1989 by the National Education Summit in which President George H. W. Bush assembled the governors of all 50 states to set education goals for
the nation. Goal 4 from the summit stated, “By the year 2000, U.S. students will be first in the world in mathematics and science achievement,” and listed the following objectives (National Education Goals Panel, 1993, p. xi):

1. Math and science education will be strengthened throughout the system, especially in the early grades.
2. The number of teachers with a substantive background in mathematics and science will increase by 50%.
3. The number of U.S. undergraduate and graduate students, especially women and minorities, who complete degrees in mathematics, science and engineering will increase significantly.

The panel determined that the United States must have world-class standards of educational performance because of modern global economic competition, and that the U.S. must know how its schools compare with the best schools in the world (National Mathematics Advisory Panel, 1993).

In 2006, President George W. Bush appointed members to the National Mathematics Advisory Panel “amid widespread alarm at the sorry state of math achievement in America” (Hechinger, 2008, para. 3). The panel analyzed survey data from 743 active algebra teachers, reviewed written commentary from 160 individuals and organizations, reviewed more than 16,000 research articles and policy reports and heard public testimony from 110 individuals (National Mathematics Advisory Panel, 2008). From their research, the panel synthesized a broad set of 45 main findings and recommendations in the following areas: curricular content, learning processes, teachers and teacher education, instructional practices, instructional
In 2009, President Obama declared improvement of science, technology, engineering and math (STEM) education to be a national priority and presented a campaign aimed at promoting STEM education. The effort was backed with about $260 million from companies such as Sony, Discovery Communication, Time Warner Cable and the Bill and Melinda Gates Foundation (Albanesius, 2009). Obama’s goals for STEM education included improving the quality of math and science education and expanding STEM education for under-represented groups such as women and minorities (Albanesius, 2009).

Because of the importance of improved mathematics education/achievement in the United States—for the well-being of the individual and the country as a whole—it is important for researchers to investigate and identify factors that may affect students’ mathematics achievement.

**State of Maryland context.** In five of the last six years, the state of Maryland was ranked number one in the *Quality Counts* report published annually by *Education Week*. The study uses more than 100 indicators that examine state-level efforts to improve public education (The College Board, 2012; Ujifusa, 2013). Maryland attributes its success to reform efforts over more than 20 years that have built a solid infrastructure.

The first wave of reform, from 1989 through 2002, created a comprehensive public assessment and accountability system that placed responsibility for student achievement on schools, LEAs and the State. The second reform wave, from 2002 through 2009,

- provided major funding increases to schools,
implemented new assessments, thereby increasing accountability for LEAs to eliminate performance gaps and improve student achievement,

created and disseminated a statewide curriculum, and

established the Science, Technology, Engineering and Mathematics Task Force (State of Maryland, 2010).

In May, 2010, the Maryland legislature passed the Education Reform Act of 2010 which required that local school boards establish teacher and principal evaluation models which include student growth data as a significant element of the evaluations. The following month, an Executive Order signed by Governor Martin O’Malley established the Maryland Council for Educator Effectiveness. This council established definitions for “effective” and “highly effective” teachers and principals and made recommendations for development of the educator evaluation system that had been mandated by the Education Reform Act of 2010 (Maryland Council for Educator Effectiveness, 2011; Maryland Council for Educator Effectiveness, 2012).

The state’s teacher/principal evaluation system evaluates educators on two general standards, with each accounting for 50% of the evaluation: professional practice and student growth. The professional practice standard includes four domains: planning and preparation, classroom environment, instruction, and professional responsibilities; LEAs can include additional measurements of professional practice that follow guidelines approved by the Maryland State Department of Education. The student growth standard includes two measures: a LEA growth measure (20%) and a statewide growth measure (30%) which must include a statewide assessment (if it exists) as one of multiple measures (Maryland Council for Educator Effectiveness, 2011).
Supported by $250 million from competitive federal Race to the Top (RT3) program, the MSDE’s mission is to “create a world-class system preparing students for college and career success in the 21st century” (State of Maryland, 2010, p. 3) through concerted reform efforts in three areas: the new Common Core State Curriculum, the Partnership for Assessment of College and Careers (PARCC) assessments, and new teacher and principle evaluations (Maryland State Department of Education, 2012). MSDE’s plan for improving student achievement and college and career readiness is built upon six strategies (State of Maryland Department of Education, 2013):

1. To better prepare students for college and careers in an era of global competition, MSDE will transition to new standards and assessments. Strategy goals include implantation of Common Core State Standards and testing students on the core standards using internationally benchmarked assessments.

2. To be able to track student progress from pre-kindergarten through higher education and into the workforce, Maryland will increase the capacity of its comprehensive statewide longitudinal data system. A goal of this strategy is to house data collected from state agencies and provide data to policy makers, educators, and the public for the purposes of enhancing education and support services and promoting transparency.

3. The strategy to close the readiness gap and improve transitions into school, higher education and the workforce includes providing children with early care and learning opportunities so they are fully ready for kindergarten according to the Maryland Model for School Readiness.

4. To sustain and enhance Maryland’s leadership status in the knowledge-based, innovation economy, Maryland will enhance science, technology, engineering and mathematics
(STEM) education. Goals included in this strategy include (a) increasing student, teacher and parent participation in Maryland’s STEM Innovation Network; (b) increasing the number of STEM college graduates; and (c) expanding enrollment in career and technical education (CTE) in STEM-related programs.

5. To help Marylanders gain skills needed for jobs in the State’s high-demand industries such as bioscience and technology, Maryland will expand career and technology education. A goal of this strategy is to increase the number of CTE graduates who earn industry certifications and/or licenses related to their high school program of study.

6. To improve the teaching and learning environment in Maryland, Maryland will provide support to recruit and train qualified and effective teachers and principals. This strategy includes providing to poor and minority students highly qualified and effective teachers at the same rate as more advantaged students and expanding professional development opportunities for teachers that incorporates interaction with industry practitioners.

Factors That May Affect Mathematics Achievement

This study investigates some student factors and some teacher factors that may affect students’ mathematics achievement. It also examines the impact of a formative mathematics benchmarking intervention put in place by the local school district.

Student characteristics. This study considers two student characteristics that may affect mathematics achievement: prior knowledge and gender.

Prior knowledge. The impact of prior knowledge on learning and performance has long been of interest to education researchers; many authors concur that prior knowledge is important to the learning process (Dochy, 1988; Dochy, Segers, & Buehl, 1999; Hailikari, Nevgi, & Lindblom-Ylanne, 2007; Shapiro, 2004). Dochy et al. (1999) cited a number of studies in which
prior knowledge significantly affected performance: Bloom’s study in pretest scores accounted for 25% - 81% of the variance in posttest scores; Dochy’s 1992 study in which up to 42% of performance variance on state tests could be explained by prior knowledge; and the study by Tobias which found that 30 - 60% of variance was explained by prior knowledge. Dochy (1999), in his review of 183 empirical studies conducted between 1978 and 1999, discovered that 91.5% of the studies showed positive effects of prior knowledge on performance. Shapiro (2004) urged researchers to include measures of prior knowledge in analyses, warning that “methodological attempts to “wipe away” prior knowledge as a variable in research can mitigate the conclusions drawn by learning outcomes research” (pp. 182-183).

In particular, domain-specific prior knowledge (prior knowledge of a specific content area) has been shown to impact achievement in, for example, psychology (e.g., R. Thompson & Zamboanga, 2003), pharmaceutical chemistry (Hailikari, Katajavuori, & Lindblom-Ylanne, 2008), (Yenilmez, Sungur, & Tekkaya, 2006) and biology (Randler, 2009).

Prior mathematics knowledge has been a strong predictor of mathematics achievement. Alban (2002) investigated the impact of prior mathematics performance using ordinary least squares (OLS) and MLM analyses and found that prior performance was statistically significant in all the models she used. Similarly, prior mathematics performance was the most influential factor on student mathematics achievement in Germuth’s (2003) assessment of teacher effectiveness. In a study about the effects of ability grouping on mathematics achievement (Bode, 1996), a high level of previous mathematics achievement was also shown to impact average mathematics achievement. Pinkham and Ansley (1996) explored prediction of mathematics performance for making student placement decisions and found that using a combination of previous mathematics grades (most highly correlated with current math grades)
and *Iowa Algebra Aptitude Test* scores produced the most accurate and efficient classification possible for their sample of 997 Iowan students in grades six through nine. Prior knowledge and strategies explained almost 50% of the variance in SAT mathematics scores of 108 parochial high school juniors and seniors (Byrnes & Takahira, 1993). Domain-specific prior knowledge was the strongest predictor of mathematics achievement in a required mathematics course of 139 math and science majors; prior knowledge and previous academic success explained 55% of the variance in their achievement (Hailikari, 2008). Prior mathematics knowledge was a significant predictor of 204 fourth- and 185 seventh-grade students’ comprehension of graphs (Curcio, 1981).

**Gender differences.** Difference in cognitive abilities related to gender is another area in which researchers have long been interested. Havelock Ellis’ book, *Man and Woman*, published in 1894, summarized early experimental evidence of gender cognitive differences (H. B. Thompson, 1903). In her monograph, *The Mental Traits of Sex: An Experimental Investigation of the Normal Mind in Men and Women*, Thompson (1903) detailed results of her gender differences studies on motor ability, skin and muscle senses, taste and smell, hearing, vision, intellectual faculties, and affective processes. Thompson also related differing conclusions of researchers in that era.

With respect to mathematics ability, Thompson’s (1903) research showed that “women as a whole have an equal capacity with the men for furnishing an original solution of a mathematical problem when it is called for unexpectedly” (p. 134). This contrasted with results of Calkins’ study (as cited in Thompson, 1903) which concluded that women were more likely than men to memorize a mathematical demonstration than to use reasoning. In measuring mental arithmetic and memory, Ebbinghaus (as cited in Thompson, 1903) found that boys outperformed
girls up to the age of fifteen, at which time the girls became superior at the tasks. In a similar study by Bellei (as cited in Thompson, 1903), girls outperformed boys.

Thompson’s (1903) conclusion to her gender differences study pointed not to differing levels of capacity or ability between males and females, but to differences in social influences:

The point to be emphasized as the outcome of this study is that, according to our present light, the psychological differences of sex seem to be largely due, not to difference of average capacity, or to difference in type of mental activity, but to differences in the social influences brought to bear on the developing individual from early infancy to adult years. The question of future development of the intellectual life of women is one of social necessities and ideals, rather than of the inborn psychological characteristics of sex (p. 182).

Subsequent research studies on gender differences in mathematics most commonly had one of two results (a) males are superior in mathematics, or (b) there is no gender difference in mathematics (Caplan & Caplan, 2005). More recent studies continue to show mixed results. For instance, Fryer and Levitt (2009) analyzed nationally representative Early Childhood Longitudinal Study data and found that although there were no mean differences between males and females upon entry to school, males gained one-fourth of a standard deviation in comparison to females over the first six years of school. This gender gap in math achievement was consistent across every strata of society. Similarly, Lee, Moon and Hegar (2011) found no significant gender gap in math achievement among entering kindergarteners of various races, but by the spring of their kindergarten year, Caucasian boys outperformed Caucasian girls. By spring of their third-grade year, significant gender gaps existed among all races considered in the study: African American, Asian American, Caucasian, and Hispanic.
Other studies suggest that the gender gap in mathematics performance has largely disappeared in the United States. Lindberg, Hyde, Petersen and Linn (2010) performed (a) a meta-analysis of data from 242 studies representing the testing of more than a million students, and (b) an analysis of samples from the National Longitudinal Surveys of Youth, the National Assessment of Educational Progress, the National Education Longitudinal Study of 1988, and the Longitudinal Study of American Youth. From these analyses, they concluded that female and male performance in mathematics is similar. A study published by the Center on Education Policy (Chudowsky & Chudowsky, 2010) in which data from all 50 states were included found that the percentage of girls and boys who attained proficiency in math was similar from state to state.

Regardless of whether a gender gap in mathematics performance still exists, in the United States women are underrepresented in STEM occupations, many of which require advanced mathematics skill. In 2011, 74% of STEM workers were male and 26% were female. About twice as many female science and engineering graduates are out of the labor force as are male graduates (20% as opposed to ten percent). Furthermore, since the 1990s, representation of women in computer science has declined (Landivar, 2013).

This data relates to findings of Congress listed in the proposed *Fulfilling the Potential of Women in Academic Science and Engineering Act of 2011* (Rep. Eddie Johnson [D-TX30], 2009-2011):

1. It is vital to the United States’ economic leadership and global competitiveness that the U.S. educate and train more scientists and engineers.
2. The National Academies believes that to remain a global leader in science and engineering, the U.S. must aggressively pursue the innovative capacity of women and men.

3. The number of women who are interested in STEM careers—science, technology, engineering and mathematics—declines at every educational transition, from high school through full professorship.

4. Women in the sciences and engineering remain underrepresented among university faculty and remain a small proportion of science and engineering faculty at major research universities.

5. No significant biological differences between men and women have been found to account for the lower representation of women in faculty and scientific leadership positions.

These findings emphasize the need in society for both women and men to pursue and persist in STEM careers. Therefore, “the question of gender differences in mathematics achievement, attitudes, and affect is a continuing concern as scientists seek to address the underrepresentation of women at the highest levels of science, technology, mathematics, and engineering” (Else-Quest, Hyde, & Linn, 2010, p. 103).

**Teacher characteristics.** Because teachers play a vital role in their students achievement, it is important to examine teacher characteristics that might contribute to student achievement (Toh, Ho, Riley, & Hoh, 2006). Over the last 15 years, research on teacher quality has moved to a focus of basing the measurement of teacher effectiveness on student achievement outcomes (Corcoran, 2010). Societal mindset changes and technology innovations have converged to
redefine teacher quality based on student outcomes. Corcoran (2010) identified some key factors that have contributed to this paradigm shift:

1. A political and philanthropic movement has shifted the notion of what teaching effectiveness is: Democrats and Republicans have endorsed associating teacher evaluation to student test scores and philanthropic foundations have funded efforts to such causes (e.g., the Bill & Melinda Gates Foundation, the Milken Family Foundation, and the Broad Foundation).

2. New high-stakes accountability policies have resulted in schools’ needs to measure and demonstrate results. Because of teachers’ close contact with students, high-stakes accountability has been extended to individual teachers.

3. Implementation of longitudinal data systems have facilitated matching students to their teachers and tracking student performance over time.

4. Data processing capacity and statistical modeling have advanced to permit isolation of teachers’ unique contributions to student outcomes.

Value-added analyses can identify teachers who consistently produce significant gains in students’ mathematics achievement. For example the National Mathematics Advisory Panel (2008) reported that “differences in teachers account for 12% to 14% of total variability in students’ mathematics achievement gains during an elementary school year” (National Mathematics Advisory Panel, 2008, p. 35). However, research is needed to identify and define factors that contribute to differences in teachers’ effectiveness (National Mathematics Advisory Panel, 2008). Traditional measures of teacher quality have been based on years of experience, professional certification, and degree attainment (Corcoran, 2010). The NCLBA reflects this traditional approach in its definition of “highly qualified” teachers as those having (a) a
bachelor’s degree, (b) full state certification or licensure, and (c) proof that they know each
subject they teach (U.S. Department of Education, 2004). It is challenging to identify teacher
characteristics that significantly contribute to student achievement; in his review of the literature,
Schacter (2001, p. 2) concluded that “teacher entrance and exit examination scores, years of
experience, advanced degrees and teacher credentials are either not related to student
achievement…or the relationship is small.”

However challenging it may be to identify teacher characteristics that affect student
achievement, it is important to continue research in this area for the sake teachers and of
students. This study examines two teacher characteristics that may affect students’ mathematics
performance: teacher years of experiences and professional certification level.

**Years of experience.** Teacher experience is a key factor in school district personnel
policies. It is a foundation of traditional single-salary schedules and it compels teacher transfer
policies (Rice, 2010). Rice pointed out the existence of the underlying assumption that, with
respect to teacher experience, more is better. However, the literature on the impact of teacher
experience on student achievement is mixed and does not necessarily support the assumption that
more experience is better overall.

Hanushek (1989) reviewed 140 studies and found that teacher experience correlated with
student achievement in 30% of those studies; however, he noted that these results appeared
strong only relative to the other school inputs. In addition, he cautioned that the positive
correlations may have resulted from senior teachers having selected schools and classrooms with
higher performing students. Likewise, a meta-analysis of 60 studies (Greenwald & Hedges,
1996) showed significance of teacher experience in relation to student achievement in 30% of the
studies.
Other studies showed no correlation of teacher experience with student achievement. Researchers who conducted a three-year study of students’ mathematics learning found no significant impact of teacher experience or teacher professional development on student mathematics achievement (Tarr et al., 2010). Huang and Moon (2009) used MLM to investigate the influence of teacher characteristics on achievement of 1,544 second grade students in high-poverty schools. They found that total years of teaching experience did not significantly predict student achievement.

Lusk (2011) analyzed Tennessee Value-Added Assessment System data for 200 students and 30 teachers from a three-year period to look for correlations between student gains on test scores and teacher characteristics. He found no significant correlation between teachers’ years of classroom experience and student achievement gains in mathematics.

Alban (2002) found little impact of teachers’ years of experience on student mathematics achievement. Two rounds of OLS and two MLM models showed no statistical significance related to teacher years of experience. However, teachers’ years of experience was significant and had a negative coefficient for one of the two large school systems when considered at the school-level of the 3-level MLM. Lazarus’ (2003) study also produced a negative correlation between teachers’ experience and students’ mathematics achievement. She examined data from all 331 Minnesota school districts with secondary students and found a statistically significant negative impact of teacher years of experience on student mathematics achievement: each additional year of teacher experience yielded a 0.46% decrease in the percentage of eighth-grade students who passed the Minnesota Basic Skills math test. She conjectured that older teachers may have less energy and enthusiasm and recently trained teachers may use more effective instructional techniques.
The literature supports the notion that teacher years of experience can impact student outcomes early in teachers’ careers. Teachers show the greatest gains in productivity during their first few years in the classroom; then their learning curves level off (Clotfelter, Ladd, & Vigdor, 2007; Goldhaber, 2007; Kane, Rockoff, & Staiger, 2007; Rice, 2010). For example, Kane et al. (2007) found that students of third-year teachers scored six percent of a standard deviation higher in mathematics than students of first-year teachers. In a study of New York City teacher qualifications, the largest gain in fourth- and fifth-graders’ mathematics achievement (.06 standard deviation) was due to the effect of teachers having progressed from total teaching inexperience to having taught one full year (Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2007).

**Professional certification.** Whereas the literature indicates that additional years of experience early in a teacher’s career can significantly impact students’ achievement, the effect of teacher professional certification on student achievement is unclear; research results are mixed. For example, Clotfelter, Ladd and Vigdor (2007), through analyses of 10 years’ worth of North Carolina state data in which they used regular teacher licensure as the base for their research, found negative effects of provisional or emergency teacher licensure on student achievement in mathematics and reading. Gimbert, Bol and Wallace (2007) found that students of alternatively certified teachers had a slightly higher overall mean score on the state-mandated Algebra I exam than those of traditionally certified teachers; however, the difference was statistically insignificant. A study of 87 alternatively certified teachers and 87 traditionally certified teachers from 63 schools in twenty districts and seven states found no significant difference in reading and math achievement between alternatively certified teachers’ students and traditionally certified teachers’ students (Constantine et al., 2009).
Goldhaber (2007) explored teacher licensure issues through examination of data that is maintained by the North Carolina Education Research Data Center. His study covered school years 1994 – 2004; the data included all teachers and students in North Carolina public schools during that period. Goldhaber found that teacher licensure testing was predictive of teacher effectiveness measured by student achievement, particularly in mathematics.

Other studies have found no effect for teacher certification status. For example, a study data from the Early Childhood Longitudinal Study—about 23,000 students in 1300 public and private schools—showed no significant difference in first-grade students’ math or reading achievement based on teacher certification status (Croninger, Rice, Rathbun, & Nishio, 2007). However, the researchers found that students whose teachers had greater levels of coursework in their teaching field (reading and mathematics) showed greater levels of achievement. Huang and Moon (2009), using a sample of 1,544 students of 154 teachers in 53 schools, found no statistical significance of the influence of teacher licensing status on gains in second-graders’ reading achievement.

