January 2012

Sensitivity of Value Added School Effect Estimates to Different Model Specifications and Outcome Measures

Bryce L. Pride
University of South Florida, bryce_pride@msn.com

Sensitivity of Value Added School Effect Estimates to Different Model Specifications and Outcome Measures

by

Bryce L. Pride

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Educational Measurement and Research
College of Education
University of South Florida

Co-Major Professor: John M. Ferron, Ph.D.
Co-Major Professor: Robert F. Dedrick, Ph.D.
Yi-Hsin Chen, Ph.D.
Danielle Dennis, Ph.D.

Date of Approval:
November 7, 2012

Keywords: sensitivity analysis, Gain Score Model, Layered Effects Model, educational accountability

Acknowledgments

First, I thank God for blessing me with countless experiences and relationships that have developed me towards the embodiment of 2 Peter 1:3-11. In addition, my loving wife deserves a great deal of praise for the assistance of her "eagle-eye" and continual dedication. She has sacrificed much more than anyone could have expected and it would not have been possible to achieve my goals and dreams without her support and encouragement. To her, I am forever indebted. I also thank my physical and spiritual family for their prayers, support, and encouragement throughout this life changing journey.

I thank my professors and committee for providing me with the knowledge, understanding, and guidance in the development of my craft, as well as my fellow colleagues on the same journey. A special thanks for my co-advisors and mentors, Dr. John Ferron and Dr. Robert Dedrick. Both have provided continual guidance, direction, and experiences that have developed my metacognitive processes as a journeyman researcher. For their patience, time, and insight into this dissertation, I am forever grateful.

Table of Contents

List of Tables

List of Figures

Abstract

The Adequate Yearly Progress (AYP) Model has been used to make many high-stakes decisions concerning schools, though it does not provide a complete assessment of student academic achievement and school effectiveness. To provide a clearer perspective, many states have implemented various Growth and Value Added Models, in addition to AYP. The purpose of this study was to examine two Value Added Model specifications, the Gain Score Model and the Layered Effects Model, to understand similarities and differences in school effect results. Specifically, this study correlated value added school effect estimates, which were derived from two model specifications and two outcome measures (mathematics and reading test scores).

Existing data were obtained from a moderately large and rural school district in Florida. The outcome measures of 7,899 unique students were examined using the Gain Score Model and the Layered Effects Model to estimate school effects. Those school effect estimates were then used to calculate and examine the relationship between school rankings.

Overall, the findings in this study indicated that the school effect estimates and school rankings were more sensitive to outcome measures than they were to model specifications. The mathematics and reading correlations from the Gain Score Model for school effects and school rankings were low (indicating high sensitivity), when

advancing from Grades 4 to 5, and were moderate in other grades. The mathematics and reading correlations from the Layered Effects Model were low at Grade 5 for school effects and school rankings, as were the correlations at Grade 7 for the school rankings. In the other grades, correlations were moderate to high (indicating lower sensitivity). Correlations between the Gain Score Model and the Layered Effects Model from mathematics were high in each grade for both school effects and school rankings. Reading correlations were also high for each of the grades.

These results were similar to the findings of previous school effects research and added to the limited body of literature. Depending upon the outcome measure used, school effects and rankings can vary significantly when using Value Added Models. These models have become a popular component in educational accountability systems, yet there is no one perfect model. If used, these models should be used cautiously, in addition to other accountability approaches.

Chapter I: Introduction

Statement of the Problem

It is no secret that politics, economics, global dominance, and education are intertwined with one another (Hershberg, 2005). In 1965, the Elementary and Secondary Education Act (ESEA) was enacted by President Lyndon B. Johnson to emphasize improving the education of low-income students. Nearly 20 years later, in 1981, President Reagan formed the National Commission on Excellence in Education as a result of the Secretary of Education, T.H. Bell's concern about "the widespread public perception that something is seriously remiss in our educational system" and the nation was at risk of losing their once unchallenged dominance in commerce, industry, science, and technological innovation (United States Department of Education [USDOE], 1983). That commission subsequently released the "A Nation at Risk" report in 1983. These events inevitably led to Congress amending the ESEA and reauthorizing it as the No Child Left Behind (NCLB) Act in 2002. This standards-based educational reform effort incorporated the strategies proposed by President George W. Bush which included (1) increased accountability for states, school districts, and schools; (2) greater choice for parents and students, particularly those attending low-performing schools; (3) more flexibility for states and local educational agencies in the use of federal education dollars;

and (4) a stronger emphasis on reading, especially for the youngest children (USDOE, 2010).