New York City has long been challenged in their recruitment of certified teachers; for example, in the 2000 academic year, approximately 60% of newly hired teachers were uncertified (Kane et al., 2007). Kane, Rockoff and Staiger (2007) used New York City Department of Education data from academic years 1999 through 2005 to explore the relationship of teacher certification and student achievement for students in grades 4 – 8. Kane et al. (2007) concluded that “a teacher’s certification status matters little for student learning. We find no difference between teaching fellows and traditionally certified teachers or between uncertified and traditionally certified teachers in their impact on math achievement” (p. 65). A different study of New-York-city teacher qualifications indicated that having a non-certified
teacher reduced student achievement measured by achievement on statewide mathematics and English language arts exams by 0.042 standard deviation (Boyd et al., 2007).

The mixed results of research on the impact of teacher certification have spurred debate on the topic among educators. With respect to mathematics achievement in particular, the National Mathematics Advisory Panel (2008, p. xx-xxi) stated that

Overall, findings about the relationship between teacher certification (i.e., licensure) and student achievement in mathematics have been mixed, even among the most rigorous and highest-quality studies. Research in this area has not provided consistent or convincing evidence that students of teachers who are certified to teach mathematics gain more than those whose teachers are not. The relationship between teacher certification status, the most inexact proxy for teachers’ content knowledge, and students’ mathematics achievement remains ambiguous.

The Abell Foundation (2001) reviewed about 150 studies of the relationship between teacher preparation and student achievement. The studies had been conducted within a 50-year time span by well-known national advocates of teacher certification. The Foundation concluded not only that “the academic research attempting to link teacher certification with student achievement is astonishingly deficient,” but also that teacher certification “is neither an efficient nor an effective means by which to ensure a competent teaching force. Worse, it is often counterproductive” (The Abell Foundation, p. iii). The Foundation highlighted common errors and deficiencies in research that advocates teacher certification, including the following:

- a reliance on research that promotes the case for certification and a neglect in citing research that does not;
a padding of the analyses with multiple references that do not truly support certification;

citation of research that is irretrievable or too old to be reliable;

heavy use of research that is had not been subjected to peer review;

use of advocate-designed measures of student achievement rather than standardized measures;

violation of sound statistical analysis such as failure to control for key variables such as poverty and prior student achievement, use of inadequate sample sizes, and reliance on inappropriately aggregated data (The Abell Foundation, p. iv).

The foundation argued for the discontinuation of teacher certification in the state of Maryland.

In response to The Abell Foundation’s report, Darling-Hammond (2002) challenged The Abell Foundation’s claims, rebutted the foundation’s arguments and discussed policy issues related to providing highly qualified teachers to students. Darling-Hammond included the following issues with The Abell Foundation’s report:

lack of consideration of student learning evidence related to teacher professional knowledge;

lack of evidence for, and failure to consider contradictory evidence for, the relative effectiveness of certified and uncertified teachers and teacher education outcomes;

misrepresentation of research (e.g., the inability to discount important studies that support certification and the omission of relevant studies); and

methodological issues and double standards in using research (e.g., citation of studies that fail to meet research criteria explained in the report and elimination of studies that explore teacher effectiveness with measures other than student achievement).
Darling-Hammond (p. 42) argued that rather than eliminate teacher certification, “the most sensible policy goal is to work to improve preparation opportunities and certification standards so that they increasingly approximate what teachers need to know and do in order to be successful with diverse students.”

In educators' quest to be successful with diverse students, examination of teacher and student characteristics is important. So is the examination of interventions that are implemented in the educational system to improve student achievement—interventions such as the use of ongoing assessments during the school year for monitoring student performance. Tests used in this way have a variety of names including interim, periodic, diagnostic, predictive, and benchmark (Council of Chief State School Officers, 2008).

**Benchmarking systems.** A benchmarking system is series of common assessments aligned with performance standards that are administered at specified times in a curriculum sequence to provide data to educators regarding students’ attainment of the standards (CRESST, n.d.; Herman, Osmundson, & Dietel, 2010). The data can then be used to identify and facilitate student attainment of longer-term learning goals not yet mastered. As such, benchmarking systems can perform one the five main types of formative assessment identified by Black and Wiliam (2009): formative use of summative tests.

The National Mathematics Advisory Panel (2008) recommended regular use of formative assessment for elementary school students. As part of a balanced assessment system, benchmark assessments bridge a gap between formative classroom assessments and annual high-stakes assessment (Herman et al., 2010). “The imperfect alignment between assessment for accountability purposes and for formative feedback to teachers is giving way to a host of
formative and interim benchmark assessments that provide finer grained information to teachers” (Supovitz, 2008, p. 461). Benchmarking systems in education can provide:

- an affirmation of acceptable learning expectations and performance standards (Farmer & Taylor, 1997; Herman et al., 2010);
- input for planning curriculum and instruction (Brown & Coughlin, 2007; Herman et al., 2010);
- a baseline from which new methods or programs can be evaluated (Farmer & Taylor, 1997),
- a measure of the effectiveness of instruction and/or programs (Farmer & Taylor, 1997; Herman et al., 2010);
- data for predicting future performance (Brown & Coughlin, 2007; Herman et al., 2010); and
- criteria for policy and decision making at classroom, school and district levels (Herman et al., 2010).

Many districts throughout the U.S. have used benchmark assessments to try to raise achievement levels of students, schools and districts (Henderson, Petrosino, Guckenburg, & Hamilton, 2007). Results have varied. After two years of implementation of quarterly benchmark assessments that were aligned with state middle school mathematics standards, researchers reported no significant difference in student achievement between schools using benchmark assessments and those not using benchmarks (Henderson et al., 2007).

Sherman (2008) surveyed school superintendents in her comparison of local benchmark testing in Texas public schools. Eighty-eight percent of respondents reported that their districts administered benchmark assessments. In comparing district pass rates on the Texas Assessment of
Sherman found no significant difference.

On the other hand, Young (2004) compared best practices of five elementary schools in California that had an increase of 50 or more Academic Index points for two consecutive school years, 2000 and 2001. Four of the five schools used math benchmarks to provide key indicators of student academic progress.

4Sight is a quarterly benchmark system designed to provide formative evaluations of student progress and to predict student performance. Stoltz (2008) studied the impact of 4Sight on students’ performance on the Pennsylvania System of School Assessment (PSSA) fifth-grade reading exam. The 4Sight benchmark system was a significant predictor of PSSA scores, accounting for 60.5% variance in scores. A similar study indicated strong correlation between the 4Sight mathematics benchmark and the PSSA mathematics exam for students in grade three and students in grade five (Potteiger, 2008).

Trimester benchmark assessments coupled with essential standards assessments were components of a quality assurance system implemented in 2005 by Vernon Middle School. At the time, 14.8% of the school’s students were proficient in mathematics; the school ranked last in the California 34 school Ontario-Montclair School District. In this system, teachers received training on how to analyze their own students’ achievement data to identify strengths and weaknesses with respect to local and state standards. The results of the analyses guided instruction: topics were arranged on an instructional calendar and the teachers taught mini-lessons, as well as maintenance and reinforcement lessons, that coincided with the calendar. After each four-lesson unit, mini-assessments were administered; the results were analyzed and
students who were unsuccessful were scheduled for re-teach sessions. The principal reported a 29 point increase in the school’s Academic Performance Index (API) to 661 (Bettger, 2007).

Black and Wiliam (2010) did an extensive search of the formative assessment literature that was published from 1988 through 1997 and included books and more than 150 journals. Black and Wiliam concluded that there is evidence that “formative assessment is an essential component of classroom work and that its development can raise standards of achievement” (p. 90).

**Statistical Models**

The statistical models of interest in this study are multiple linear regression and multilevel models. Following are brief backgrounds of the two models, a discussion of theoretical matters related to modeling of nested data and a review of comparative analyses of statistical models.

**Multiple linear regression (MLR).** Cohen and Cohen (1983) describe multiple regression/correlation (MRC) as a highly general and flexible data-analytics system that is useful whenever a quantitative variable is to be examined as a function of any factors of interest. Linear regression beginnings include a study conducted in 1875 by Charles Darwin's cousin, Sir Francis Galton, in which he plotted the weights of daughter pea seeds against the weight of mother pea seeds. For a particular size of mother seed, he discovered that the median weights of the daughter seeds approximated a straight line with a positive slope less than 1 (Stanton, 2001).

MRC emerged in the behavioral and biological sciences around the turn of the twentieth century as researchers studied covariation of observed properties of subjects (Cohen & Cohen, 1983). The seminal first edition of the classic 1959 text by Ezekiel and Fox, *Methods of
Correlation and Regression Analysis, was published by Fox in 1930 (Bickel, 2007). Over time, the research emphasis moved from correlation to regression (Bickel, 2007).

A traditional application of regression analysis in educational research is the prediction of outcomes: the dependent (also called criterion, response or outcome) variables are predicted by the independent (also known as predictor or explanatory) variables. This application of MRC was typically restricted to models for which the relationship between the dependent variable and the independent variables can be expressed by a linear equation and for which the observations are normally distributed. This specialized model is now known as linear regression (one independent variable) or multiple linear regression (more than one independent variable).

The flexibility of multiple linear regression (MLR) becomes apparent when one looks at the potential variety of characteristics of the independent variables. The independent variables may be quantitative or qualitative; main effects, interactions or covariates; characterized by missing data; correlated or uncorrelated with one another; naturally occurring or outcomes of an experimental design; a single variable or groups of variables (Cohen & Cohen, 1983). This flexibility has likely contributed to the wide use of MLR in social science research, particularly in settings such as schools, where controlled experiments are not always possible. Many applications of MLR exist in the literature; some examples related to student achievement follow.

Bohn, Butts and Raun (1968) used MLR to study teacher characteristics as predictors of success (indexed by student achievement) in teaching science curriculum. They found that 45% of the variance in student achievement score was explained by the predictors: sex, grade level, school district, years of experience, and hours of science, with grade level and years of teaching experience contributing significantly to predicting teaching success (Bohn et al.).
Nash (2010) explored the relationship between transformational leadership dimensions and student performance in reading and mathematics with a multiple linear model. This model suggested that idealized behavior (or charisma), intellectual stimulation, and inspirational motivation of principals were predictors of third- and fifth-grade students’ achievement.

In another study, Du and Hu (2008) used MLR to model the relationship between the allocation of educational resources and student academic performance in rural areas of China. In one of the models developed in this study, the dependent variable was third-grade students’ mean math score. The independent variables—previous semester’s math final score, ratio of full-time minority teachers, ratio of full-time qualified teachers, student-teacher ratio, teachers’ years of experience, class size, mean years of mothers’ education, number of books per student, area of floor space per student, and public expenses per student—accounted for 54% of the variance in students’ mathematics performance. Of those variables, previous math performance, floor space and mother’s education contributed most to the explained variance in mathematics performance (Du & Hu, 2008).

Erbe (2000) used MLR to explore correlates of school achievement in Chicago Elementary Schools. While controlling for previous mathematics achievement, she found that three categories of school variables contributed significantly to the overall 78% explained variance in mathematics performance: (a) school climate (19% of the explained variance) – school commitment, focus on student learning, and inclusive leadership; (b) teacher beliefs (24%) – limits on students’ capability to learn and parent involvement, and (c) socioeconomic variables (57%) – low income and mobility.

The traditional use of MLR models in exploring the impact of student, teacher and school characteristics on student achievement has recently given way to new statistical models. The
development in recent years of more advanced statistical methods, such as hierarchical linear modeling, gives researchers access to more sophisticated tools for analysis of education support system data (Chaplin, 2003; Hox, 2002; Tabachnick & Fidell, 2007).

**Multilevel modeling (MLM).** Many data acquired from naturally occurring settings have a hierarchical or nested structure. In education settings, students are nested in classes, under teachers who are nested in schools, which in turn, are nested in counties, states and countries. Goldstein (1999) pointed out that “the existence of such data hierarchies is neither accidental nor ignorable” (p. 1). Once group structures are established (even if randomly established), “the group and its members both influence and are influenced by the group membership” (Goldstein, 1999, p. 1). Goldstein cautions that ignoring these relationships risks oversight of group effects.

MLM, also known as hierarchical linear modeling (HLM), was developed because of researchers’ concerns about conceptual and methodological issues when conducting research with a hierarchical (nested) data structure (Bryk & Raudenbush, 1992; Heck & Thomas, 2000). MLM is used specifically for analyzing data that is nested, for example, data that include measures from the individual level (e.g., student) and the aggregate level (e.g., school) (Astin & Denson, 2009; Bickel, 2007; Bryk & Raudenbush; Hox, 2002). Goldstein (1999) noted that researchers are interested in comparing educational institutions in terms of the achievement of their students, in part, for public accountability, but also to investigate the factors that explain differences among schools. Researchers also use MLM to explore the impact of classroom and teacher characteristics on student achievement. Other applications of MLM include (a) analysis of growth using longitudinal data in which repeated measurements are nested under individuals (Bryk & Raudenbush; de Leeuw & Meijer, 2008; Goldstein, 1999; Holt, 2008; Hox; Paterson & Goldstein, 1991); (b) analysis of cross-classified data structures; for example, studying student
achievement based on students’ cross-classification in two groups, such as neighborhood and school (Beretvas, 2004; Bickel; Goldstein; Hox; Paterson & Goldstein); and (c) meta-analysis of topically related studies in which the level-1 unit is subjects and the level-2 unit is studies (Bryk & Raudenbush; de Leeuw & Meijer; Goldstein; Hox; S. A. Kalaian & Kasim, 2008; H. A. Kalaian, 1994).

Bode (1996) used MLM to explore the impact of ability grouping on 3,991 eighth-grade students’ mathematics performance and found that within-class ability grouping had no effect on average mathematics achievement, nor did the grouping impact the relationship between prior and subsequent mathematics achievement.

Chouinard and Roy (2008) used MLM for a longitudinal study that examined 1,130 Canadian high-school students’ motivation in mathematics to investigate motivation changes over time with respect to student’s gender. They chose MLM over traditional models such as ANOVA (a) to avoid reduction of statistical power because of missing data (MLM procedures make generalizations from the data set without having to excluded a respondent for a missing data point); and (b) because MLM facilitates modeling the initial outcome and the degree of change over time for each participant. The results showed a regular decline of motivation in mathematics for both genders during high school (grades 7-11 in Quebec) which was heightened in grades 9 and 11, and that boys were more affected than girls.

The impact of school privatization and K-8 reforms on mathematics achievement growth at high-poverty middle schools in Philadelphia was the focus of a study by Mac Iver and Mac Iver (2009). They used a three-level MLM. Level 1 modeled student achievement. Level 2 accounted for students’ differences in prior mathematics achievement and modeled achievement differences during middle school with respect to experiences and characteristics that varied
among students within the same school. The third level measured interactive effects on mathematics achievement growth of school management and grade span interventions while controlling for differences in school poverty level, each school’s average achievement of its incoming sixth-grade cohort, and differences between schools that existed before assignment of schools to educational management organizations. Mac Iver and Mac Iver concluded that the expensive experiment of privatization of schools did not pay off in that the privatized schools did not produce better mathematics achievement gains than the district-managed schools.

Schreiber and Chambers (2003) analyzed TIMMS data using a two-level MLM to investigate student characteristics that impact mathematics literacy achievement and to determine whether differences exist among schools that contribute to differences in mathematics achievement. They found that the following student-level variables had a negative impact on mathematics literacy achievement: poor math attitude, parent’s lower formal education level, and greater after-school work hours. At the school level, the following variables were positively related to mean school mathematics literacy achievement: higher level of parental education, positive school climate, higher average of students perceiving the class as active, and a higher average of students believing that math success is based on hard work.

Multiple linear regression and multilevel models have been used frequently to analyze education data that is nested. MLM was designed specifically to analyze nested data sets and offers some advantages over MLR.

**Theoretical matters.** When the number of groups under consideration is small (less than five), MLM offers little advantage over MLR as there is not enough information to accurately discern group-level variation (Gelman & Hill, 2009). However, in education settings, MLM offers potential benefits in the data analysis process when compared to a standard linear
regression technique (Chaplin et al., 2009; Tabachnick & Fidell, 2007). In the last few decades, researchers have gravitated to using MLM to analyze data sets with large numbers of groups to take advantage of what they see as advantages of multilevel models over classic regression models.

For example, standard linear regression models assume independence of observations. The very nature of hierarchical data violates the assumption of independence of observations as there will be similarities of individuals based on belonging to the same group. In traditional regression models, violation of the independence of observations assumption can lead to excessive Type I errors and biased parameter estimates (Peugh, 2010). By design, MLM accounts for the individual and group relationships that are present in nested data (Heck & Thomas, 2000).

Another theoretical matter is the unit of analysis. Traditionally, MLR analyses of data from hierarchical systems—such as schools in which students are nested under teachers—were performed by aggregating or disaggregating all variables to a single level of interest (Bryk & Raudenbush, 1992; Heck & Thomas, 2000; Hox, 2002; Leeuw & Meijer, 2008; Pollack, 1998; Tabachnick & Fidell, 2007). For example, a researcher might choose to make students the unit of analysis by assigning teacher characteristic data (e.g., degree attained or number of hours of continuing education) to each student to perform the regression analysis on the student level. Because in effect, the teacher data is basically analyzed as a student-based variable (all students with the same teacher have the same value on teacher characteristics), the assumption of independence of observations in traditional MLR is violated (Bryk & Raudenbush, 1992; Heck & Thomas, 2000; Hox, 2002). This approach results in estimates of standard errors that are too small, giving spurious statistically significant results (Hox, 2002).
Conversely, a researcher might make teachers the unit of analysis by using the mean student achievement score by teacher to determine if groups of students that differ by their teachers’ characteristics (such as degree attained or number of continuing education units) have differing achievement means. This aggregation of data facilitates a teacher-level analysis, but failing to consider within-group variability can distort relationships between the teacher units (Heck & Thomas, 2000). Cress (2008) pointed out that in this case, analysis is conducted on a much smaller number of units and that data is wasted. Furthermore, the loss of individual student information due to aggregation leads to loss of power in the statistical analysis (Hox, 2002; Tabachnick & Fidell, 2007). Paterson and Goldstein (1991) related that in aggregation models, slight changes to the model or the data can produce unreliable statistical estimates. Paterson and Goldstein also pointed out that using aggregated data is not helpful for explaining variance at the individual student level; as they put it, “learning is done by children, not schools or LEAs, and we cannot offer explanations based upon aggregate level analyses that would make explanatory sense” (p. 3).

These two traditional regression approaches—data aggregation and data disaggregation—have drawbacks, and the two approaches may produce different results. MLM was designed to address interdependence of individuals within a group and to model individual-level variance as well as group-level variance (Bryk & Raudenbush, 1992). MLM can better accommodate the greater complexity of contemporary data analysis of hierarchical education settings, because MLM permits researchers to avoid aggregated and disaggregated data analysis. Considering the range of analysis units now made available to researchers by education data systems and online learning management systems (see Figure 2, from S. R. Díaz, personal communication, December 13, 2011), avoidance of data aggregation and disaggregation is an important
contribution of MLM to the research community. It behooves the research community to compare and contrast statistical models to gain further understanding of the implications the models’ differences.

**Figure 2.** Education units of analysis continuum

**Comparative studies.** Astin and Denson (2009) compared ordinary least squares (OLS) regression and MLM in multi-campus studies of students to assess the relative fit of the models and to determine the extent to which the two models might suggest different conclusions about the effects of institutional-level variables. The OLS and the MLMs used to explore students’ political identification yielded essentially the same results. Astin and Denson concluded that requiring authors to use MLM rather than OLS for nested data sets is not reasonable. However, they did encourage re-running the analysis using MLM when significant institutional-level effects are found using OLS or when a more conservative estimate of institution-level significance is desired.
Alban (2002) compared OLS models with a 2-level MLM and a 3-level MLM in a study designed to investigate the impact of student, teacher and school characteristics on students’ achievement in two large school systems in Maryland. Signs of the coefficients of student-level variables and the significance or non-significance of student-level variables were consistent across all models in all content areas that were investigated. Results varied for school-level variables. The significance of school-level and teacher-level variables was not consistent between the OLS models and the MLMs. Although all the OLS analyses showed statistical significance of school-level variables in one of the school districts, the 2-level MLM showed no statistical significance of school-level variables. The 3-level model results varied across content areas. In addition, OLS analysis and MLM analysis produced opposite signs for some of the school variables. For teacher-level variables, OLS techniques resulted in more findings of significant impact of the variables that did the MLM models.

Using student-level data and school-level data, Shafer, Yen, and Rahman (2000) compared agreement and stability of MLM, OLS and weighted least squares (WLS) models to evaluate school effects on 23,461 third- and 21,226 fifth-grade students’ achievement on the 1998 Maryland school Performance Assessment Program. Yen et al. studied agreement by comparing inter-correlations among the methods, using the average of 12 correlation coefficients (six at grade level three and six at grade level five). When using the same independent measures, the agreement between OLS and WLS was greater than that between OLS and MLM or WLS and MLM. When different independent measures were used, there was little difference among the correlations across method pairs. Yen et al. examined stability by (a) assessing randomly equivalent subgroups on the independent measures, and (b) looking at the consistency of school effects between grades three and five by finding a single school effect at each grade level. Both
explorations produced the same result: MLM was the most stable. The authors concluded that (a) based on stability, MLM should be used for school effects measures (especially for small schools); and (b) if only school-level data is available, OLS and WLS are viable methods.

In a study that assessed the effect of mathematics efficacy on mathematics achievement measured by the 2005 PISA, Kitsantas, Ware and Cheema (2010) compared three sets of statistical models that predicted mathematics achievement: (a) a multiple regression model with only student context variables as predictors; (b) an aggregate multiple regression model using school context predictor variables; and (c) a set of five MLMs. Regardless of the analysis method, mathematics self-efficacy was an important predictor of mathematics achievement. MLM models showed that the ratio of white to non-white students at school significantly influenced mathematics achievement of individual students; the MLR models did not suggest that the ratio impacted mathematics achievement. Kitsantas et al. concluded that the differences in results from the MRL and MLM approaches suggested that prior studies that used a single level analysis method should be interpreted with caution because they are very likely biased. They also cautioned that even though disaggregated multiple regression models sometimes produces results similar to MLM, researchers should not assume that disaggregated MLR is an adequate substitute for models that are designed for analysis of nested data; similarity of results is not guaranteed.

Kennedy, Teddlie and Stringfield (1993) used data from Phase II of the Louisiana School Effectiveness Study, “one of the more systematic and comprehensive attempts in America to explore the dynamics of school effects on student achievement outcomes” (p. 5) to provide empirical insight into how studies that have used older single-level regression strategies should be viewed in light of technological advances in statistical modeling. An MLM and traditional
student-level and school-level models of school effects on student achievement were compared to identify schools that appeared unusually effective or ineffective at nurturing student achievement. School rankings and correlates of achievement were quite consistent across methods; the results relating to SES were less consistent. The authors noted that school rankings for mathematics were in disharmony with other measures and contended that “mathematics achievement warrants special attention as distinct from the more verbal subject areas” (Kennedy et al., p. 24). They concluded that results from older statistical methods regarding the distribution of correlates of achievement “may wisely be viewed with some caution” (Kennedy et al., p. 24).

Rethinam, Pyke and Lynch (2008) compared MLM and ANOVA statistical techniques to investigate the effectiveness of science curriculum materials for sixth- and seventh-graders in a large, diverse, metropolitan U.S. school district. Similar findings related to the overall effectiveness of each curriculum unit resulted with the two models. However, the MLM model revealed larger effect sizes, significant variability at the classroom level, and significant individual and classroom variables. Rethinam et al. concluded, “when the students are nested in classrooms and schools, researchers should consider multi-level analysis and account for classroom/school contexts” (p. 31).

Germuth (2003) assessed teacher effectiveness using OLS and MLM models. She found that teacher and school characteristics had very little to do with the differences in students’ end-of-grade mathematics scores as most of the variance in student performance was explained by measures of past achievement. She concluded that the MLM models used in her study offered few advantages over the OLS models, and that for most school systems, use of OLS models can help distinguish among effective and ineffective teachers.
Comparative studies of methods for analyzing nested data in education settings may show that the results from traditional regression models and multilevel models do not differ drastically. Such studies often show small coefficient differences but typically larger standard error estimates in the multilevel models, leading to fewer statistically significant results with MLMs. However, in light of the relationships among education, technology and society, “we should ask ourselves just how important it is to gain a bit of precision in coefficient estimation, to avoid deflated standard errors, and to avoid misleading results from tests of significance” (Bickel, 2007, p. 12). “The growing use of MLM suggests that researchers are becoming increasingly aware of and willing to deal with the important issue of clustering data within groups. This implies that the conclusions reached in these studies can be taken more seriously than those of many studies of the past that ignore clustering as the estimated standard errors are less likely to be biased” (Chaplin, 2003, p. 12).
Chapter 3 – Methodology

With a goal of exploring connections among society, education and technology, the focus of the study was to investigate how different methods of analyzing data made available through contemporary information systems potentially impact mathematics education. The approach was to estimate models in which a student’s mathematics achievement in one year is predicted as a function of the student’s gender, his/her performance on district-wide mathematics benchmark exams, his/her previous mathematics achievement, and his/her teacher’s professional certification level and number of years of teaching experience. Specifically, the study compared two predictive models, a multiple linear regression (MLR) model and a series of multilevel models (MLMs), to determine the effects of student characteristics and teacher characteristics on mathematics achievement of 4th- through 8th-grade students in academic years 2007, 2008 and 2009 as measured by the Maryland School Assessment (MSA) mathematics exam.

Research Questions

Three research questions (RQs) were investigated.

RQ1: How does multiple linear regression inform the prediction of mathematics achievement from a linear combination of student characteristics—gender, previous mathematics achievement, district mathematics benchmark exam performance—and teacher characteristics—professional certification level and years of experience?

RQ2: How does hierarchical linear modeling, with students and teachers as the respective level-1 and level-2 sampling units, inform the prediction of mathematics achievement from linear combinations of the same student characteristics—gender, previous mathematics achievement, district benchmark exam performance—and the same teacher characteristics—professional certification level and years of experience?
RQ3: How do MLR and MLM results compare for informing K-12 practice and policy with respect to student mathematics achievement?

Data Acquisition and Preparation

The study used pre-existing data that were routinely collected and stored by the school district. Before releasing the data for use in this research, the school district assigned randomly generated identification codes (IDs) to all students, teachers and schools represented in the data, thereby de-identifying schools, teachers and students.

The student data obtained from the school district were contained in electronic files by year: 2006, 2007, 2008 and 2009. The data was in a univariate format; for each given year, each student had a record for every mathematics benchmark exam that s/he completed in that year – as many as 29 in 2007, 9 in 2008, and 8 in 2009. The relevant student data consisted of 63,730 records representing 4,757 students in grades 4 through 8 who were taught by 129 teachers. Each record in the data set included a student ID, the benchmark exam ID, the ID of the teacher under which the student took the benchmark exam, as well as the student’s gender, grade, special education status, raw benchmark exam score, and raw MSA mathematics exam score (from March of the given academic year). The original data files represented 4,757 unique students in grades four through eight who were taught by 129 unique teachers during the 2007, 2008 and 2009 academic years. The 2006 data were requested only for the purpose of pulling 2006 math MSA scores into subsequent years’ records as a measure of previous mathematics achievement.

The teacher data obtained from the school district were contained in one electronic file. Each teacher’s record in the file included the teacher’s ID, his/her professional certification level, and his/her number of years of teaching experience for each year included in the study.
Student records were aggregated to include, at most, one record per year per student for the 2007, 2008, and 2009 academic years. Working with each year’s student data, the records were formatted to include a field for the academic year, the student’s ID, his/her teacher’s ID, and a field for each variable used in the study: the MSA mathematics score, student gender, mean of the student’s standardized benchmark exam scores for each academic year, the student’s prior standardized mathematics MSA score, the teacher’s years of experience and level of professional certification, mean of teacher years of experience (for L2 of the MLM) and maximum teacher certification level attained (for L2 of the MLM). Data preparation continued with calculation, recoding, inspection and cleaning of the data as follows:

1. For each case, the math benchmark exam score was standardized. Then for each student, benchmark exam z-scores for a given year were aggregated, and the mean of the student’s benchmark z-scores was calculated. Referencing the student ID in each case (in each year’s data), the mean of the student’s benchmark z-scores for the given year was added to each case.

2. For each case, math MSA scores were standardized. Then, referencing student ID in each case, the students’ prior mathematics MSA z-score was added to each case.

3. Teacher data—years of experience and professional certification level—were added to the appropriate fields in each record. In addition, the mean of each teacher’s years of experience from 2007 through 2009 and the maximum level of the teacher’s professional certification between 2007 and 2009 were determined and added to each record for use in the MLM. Teacher certification level was recoded from an alpha format to a numeric format, with standard professional certification equal to 0 and advanced professional certification equal to 1.
4. Student gender was recoded from an alpha format to a numeric format, with female equal to 0 and male equal to 1.

5. Records with no mathematics MSA score were deleted, as were records with invalid MSA scores (those that were less than 240, the minimum attainable score).

6. Since students had a record in each year’s data for every benchmark exam they took (up to 29), the columns containing the benchmark exam IDs and the benchmark exam scores (raw and standardized) were deleted, leaving the column containing the mean of a student’s benchmark exams for a given year. Duplicate records were then identified and discarded.

7. Some students still had more than one record per year in the data set due to the student having changed teachers during a year or the student’s gender having been miscoded in one or more records in a given year. If a student changed teachers in an academic year, only the record that was associated with the teacher under which the student took the greatest number of benchmark exams was retained in the data set. If a student took an equal number of exams under each teacher s/he had in a year, all of the student’s records in the given year were deleted. For coding inconsistencies in the student gender field, a record with the correct gender was retained when gender could be determined from the original data; otherwise, all the student’s records in the given year were deleted.

8. Because data obtained from the school district did not include information on whether special education students took the standard version or the alternate version of the mathematics MSA exam, cases of students who were designated as special education students were removed from the data set.
9. The prepared data for each year were merged into one file.

The data cleaning and aggregation process yielded a file of 5,224 students’ cases with complete data representing 3,514 students (74% of those represented in the original data) taught by 99 teachers (77% of those represented in the original data set) in 2007, 2008 and 2009. The data was imported into PASW Statistics 18 for development of the statistical models used in the study.

**Population and Sample**

The population for the study consisted of elementary and middle-school students enrolled in the Allegany County Public School district in western Maryland. Total annual fourth-grade and fifth-grade enrollment during the study’s time frame ranged from 1,327 students to 1,357 students; annual enrollment in grades six through eight ranged from 1,994 students to 2,078 students (Maryland State Department of Education, 2013). During the time frame of the study, the attendance rate at each elementary and middle school was at least 94% each year. In the elementary schools, the percentage of students who were absent less than five days ranged from 32% to 37% in any given year; the percentage of students who were absent more than 20 days was less than 6% in any given year. In the middle schools, the percentage of students who were absent less than five days ranged from 27% to 32% in any given year; the percentage of students who were absent more than 20 days ranged from 9% to 12% in any given year (Maryland State Department of Education). The percentage of students who received free/reduced meals ranged from 28% to 81% among the elementary schools and from 34% to 54% among the middle schools (S. Milburn, personal communication, November 11, 2009).

The annual percentage of elementary-school teachers holding an advanced professional certificate ranged from 70% to 74% during the study’s time frame; the annual percentage of
those holding a standard professional certificate ranged from 23% to 32%. Approximately 80% of the district’s middle-school teachers held an advanced professional certificate during the study’s time frame; the annual percentage of middle-school teachers holding a standard professional certificate ranged from 15% to 18% (Maryland State Department of Education, 2013).

The sample for the study consisted of 5,224 records of 3,514 students in grades 4 through 8 who were taught by 99 teachers in Allegany County, Maryland during the 2007, 2008 and 2009 academic years. Students who met the following criteria constituted the sample:

1. The student was enrolled in grade 4, 5, 6, 7, or 8 in an Allegany County, Maryland public school during the 2007, 2008, or 2009 school year.
2. The student completed at least one district mathematics benchmark exam and the MSA mathematics exam in 2007, 2008 or 2009.
3. The student did not receive special education services. These records were removed from the data set to avoid confounding the analyses, as the students’ records did not include information on whether the student had taken the alternate version of the MSA.
4. The student’s record had no missing data for the variables used in the study.

Variables

Math performance, as measured by the math MSA was the criterion (outcome, or dependent) variable for the MLR model and the MLMs. Variations of five predictor variables were used in the models. Table 5 provides details regarding the statistical functions of the variables.
Variables and Their Statistical Functions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Function</th>
<th>Format/Coding</th>
<th>MLM Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math MSA score for a given year</td>
<td>Criterion</td>
<td>Raw score</td>
<td>Criterion</td>
</tr>
<tr>
<td>Math MSA score: prior academic year</td>
<td>Predictor</td>
<td>Standardized raw score</td>
<td>L1: student level</td>
</tr>
<tr>
<td>Student’s gender</td>
<td>Predictor</td>
<td>0 = female; 1 = male</td>
<td>L1: student level</td>
</tr>
<tr>
<td>Mean of student’s standardized benchmark exams in a given year</td>
<td>Predictor</td>
<td>Mean of standardized raw scores for given year</td>
<td>L1: student level</td>
</tr>
<tr>
<td>Teacher’s years of experience</td>
<td>Predictor</td>
<td>MLR: number of years of teaching experience; MLM: mean number of years of teaching experience from 2007-2009</td>
<td>L2: teacher level</td>
</tr>
<tr>
<td>Teacher’s professional certification level</td>
<td>Predictor</td>
<td>0 = standard; 1 = advanced; MLM: maximum certification attained from 2007-2009</td>
<td>L2: teacher level</td>
</tr>
</tbody>
</table>

Criterion variable. The criterion variable was academic performance on the Maryland School Assessment (MSA) exam in school years 2006-7 through 2008-9. The mathematics MSA exam is a state-mandated test that meets the testing requirements of the *NCLBA*. The MSA assesses student achievement with respect to the state mathematics curriculum standards. Students in grades 3 through 8 take the exam annually in March. The mathematics MSA is completed in two days, with approximately ninety minutes allotted for the test each day. Basic and higher level skills are tested with multiple-choice questions and questions that require written responses (Maryland State Department of Education, n.d.b).
**Predictor variables.** Five predictor variables were used in the study:

1. Student gender was included because of the disparity in the number of men and the number of women in mathematics, science, engineering and technology careers.

2. Student performance on mathematics benchmark exams was included to explore the impact on math achievement of the school district’s implementation of regular benchmark testing that aligns with state mathematics curriculum standards. From year to year, the set of benchmark exams administered differed in number; therefore, the means of students’ standardized benchmark exam scores for an academic year were used in the analyses.

3. Prior mathematics achievement, measured by a student’s previous MSA mathematics z-score, was selected (a) to account for varying mathematics skill levels of students as they enter a new grade, and (b) because of the correlation between prior mathematics knowledge and current mathematic performance that is noted in the literature.

4. Teacher professional certification level was included to explore the relationship between a teacher’s post-baccalaureate education and student achievement. Two levels of Maryland teacher certification were included in the study – standard professional certification and advanced professional certification. Minimally, for standard professional certification, a teacher must meet all state certification requirements and be employed by a Maryland local school system or an accredited nonpublic school. For advanced professional certification, a teacher must have three years of full-time professional school-related experience, six semester hours of acceptable credit, and a master’s degree or a minimum of 36 semester hours of post baccalaureate course work which must include at least 21 semester hours of graduate
credit. Teacher certification is valid for five years in the state of Maryland; permanent certification is not granted (Maryland State Department of Education, n.d.a).

5. Teachers’ years of experience was included to investigate the relationship between a teacher’s years of experience and student achievement. Even though school districts often default to using seniority as an indicator of effectiveness and as a criterion for most high-stakes decisions (Weisberg et al., 2009), the literature does not necessarily support the assumption that more experience contributes significantly to student achievement.

Models

Two commonly used statistical approaches to building predictive models were used to analyze student and teacher effects on mathematics achievement: multiple linear regression and multilevel modeling.

Multiple linear regression model (MLR). To address RQ1, a multiple linear regression between MSA mathematics score as the criterion variable and student and teacher characteristics as the predictor variables was conducted with Predictive Analytics SoftWare (PASW), version 18, to determine significant contributions ($p < 0.05$) to student mathematics achievement of the student’s gender, the student’s prior math MSA score, the mean of the student’s standardized benchmark scores for the given year, the teacher’s professional certificate level for the given year, and the teacher’s years of experience for the given year. The PASW ENTER method was used, and given the large sample size, only cases with complete data were used.

MLR equation. The model disaggregated teacher data to analyze math achievement at the student level. The model is represented by the equation:
\[(mMSA)_i = b_0 + b_1(stuGend)_i + b_2(BMzMean)_i + b_3(PmMSAz)_i + b_4(tCert)_i + b_5(tExp)_i + r_i, \] where

\(i\) represents the case number, \(i \in [1, 2, \ldots, 5224]\);

\((MSA)_i\) is the predicted MSA score for the student represented by case \(i\);

\(b_k (k = 0, 1\ldots 5)\) are the regression coefficients;

\((stuGend)_i\) signifies the gender of the student represented by case \(i\);

\((PmMSAz)_i\) denotes the prior math MSA score of the student represented by case \(i\);

\((BMzMean)_i\) is the mean of the mathematics benchmark exam scores for a given year for the student represented by case \(i\);

\((tExp)_i\) signifies, for the student represented in case \(i\), the number of years of professional experience of the student’s teacher;

\((tCert)_i\) signifies, for the student represented in case \(i\), the professional certification level of the student’s teacher; and

\(r_i\) signifies the residual error for case \(i\).

**MLR Null and Alternate Hypotheses.** The null and alternate hypotheses for the MLR model were:

\[H_0: R = 0,\] or, the multiple correlation coefficient, \(R\), is zero; i.e., there is no significant relationship between MSA score and a linear combination of the predictor variables; and

\[H_a: R \neq 0,\] i.e., a significant relationship exists between MSA score and a linear combination of the predictor variables.
Rejection of the null hypothesis would warrant further exploration as to the significance of the individual predictors in the model. The null and alternate hypotheses for determining significance of the individual predictor variables were:

\[ H_{0k}: \beta_k = 0, \text{ with } k = 1, 2 \ldots 5; \text{ or, the individual regression coefficients in the population equal zero; and} \]

\[ H_{ak}: \beta_k \neq 0; \text{ i.e., one or more of the predictor variables significantly contributed to the regression when used in combination with the other predictors.} \]

**Multilevel Models (MLMs).** To address RQ2, PASW (version 18) was used to analyze a progression of increasingly complex two-level linear mixed models in which student cases were nested under teachers. Of interest in the MLM was the relationship between a student’s math achievement and a linear combination of student and teacher characteristics that accounts for the hierarchical structure of the educational setting.

The development of the MLM model progressed with specification of the following models: (a) Model 0, the null model, which had no predictors; (b) Model 1, a student-level random intercept model that added three L1 predictors to the model; (c) Model 2, a teacher-level random intercept model that added four L2 predictors: two teacher-characteristic predictors and two aggregate student characteristic predictors; and Model 3, a random slope and intercept model in which the slopes of the L1 predictors were allowed to vary across teachers. Restricted maximum likelihood estimation was used in each MLM.

The data set of 5,224 cases representing 3,514 students who were taught by 99 teachers that was used in the MLR model was aggregated in the following manner for use in MLM modeling: (a) the mean of each teacher’s years of experience from 2007-2008 was used in place of teacher experience data by year, (b) the highest level of professional certification attained by
each teacher during the 2007-2009 time frame was used in place of teacher certification data by year, and (c) student characteristics data were aggregated by teacher for use in the L2 analyses. The sample size was adequate for MLM, exceeding the recommended 30 teachers who taught a minimum of 30 students (Kreft, 1996, as cited in Hox, 2002). Because non-categorical student-level predictors were standardized, centering was not used in the MLMs. The MLMs violate the assumption of independence of observations because one teacher might contribute to multiple student data.