The NCLB Act requires that all students show Adequate Yearly Progress (AYP) toward the goal of 100% proficiency in reading and mathematics by the year 2014 (Choi, Seltzer, Herman, & Yamashiro, 2007). AYP is a status-based accountability model that examines the academic performance of student cohorts from year to year in terms of the proportion of students attaining proficiency or advanced levels on state standards-based tests (Choi et al., 2007). Each state provides its own definition of AYP for state standards against which each school district and school must compare itself (USDOE, 2002). According to Hershberg (2005), the NCLB requirement which mandates schools to bring all students to high standards by 2014, was a worthy goal but a problem was found in determining how to identify which schools were on target to meet those requirements. He also indicated that in most cases, AYP measures can distinguish successful and unsuccessful schools, but for many, the AYP measures fail to depict fair and complete assessments of school performance. One reason for this dilemma was that AYP focuses on overall proficiency to the exclusion of individual students' academic growth (Hershberg, 2005). A popular approach to address this dilemma is the use of Growth Models.

Growth Models are one approach states can use to measure students' academic achievement longitudinally (e.g., 3rd, 4th, 5th grade mathematics scores on academic achievement tests). Longitudinal analysis looks at data with two or more points in time, as opposed to cross-sectional analysis, which considers only one point in time. In 2005,

Secretary of Education Margaret Spellings announced a pilot program that allowed up to 10 states to develop and implement Growth Models into their accountability systems if they followed the seven "bright-line" principles of NCLB (USDOE, 2009). The required seven bright-line NCLB principles were (1) ensuring that all students become proficient by 2014, (2) making grade-level proficiency the expected standard of achievement as opposed to student or school characteristics, (3) holding schools and districts accountable for all student subgroups' achievement in reading and mathematics, (4) assessing students in Grades 3 through 8 in reading and mathematics, (5) developing assessment systems that produce comparable results from grade to grade and year to year, (6) monitoring student progress as part of the state data system, and (7) including student participation and achievement as separate academic indicators in the state accountability system (USDOE, 2009). By 2009, 15 states (Tennessee, North Carolina, Delaware, Arkansas, Florida, Iowa, Ohio, Alaska, Arizona, Michigan, Missouri, Colorado, Minnesota, Pennsylvania and Texas) were approved to implement some type of Growth Model (USDOE, 2009).

One type of Growth Model used in educational accountability systems is the Value Added Model (VAM). VAMs attempt to estimate to what extent changes in student academic achievement performance, or outputs (e.g., mathematics achievement scores), can be attributed to particular inputs (especially teachers, schools, or educational reforms) "received" by the student over a specified period of time (Wiley, 2006). These estimates of the effects of a particular teacher or classroom on student learning would be analogous to the estimated effects of a particular worker's efforts on a firm's output, such

as net profits or return on investments (Wiley, 2006). There are a variety of VAMs that are currently being used and investigated for use in educational accountability decisions.

The problem with the various accountability models is the difficulty in determining which model provides the most accurate picture of student academic achievement and school effectiveness. Each model has strengths and weaknesses and provides a varied perspective of student achievement and school effectiveness. Research and collaboration among education stakeholders will likely assist in determining the best accountability modeling approach.

Rationale for the Study

Given the interest in improving America's schools and holding states and districts accountable for student achievement, there have been disagreements about which accountability model(s) should be used to evaluate school effectiveness and compare schools. The Status Model (AYP) provides a perspective of student proficiency and the Growth Model provides a perspective of student improvement from one year to the next; further, VAMs provide a perspective of the degree to which a school contributes to a students' academic achievement. Decisions regarding the best model will require additional methodological investigations. This study will investigate methodological issues related to using VAMs for educational accountability purposes.