**MLM Null and Alternate Hypotheses.** Of key interest in the MLM is the impact of the hierarchical structure of the educational setting (students nested under teachers). Therefore, the model hypotheses focus on variability at the teacher level:

- \( H_0 \): the means of students’ math MSA scores by teacher (the intercepts) are equal.
- \( H_a \): there is significant variability in the means of students’ math MSA scores (intercepts) across teachers.

**Model 0: Null Model.** A linear mixed model with no predictors was run to assess how the variation in math achievement fell across the levels of the model. The null model for \( i \) student cases nested under \( j \) teachers is represented by the equation:

\[
(m_{MSA})_{ij} = \gamma_{00} + u_{0j} + r_{ij}, \text{ where}
\]

\( (m_{MSA})_{ij} \) is the predicted math MSA score of student-case \( i \) of teacher \( j \);
\( \gamma_{00} \) is the teacher-level intercept, or mean math MSA value across all teachers;
\( u_{0j} \) is the deviation from the math MSA grand mean of teacher \( j \)’s student cases; i.e., the unique effect of teacher \( j \) on the intercept; and
\( r_{ij} \) is student-case \( i \)’s deviation from the math MSA mean of teacher \( j \)’s student cases.
Specifically, the model was used to partition the variance in math achievement into two portions: $r_{ij}$, the variance in individual students’ math MSA scores for students who had the same teacher (will be referred to within-teacher variance throughout this document) and $u_{0j}$, the variance in the mean math MSA score of students grouped by teachers (will be referred to between-teacher variance). The model was also used to determine how much of the variance in math achievement lay between teachers.

**Model 1: Student-Level (L1) random intercept model.** In Model 1, fixed student-level predictors—student gender, mean of benchmark exam z-scores, and prior math MSA z-score—were added to the model to predict students’ mathematics achievement and to examine variability in the mean of students’ math MSA scores (intercepts) across teachers. The student-level random intercept model is represented by the equation:

\[
(m\text{MSA})_{ij} = \gamma_{00} + \gamma_{10}(stuGend)_{ij} + \gamma_{20}(BMz\text{Mean})_{ij} + \\
\gamma_{30}(Pm\text{MSA}z)_{ij} + u_{0j} + r_{ij}, \text{ where}
\]

\[(m\text{MSA})_{ij} \] is the predicted math MSA score of student-case i of teacher j;

\[(stuGend)_{ij} \] represents the gender of student-case i of teacher j;

\[(BMz\text{Mean})_{ij} \] is the mean of the benchmark exam z-scores of student-case i of teacher j;

\[(Pm\text{MSA}z)_{ij} \] is the prior math MSA z-score of student-case i of teacher j;

\[\gamma_{00} \] is the teacher-level intercept, or math MSA grand mean;

\[\gamma_{k0}, \text{ with } k = 1, 2, 3, \] is the slope associated with the accompanying L1 predictor; i.e. the effect of the accompanying predictor. E.g., \[\gamma_{10} \] is the student-gender slope, or, the L1 effect of student gender;

\[u_{0j} \] indicates deviation from the math MSA grand mean of teacher j’s student cases; and

\[r_{ij} \]
\( r_{ij} \) represents student \( i \)'s deviation from the math MSA mean of teacher \( j \)'s student cases.

**Model 2: Teacher-level (L2) random intercept model.** The level-two random intercept model introduced teacher-level predictors to the model to further explain variability in mean math MSA scores (intercepts) across teachers. The model included, for each teacher, the mean of his/her years of experience from 2007 through 2009 and the maximum level of certification held during the 2007-2009 time frame. L1 predictor values were aggregated by teacher and added to the model to account for the impact of student characteristics at the teacher level. The L2 random intercept model is represented by the equation:

\[
(m_{MSA})_{ij} = \gamma_0 + \gamma_{10} (stuGend)_{ij} + \gamma_{20} (BMzMean)_{ij} + \\
\gamma_{30} (PmMSAz)_{ij} + \gamma_{01} (max tCert)_j + \\
\gamma_{02} (tExpMean)_j + \gamma_{03} (BMzMean\_mean)_j + \\
\gamma_{04} (PmMSAz\_mean)_j + u_{0j} + r_{ij}, \text{ where}
\]

\( (m_{MSA})_{ij} \) is the predicted math MSA score of student-case \( i \) of teacher \( j \);

\( (stuGend)_{ij} \) represents the gender of student-case \( i \) of teacher \( j \);

\( (BMzMean)_{ij} \) is the mean of the benchmark exam z-scores of student-case \( i \) of teacher \( j \);

\( (PmMSAz)_{ij} \) is the prior math MSA z-score of student-case \( i \) of teacher \( j \);

\( (tExpMean)_j \) is the maximum certification level attained by teacher \( j \) during the 2007-2009 time frame;

\( (PmMSAz\_mean)_j \) is the mean of the prior math MSA z-scores of teacher \( j \)'s student cases;
$\bar{z}_{ij}$ is the mean, for teacher $j$, of his/her student-cases’ benchmark $z$-score average;

$\gamma_{00}$ is the teacher-level intercept, or math MSA grand mean;

$\gamma_{kn}$ (with $k = 1, 2, 3$) is the slope associated with the accompanying L1 predictor;

$\gamma_{0l}$ (with $l = 1, 2, 3, 4, 5$) is the slope associated with the accompanying L2 predictor;

e.g., $\gamma_{01}$ is the teacher certification slope, i.e., the effect of teacher certification;

$u_{0j}$ indicates deviation from the math MSA grand mean of teacher $j$’s student cases; and

$r_{ij}$ represents student $i$’s deviation from the math MSA mean of teacher $j$’s student cases.

**Model 3: Random Slope and Intercept Model.** A random slope and random intercept model was developed to examine whether slopes varied randomly across teachers, and if so, to determine whether relationships exist between teacher-level parameters and slope strength. The random slope and intercept model is represented by the equation:

$$(m_{MSA})_{ij} = \gamma_{00} + \gamma_{10} (stuGend)_{ij} + \gamma_{20} (BMzMean)_{ij} +$$

$$\gamma_{30} (PmMSAz)_{ij} + \gamma_{01} (max tCert)_{j} +$$

$$\gamma_{02} (tExpMean)_{j} + \gamma_{03} (PmMSAz\_mean)_{j} +$$

$$\gamma_{04} (BMzMean\_mean)_{j} + u_{0j} + u_{1j} (BMzMean)_{ij} +$$

$$u_{2j} (PmMSAz)_{ij} + r_{ij}, \ \text{where}$$

$(m_{MSA})_{ij}$ is the predicted math MSA score of student-case $i$ of teacher $j$;

$(stuGend)_{ij}$ represents the gender of student-case $i$ of teacher $j$;

$(BMzMean)_{ij}$ is the mean of the benchmark exam $z$-scores of student-case $i$ of teacher $j$;

$(PmMSAz)_{ij}$ is the prior math MSA $z$-score of student-case $i$ of teacher $j$;
(maxCert)\(_j\) is the maximum certification level attained by teacher \(j\) during the 2007-2009 time frame;

\((tExpMean)\)\(_j\) is teacher \(j\)’s mean years of experience from 2007 through 2009;

\((PmMSA_{\text{z-mean}})\)\(_j\) is the mean of the prior math MSA z-scores of teacher \(j\)’s student cases;

\((BM_{\text{z-mean}})\)\(_j\) is the mean, for teacher \(j\), of his/her student-cases’ benchmark z-score average;

\(\gamma_{00}\) is the teacher-level intercept, or math MSA grand mean;

\(\gamma_{k0}\), with \(k = 1, 2, 3\), is the slope associated with the accompanying L1 predictor;

\(\gamma_{0l}\), with \(l = 1, 2, 3, 4, 5\), is the slope associated with the accompanying L2 predictor;

\(\gamma_{01}\) is the teacher certification slope, i.e., the effect of teacher certification;

\(u_{0j}\) indicates deviation from the math MSA grand mean of teacher \(j\)’s student cases;

\(u_{kj}\), with \(k = 1, 2, 3\), indicates, for L1 predictor \(k\), teacher \(j\) slope deviation from the overall slope; e.g., \(u_{3j}\) indicates the deviation in the prior math achievement slope of teacher \(j\) from the overall prior math achievement slope; and

\(t_{ij}\) represents student \(i\)’s deviation from the math MSA mean of teacher \(j\)’s student cases.

In summary, two predictive modeling approaches were utilized to explore the impact of student characteristics and teacher characteristics on mathematics achievement of 3,513 fourth-through eighth-grade students who were taught by 99 teachers in a small school district in Appalachia. The first approach used a multiple linear regression model with individual student cases as the level of analysis. The second approach used a succession of 2-level multilevel models in which individual student cases at level 1 were nested under teachers as the level 2 unit.
Limitations of the Study

The MLR and MLM models are specific to a restricted population of elementary and middle-school students in one Appalachian school district, and are not necessarily expected to serve as a definitive predictive model for other populations. What can be generalized beyond this particular sample are the implications for conducting MLR versus MLM in educational research and practice.

The MLR and MLM regression models are designed to demonstrate relationship, not causality. As with any regression model, these models are extremely sensitive to the predictor variables that are included in the models (Tabachnick & Fidell, 2007). It is possible that the inclusion of more or different student and/or teacher characteristics and/or school characteristics as predictor variables would change the models significantly. For example, these models do not take into account school characteristics nor do they account for other variables—such as those related to students’ home lives, their socioeconomic status, their parents’ education level, their special education status, student mobility, the method of student placement in classrooms, etc.—some of which may affect student achievement outcomes. Therefore, the models should not be viewed as having addressed all the relevant factors comprehensively.

The sampling process limits generalization of the study with respect to students who receive special education services, as special education students were removed from the sample. Although an attempt was made to retain records in the data set for students who changed schools and/or teachers, generalization of the results with respect to students with high mobility may be limited, as some records for students who had more than one mathematics teacher in a given year were not included in the sample. The sampling process also violates the assumption of independence of samples as each student potentially counts once per year.
In any given year, and from year to year, the set of benchmark exams administered differed in number, and therefore in structure. Regardless, the combined set of benchmark exams were designed to measure math achievement overall for a given year. For this reason, this study examines students’ mean benchmark scores per given year as an independent variable of choice.

The study methodology and limitations have been detailed in Chapter 3. Chapter 4 presents the results of the MLR and the MLM models; and Chapter 5 addresses the research questions associated with the models.
Chapter 4 - Results

Research Questions

Two statistical approaches for building predictive models were used to address the following research questions:

RQ1: How does multiple linear regression inform the prediction of mathematics achievement from a linear combination of student characteristics—gender, previous mathematics achievement, district benchmark exam performance—and teacher characteristics—professional certification level and years of experience?

RQ2: How does multilevel modeling, with students and teachers as the respective level-1 and level-2 sampling units, inform the prediction of mathematics achievement from linear combinations of the same student characteristics—gender, previous mathematics achievement, district benchmark exam performance—and the same teacher characteristics—professional certification level and years of experience?

RQ3: How do MLR and MLM results compare for informing K-12 practice and policy with respect to student mathematics achievement?

Chapter 4 presents the predictive modeling results. Chapter 5 compares the two statistical approaches and discusses implications.

Multiple Linear Regression Model

To address RQ1, PASW (Predictive Analytics SoftWare, version 18) was used to conduct a multiple linear regression between MSA mathematics score as the criterion variable and student and teacher characteristics as the predictor variables to determine significant contributions ($p < 0.05$) to student mathematics achievement of the student’s gender, the student’s prior math MSA score, the mean of the student’s standardized benchmark scores for
the given year, the teacher’s professional certificate level for the given year, and the teacher’s years of experience for the given year. Table 6 displays the descriptive statistics for the variables incorporated in the MLR model. Analysis of variance results are shown in Table 7 and Table 8 presents the model summary.

**Null and alternate hypotheses.** The hypotheses for the MLR model were:

\[ H_0: R = 0 \], or, the multiple correlation coefficient, \( R \), is zero; i.e., in the population there is no significant relationship between math MSA score and a linear combination of the predictor variables; and

\[ H_a: R \neq 0 \], i.e., a significant relationship exists in the population between MSA score and a linear combination of the predictor variables.

Table 6

**MLR Model Descriptive Statistics**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math MSA score</td>
<td>426.15</td>
<td>35.80</td>
<td>.32</td>
<td>.10</td>
<td>5224</td>
</tr>
<tr>
<td>Student gender</td>
<td>.50</td>
<td>.50</td>
<td>.02</td>
<td>-2.00</td>
<td>5224</td>
</tr>
<tr>
<td>Mean of benchmark ( z )-scores</td>
<td>.10</td>
<td>.79</td>
<td>-.77</td>
<td>.23</td>
<td>5224</td>
</tr>
<tr>
<td>Prior math MSA ( z )-score</td>
<td>.09</td>
<td>.89</td>
<td>.13</td>
<td>.15</td>
<td>5224</td>
</tr>
<tr>
<td>Teacher certification level</td>
<td>.73</td>
<td>.45</td>
<td>-1.01</td>
<td>-.97</td>
<td>5224</td>
</tr>
<tr>
<td>Teacher years of experience</td>
<td>14.22</td>
<td>10.36</td>
<td>.72</td>
<td>-.62</td>
<td>5224</td>
</tr>
</tbody>
</table>
Table 7

*Analysis of Variance*

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>469,462,051</td>
<td>5</td>
<td>93,8924.103</td>
<td>2452.845</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>199,739,762</td>
<td>218</td>
<td>382.790</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>669,201,814</td>
<td>223</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a Predictors: (Constant), tExp, BMzMean, stuGend, tCert, PmMSAz

*b Dependent Variable: mMSA

Table 8

*MLR Model Summary*

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>Std. Error of the Estimate</th>
<th>R² Change</th>
<th>F Change</th>
<th>df1</th>
<th>df2</th>
<th>Sig. F Change</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.838</td>
<td>.702</td>
<td>.701</td>
<td>19.565</td>
<td>.702</td>
<td>2452.845</td>
<td>5</td>
<td>5218</td>
<td>.000</td>
<td>1.398</td>
</tr>
</tbody>
</table>

*a Predictors: (Constant), stuGend = student gender; BMzMean = mean of student’s benchmark z-scores; PmMSAz = student’s prior math MSA z-score; tCert = teacher certification level; tExp = teacher years of experience.

*b Dependent Variable: mMSA = math MSA score.

**Model fit and parameters.** The null hypothesis, $H_0$, was rejected as $R = .838$ was significantly different from zero, $F(5, 5218) = 2452.845, p < .001$. The adjusted $R^2$ value of .701 indicates that more than 70% of the variability in math MSA scores is predicted by a linear combination of student gender, student performance on math benchmark exams and on a prior math MSA exam, teacher years of experience and teacher certification level.
Pearson coefficients were calculated for each pair of variables (see Table 9). Each predictor had a significant, positive correlation with the criterion variable, math MSA score. Of all the predictors, performance on the benchmark exams ($r = .782, p < .001$) and prior math achievement ($r = .769, p < .001$) correlated most highly with student math achievement.

Table 9

*Pearson Correlation Matrix*

<table>
<thead>
<tr>
<th>Pearson Correlation</th>
<th>mMSA</th>
<th>stuGend</th>
<th>BMzMean</th>
<th>PmMSAz</th>
<th>tCert</th>
<th>tExp</th>
</tr>
</thead>
<tbody>
<tr>
<td>mMSA</td>
<td>1.00</td>
<td>.026</td>
<td>.782</td>
<td>.769</td>
<td>.091</td>
<td>.058</td>
</tr>
<tr>
<td>stuGend</td>
<td>.026</td>
<td>1.000</td>
<td>-.022</td>
<td>.035</td>
<td>-.032</td>
<td>-.033</td>
</tr>
<tr>
<td>BMzMean</td>
<td>.782</td>
<td>-.022</td>
<td>1.000</td>
<td>.730</td>
<td>.025</td>
<td>-.017</td>
</tr>
<tr>
<td>PmMSAz</td>
<td>.769</td>
<td>.035</td>
<td>.730</td>
<td>1.000</td>
<td>.024</td>
<td>.020</td>
</tr>
<tr>
<td>tCert</td>
<td>.091</td>
<td>-.032</td>
<td>.025</td>
<td>.024</td>
<td>1.000</td>
<td>.625</td>
</tr>
<tr>
<td>tExp</td>
<td>.058</td>
<td>-.033</td>
<td>-.017</td>
<td>.020</td>
<td>.625</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sig. (1-tailed)</th>
<th>mMSA</th>
<th>stuGend</th>
<th>BMzMean</th>
<th>PmMSAz</th>
<th>tCert</th>
<th>tExp</th>
</tr>
</thead>
<tbody>
<tr>
<td>mMSA</td>
<td>.</td>
<td>.031</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>stuGend</td>
<td>.031</td>
<td>.</td>
<td>.059</td>
<td>.006</td>
<td>.009</td>
<td>.019</td>
</tr>
<tr>
<td>BMzMean</td>
<td>.000</td>
<td>.059</td>
<td>.</td>
<td>.000</td>
<td>.037</td>
<td>.110</td>
</tr>
<tr>
<td>PmMSAz</td>
<td>.000</td>
<td>.006</td>
<td>.000</td>
<td>.</td>
<td>.043</td>
<td>.073</td>
</tr>
<tr>
<td>tCert</td>
<td>.000</td>
<td>.009</td>
<td>.037</td>
<td>.043</td>
<td>.</td>
<td>.000</td>
</tr>
<tr>
<td>tExp</td>
<td>.000</td>
<td>.009</td>
<td>.110</td>
<td>.073</td>
<td>.000</td>
<td>.</td>
</tr>
</tbody>
</table>

*Note.* N = 5,224 for each variable: mMSA = math MSA score; stuGend = student gender; BMzMean = mean of student’s benchmark z-scores; PmMSAz = student’s prior math MSA z-score; tCert = teacher certification level; tExp = teacher years of experience.
Among the predictors, there was a moderate and positive correlation between prior math achievement and performance on the benchmark exams \((r = .730, p < .001)\) and between teacher certification level and teacher years of experience \(r = .625, p < .001\). Student gender was positively related to prior math performance \((r = .035, p < .05)\); however, there was a small but significant negative correlation between student gender and both teacher characteristics: teacher years of experience \((r = -.033, p < .05)\) and teacher certification level \((r = -.032, p < .05)\). On the other hand, teacher certification was positively related to benchmark exam performance \((r = .025, p < .05)\) and prior math achievement \((r = .024, p < .05)\).

The significance of each of the predictor variables in the regression model was considered by testing the null hypotheses, \(H_0^k: \beta_k = 0\), with \(k = 1, 2\ldots 5\); or, the individual regression coefficients in the population equal zero. The alternate hypothesis was \(H_{\alpha k}: \beta_k \neq 0\); i.e., that one or more of the predictor variables significantly contributed to the regression when used in combination with the other predictors.

The null hypotheses, \(H_{0k}\), regarding statistical significance of benchmark exam performance, prior math MSA performance, teacher years of experience, and teacher certification level were rejected, as these variables contributed significantly to the prediction of math MSA score (regression coefficients, standard errors and significance statistics can be found in Table 10). Although the bivariate correlation between math MSA score and student gender was statistically significant \((r = .03, p < .05)\), student gender did not significantly contribute to the regression model; therefore, the null hypothesis, \(H_{0k}\), was not rejected for student gender. Evidently, the relationship between math achievement and student gender is mediated by the relationships between gender and the other predictors: prior math achievement, performance on the benchmark exams, teacher years of experience, and teacher certification level.
Evaluation of model assumptions. The following multiple regression assumptions and issues were assessed:

- adequate sample size (Field, 2005; Tabachnick & Fidell, 2007);
- variable characteristics assumptions (Berry, 1993; Field, 2005),
- nonzero variance of predictor variables (Berry, 1993; Field, 2005),
- independence of residuals (Berry, 1993; Field, 2005; Tabachnick & Fidell, 2007),
- linearity of the relationship between the criterion variable and the predictor variables (Berry, 1993; Field, 2005),
- homoscedacity of residuals (Berry, 1993; Field, 2005; Tabachnick & Fidell, 2007),
- no perfect multicollinearity (Berry, 1993; Field, 2005; Tabachnick & Fidell, 2007),
- no significant outliers or influential data points (Field, 2005; Tabachnick & Fidell, 2007),
- normality of the distribution of the residuals (Berry, 1993; Field, 2005; Tabachnick & Fidell, 2007), and
- independence of all values of the criterion variable (Field, 2005).

With five predictor variables, the 5,224 cases used in the study well exceeded established sample-size guidelines for MLR:

1. For overall fit of a regression model with five predictors, a minimum sample size of 90 (50 + 8k, where k is the number of predictors) is required (Green, as cited in Field, 2005 and in Tabachnick & Fidell, 2007).
2. For testing individual predictors, a minimum of 109 (104 + k) cases are needed (Green, as cited in Field, 2005 and in Tabachnick & Fidell, 2007).
3. To achieve a high level of power, 80, 200 and 600 cases are needed for large, medium and small effect sizes, respectively (Miles & Shevlin, as cited in Field, 2005).

The predictor variables were quantitative or dichotomous and as shown in Table 11, had non-zero variance (Berry, 1993). The criterion variable, although discrete, had numerous categories and represented a quantitative attribute (Field, 2005); however, the outcome variable was bounded (Berry, 1993) as attainable values ranged from 240 through 650 but the range of values in the data set was 307 through 574.

The Durbin-Watson statistic, $d = 1.398$, fell within a conservative range of 1 to 3 for tenability of the assumption of independence of residuals (Field, 2005). The overall lack of curvature in the scatterplot of the standardized predicted value against the standardized residual (see Figure 3) suggested that the relationship between the math MSA score and the predictors collectively was approximately linear. Partial regression plots showed approximate linear relationships between math performance (math MSA score) and each of the variables teachers years of experience (see Figure 4), prior math performance (see Figure 5), and benchmark exam performance (see Figure 6).
Table 10

*MLR Coefficients and Their Confidence Intervals, Correlations and Collinearity Statistics*

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized coefficients</th>
<th>Standardized coefficients</th>
<th>95.0% confidence interval for B</th>
<th>Correlations</th>
<th>Collinearity statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td>t</td>
<td>Sig.</td>
</tr>
<tr>
<td>(Constant)</td>
<td>417.293</td>
<td>.605</td>
<td>690.075</td>
<td>.000</td>
<td>416.108</td>
</tr>
<tr>
<td>stuGend</td>
<td>1.711</td>
<td>.543</td>
<td>.024</td>
<td>3.148</td>
<td>.002</td>
</tr>
<tr>
<td>BMzMean</td>
<td>21.519</td>
<td>.504</td>
<td>.475</td>
<td>42.717</td>
<td>.000</td>
</tr>
<tr>
<td>PmMSAz</td>
<td>16.815</td>
<td>.445</td>
<td>.420</td>
<td>37.794</td>
<td>.000</td>
</tr>
<tr>
<td>tCert</td>
<td>4.453</td>
<td>.779</td>
<td>.055</td>
<td>5.717</td>
<td>.000</td>
</tr>
<tr>
<td>tExp</td>
<td>.082</td>
<td>.034</td>
<td>.024</td>
<td>2.438</td>
<td>.015</td>
</tr>
</tbody>
</table>

*Note.* Dependent Variable: mMSA = math MSA score; stuGend = student gender; BMzMean = mean of student’s benchmark z-scores; PmMSAz = student’s prior math MSA z-score; tCert = teacher certification level; tExp = teacher years of experience.
Table 11

Variance of Predictor Variables

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student gender</td>
<td>.250</td>
</tr>
<tr>
<td>Mean of benchmark z-scores</td>
<td>.624</td>
</tr>
<tr>
<td>Prior math MSA z-score</td>
<td>.799</td>
</tr>
<tr>
<td>Teacher certification level</td>
<td>.199</td>
</tr>
<tr>
<td>Teacher years of experience</td>
<td>107.408</td>
</tr>
</tbody>
</table>

Figure 3. Scatterplot of standardized predicted value against standardized residual.
Figure 4. Partial regression plot: teacher years of experience.

Figure 5. Partial regression plot: prior math MSA z-score
The assumption of homoscedasticity of residuals (i.e., the assumption that the spread of residuals is fairly consistent at each point along any predictor variable) was assessed using the scatterplot of the standardized predicted value against standardized residual. The slight funnel shape of the scatterplot (see Figure 3) indicated that heteroscedasticity may be present in the data. Moreover, the quantitative predictors’ partial regression plots indicated relative homoscedasticity for teacher experience data (see Figure 4) and student prior math achievement data (see Figure 5), but heteroscedasticity may be present in the benchmark exam data (see Figure 6).

With respect to the assumption of no perfect multicollinearity (i.e., that two or more variables are not very closely linearly related), (a) correlation between any two predictors (see Table 9) was less than .8 (Field, 2005), (b) the largest variance inflation factor (see VIFs in Table 10) was well below 10 (Myers, as cited in Field, 2005; Bowerman & O’Connell, as cited in Field, 2005), and (c) tolerance values (see Table 10) were well above .1 (Menard, as cited in
Field, 2005). Therefore, it was concluded that there was no perfect multicollinearity within the data.

Cases were examined to determine whether there were any significant outliers or data points cases that substantially influenced the regression model. The following evidence suggested that the model is fairly reliable and was not unduly influenced by any subset of cases (Field, 2005):

1. Standardized residuals were examined and it was determined, as would be expected, that 5% of the cases ($n = 262$) had standardized residuals outside the interval $[-2, 2]$. However, there was a disproportionate number of cases (109, or 2%) that had standardized residuals outside the interval $[-2.5, 2.5]$. This is twice that which would be expected.

2. The largest Cook’s value for the cases in the data set was .001 which was well below 1, the accepted cutoff for identifying influencers.

3. The centered leverage of each case fell below Stevens’ (as cited in Field, 2005) recommended cutoff of three times the average leverage: $3(k + 1)/n = 3(5 +1)/5224 = 18/5224 = .004$.

4. No case had a Mahalanobis distance greater than 15.5; so they do not come close to exceeding the guideline of a value of 25 for large samples with five predictors (Barnett & Lewis, as cited in Field, 2005).

5. The standardized DFBeta values for each case and for each parameter fell with the range of -1 to +1.

6. The covariance ratio (CVR) for each case was greater than the lower limit of acceptable values for the CVR: $1-3(k + 1)/n = 1-3(5 +1)/5224 = -18/5224 = 0.997$. To
the contrary, 1% of the cases had CVRs greater than the upper limit of acceptable CVR values: $1 + 3(k + 1)/n = 1 + 3(5 + 1)/5224 = 1 + 18/5224 = 0.997$; however, the CVRs of these cases were close to the upper limit, with the largest CVR exceeding the limit by only 0.004.

The assumption of normality of the distribution of the residuals (Barry, 1993; Field, 2005) was evaluated by inspection of the histogram of the standardized residuals (see Figure 7) and the normal probability plot of the standardized residuals (see Figure 8). These figures indicated that the residuals were approximately normally distributed, with a mean of zero and a standard deviation of one.

The assumption of independence of all values of the criterion variable (Field, 2005) was violated as the data set contained multiple records for some of the students (maximum of three per student, one record for each of the three years of the study’s time frame).

Figure 7. MLR standardized residual histogram.
**Figure 8.** Normal probability plot of MLR standardized residual.

**MLR summary and model equation.** A multiple linear regression was run to predict math achievement from student gender, prior math performance, math benchmark exam performance, teacher years of experience, and teacher certification level. The assumptions of nonzero variance of predictors, linearity, independence of residuals, lack of significant outliers/influencers, no perfect multicollinearity, and normality of residuals were met. Slight heteroscedacity may be present in the data. The assumption of independence of criterion values was violated because the data contained up to three cases per student—one case for each year of the study.

The predictor variables significantly predicted math MSA score, $F(5, 5218) = 2452.845$, $p < .001$. Greater than 70% of the variance in math MSA scores is attributable to the combined effect of these predictors. Student performance on the math benchmark exams had the greatest impact on math achievement. The MLR model is represented by the equation:
\[
(mMSA)_i = 417.30 + 1.71(stuGend)_i + 21.52(BMzMean)_i + \\
16.82(PmMSAz)_i + 4.45(tCert)_i + .08(tExp)_i + \epsilon_i,
\]

where \(i\) represents the student-case number, \(i \in [1, 2...5224]\);

\((mMSA)_i\) is the predicted MSA score for the student represented by case \(i\);

\((BMzMean)_i\) is the mean of the mathematics benchmark exam scores for a given year for the student represented by case \(i\);

\((PmMSAz)_i\) denotes the prior math MSA score of the student represented by case \(i\);

\((tCert)_i\) signifies, for the student represented in case \(i\), the professional certification level of the student’s teacher;

\((tExp)_i\) signifies, for the student represented in case \(i\), the number of years of professional experience of the student’s teacher; and

\(\epsilon_i\) signifies the residual error for student-case \(i\).

**Multilevel Models**

**Overview.** To address RQ2, a series of linear mixed models were run using PASW (version 18). The L1 unit of analysis was individual student performance on the math MSA; the L1 variance estimate is an estimate of how differently students with the same teacher perform on the math MSA exam (within-teacher variance). The L2 unit of analysis was the mean math MSA score of students grouped by teacher. The L2 variance estimate indicates how math achievement varies across students grouped by teacher (between-teacher variance). The hypotheses for the MLM were:

\(H_0: \) the means of students’ math MSA scores by teacher are equal.

\(H_a: \) there is significant variability in the means of students’ math MSA scores across teachers.
The multilevel model was developed by fitting the following sequence of MLMs to the data to investigate the impact of student and teacher characteristics on mathematics achievement while accounting for the hierarchical structure of the data:

1. Model 0, the null model, partitioned the variance in math achievement into within- and between-teacher components with no predictors present in the model.

2. Model 1, a student-level random intercept model, added three student-level predictors to the null model: gender, performance on benchmark exams and prior math achievement.

3. Model 2, a teacher-level random intercept model, added four teacher-level predictors to the model. Two were teacher characteristics: professional certification level and years of experiences; the others were student characteristics aggregated by teacher: mean of benchmark exam performance and mean prior math achievement.

4. Model 3, a random slope and intercept model, allowed the L1 slopes associated with student characteristics to vary randomly across teachers.

For each successive MLM, there was significant variability in the means of students’ math MSA scores across teachers (see Wald Z statistics for intercept variance in Table 12), so the null hypothesis was rejected for each model. The variance in math MSA scores within teachers was significant in each model as well (see residual Wald Z statistics in Table 12). Therefore, at each step of the MLM model analyses, development of the subsequent model was warranted to try to explain the remaining variance within and between teachers.
Table 12

Wald Z Statistic and 95% Confidence Intervals for Variance in the MLM Models

<table>
<thead>
<tr>
<th></th>
<th>Residual Intercept variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wald Z ((p/2 = \text{one-tailed sig. in parentheses}))</td>
</tr>
<tr>
<td>Model 0</td>
<td>50.61 (.000) 6.45 (.000)</td>
</tr>
<tr>
<td>Model 1</td>
<td>50.62 (.000) 6.58 (.000)</td>
</tr>
<tr>
<td>Model 2</td>
<td>50.62 (.000) 6.43 (.000)</td>
</tr>
<tr>
<td>Model 3</td>
<td>49.37 (.000) 6.16 (.000)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>[893.681, 965.657] [371.510, 682.117]</td>
</tr>
<tr>
<td>Model 1</td>
<td>[251.04, 271.25] [92.80, 168.37]</td>
</tr>
<tr>
<td>Model 2</td>
<td>[251.03, 271.24] [83.86, 154.33]</td>
</tr>
<tr>
<td>Model 3</td>
<td>[229.47, 248.43] [77.99, 147.41]</td>
</tr>
</tbody>
</table>

Table 13 shows that with the development of each successive MLM model, the residual (within-teacher variance) and the intercept variance (between-teacher variance) decreased; and Table 12 shows that the 95% confidence interval for the residual and intercept variance decreased with each successive model. Hence, there was a better fit of the data with each successive model. The mean within-teacher reliability estimate for each of the four models (see Table 13) was greater than .92; therefore, the estimated differences across teachers are reliable indicators of actual differences among teachers’ population means.
Table 13

**MLM Estimates: Intraclass Correlation, Reliability, and Reduction in Variance**

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Model 0</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>928.972</td>
<td>260.950</td>
<td>260.940</td>
<td>238.762</td>
</tr>
<tr>
<td>Intercept variance</td>
<td>503.402</td>
<td>124.998</td>
<td>113.763</td>
<td>107.219</td>
</tr>
<tr>
<td>Reduction in variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-teacher portion</td>
<td>0.719</td>
<td>0.000</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>Between-teacher portion</td>
<td>0.752</td>
<td>0.090</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>Intraclass correlation</td>
<td>.351</td>
<td>.324</td>
<td>.304</td>
<td></td>
</tr>
<tr>
<td>Mean within-teacher reliability</td>
<td>.936</td>
<td>.928</td>
<td>.922</td>
<td>.924</td>
</tr>
</tbody>
</table>

As shown in Table 14, the L1 student characteristics—student gender, benchmark exam performance and prior math performance—were significant predictors of math achievement in all the MLMs. To the contrary, the L2 teacher characteristics—certification level and years of experience—were not significant predictors in any of the models. However, at the teacher level, aggregate student predictors did moderate the relationship between math MSA score and benchmark exam performance (Models 2 and 3) and between math MSA score and prior math achievement (Model 2). In other words, in Model 2 the relationship between math achievement and prior math MSA score varies significantly according to the mean (by teacher) of students’ prior math MSA z-scores; and the relationship between math achievement and benchmark exam performance varies significantly according to the value of the aggregate student benchmark measure.
### Table 14: MLMs' Fixed Effects Estimates and Variance-Covariance Estimates for Math Achievement Predictors

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 0</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ( (\gamma_{00}) )</td>
<td>428.13 (2.33)</td>
<td>425.54 (1.19)</td>
<td>425.02** (2.21)</td>
<td>423.39** (2.21)</td>
</tr>
<tr>
<td>Level 1 (student-case specific)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student gender ( (\gamma_{10}) )</td>
<td>1.86** (.45)</td>
<td>1.86** (.45)</td>
<td>1.88* (.54)</td>
<td></td>
</tr>
<tr>
<td>Benchmark exam performance ( (\gamma_{20}) )</td>
<td>20.11** (.47)</td>
<td>19.91** (.47)</td>
<td>23.06** (.90)</td>
<td></td>
</tr>
<tr>
<td>Prior math achievement ( (\gamma_{30}) )</td>
<td>17.48** (.41)</td>
<td>17.58** (.41)</td>
<td>15.61** (.60)</td>
<td></td>
</tr>
<tr>
<td>Level 2 (teacher specific)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum level of certification ( (\gamma_{01}) )</td>
<td>-0.60 (3.20)</td>
<td>-0.17 (3.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean years of experience ( (\gamma_{02}) )</td>
<td>0.04 (.15)</td>
<td>0.04 (.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students' prior math achievement ( (\gamma_{03}) )</td>
<td>-7.33* (3.21)</td>
<td>-4.53 (3.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students’ benchmark exam performance ( (\gamma_{04}) )</td>
<td>12.91* (3.66)</td>
<td>10.72* (3.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Random parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ( (u_{0j}) )</td>
<td>503.40** (78.03)</td>
<td>125.00** (19.00)</td>
<td>113.76** (17.94)</td>
<td>107.22** (17.41)</td>
</tr>
<tr>
<td>Mean of benchmark z-scores ( (u_{1j}) )</td>
<td></td>
<td></td>
<td>45.79** (10.20)</td>
<td></td>
</tr>
<tr>
<td>Prior math MSA score ( (u_{2j}) )</td>
<td></td>
<td></td>
<td>11.03* (4.01)</td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual ( (r_{ij}) )</td>
<td>928.97** (18.36)</td>
<td>260.95** (5.16)</td>
<td>260.94** (5.16)</td>
<td>238.76** (4.84)</td>
</tr>
</tbody>
</table>

*Note: Standard errors are in parentheses.  
*\( p < .05 \), **\( p < .001 \)
Model 0: Null model. A linear mixed model with no predictors was developed to partition the variance in math achievement into its within- and between-teacher components and determine how much of the variance in math achievement lay between teachers. The equation for the null model is:

\[
(mMSA)_{ij} = (mMSA)_{ij} = 428.13 + u_{0j} + r_{ij}, \text{ where}
\]

\[
(mMSA)_{ij} \text{ is the predicted math MSA score of student-case } i \text{ of teacher } j;
\]

\[
u_{0j} \text{ is the deviation from the math MSA grand mean for student-cases of teacher } j; \text{ i.e., the unique effect of teacher } j \text{ on the intercept; and}
\]

\[
r_{ij} \text{ is student-case } i'\text{ s deviation from the math MSA mean of teacher } j'\text{'s student cases.}
\]

The null model estimated the math MSA grand mean as 428 (see related estimation statistics in Table 15). There was significant variance in students’ math achievement between teachers (Wald \( Z = 6.451 \), one-tailed \( p = .000/2 < .001 \)) and within teachers (Wald \( Z = 50.606 \), one-tailed \( p < .001 \)). Table 16 presents statistics relevant to the model’s covariance parameter estimates.

Intraclass correlation was calculated using covariance parameter estimates:

\[
\rho = 503.402/(503.402 + 928.972) = .351. \text{ This indicates that approximately 35\% of the total variability in math MSA scores was between teachers.} \]
Table 15

**MLM Null Model Estimates of Fixed Effects**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>428.13</td>
<td>2.33</td>
<td>94.61</td>
<td>83.68</td>
<td>.000</td>
<td>423.50 - 432.75</td>
</tr>
</tbody>
</table>

*Note.* Dependent Variable: mMSA = math MSA score.

Table 16

**MLM Null Model Covariance Parameters Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. error</th>
<th>Wald Z</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>928.972</td>
<td>18.357</td>
<td>50.606</td>
<td>.000</td>
<td>893.681 - 965.657</td>
</tr>
<tr>
<td>Intercept Variance [subject = tID]</td>
<td>503.402</td>
<td>78.032</td>
<td>6.451</td>
<td>.000</td>
<td>371.510 - 682.117</td>
</tr>
</tbody>
</table>

*Note:* Dependent Variable: math MSA score. tID = teacher ID.

**Model 1: Student-Level (L1) Random Intercept Model.** A linear mixed model with student-level predictors—student gender, prior math achievement, and benchmark exam performance—was developed to examine variability in the intercepts across teachers. As seen in Table 17, each of the student-level predictors contributed significantly to the model. The equation for the student-level random intercept model is:

\[
(mMSA)_{ij} = 425.54 + 1.86(stuGend)_{ij} + 20.11(BMzMean)_{ij} + 17.48(PmMSAz)_{ij} + u_{0j} + r_{ij}, \tag{8}
\]

*(mMSA)_{ij}* is the predicted math MSA score of student-case *i* of teacher *j*;

*(stuGend)_{ij}* represents the gender of student-case *i* of teacher *j*;

*(BMzMean)_{ij}* is the mean of the benchmark exam *z*-scores of student-case *i* of teacher *j*;
\((Pm_{MSAz})_{ij}\) is the prior math MSA \(z\)-score of student-case \(i\) of teacher \(j\);

\(u_{0j}\) indicates deviation from the math MSA grand mean of teacher \(j\)’s student cases; and \(r_{ij}\) represents student \(i\)'s deviation from the math MSA mean of teacher \(j\)’s student cases.

The addition of the L1 predictors—student gender, mean of benchmark \(z\)-scores, and prior math MSA \(z\)-score—reduced the within- and between-teacher variance (see estimated covariance parameters in Table 18). The calculation of the reduction in within-teacher variance from the null model to model 1 \(((928.978 - 260.950)/928.972 = .719)\) indicated that students’ gender, prior math achievement and performance on the benchmark exams accounted for 72% of the within-teacher variability.