VAMs are used to estimate the unique effects schools and/or teachers add to students' learning. With a VAM, a district, school, teacher, or "program effect" can be examined; however, the two most common effects examined are schools and teachers. A

"school effect" can be described as the difference in one particular school's average student achievement growth for a given year, compared to the expected average achievement growth of all schools. The same principle applies for "teacher effects," which is the average achievement growth estimated for a particular teacher's classroom compared to all teachers in that particular grade and subject. Within the variety of VAMs, several models utilize background control variables such as race, parent's highest education level, and previous academic performance on standardized tests (Ballou, Sanders, & Wright, 2004).The inclusion of control variables removes variation and allows more precise comparisons between schools (when examining school effects) and teachers (when examining teacher effects). In both instances, schools or teachers are compared to the overall average growth rate in the category of interest.

There have been a variety of models proposed to address different study contexts. These models are often distinguished by different factors such as (a) number of data points, (b) how growth is conceptualized, (c) whether they include covariates (e.g., socioeconomic status, race), and (d) the persistence of prior teacher or school effects on future outcomes (McCaffrey, Lockwood, Loretz, Louis, & Hamilton, 2004). VAMs that focus on growth can range from simple models that measure change from one year to the next or predict a future score from a previous score in one context (e.g., one school), to more complex models where the contexts change within a study such as students who move to different schools or change teachers over the specified study time (Ballou et al., 2004; Briggs & Weeks, 2007; Lockwood, Doran, & McCaffrey, 2003; Lockwood,

McCaffrey, Hamilton, Stecher, Le, & Martinez, 2007; McCaffrey et al., 2004; Schmitz & Raymond, 2008).

Numerous VAM debates exist for a variety of issues including the validity and robustness of the estimates from these models; however, sensitivity analysis studies can be applied to address these issues. One example of this can be observed by investigating how results and implications of a study may differ due to changes in the conditions of a study (e.g., varied models used, outcome measures, or inclusion of covariates). Some degree of caution should be considered with these studies because even though the results may be stable, they may incorrect. These types of studies have been conducted and continue to be recommended for future research (McCaffrey, Lockwood, Koretz, & Hamilton, 2003; McCaffrey et al., 2004). Until the various VAM issues are rectified, the wide-scale use of these models will likely be limited (Beardsley, 2008). As previously noted, the value added to a student's achievement growth can be viewed from a number of perspectives; however, the majority of studies in the literature have focused on school and/or teacher effects. McCaffrey et al. (2003) focused on sensitivity issues related to teacher effectiveness, but mentioned that the sensitivity issues raised in their study also apply to school effect studies.

A limited number of sensitivity studies have been conducted to investigate value added effects. These past studies found value added effects to be sensitive to a variety of conditions, particularly varied model specifications in the studies of Briggs and Weeks (2007, 2011), McCaffrey et al. (2003, 2004), and Tekwe et al. (2004); and varied model specifications and outcome measures in the Lockwood et al. (2007) and Schmitz (2007)

studies. Additional studies under similar and varied conditions are still warranted to determine if results are consistent in other contexts. This study sought to build upon the research of Schmitz (2007) and Lockwood et al. (2007), as well as, Tekwe et al. (2004) and Briggs and Weeks (2011). The Schmitz (2007) and Lockwood et al. (2007) studies address sensitivity analysis issues related to teacher effects under varied models and outcome measures. The study by Tekwe et al. (2004) and Briggs and Weeks (2011) addressed sensitivity analysis issues related to school effects under varied model specifications.

Lockwood et al. (2007) found that the sensitivity of the teacher effect estimates to models and controls was low (i.e., highly correlated, having a similar relationship) in comparison to the sensitivity of the teacher effect estimates to achievement outcome measures. This indicated that changes in outcome measures were more significant than changes in the models used. Schmitz (2007) also found that there was little difference in the models' estimation of teachers' effects with the exception of one model, which attributed more variability in student gains to teacher effects than the others. He found that the correlations between mathematics and reading from three of the models were only moderate (correlational findings are usually categorized as "low," "moderate," or "high"). Tekwe et al. (2004) found that the global impact of one model, compared to another was low in their study, which used two years of data. The study also found consistent disagreement between two models, but they speculated the disagreement would decrease when analyzing three or more points in time (Tekwe et al., 2004). However, when only two points in time were considered they recommended two models

in favor of the others (Tekwe et al., 2004). Briggs and Weeks (2011) found that estimated school effects across models were moderately to highly correlated irrespective of the specific test subject or pair of models considered; however, they found considerable variability in the correlations.

In light of the national attention given to educational accountability systems, it behooves educational stakeholders to consider the strengths and weaknesses of all accountability models and select the most appropriate combination of approaches. This study contributed to the research that supports the evaluation of VAMS for educational accountability, thus enabling provision of a more holistic perspective of student achievement and school effectiveness.

Purpose of the Study

The purpose of this study was to examine two Value Added Model specifications to understand similarities and differences in school effect results. Specifically, this study correlated value added school effect estimates, which were derived from two model specifications (Gain Score and Layered Effects) and two outcome measures (mathematics and reading). Next, the school rankings were compared and correlated using the same models and outcome measures (or outcomes). Conducting these analyses (1) demonstrated how stable the value added school effect estimates were when the models and the outcomes were altered, (2) determined whether these models could be used interchangeably to compare schools, (3) examined whether school effect results were similar to the previous findings for teacher effects, and (4) added to the methodological

research literature required for a fuller understanding of the implications of using VAMs for educational accountability. This study used secondary data collected from a moderately large and rural school district, which included test scores of students in elementary school (Grades 3 through 5) and middle school (Grades 6 through 8). Four research questions were used to guide the analyses.

Research Questions

Research Question 1. What is the relationship between school effect estimates from the Gain Score Model when mathematics achievement scores are used versus reading achievement scores? What is the relationship between school effect estimates from the Layered Effects Model when mathematics achievement scores are used versus reading achievement scores?

Research Question 2. What is the relationship between school effect estimates based on mathematics achievement scores when the Gain Score Model is used versus the Layered Effects Model? What is the relationship between school effect estimates based on reading achievement scores when the Gain Score Model is used versus the Layered Effects Model?

Research Question 3. What is the relationship between school rankings from the Gain Score Model when mathematics achievement scores are used versus reading achievement scores? What is the relationship between school rankings from the

Layered Effects Model when mathematics achievement scores are used versus reading achievement scores?

Research Question 4. What is the relationship between school rankings based on mathematics achievement scores when the Gain Score Model is used versus the Layered Effects Model? What is the relationship between school rankings based on reading achievement scores when the Gain Score Model is used versus the Layered Effects Model?

Significance of the Study

This study was significant given the interest in the academic growth of youth in the United States and the desire to hold schools accountable for students' academic growth. Accountability approaches, in particular the Status, Growth, and VAM models, each have strengths and weaknesses to consider when they are used to identify the most or least effective schools, programs, or teachers for students. Sensitivity analysis studies are an avenue that can explore consistency in value added school effect results before wide-scale policy implementations take place. These studies are important considering the consequences students, teachers, schools, and districts face when students' growth is inadequate. Value added estimates should be used as one part of a holistic approach to educational accountability and should not be used as a sole measure from which to make policy decisions. This study adds to the literature and our understanding of value added

estimates, which may aid in the efforts to make responsible inferences regarding research and evaluation findings and educational policies.

Definition of Terms

The terms defined below were used throughout this study:

Academic Achievement. The degree to which students "meet" or "exceed" academic standards established for a particular subject, such as reading, mathematics, or science (USDOE, 2009).

Accountability. The idea of holding states, districts, schools, educators, and students responsible for results (Education Week, 2004).

Adequate Yearly Progress. A status-based model defined by the NCLB Act. Each state establishes a definition of "Adequate Yearly Progress" (AYP) to use each year to determine the achievement of each school district and school (USDOE, 2002).