The reduction in variance estimate for between-teacher variance suggested that 75% of the variation in math MSA means across teachers could be attributed to differences in the student characteristics in the model. Calculation of the intraclass correlation indicated that approximately 32% of the total variability in math MSA scores was between teachers. See Table 13 for subsequent changes in within- and between-teacher variance and in intraclass correlation in subsequent MLM models.

As shown in Table 17, after controlling for student gender, mean of benchmark exam \(z\)-scores and prior math MSA \(z\)-scores in the model, a statistically significant amount of variation remained between teachers \((Wald Z = 6.580,\ \text{one-tailed} \ p = < .001)\) and within teachers \((Wald Z = 50.619,\ \text{one-tailed} \ p < .001)\). Therefore, the addition of other predictors to the model that might explain the variability was warranted.
Table 17

*Level 1 Random Intercept Model Estimates of Fixed Effects*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t</th>
<th>Sig.</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>425.54</td>
<td>1.19</td>
<td>107.52</td>
<td>358.28</td>
<td>.000</td>
<td>423.19 - 427.90</td>
</tr>
<tr>
<td>stuGend</td>
<td>1.86</td>
<td>.45</td>
<td>5137.07</td>
<td>4.12</td>
<td>.000</td>
<td>.98 - 2.75</td>
</tr>
<tr>
<td>BMzMean</td>
<td>20.11</td>
<td>.47</td>
<td>5216.13</td>
<td>43.00</td>
<td>.000</td>
<td>19.19 - 21.03</td>
</tr>
<tr>
<td>PmMSAz</td>
<td>17.48</td>
<td>.41</td>
<td>5215.71</td>
<td>42.74</td>
<td>.000</td>
<td>16.67 - 18.28</td>
</tr>
</tbody>
</table>

*Note.* Dependent Variable: mMSA = math MSA score; stuGend = student gender; BMzMean = mean of student’s benchmark z-scores; PmMSAz = student’s prior math MSA z-score.

Table 18

*Level 1 Random Intercept Model Estimates of Covariance Parameters*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. error</th>
<th>Wald Z</th>
<th>Sig.</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>260.95</td>
<td>5.16</td>
<td>0.62</td>
<td>.000</td>
<td>251.04 - 271.25</td>
</tr>
<tr>
<td>Intercept Variance [subject = tID]</td>
<td>125.00</td>
<td>19.00</td>
<td>6.58</td>
<td>.000</td>
<td>92.80 - 168.37</td>
</tr>
</tbody>
</table>

*Note.* Dependent Variable: math MSA score. tID = teacher ID.

**Model 2: Teacher-level (L2) random intercept model.** Model 2 added teacher-level predictors to the model to further explain variability in mean math MSA scores (intercepts) across teachers. For each teacher, the model included the mean of his/her years of experience from 2007 through 2009 and the maximum level of certification held during 2007-2009. Two L1 predictor values were aggregated by teacher and added to the model to account for the impact of student characteristics at the teacher level: prior math achievement and benchmark exam.
performance. The model’s fixed effects estimates are presented in Table 19; covariate parameter estimates are in Table 20.

In Model 2, each of the student-level predictors—student gender, prior mathematics achievement and benchmark exam performance—contributed significantly \( (p < .001) \) to math achievement (see Table 19). As for teacher-level predictors, advanced teacher certification had a negative but insignificant impact on mathematics achievement \( (\gamma_{01} = -0.60, p > .05) \). Teacher years of experiences was insignificant as well, but positive \( (\gamma_{02} = 0.04, p > .05) \). After controlling for individual student performance on prior math MSA and benchmark exams within teachers, aggregate student benchmark performance significantly and positively contributed to math achievement \( (\gamma_{04} = 12.91, p < .05) \); and aggregate prior student math MSA performance significantly and negatively contributes to math achievement \( (\gamma_{03} = -7.33, p < .05) \). The equation for the teacher-level random intercept model is:

\[
(m_{MSA})_{ij} = 425.02 + 1.86(stuGend)_{ij} + 19.91(BMzMean)_{ij} + \\
17.58(PmMSAz)_{ij} - 0.60(max tCert)_{j} + \\
0.05(tExpMean)_{j} - 7.33(BMzMean_{mean})_{j} + \\
12.91(BMzMean_{mean})_{j} + u_{0j} + r_{ij},
\]

where

\((m_{MSA})_{ij}\) is the predicted math MSA score of student-case \( i \) of teacher \( j \);

\((stuGend)_{ij}\) represents the gender of student-case \( i \) of teacher \( j \);

\((BMzMean)_{ij}\) is the mean of the benchmark exam \( z \)-scores of student-case \( i \) of teacher \( j \);

\((PmMSAz)_{ij}\) is the prior math MSA \( z \)-score of student-case \( i \) of teacher \( j \);

\((max tCert)_{j}\) is the maximum certification level attained by teacher \( j \) during the 2007-2009 time frame;
(tExpMean)\_j is teacher j’s mean years of experience from 2007 through 2009;

(PmMSAz\_mean)\_j is the mean of the prior math MSA z-scores of teacher j’s student cases;

(BMzMean\_mean)\_j is the mean, for teacher j, of his/her student-cases’ benchmark z-score average;

u\_{0j} indicates deviation from the math MSA grand mean of teacher j’s student cases; and

r_{ij} represents student i’s deviation from the math MSA mean of teacher j’s student cases.

Table 19

**Level 2 Random-Intercept Model Fixed Effects Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t</th>
<th>Sig.</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>425.02</td>
<td>2.21</td>
<td>100.33</td>
<td>192.32</td>
<td>.000</td>
<td>420.63</td>
<td>429.40</td>
</tr>
<tr>
<td>stuGend</td>
<td>1.86</td>
<td>0.45</td>
<td>5137.29</td>
<td>4.12</td>
<td>.000</td>
<td>0.98</td>
<td>2.75</td>
</tr>
<tr>
<td>BMzMean</td>
<td>19.91</td>
<td>0.47</td>
<td>5124.99</td>
<td>42.25</td>
<td>.000</td>
<td>18.99</td>
<td>20.84</td>
</tr>
<tr>
<td>PmMSAz</td>
<td>17.58</td>
<td>0.41</td>
<td>5125.02</td>
<td>42.66</td>
<td>.000</td>
<td>16.77</td>
<td>18.38</td>
</tr>
<tr>
<td>maxtCert</td>
<td>-0.60</td>
<td>3.20</td>
<td>97.09</td>
<td>-0.19</td>
<td>.852</td>
<td>-6.94</td>
<td>5.75</td>
</tr>
<tr>
<td>tExpMean</td>
<td>0.05</td>
<td>0.15</td>
<td>95.36</td>
<td>0.34</td>
<td>.732</td>
<td>-0.25</td>
<td>0.35</td>
</tr>
<tr>
<td>BMzMean_mean</td>
<td>12.91</td>
<td>3.66</td>
<td>103.67</td>
<td>3.52</td>
<td>.001</td>
<td>5.65</td>
<td>20.18</td>
</tr>
<tr>
<td>PmMSAz_mean</td>
<td>-7.33</td>
<td>3.21</td>
<td>106.35</td>
<td>-2.29</td>
<td>.024</td>
<td>-13.68</td>
<td>-0.97</td>
</tr>
</tbody>
</table>

a. Dependent Variable: mMSA = math MSA score; stuGend = student gender; BMzMean = mean of student’s benchmark z-scores; PmMSAz = student’s prior math MSA z-score; maxtCert = maximum teacher certification level attained; tExpMean = mean of teacher years of experience; BMzMean\_mean = aggregate mean by teacher of students’ benchmark exam z-scores; PmMSAz\_mean = aggregate mean by teacher of students’ prior math MSA z-score.
The addition of the teacher-level predictors—teacher certification, teacher years of experience, aggregate mean of students’ benchmark z-scores, and mean of students’ prior math MSA z-scores—did not affect within-teacher variance in math achievement. However, the addition reduced the between-teacher variance from that in Model 1 (see a comparison of the variances in the MLMs in Table 13); an additional 9% of the variability in math achievement was accounted for in Model 2. The intraclass correlation estimate for Model 2 suggests that about 30% of the total variability in math MSA scores was between teachers.

As shown in Table 20, after controlling for the L1 student characteristics and the additional teacher-level predictors, a statistically significant variability in math achievement remained between teachers (Wald Z = 6.43, one-tailed \( p < .001 \)) and within teachers (Wald Z = 50.62, one-tailed \( p < .001 \)). Therefore, the addition of other predictors to the model that might explain the variability was warranted.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. error</th>
<th>Wald Z</th>
<th>Sig.</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>260.94</td>
<td>5.16</td>
<td>50.62</td>
<td>.000</td>
<td>251.03 271.24</td>
</tr>
<tr>
<td>Intercept Variance</td>
<td>113.76</td>
<td>17.70</td>
<td>6.43</td>
<td>.000</td>
<td>83.86 154.33</td>
</tr>
</tbody>
</table>

Note. Dependent Variable: math MSA score.

**Model 3: Random slope and intercept model.** Since in the previous MLMs, individual student gender, prior math achievement and benchmark exam performance each were significantly related to math MSA score, the slopes of these predictors were allowed to vary
across teachers in Model 3 to determine whether they vary randomly, and if they do, to examine relationships between L2 predictors and the strength of those relationships. Table 21 presents the fixed effects estimates for the model; the estimates for covariance parameters can be found in Table 22. As seen in Table 21, aggregate benchmark exam performance and students’ gender, their benchmark exam performance, and prior math achievement each significantly impact math achievement in this model.

Allowing individual student characteristic slopes to vary randomly across teachers resulted in an increase in the coefficient for benchmark exam performance, from 19.91 to 23.06 and a decrease in the prior math achievement coefficient, from 17.58 to 15.61. It also showed a decrease in the variance remaining at the student level, from 260.94 in to 238.76, as well as at the teacher level (from 113.76 to 107.22. Even so, remaining variance at the teacher level was still significant ($Wald Z = 6.16$, one-tailed $p < .001$), as was remaining variance at the student level ($Wald Z = 49.37$, one-tailed $p < .001$). Furthermore, the benchmark-performance slope variance was significant ($Wald Z = 4.49$, one-tailed $p < .001$), as was the prior-math-achievement slope variance ($Wald Z = 2.75$, one-tailed $p/2 = .003 < .05$), indicating that that these slopes vary across teachers in the sample.
Table 21

*Random Slope and Intercept Model Fixed Effects Estimates*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>$t$</th>
<th>Sig.</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>423.39</td>
<td>2.21</td>
<td>101.15</td>
<td>191.75</td>
<td>.000</td>
<td>419.01</td>
<td>427.77</td>
</tr>
<tr>
<td>stuGend</td>
<td>1.89</td>
<td>.54</td>
<td>59.78</td>
<td>3.48</td>
<td>.001</td>
<td>.80</td>
<td>2.96</td>
</tr>
<tr>
<td>BMzMean</td>
<td>23.06</td>
<td>.90</td>
<td>105.32</td>
<td>25.57</td>
<td>.000</td>
<td>21.27</td>
<td>24.85</td>
</tr>
<tr>
<td>PmMSAz</td>
<td>15.61</td>
<td>.60</td>
<td>89.95</td>
<td>25.95</td>
<td>.000</td>
<td>14.42</td>
<td>16.81</td>
</tr>
<tr>
<td>maxtCert</td>
<td>-0.17</td>
<td>3.20</td>
<td>98.51</td>
<td>-0.05</td>
<td>.958</td>
<td>-6.52</td>
<td>6.18</td>
</tr>
<tr>
<td>tExpMean</td>
<td>.04</td>
<td>.15</td>
<td>96.08</td>
<td>.29</td>
<td>.773</td>
<td>-.26</td>
<td>.35</td>
</tr>
<tr>
<td>PmMSAz_mean</td>
<td>-4.53</td>
<td>3.25</td>
<td>111.99</td>
<td>-1.40</td>
<td>.133</td>
<td>-10.97</td>
<td>1.90</td>
</tr>
<tr>
<td>BMzMean_mean</td>
<td>10.07</td>
<td>3.72</td>
<td>109.67</td>
<td>2.88</td>
<td>.005</td>
<td>3.35</td>
<td>18.09</td>
</tr>
</tbody>
</table>

*Note.* Dependent Variable: mMSA = math MSA score; stuGend = student gender; BMzMean = mean of student’s benchmark z-scores; PmMSAz = student’s prior math MSA z-score; maxtCert = maximum teacher certification level attained; tExpMean = mean of teacher years of experience; PmMSAz_mean = aggregate mean by teacher of students’ prior math MSA z-score; BMzMean_mean = aggregate mean by teacher of students’ benchmark exam z-scores.
### Table 22

**Random Slope and Intercept Model Estimates of Covariance Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. error</th>
<th>Wald Z</th>
<th>Sig.</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>238.76</td>
<td>4.84</td>
<td>49.37</td>
<td>.000</td>
<td>229.47</td>
<td>248.43</td>
</tr>
<tr>
<td>Intercept variance</td>
<td>107.22</td>
<td>17.41</td>
<td>6.16</td>
<td>.000</td>
<td>77.99</td>
<td>147.41</td>
</tr>
<tr>
<td>stuGend variance</td>
<td>5.45</td>
<td>3.38</td>
<td>1.61</td>
<td>.107</td>
<td>1.62</td>
<td>18.39</td>
</tr>
<tr>
<td>BMzMean variance</td>
<td>45.79</td>
<td>10.20</td>
<td>4.49</td>
<td>.000</td>
<td>29.59</td>
<td>70.85</td>
</tr>
<tr>
<td>PmMSAz variance</td>
<td>11.03</td>
<td>4.01</td>
<td>2.75</td>
<td>.006</td>
<td>5.42</td>
<td>22.48</td>
</tr>
</tbody>
</table>

*Note.* Dependent Variable: math MSA score; stuGend = student gender; BMzMean = mean of student’s benchmark z-scores; PmMSAz = student’s prior math MSA z-score.

The equation for Model 3 is:

\[
(mMSA)_{ij} = 423.39 + 1.89(stuGend)_{ij} + 23.06(BMzMean\_mean)_{ij} + 15.61(PmMSAz)_{ij} - .17(max\_tCert)_{j} + .04(tExpMean)_{j} - 4.53(PmMSAz\_mean)_{j} + 10.72(BMzMean\_mean)_{j} + u_{0j} + u_{1j}(BMzMean)_{ij} + u_{2j}(PmMSAz)_{ij} + r_{ij}, \quad (10)
\]

\[(mMSA)_{ij}\] is the predicted math MSA score of student-case \(i\) of teacher \(j\);

\[(stuGend)_{ij}\] represents the gender of student-case \(i\) of teacher \(j\);

\[(BMzMean)_{ij}\] is the mean of the benchmark exam z-scores of student-case \(i\) of teacher \(j\);

\[(PmMSAz)_{ij}\] is the prior math MSA z-score of student-case \(i\) of teacher \(j\);

\[(maxtCert)_{j}\] is the maximum certification level attained by teacher \(j\) during the 2007-2009 time frame;
(tExpMean)\textsubscript{j} is teacher j’s mean years of experience from 2007 through 2009; 
(PmMSAz\textunderscore mean)\textsubscript{j} is the mean of the prior math MSA z-scores of teacher j’s student cases; 
(BMzMean\textunderscore mean)\textsubscript{j} is the mean, for teacher j, of his/her student-cases’ benchmark z-score average; 
u\textsubscript{0j} indicates deviation from the math MSA grand mean of teacher j’s student cases; 
u\textsubscript{kj}, with k = 1, 2, indicates, for L1 predictor k, teacher j slope deviation from the overall slope; e.g., u\textsubscript{2j} indicates the deviation in the prior math achievement slope of teacher j from the overall prior math achievement slope; and 
ri\textsubscript{ij} represents student i’s deviation from the math MSA mean of teacher j’s student cases.

**MLM summary.** MLM analysis began with development of a linear mixed model with no predictors that determined that significant levels of variance in math achievement existed between teachers and within teachers. Subsequently, a progression of three increasingly complex two-level multilevel models were fitted to the data to predict math achievement from student-level (L1) and teacher-level (L2) predictors: student gender, prior math performance, math benchmark exam performance at L1, and teacher years of experience, teacher certification level, and aggregates of students’ prior math achievement and benchmark exam performance at L2.

The student-level predictors contributed significantly to all (non-null) models. Performance on benchmark exams had the greatest effect on math achievement in each of the (non-null) models; prior math achievement had the second-highest effect in each of the (non-null) models. In each of the (non-null) models, boys were expected to outperform girls on the math MSA exam by about 2 points.
Teacher characteristics (certification level and years of experience) did not contribute significantly in any of the (non-null) models. However, student characteristics aggregated at the teacher level did significantly contribute—in particular, benchmark performance (in Models 2 and 3) and prior math achievement (in Model 2) aggregates. Model fit improved through the successive models, as evidenced by a reduction in residuals and intercept variances (see Table 13) and smaller 95% confidence intervals (Table 12) from model to model. However, each model had significant variability in math achievement between and within teachers that were not explained by the model (see Table 12).
Chapter 5 – Discussion

The purpose of this study was to explore relationships among education, society and technology by comparing traditional and more recently developed math achievement predictive models and considering how the results of the models might impact mathematics education. The study focused on three research questions: RQ1 and RQ2 addressed the question of how multiple linear regression and multilevel modeling inform the prediction of mathematics achievement from linear combinations of student characteristics and teacher characteristics. RQ3 compared the MLR model results and the MLM model results with respect to K-12 policy and practice.

Research Question 1

RQ1: How does multiple linear regression inform the prediction of mathematics achievement from a linear combination of student characteristics—gender, previous mathematics achievement, district benchmark exam performance—and teacher characteristics—professional certification level and years of experience?

Multiple linear regression produced a model in which the combination of student and teacher characteristics significantly predicts mathematics achievement, $F(5,5218) = 2452.845$, $p < .001$. Greater than 70% of the variance in math achievement as measured by the state-mandated math MSA exam is attributable to the combined effect of these predictors. Each predictor in the model contributed significantly to math achievement prediction in the following order of impact (highest to lowest, as the standardized coefficients in parentheses show): student benchmark exam performance ($\beta = .475$, $p < .001$), student prior math achievement ($\beta = .420$, $p < .001$), teacher professional certification level ($\beta = .055$, $p < .001$), student gender ($\beta = .024$, $p = .002 < .05$), and teacher years of experience ($\beta = .024$, $p = .015 < .05$). The ordering of the beta values associated with benchmark exam performance and prior math achievement suggests that
students are responsive to instruction; otherwise, prior achievement would have risen to be the most influential predictor. The MLR model (see equation 6) suggested that

1. After adjusting for student and teacher characteristics, the expected average math MSA score over all students is 417.

2. Boys are expected to outperform girls on the math MSA by about 2 points.

3. Of the predictors in the model, benchmark exam performance has the greatest impact on math achievement, followed closely by prior mathematics achievement. An increase of one standard deviation in a student’s mean benchmark performance results in a 22-point increase on the math MSA exam. An increase of one standard deviation on a student’s prior math MSA exam produces an increase of 17 points on the math MSA exam.

4. Students whose teachers have advanced professional certification are expected to score 5 points higher on the math MSA exam.

5. For each additional year of a teacher’s experience, his/her students are expected to score 0.8 point higher on the math MSA exam; i.e., students are expected to score 4 points higher for every five years of teacher experience.

6. Although each predictor contributes significantly to math achievement in the MLR model, when compared to prior math achievement and benchmark exam performance, teacher certification and experience are relatively weak contributors to math achievement.

Research Question 2

RQ2: How does hierarchical linear modeling, with students and teachers as the respective level-1 and level-2 sampling units, inform the prediction of mathematics achievement from
linear combinations of the same student characteristics—gender, previous mathematics achievement, district benchmark exam performance—and the same teacher characteristics—professional certification level and years of experience?