Gain Score Model. Models that specify a one-year gain score (current score minus the previous score) separately for each year and link student gains to their current-year teacher or school effects (McCaffrey et al., 2003).

Layered Effects Model. A mixed effects model that estimates the specific effect that systems, schools, and teachers have on student academic achievement gains from standardized assessments each year (Sanders & Horn, 1994).

Sensitivity Analysis. Analyses that are used to determine whether the results and implications of a study differ when modifications are made to the study (Forrester, Breierova, & Choudhari, 2001). For example, when a school's ranking changes while comparing two models, then the school's ranking would be considered sensitive to the model specifications. Otherwise, the rankings would be considered insensitive to model specifications.

School Effect Estimates. An estimated measure of the difference between a school's actual value or average achievement growth rate for a particular grade, compared to the expected value or growth rate for an average schooling experience or year's growth for the same grade (Goldschmidt et al., 2005; Schmitz, Raymond, 2008).

Standards-Based Tests. Academic assessments that are based upon established academic standards (USDOE, n.d.).

Value Added Model (VAM). A type of Growth Model which may use statistical controls (e.g., student demographics, prior achievement, etc.) in order to isolate the specific

effects a particular school, program, or teacher have on students' academic achievement each year (Goldschmidt et al., 2005).

Value Added School Effect Estimate. An estimate of the difference between a school's actual average academic achievement growth rate compared to the expected growth rate for an average school year's growth rate (Goldschmidt et al., 2005).

Delimitations

The results of this study were delimited such that only two levels of outcome measures and two levels of model specifications were examined. The models were two-level models, with repeated measures nested in students, and students nested in schools. The teacher level was omitted from this analysis to maintain a limited focus on examining school effects. In addition, the sample represented students who tested in one moderately large and rural school district for Grades 3 through 5 and Grades 6 through 8 from 2005 to 2010.

Limitations

This study was limited in two ways. First, the sample consisted of students' standardized achievement test scores in Grades 3 through 5 and Grades 6 through 8 from one moderately large and rural school district in Florida. Within this limitation, the student sample may or may not reflect the U.S. population, but most likely will reflect the population of the school district from which the sample was taken. The second limitation

was found in using existing district data for secondary analysis. When using existing data, an analyst may be limited in the knowledge of errors such as matching student IDs with test scores or accurate chronicles of students changing schools within a school year. When using secondary data, the researcher has no control of the design conditions of the original study under which the data collected, the variables investigated, coding methods for variables, instruments used, reliability and validity of scores from the instruments, or how the data are arranged in the dataset.

Organization of Remaining Chapters

The remaining chapters present previous research and the methods, results, and the discussion of this study. Chapter II, Literature Review, begins with a description of the contextual framework of educational accountability models. It then includes research regarding various VAM characteristics and value added sensitivity studies. Chapter III, Method, provides a discussion of the research methods used to investigate the research questions posed. The chapter presents the research design, population and sample investigated, variables used in this study, instruments for outcome measures, data collection procedures, data analysis, and software/technology used for analyses. Chapter IV, Results, presents the findings from the preliminary and primary data analyses. Lastly, Chapter V, Discussion, presents a final discussion of the results, limitations of the study, directions for future research, and closing remarks.

Chapter II: Literature Review

Overview

The No Child Left Behind (NCLB) Act and Adequate Yearly Progress (AYP) were new paradigms in that accountability became the standard in the arena of education. Schools, districts, and states became accountable for their students to meet their Annual Measurable Objectives. Previous researchers have alluded to unintended consequences of NCLB and AYP as being somewhat unfair to the students in the most need because it encourages schools to focus on moving students who are closest to the threshold over it, virtually ignoring their lowest performing students (Choi, Seltzer, Herman, & Yamashiro, 2007). To address these concerns, Growth Models and Value Added Models (VAMs) have been proposed as supplements for use in educational accountability systems; however, VAMs are still somewhat controversial and have not gained nationwide acceptance for a variety of reasons. There is a need for further investigations of these models to determine if they are suitable for accountability or school improvement efforts. This study built upon previous value added studies that compared results from different model specifications and outcome measures.

This chapter presents a review of previous literature in three segments providing the background for this study: (1) a description of the contextual framework of educational accountability models, (2) VAM characteristics and approaches, and (3)

value added effects sensitivity studies. The first section is brief; the majority of the literature focuses on various aspects of VAMs. Lastly, the chapter closes with a literature review summary.

Due to the breadth of VAM research topics as a whole, it is necessary to denote which literature was included and which literature was excluded as a major focus. The studies presented in this chapter were confined to those relating to value added modeling of teacher and/or school effects. Although this study was conducted with school effects as the focus, literature on teacher effects was included as a reference and bridge to the investigation of school effects. The reason for this was due to the limited number of value added studies that focused on the sensitivity of estimates (McCaffrey et al., 2003; Briggs & Weeks, 2007, 2011). In addition, the studies in this chapter were limited to those conducted in the United States. These inclusions and exclusions are important in conveying the context used to guide this study.

Contextual Framework: Educational Accountability Models

The contextual framework further clarifies the focus of this study and the various challenges of selecting the best methods to use in an accountability system. According to Hershberg (2005), the NCLB requirement that schools bring all their children to high standards by 2014 was a worthy goal, but a problem was found in determining how to identify which schools were on target to meet those requirements. He indicated that in most cases, AYP measures can distinguish successful and unsuccessful schools, but for many, the AYP measures fail to depict fair and complete assessments of school

performance. One reason for this dilemma is because AYP focuses on achievement to the exclusion of academic growth (Hershberg, 2005). Growth solely does not provide the information on status or distance to proficiency, which is most important for the NCLB Act and meeting AYP.

Status Models and Growth Models are the main approaches used in educational accountability within the United States and they both have strengths and weaknesses. Status Models and Improvement Models both provide one perspective of the academic performance of students. Growth Models provide another perspective of student performance in regard to how individual students progress from year to year and how schools compare to this growth. Growth Models can be used to monitor individual students' academic progress; however, VAMs, which are a type of growth model, go a step further in attempting to estimate the extent to which schools and/or teachers add to students' academic achievement, or growth. When utilized independently, the Status and Growth Models provide a limited perspective of student and school performance. However, when used in conjunction with one another, these models can provide a more holistic perspective of student academic achievement and the effectiveness of schools. It is important to continually study and monitor each of these modeling approaches to determine how accurate they are in providing student proficiency, student progress, and school effectiveness information.

Status and Improvement Models. Goldschmidt et al. (2005) provided a description of the distinctions between different accountability models (e.g., Status, Improvement, and Growth). According to Goldschmidt et al. (2005), Status Models were

oftentimes contrasted with Growth Models. Also, they indicated that Status Models took snapshots of a subgroup's or school's level of student proficiency at one point in time or an average of two or more points in time. Goldschmidt et al. explained that this proficiency level was then compared to an established target that can vary between states. They defined progress, or growth, under this type of model as the percentage of students achieving proficiency for a particular year and the school was evaluated based on whether the student group met or failed to meet the established target. Lastly, they identified another type of Status Model, which was an Improvement Model that measured change between different cohorts of students (e.g., 2009 7th graders to 2010 7th graders) (Goldschmidt et al., 2005).

Growth Models. Growth Models are models of education accountability that measure academic progress by tracking the change in achievement scores of individual students or cohorts of students from year to year (e.g., Cohort 1 in 3rd Grade to Cohort 1 in 4th Grade, etc.) with one of the goals of determining the average growth made among students and schools (Goldschmidt et al., 2005). Achievement growth comparisons over time at the school level, indicated by Goldschmidt et al., would be determined from the aggregated growth of individual students in the school (e.g., comparing a three-year average school growth between schools A, B, or a state average) after controlling for each student's background and prior achievement. The researchers also determined that a school's ability to facilitate academic achievement growth over time is a better indicator of academic performance than the Status Models that look only at one point in time; schools could then be ranked based upon their average growth estimates.