Modeling a sequence of increasingly complex multilevel models began with a null model in which there were no predictors and culminated in the development of a two-level random slope and intercept model that used student characteristics in the L1 analysis and used teacher characteristics as well as student characteristics that were aggregated by teacher in the L2 analysis. The intraclass correlation of the null model, $\rho = .351$, indicates that 35% of the total variability in math achievement lies between teachers and 65% of the total variability lies within teachers.

**MLM Model 1: Student-level random intercept model.** Model 1 of the sequence is a model in which only student-level predictors are included. Each of the predictors in this model is statistically significant: benchmark exam performance ($\gamma_{30} = 20.11$, $t = 43.004$, $p < .001$), prior math achievement ($\gamma_{20} = 17.48$, $t = 42.738$, $p < .001$), and gender ($\gamma_{10} = 1.86$, $t = 4.119$, $p < .001$). In Model 1, 32% of the total variability in math achievement lies between teachers and 68% lies within teachers. Addition of the student-level predictors produced a model that accounts for 72% of the within-teacher variance in math achievement and 75% of the between-teacher variance.

The L1 random intercept model (see equation 8) suggests that after adjusting for student gender, benchmark exam performance and prior math achievement,

1. The estimated average mean math MSA score by teacher (i.e., the teacher-level intercept) is 426.
2. Boys are expected to outperform girls on the math MSA by about 2 points.
3. Of the predictors in the model, benchmark exam performance has the greatest impact on math achievement. An increase of one standard deviation in a student’s mean benchmark exam performance is expected to result in a 20-point increase in the math MSA exam score. An increase of one standard deviation on a student’s prior math MSA exam score is expected to produce an increase of 18 points on the math MSA exam.

**MLM Model 2: Teacher-level random intercept model.** Model 2 built upon Model 1 by adding teacher-level predictors. As in Model 1, the student-level predictors contribute significantly to the model: benchmark exam performance ($\gamma_{30} = 19.91$, $t = 42.253$, $p < .001$), prior math achievement ($\gamma_{20} = 17.58$, $t = 42.655$, $p < .001$), and gender ($\gamma_{10} = 1.86$, $t = 4.122$, $p < .001$). Of the additional teacher-level predictors, only those that are aggregates of L1 student characteristics significantly contribute to the model: benchmark exam performance ($\gamma_{03} = 12.91$, $t = 3.525$, $p = .001 < .05$) and prior math achievement ($\gamma_{03} = -7.33$, $t = -2.287$, $p = .024 < .05$).

In Model 2, neither teacher certification level ($\gamma_{01} = -.60$, $t = -.187$, $p = .852 > .05$) nor teacher years of experience ($\gamma_{02} = .05$, $t = .344$, $p = .732 > .05$) contribute significantly to prediction of math achievement.

In Model 2, 30% of the total variability in math achievement lies between teachers and 70% lies within teachers. Addition of the teacher-level predictors produces a model that accounts for 72% of the within-teacher variance (no change) in math achievement and 84% of the between-teacher variance (an increase of 9%).

The teacher-level random intercept model (see equation 9) suggests that after adjusting for student- and teacher-level characteristics included in the model,

1. The estimated average mean math MSA score by teacher is 425.
2. Boys are expected to outperform girls on the math MSA by about 2 points.

3. A 20-point increase in math MSA score is expected for an increase by a standard deviation in the mean of a student’s benchmark exam performance.

4. An increase of one standard deviation in prior math MSA performance is expected to produce an 18-point increase in math MSA score.

5. For an increase of one standard deviation in the average by teacher of his/her students’ mean benchmark scores, a 13-point increase in math MSA score is expected.

6. A seven-point decrease is expected in math MSA score for one standard deviation increase in the average by teacher of his/her students’ prior math MSA scores.

7. Teacher certification and experience are weak contributors to math achievement; and students whose teachers have advanced certification are expected to score slightly lower on the math MSA than those whose teachers have standard certification.

**MLM Model 3: Random slope and intercept model.** Model 3 uses the same L1 and L2 predictors as Model 2, but allows the benchmark performance-achievement slope and the prior math performance-achievement slope to vary across teachers. As with all the other (non-null) MLMs, student-level predictors contribute significantly to the model: benchmark exam performance \((\gamma_{30} = 23.06, \ t = 25.75, p < .001)\), prior math achievement \((\gamma_{20} = 15.61, \ t = 23.95, \ p < .001)\), and gender \((\gamma_{10} = 1.89, \ t = 6.48, p = .001 < .05)\). Again, teacher years of experience \((\gamma_{02} = -.17, \ t = -.05, p = .958 > .05)\) and teacher certification level \((\gamma_{01} = .04, \ t = .29, \ p = .773 > .05)\) do not contribute significantly to the model. In Model 3, the only teacher-level predictor that is statistically significant in the model is the aggregate benchmark exam predictor \((\gamma_{03} = 10.07, \ t = 2.88, p = .005 < .001)\). Allowing the slopes to vary across teachers produces a
model that accounts for 80% of the within-teacher variance in math achievement (an increase of 8%) and 93% of the between-teacher variance (an increase of 9%).

The random slope and intercept model (see equation 10) suggests that after adjusting for student and teacher characteristics and allowing individual student characteristic slopes to vary randomly in the model,

1. The estimated average mean math MSA score by teacher is 423.
2. Boys are expected to outperform girls on the math MSA by about 2 points.
3. A 23-point increase in math MSA score is expected for each standard deviation increase in the mean of a student’s benchmark exam z-scores.
4. A standard deviation increase in prior math MSA z-score is expected to produce a 16-point increase in math MSA score.
5. For each standard deviation increase in the in the average by teacher of his/her students’ mean benchmark z-scores, an 11-point increase in math MSA score is expected.
6. A five-point decrease is expected in math MSA score for each standard deviation increase in the average by teacher of his/her students’ prior math MSA scores.
7. Teacher certification and experience are weak contributors to math achievement; and students whose teachers have advanced certification are expected to score slightly lower on the math MSA than those whose teachers have standard certification.

Research Question 3

RQ3: How do MLR and MLM results compare for informing K-12 practice and policy with respect to student mathematics achievement?
Importance of the study’s larger context. Before discussing implications of this study’s results for education practices and policy, it is important to go back to the larger context of this study: the relationships among society, education, and technology. Within these connections lie motivating factors for improving math education and ensuring student success in mathematics.

First, the well-being of U.S. society is dependent on the outcomes of mathematics education (National Research Council, 2007; National Research Council, 2010), including: U.S. national security (National Mathematics Advisory Panel, 2008), the United States’ ability to innovate and to compete in global markets (National Mathematics Advisory Panel, 2008), the prosperity as a nation (National Research Council, 2007), and the capacity to meet workforce skill demands (Hanushek et al., 2010; National Mathematics Advisory Panel, 2008; National Research Council, 2007), particularly in STEM fields.

Additionally, the well-being of individuals in U.S. society is reliant on math education outcomes because math achievement impacts

- school readiness (Culpepper et al., 2010);
- college readiness (Achieve, 2009; Culpepper et al., 2010; Zelkowski, 2011);
- career readiness (Achieve, 2004; Achieve, 2009; Achieve, 2011; Culpepper et al., 2010);
- graduation from high school (Baum & Payea, 2005);
- access to and/or enrollment in college (Achieve, 2011; National Mathematics Advisory Panel, 2008; Spielhagen, 2006);
- retention in college (Achieve, 2011; Moses et al., 2011; Scott et al., 2009);
- success in college (Achieve, 2011; Culpepper et al., 2010; National Mathematics Advisory Panel, 2008; Sadler & Tai, 2007; Scott et al., 2009); and
bachelor’s degree completion (Adelman, 2006; Baum & Payea, 2005; Zelkowski, 2011).

Math achievement contributes to productive adult life (National Mathematics Advisory Panel, 2008), to an individual’s capacity for higher-order thinking (Achieve, 2011), and to one’s earning potential (Badenhausen, 2008; Carnevale & Desrochers, 2002; National Mathematics Advisory Panel, 2008; Rose & Betts, 2001). Furthermore, in the current information age, policy leaders and citizens in general need quantitative skills (National Mathematics Advisory Panel, 2008) to be able to process data-based reports in the workplace and in local, regional, federal and world news and to understand potential consequences of decisions that are made with respect to data.

Realizing the value of education in general, U.S. society has made and is making efforts to improve education: the federal government has passed legislation and appropriated funds to programs that improve education. Such programs include those that focus on school improvement, teacher and principal quality, development of academic standards and assessment instruments, measurement of academic achievement, development of education data systems, and education research (U. S. Department of Education, Office of Communications and Outreach, 2010; U.S. Department of Education, Office of Communications and Outreach, 2012). Specific to math education, the federally funded Mathematics and Science Partnerships program was established to improve content knowledge of teachers and improve student performance in math and science. Specific to teacher certification, the goal of the Advanced Certification or Advanced Credentialing program was to improve teaching and learning and school leadership through advanced certification efforts (U. S. Department of Education, Office of Communications and Outreach, 2010).
Along with efforts to improve education, U.S. society has a renewed interest in education accountability. The *No Child Left Behind Act of 2001* placed accountability for academic achievement of all students on states, LEAs, and schools (*No Child Left Behind Act of 2001*, 2002). More recently, with programs such as *Race to the Top*, evaluation of and/or compensation to educators are being based in part on student academic growth (Teachers reach agreement on pay raises.2013; *Race to the top bonus payments to cetified (sic) personnel in persistently low-performing schools*.2013; Azuz, 2012; Gadsden County Schools, n.d.). Individual educators are now being held accountable for students’ academic growth.

Technology is now critically tied to society and education as it plays a major role in many education improvement efforts. Benchmarking systems provide student performance data that can be used for formative assessment. Data systems facilitate the collection and storage of data and provide access to educators and researchers who can use the data to improve education and to communicate to stakeholders. State longitudinal data systems enable analyses of student growth from kindergarten through college and enable tracking of students as they enter the workforce. Statistical modeling contributes to the identification of low- and high-achieving schools and factors that contribute to student, teacher, principal and school success.

But why should stakeholders care about differences that are due to technological advances in statistical modeling? How might policy and practice in mathematics education differ based on differences in statistical modeling results? What do the study results imply for policy and practice in K-12 mathematics education?

To address these questions,

1. Key differences and similarities of the results from the two modeling approaches are summarized;
2. Policy/practice differences due to statistical modeling choice are illustrated; and

3. Implications for K-12 policy and practice, based on MLM results and the similarities of the MLR and MLM results, are presented.

**Key differences and similarities in modeling results.** The MLR and MLM models’ estimated parameters are shown in Table 23 for the purpose of comparing modeling results. As the models’ estimated parameters show, the two modeling approaches used in the study revealed the following similarities among the MLR and MLM results:

1. Student benchmark exam performance is a positive and significant predictor of math achievement in the MLR and MLM models.

2. Student prior mathematics performance is a positive and significant predictor of math achievement in both modeling approaches.

3. Student gender is a positive, albeit weak, predictor of math achievement in the MLR and MLM models, accounting for an expected two-point higher math MSA score for boys than for girls in all the models.

There is a substantial difference in the results of the two modeling approaches: In the traditional MLR model, all student-based predictors are significant and all teacher-based predictors are significant: student gender, prior math achievement, math benchmark exam performance, teacher years of experience, and teacher level of professional certification (see estimated parameters in Table 23). However, only student characteristics contribute significantly to the prediction of mathematics achievement in the more recently developed MLM approach.
### Table 23

*Models’ Estimated Parameters*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLR Model</th>
<th>MLM Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model 1</td>
</tr>
<tr>
<td>Intercept</td>
<td>417.29** (.605)</td>
<td>425.54** (1.19)</td>
</tr>
<tr>
<td>Student gender</td>
<td>1.71* (.54)</td>
<td>1.86** (.45)</td>
</tr>
<tr>
<td>Benchmark exam performance</td>
<td>21.52** (.50)</td>
<td>20.11** (.47)</td>
</tr>
<tr>
<td>Prior math achievement</td>
<td>16.82** (.45)</td>
<td>17.48** (.41)</td>
</tr>
<tr>
<td>Teacher certification level</td>
<td>4.53** (.78)</td>
<td>-0.60 (3.20)</td>
</tr>
<tr>
<td>Teacher years of experience</td>
<td>.08* (.03)</td>
<td>0.04 (.15)</td>
</tr>
<tr>
<td>Students' aggregate prior math achievement</td>
<td></td>
<td>-7.33* (3.21)</td>
</tr>
<tr>
<td>Students’ aggregate benchmark exam performance</td>
<td></td>
<td>12.91* (3.66)</td>
</tr>
</tbody>
</table>

*Note:* Standard errors are in parentheses.

* *p < .05, **p < .001
Implications for K-12 policy and practice. Because the data set used in this study is nested and significant variation in math performance exists between teachers’ groups of students, implications for K-12 policy and practice are based primarily on the MLM results (and implicitly on the MLR results that are consistent with the MLM results).

Statistical modeling in education. A major point to be taken from this study is that the choice of statistical model for the analyses of education data can impact K-12 policy and practice. The fact that the MLR and the MLM statistical models in this study, while using the same data set, yielded some substantially different results highlights the importance of using appropriate statistical methods/models in education research.

Given the importance of math education to U.S. society, the societal demand for education accountability, and the emphasis on data-driven decision making, education stakeholders will benefit from the use of statistical models that are as accurate as possible and that reduce the potential for errors. Given the recent practice of basing teacher and/or principal evaluation in part on student academic growth, it is important to use statistical models that account for student characteristics at the point a student enters a school or classroom and that provide conservative estimates with respect to the statistical significance of variables that influence academic achievement. It is also important to use statistical models that are theoretically sound for the particular structure of the data that is being modeled.

To illustrate the potential impact of statistical modeling practices on education policy and practice, differences in decisions that could very well be made based on the different results of the two modeling approaches in this study are discussed next.

Practice and policy differences relating to teacher experience. The significance of teacher years of experience as a predictor of mathematics achievement in this study is
inconsistent across models. Of the three models in which it is included, teacher experience is a significant predictor of math achievement in the MLR model; however, teacher experience is not a significant predictor of math achievement in any of the MLMs. Therefore, K-12 policies and practices related to teacher experience would likely differ based on which statistical model is used.

For example, basing policy and practice decisions on MLR results, which deem teacher experience to be a positive and significant predictor of math achievement, administrators would try to recruit and retain teachers with a greater number of years of teaching experience; and it makes sense to include years of experience in teacher compensation models that are based in part on student achievement (though teacher experience would not be weighted heavily in the model because it is a weak predictor). Administrators might also place students who are not performing well in math with teachers who have more experience. Decision and policy makers at state and federal levels might institute incentive programs with a goal of improving learning outcomes at poorly performing schools by relocating more experienced teachers to such schools. Given the advantages that MLM offers over MLR for analysis of hierarchical data sets, the significance of teacher experience as a predictor in the MLR model should be viewed with caution.

Conversely, based on MLM results, which deem teacher experience to be a statistically insignificant predictor of math achievement, administrators would not necessarily consider years of experience in hiring decisions, compensation policies or tenure decisions, nor would they have expectations of improved performance of students due to a placement with a teacher who has a greater number of years of teaching experience. If the focus at the state education level is on student achievement, policy makers would question whether teacher experience should be included as a criterion for advanced certification, tenure, and teacher compensation models.
**Practice and policy differences relating to teacher certification.** As is the case with teacher years of experience, advanced teacher professional certification as a predictor of math achievement is significant only in the MLR model. However, another inconsistency exists: advanced teacher certification has a positive effect on math achievement in the MLR model and a negative effect on math achievement in the two MLMs in which it is included. Again, the difference in modeling results will produce different impacts on K-12 policy and practice.

Based on the MLR results, administrators would likely try to recruit and retain teachers with advanced certification, because the MLR model suggests that teachers with advanced certification will produce positive and significant results in their students’ math achievement—about five points higher on the math MSA exam than students whose teachers have standard certification. An administrator might conclude that advanced teacher certification will contribute significantly to attainment of a school’s AYP. Therefore, based on the MLR results, administrators may also strive to provide incentives for teachers to acquire advanced certification, to pay higher wages to teachers with advanced certification, and to place low-achieving students with teachers who have advanced certification. Administrators may also look to teachers with advanced certification to mentor teachers with standard or provisionary certification. It is feasible that state and federal decision makers would continue allocating significant funding to promote teacher attainment of advanced certification, as the federal government did in 2008 through 2010 with allocations of more than $30 million to advanced certification programs (see Table 2).

On the other hand, based on the MLM results, administrators would likely be much less inclined to hire and retain teachers with advanced certification—especially if salary structures require higher pay for teachers with advanced certification. The expectation would be that
students whose teachers have advanced certification would slightly underperform in math compared to students whose teachers have standard certification. Furthermore, advanced certification would not be considered in compensation models or tenure decisions. At the federal level, funding of education programs to improve student performance would not focus on helping teachers acquire advanced certification.

These examples of policy/practice differences due to choice of statistical approach in the analysis of education data show the value of emerging, more sophisticated technologies as tools to inform education practice and policy. As leaders and educators strive to make data-informed decisions and policies to improve education and to allocate resources where the resources will have the greatest impact, it is critical—when significant variance by groups exists—to use statistical models such as MLM that are designed for dealing with nested data and are thereby less likely to produce spurious significant results. For purposes of recruitment, hiring, compensation, resource allocation, reform initiatives, policymaking, decision making and effective practice in K-12 math education, it is vital that leaders, educators and researchers use data and appropriate statistical models to guide U.S. schools in the delivery of the best mathematics education possible for the nation’s students. Furthermore, in fairness to educators who are increasingly being held accountable for student achievement, it is important to use statistical models that moderate for student characteristics and school characteristics.

**Gender gap in mathematics performance.** Student gender is a positive and significant, albeit weak, predictor of math achievement in the MLR model and in each MLM. The effect of gender is consistent across models: boys are expected to outperform girls on the math MSA exam by about two points. The significance of gender as a predictor of math achievement in the
models in this study concurs with the literature that suggests a gender gap exists in mathematics performance, with males outperforming females (e.g., Fryer & Levitt, 2009; Lee et al., 2011).

Because the United States’ economic leadership, national security, and global competitiveness is dependent upon training more scientists and engineers, which in turn is dependent on mathematics achievement, gender gaps in mathematics performance must be eliminated to fulfill the potential of all students in mathematics. For K-12 practice and policy, this means working with researchers to identify gaps as early as possible, to identify factors—including social, geographic, and cultural factors—that contribute to those gaps, to identify practices that can help eliminate the gaps, to put those practices in place, and to assess the outcomes of those practices.

**Early mathematics education.** Prior math performance at the individual student level is a positive and significant predictor of math achievement in the MLR model and in each MLM in which it was included. It is the second-strongest predictor in all the models. This adds to the body of literature that indicates that prior math knowledge significantly contributes to mathematics achievement (e.g., Alban, 2002; Bode, 1996; Curcio, 1981; Germuth, 2003; Pinkham & Ansley, 1996). As a teacher-level predictor of math achievement, prior math performance contributes significantly in one of the two MLMs in which it is included. An unexpected result is that although prior math performance at the student level has a positive effect on math achievement, prior math performance has a negative impact as a teacher-level predictor in the MLMs. More investigation is needed to determine why this aggregate phenomenon occurs across teachers. The negative effect of aggregate prior math achievement may have to do with the manner in which students are assigned to teachers. One possibility is that groups of students who performed very well on their previous math MSA exams may have
been assigned to a teacher that seems to have a low or negative impact on their current achievement. On the other hand, the situation could be that students who underperformed and/or performed poorly on their previous math MSA exams were assigned to a particularly effective teacher. Or, a particular teacher or two may be outliers who have a huge impact on their students’ math achievement. Regardless of the negative aggregate impact of prior math performance, the net effect of prior math achievement is positive in both MLMs.

What does this relationship between prior math performance and math achievement imply for policy and practice in K-12 mathematics education? Mathematics is a discipline in which new skills are built upon prior skills; e.g., a student will likely not be successful in solving an equation for $x$ before the student is able to add, subtract, multiply and/or divide. The relationship between prior math knowledge and math achievement revealed in the literature and in this study emphasizes the following with respect to K-12 policy and practice in mathematics education:

1. Math knowledge and skill gained in early years of education impact students’ future success in math. Therefore, it is critical that effective mathematics education begin in kindergarten for students who are school-ready and in pre-kindergarten programs for students who are not school-ready.

2. It is vital that effective math education continue through students’ successful completion of Algebra II, not only because of the relationship between prior math performance and math achievement, but also because of the specific relationship between success in algebra and
   - students staying in math longer (Spielhagen, 2006),
   - improved college readiness (Davis, 2010; Edge, 2009),
greater rates of college attendance (Adelman, 1999; National Mathematics Advisory Panel, 2008; Spielhagen, 2006), and

- increased rates of graduation from high school and college (Adelman, 1999; Baum & Payea, 2005; Davis, 2010).

Ensuring the success of students through Algebra II builds for them a strong foundation for continued success and increases the nation’s potential to meet workforce demands for positions that are dependent on math knowledge and skill and to compete in a global economy.

3. As funding and resource allocation decisions for math education are made at federal, state, and local levels and by philanthropic institutions, serious consideration should be given to fund programs and practices that ensure early and continued mathematics success, minimally through student success in Algebra II.

4. Educators should consider instituting systems of formative assessment of students’ math skills with follow-up interventions to eliminate skill gaps early in students’ education. Waiting until third or fourth grade can seriously disadvantage students who struggle with math in the early grades.

*The promise of benchmarking systems.* Not only is benchmark exam performance a positive and significant predictor of math achievement in the MLR model and in each MLM, it is the strongest predictor in all the models in which it is included. Furthermore, benchmark exam performance is the only teacher-level predictor that contributes significantly to each of the two MLMs in which it is included. The significant, strong, and positive effect of student benchmark exam performance in every model adds to the literature that suggests benchmarking systems positively impact mathematics achievement (e.g., Bettger, 2007; Potteiger, 2008; Stoltz, 2008;
The strength of benchmark exam performance as a predictor of math achievement also supports the National Mathematics Advisory Panel’s (2008) recommendation for schools to regularly use formative assessment for elementary school students.

As state departments of education, K-12 decision makers, and teachers consider implementation of benchmarking systems, it is vital to realize that merely adding more assessments to the classroom experience is not the key to improving math achievement. In order to effectively inform quality improvement, the assessments must be aligned to desired learning outcomes. The recently developed Common Core State Standards for mathematics are the logical choice as outcomes for alignment with benchmarking systems as 45 states, Washington, DC, four territories and the Department of Defense Education Activity have adopted the Common Core State Standards (Common Core State Standards Initiative, 2012).

Impacts of implementing and sustaining district-wide and state-wide benchmarking systems aligned with the Common Core State Standards to be considered by decision makers include the following:

1. A benchmarking system provides to educators timely data to inform teaching and learning practices in their classrooms and schools. Benchmark exam data can be used not only to identify learning gaps of individual students, but also to help identify particular math education strengths and weakness of schools within a district or the state as well as math topics or teaching methods for which teachers may benefit from additional professional development.

2. A benchmarking system creates potential for improved individual student learning across the school district throughout the school year. Short-term benefits to students of greater learning include improved readiness for subsequent math classes, greater
self-confidence, and improved achievement on state mathematics assessments. In the long term, student benefits may include greater levels of college and career readiness (Davis, 2010; Edge, 2009), enhanced ability for abstract reasoning (Achieve, 2011), success in college and the workplace, and well-being in society (Achieve, 2011).

3. District-wide and state-wide benchmark exams that are aligned with the Common Core State Standards will likely benefit students who frequently change schools within a district or the state because (a) the targeted grade-level benchmark learning outcomes are consistent across the district and state, and (b) the intake school and teacher have access to data that identifies the students’ specific skill gaps.

4. Benchmark data can also be used to help identify teachers who seem to be particularly effective at teaching particular topics and teachers whose students consistently show strong growth in math skills. Further examination of characteristics, methodologies, and classroom atmospheres of these teachers while controlling for student and schools characteristics can provide more insight into what makes an effective teacher. Administrators may want to assign teachers who are identified as highly effective as peer mentors to teachers who are struggling or have the highly effective teachers provide professional development sessions for the district.

5. Improved individual student learning illuminates the success of teachers, principals, school districts and the state education departments that support schools.

To effectively use a benchmarking system for formative purposes, decision makers must ensure that the system provides data to educators and researchers in a format that can be used to improve teaching and learning. Therefore, education data systems with user-friendly dashboards
are needed along with training for administrators, teachers, and researchers that help them use data to plan and provide learning opportunities for students as well as professional development for teachers that will bolster their weak areas.

**Teacher effectiveness.** The teacher characteristics included in this study—years of experience and level of certification—fared poorly as predictors of mathematics achievement: neither characteristic significantly contributes to the prediction of math achievement in the MLMs. Moreover, advanced teacher certification is expected to have a slightly negative impact on student mathematics achievement; and teacher experience is the weakest predictor of math achievement in all the models in which it was included—MLR and MLMs. These results challenge current, long-standing recruitment, hiring, compensation and tenure policies and practices in education that are based on underlying assumptions that more teaching experience and/or advanced certification equate to better teaching.

The results of this study with respect to teacher characteristics cause one to question how to define teacher effectiveness. As an example, Maryland’s teachers are to be evaluated on student growth and professional practice, each accounting for 50% of the evaluation. The professional practice component includes four domains: planning and preparation, classroom environment, instruction, and professional responsibilities. One might ask: How are these domains measured? How do these domains relate to student achievement? What planning and preparation practices have been identified as significant contributors to students’ math achievement? What instructional methods and what practices in establishing classroom environment have been shown to contribute to student achievement? How do a teacher’s professional responsibilities relate to student achievement? How should levels of teacher effectiveness defined, measured, and rewarded?
Beyond the need to define, measure, and reward teacher effectiveness in a meaningful way, there is need to evaluate the value of teacher certification. Certainly, because teacher certification requirements and levels of certification vary from state to state, more research is needed regarding the impact of teacher certification on mathematics education. However, the mixed results in the literature and the results of this study with respect the influence of teacher certification on math achievement call to question the value of teacher certification and advanced levels of teacher certification.

At the state level, if improved student achievement is the goal, then policy makers will want to consider questions related to teacher certification such as:

1. If teacher years of experience do not significantly contribute to math achievement, should teacher years of experience be used as a criterion for advanced certification of math teachers?

2. What criteria could be used for varying levels teacher certification to ensure the certification adds value to educational systems: content knowledge, evidence of student academic growth, pedagogy, classroom management?

3. What teacher characteristics, professional development programs, teaching approaches, classroom management techniques, etc., have been shown through research to have a positive impact on student achievement?

4. Given mixed reports in the literature of the impact of teacher certification on student achievement, is teacher certification, as it is now awarded, relevant and necessary for quality education?

At an inter-state level, decision makers would investigate similar issues due to inter-state teacher certification reciprocity agreements. With the Common Core State Standards as targeted
learning objectives across states, inter-state teacher certification criteria could be based in part on student attainment of the Common Core State Standards.

At the broader national level, decision makers may want to initiate further research programs that control for student characteristics and investigate whether and to what extent relationships exist between student mathematics achievement and teacher characteristics, e.g., level of certification, years of experience, college GPA in math courses, college GPA in education courses, subject-exam scores in subjects in which teachers are certified to teach, classroom practices, involvement with students in extra-curricular activities, particular professional development activities, etc. If sufficient research-based evidence can be gathered to associate particular teacher characteristics/accomplishments with students’ math achievement, then education leaders should appropriate significant portions of federal education funding to education initiatives that promote attainment of those specific characteristics.

**Teacher preparation.** The needs to define teacher effectiveness and to evaluate the value and relevance of teacher certification directly relate to a need for determining the efficacy and relevance of teacher education programs. Teacher education programs in the United States have been “undergoing an unprecedented degree of scrutiny and challenge” (Russell & Wineburg, 2007, p. 1). To improve the effectiveness of teacher education programs, the American Association of State Colleges and Universities (AASCU) has proposed a national framework—“a set of guidelines that institutions might use to guide data collection as they audit their programs to provide evidence of effectiveness” The framework is based on two premises:

…first, that teacher education accountability is an important and legitimate goal, particularly for state institutions that have an obligation to be accountable to the public; and, second, that robust, evidence-based systems demonstrating effectiveness must be in
place to achieve educational outcomes, guide program improvement, and assure and protect the public” (Russell & Wineburg, 2007, p. 2).

Russell and Wineburg (2007) propose building upon what has already been learned from efforts such as measuring content knowledge and teacher performance, measuring student achievement, measuring retention and support of teachers, and developing systems for collecting evidence of the effectiveness of teacher education programs.

Given (a) the significant impact of students’ prior math performance on math achievement and (b) the importance of effective mathematics education through Algebra II to the individual and the nation, what should be done to better prepare 21st-century math teachers? Questions for leaders, policy makers, and researchers to consider include:


2. Do differences in entrance standards for teacher education programs contribute to teacher effectiveness? To what extent?

3. Do differences in teacher education curricula contribute to teacher effectiveness? To what extent? Should teacher preparation programs for math teachers require more math courses in the teacher education curriculum?

4. Do differences in graduation requirements from teacher preparation programs contribute to teacher effectiveness? To what extent? Should graduation requirements depend more heavily upon demonstrated excellence of a more rigorous set of math skills and understanding?

From another perspective, leaders, policy makers, and researchers may consider the following:
5. Are four- or five-year teacher education programs necessary to produce effective mathematics teachers?

6. What teacher characteristics and skills sets correlate with student math achievement, and what alternate routes may be as effective or more effective in bringing high quality math teachers into classrooms across the nation?

**Student-centered funding: put the student back in the education equation.** The facts that (a) in both modeling approaches, all of the student characteristics are significant predictors of math achievement; and (b) in the MLM model, which is designed to account for the nesting of data, the teacher characteristics are not significant predictors of math achievement suggest that an examination of resource allocation in education is in order.

Almost 15 years ago, the state of Maryland established *Every Child Achieving: A Plan for Meeting the Needs of the Individual Learner* in response to achievement gaps among students. This plan focused in part on educators, recommending strategies for “building and improving the skills of teachers and the leadership of administrators” (U.S. Commission on Civil Rights - Office of the General Counsel, 2004, p. 118). The plan did not significantly improve test performance or reduce achievement gaps (U.S. Commission on Civil Rights - Office of the General Counsel, 2004).

Moreover, the overarching goals of many federal education programs are to improve the quality of education and to increase students’ academic achievement. However, a good number of U.S. education programs target student achievement indirectly by focusing directly on educators. Referring to Table 2, one can see that in a five-year period, the federal government allocated more than $16.2 billion to

- encourage and support teachers seeking advanced certification,
- improve teacher and principal quality,
- increase the number of highly qualified teachers,
- increase teacher and principal effectiveness,
- improve teacher preparation,
- recruit highly qualified individuals into the teaching force from other occupations,
  and
- create performance-based compensation systems to reward educators for student achievement.

If decision makers and legislators continue trying to improve student achievement by channeling funds to educators as an indirect benefit to student learning, will students realize sufficient mathematics achievement to spur their success in college and life in the 21st century? Will the nation’s education systems produce sufficient numbers of mathematically skilled high-school and college graduates who can contribute to and maintain the United States’ eminence, well-being, safety, innovation, and global competitiveness? Where might federal and state education funding be shifted to more directly promote student success, thereby indirectly effecting teacher and principal success? How should resources be allocated in education to most efficiently and directly improve student achievement?

The results of this study and the mixed results in the literature with respect to student mathematics achievement, viz., the significance of student characteristics and the non-significance of teacher characteristics, suggest that it is time to put students back in the education equation: to make federal and state education policy and practice decisions based on students’ needs and to modify allocation practices and policy to directly promote student success.
Increasing numbers of schools are adopting student-centered learning (SCL) principles and practices in response to student learning expectations instituted through legislation and programs such as the *No Child Left Behind Act* and the *Race to the Top* program (Miller, Gross, & Ouijdani, 2012). The focus of SCL is learning as opposed to teaching; and SCL approaches are driven by the “knowledge, skills, interests, and needs of the students” (Randolph Public Schools, para. 4). SCL practices include personalized learning, authentic instruction, mastery-based assessment, learning beyond the school walls (e.g., personalized learning through online/blended programs, dual enrollment, project-based or community-based learning (U.S. Department of Education, n.d.a)), and models that change the school schedule (Miller, Gross, & Ouijdani, 2012).

Student-centered allocation models are based on equity of funding for individual students to meet all students’ learning needs (Western Australia Department of Education, 2013). These models concentrate resources on learning and on providing new programs for students. For example, a school might reallocate support and administrative resources—based on students’ needs—to differentiate instruction, use mastery-based assessment (Miller et al., 2012), or implement flexible scheduling, an internship program, student choice in the curriculum, new attendance policies, or virtual schooling (Levin, Datnow, & Carrier, 2012). Student-centered allocation models do not necessarily require greater funding (Levin, Datnow, & Carrier, 2012; Miller et al., 2012).

To more directly impact student achievement, federal education programs should focus on the development and implementation of student-centered funding policy and programs based on principles, practices, and/or recommendations such as the following:
1. Efficiency, equity, autonomy, transparency and accountability are valued characteristics of student-centered funding (Boston Public Schools, n.d.; Doyle, Hassel, & Locke, 2012; Fermanich & Hupfeld, 2009; New York City Department of Education, Division of Finance, 2012; Taylor, 2009).

2. School funding is based on student needs (Boston Public Schools, n.d.; Doyle et al., 2012; Western Australia Department of Education, 2013).

3. Funds are allocated on a per-student basis. The funding follows the student to his/her school of attendance (Boston Public Schools, n.d.; New York City Department of Education, Division of Finance, 2012; Snell, 2013; Taylor, 2009; Western Australia Department of Education, 2013).

4. Per-student funding levels vary based on
   - grade span (Boston Public Schools, n.d.; New York City Department of Education, Division of Finance, 2012; Perry, 2012; Western Australia Department of Education, 2013);
   - other student needs, e.g., students with disabilities, social/economic disadvantage, ESOL students, gifted and talented students, pregnant students (Boston Public Schools, n.d.; Perry, 2012; Taylor, 2009; Western Australia Department of Education, 2013); and/or
   - school size, e.g., small and/or rural schools receive additional funding (Western Australia Department of Education, 2013).

5. Per-student funding should be in the form of real dollars that can be spent flexibly to meet student need, as opposed to coming in the form teaching positions, student-teacher ratios, etc. (Thomas B. Fordham Institute, 2006).

7. Data and reporting systems that link funding, expenditures and student outcomes are needed; e.g., state longitudinal systems that house student and teacher information should be linked to a financial data repository that holds revenue and expenditure data (Doyle et al., 2012; Fermanich & Hupfeld, 2009).

Summary

In this study, traditional and more recently developed predictive models in the analyses of mathematics achievement were compared and contrasted. The overarching goal of the study was to examine relationships among society, education, and technology. The relationships were investigated through (a) reviewing the literature to establish a relational context of these three components; (b) conducting a multiple linear regression and a series of multilevel models using the same data set to examine the relationship between students’ mathematics achievement and a combination of student- and teacher-based predictors; (c) comparing and contrasting the models’ results; and (d) connecting the technology of statistical modeling back to relationships with society and education through discussion of implications and questions regarding K-12 policy and practice raised by the study’s results. To summarize,

1. The two modeling approaches were similar in that (a) all student-based predictors were statistically significant across models; (b) student performance on the benchmark exams contributed most to the prediction of math achievement in all the models; (c) prior math performance was the second-strongest predictor of math achievement in all the models; and (d) student gender, teacher years of experience, and teacher level of certification were weak predictive contributors in all the models.
The two results of the two modeling approaches had some substantial differences:
(a) teacher-based predictors were statistically significant only in the MLR model; and
(b) advanced teacher certification had a positive effect on math achievement in the MLR model, but had a slightly negative effect on math achievement in the MLM models.

The study’s implications include the following:

1. Given the insignificance in this study of teacher experience and of advanced teacher certification in predicting math achievement implies that there is a need to
   - identify teacher characteristics that contribute to student achievement of targeted learning outcomes,
   - define and measure effective teaching,
   - reform education policy and practice to reward effective teachers.

2. Given the significance of student-based predictors and the insignificance of teacher-based predictors in this study, decision makers, administrators, legislators, educators, and philanthropists who want to effect improved mathematics education should consider moving to student-centered funding models that focus on
   - providing effective mathematics education early in students’ schooling,
   - ensuring student success in math, at a minimum, through Algebra II,
   - using benchmarking systems aligned with Common Core State Standards for formative assessment that will identify individual student’s learning needs and inform teaching practices.

3. Given the differences in results based on statistical modeling approach in this study, education stakeholders—decision makers, policy makers, administrators, educators,
and researcher—who want to improve education and allocate resources where the resources will have the greatest impact should insist on appropriate statistical models when group affects are present in a research setting.

**Suggestions for Further Research**

With respect to predictive modelling in education research, research suggestions include:

- the development of multilevel models that include the classroom level to assess the effects on student achievement of factors such as placement of students in classes by mathematical ability, time on task in the classroom, male-female ratio in the classroom, teaching methods and practices, discipline issues, teacher absences, and team-teaching approaches;

- the addition of teacher-level predictors to MLMs—such as subject knowledge, gender, interactions with students outside the classroom (clubs, sports, etc.), self-efficacy, self-assessment of ability to teach particular curriculum topics, and participation in professional development—to further examine teacher characteristics that may contribute to student success;

- inclusion of additional student-level predictors in MLMs—such as student absences, socio-economic status, level of parent education, time spent on math outside the classroom (after-school math programs, tutoring, time on homework), effect of reading ability on math performance, assessment of year-to-year changes in math achievement—to further examine student characteristics that may contribute to student success;

- use of Diaz’s education units of analysis continuum (see Figure 2) to build finely grained models with units of measurement as small as student interaction with content.
and to build MLMs with large scopes using state, national and international data as measurement units.

- development of MLMs with a school-level unit of analysis to account for school effects such as curriculum, overall socio-economic status for the school, teaching incentives, or principal effectiveness.

Relevant to the literature review, this study’s results, and relationships that exist among society, mathematics education, and technology, topics for further research include

- identification of factors that contribute to student achievement in mathematics;
- identification of elements that contribute to educator effectiveness in mathematics;
- identification of components that contribute to gender-based performance gaps in math achievement as well as identification of factors to mitigate those gaps;
- development of a research-based, measurable definition of teacher effectiveness;
- development of data and reporting systems that enable solid, data-informed decision making for K-12 policy and practice;
- examination of state and inter-state teacher certification criteria, including exploration of teacher effectiveness measures related to student attainment of the Common Core State Standards;
- development, implementation and evaluation of student-centered funding models.
References


Bode, R. K. (1996). *Is it ability grouping or the tailoring of instruction that makes a difference in student achievement?*


Carnegie Foundation for the Advancement of Teaching. (2010). *Five foundations fund initiative to improve student success in community colleges*


Doyle, D., Hassel, B. C., & Locke, G. (2012). *Smarter funding, better outcomes: Georgia’s roadmap for K-12 finance reform* Public Impact; and Atlanta, GA: Georgia Chamber of Commerce.


(Supplemental)

readiness. (Doctor of Education, University of North Texas). ProQuest Dissertations and Theses,


Education. Retrieved from


Lusk, L. (2011). *Teacher impact on student achievement in math and reading/language arts as moderated by teacher characteristics.* (Doctor of Philosophy, Capella University). (ProQuest Dissertations and Theses)


Organisation for Economic Cooperation and Development. (n.d.). Programme for international student assessment (PISA) Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd.org/department/0,3355,en_2649_35845621_1_1_1_1_1,00.html


Race to the top bonus payments to certified (sic) personnel in persistently low-performing schools. (2013). (Policy ID No. TCS-T-001).North Carolina State Board of Education.


Sherman, S. K. (2008). *Preparing the Texas assessment of knowledge and skills (TAKS): A study of local benchmark testing in Texas public schools*. (Ed.D., Texas A&M University -