Value Added Models. A commonly mentioned type of Growth Model and application in education is a Value Added Model (VAM). VAMs are the most recent methods used in education to estimate the unique contributions that teachers, schools, and districts make upon students' academic performance or achievement growth (Goldschmidt et al., 2005). Some models use covariates to separate the effects of non-school-related factors (e.g., family, peer, and individual influence) from a school's performance so appropriate comparisons can take place. Goldschmidt et al. also determined that schools using VAMs can have positive achievement growth and a negative value added estimate (e.g., School A gained an average of 25 points over three years, but the district average was 40 points greater than School A's average over the same three years).

The concept of value added modeling is not totally unique nor is it new, but rather it is a method that has been used in the business sector and applied in education as a means to estimate the effects that schools, teachers, or programs have upon student academic achievement and growth. According to Wiley (2006), this approach took its roots from econometrics and educational statistics. He noted that economists used "production function" models to mathematically describe how a firm created output from its inputs or how its resources and procedures were used to produce products. He also stated that the production function measured productivity (value created) from a specific collection of inputs and the more valuable inputs were those that were more productive and provided greater output per unit. Wiley (2006) stated that economists interested in education used the input/output model to estimate how factors affect the outcomes of

schooling. When using "production function" models in education, also known as Education Production Functions (EPF), a central question Wiley (2006) posed was, "To what extent could changes in student performance or output (i.e., mathematics achievement scores) be attributed to particular inputs (i.e., teachers, schools, or educational reforms) 'received' by the student over a specified period of time?" He indicated that the EPF estimates of the effects of a particular teacher on student learning were analogous to the estimated effects of a particular worker's efforts on a firm's output.

Wiley (2006) stated that educational researchers developed approaches for investigating teacher and school effects similar to that of economists through longitudinal analysis of student assessment data. He noted that some of the early models were simple year-to-year changes in scores or predictions of current-year scores using the previous-year scores in hierarchical and non-hierarchical formats. The more complex statistical models used by educational statisticians became known under a number of names (e.g., Hierarchical Linear Models, Multilevel Models, or Random Effects Models).

In 2009, states that used some form of VAM in their assessment programs included Arkansas, Delaware, Florida, Louisiana, Minnesota, Ohio, Pennsylvania, South Carolina, and Tennessee, with Tennessee ranking as a leader with the best example to date (Wisconsin Policy Research Institute Inc., 2009). VAMs were mandated in Tennessee, Ohio, Pennsylvania, and several hundred school districts in 21 states (Wisconsin Policy Research Institute Inc., 2009). Dr. William Sanders and his colleagues at the University of Tennessee-Knoxville were the key developers of the Tennessee Value Added Assessment System (TVAAS) (Sanders, Saxton, & Horn, 1997), which is

one of the most widely known and complex VAMs. Dr. Sanders has been most credited with introducing the combination of value added assessment with Mixed Model methodology to education and its policy makers (Raudenbush & Bryk, 1986; Raudenbush, & Bryk, 2002). He worked specifically with Tennessee policy makers, and legislative actions resulted in the implementation of TVAAS in all public schools in the state of Tennessee (Sanders & Horn, 1994, 1998; Wright, Horn, & Sanders, 1997). In Tennessee, TVAAS is the process of estimating the district effects, school effects, and teacher effects on the academic growth of students in Grades 3 through 8 in science, mathematics, social studies, language arts, and reading (Sanders et al., 1997). A number of other researchers have continued to expand the work done in this area including Ballou et al. (2004); Briggs and Weeks (2007, 2011); McCaffrey et al. (2003); McCaffrey et al. (2004); Schmitz (2007, 2008); and Tekwe, Carter, Ma, Algina, Lucas, Roth, Ariet, Fisher, and Resnick (2004).

Educational accountability is an area with many facets that are applicable to a variety of stakeholders such as politicians, methodologists, practitioners, and the public as a whole. It contains a broad area of research that has not thoroughly been explored and still has room for improvements. There are a variety of modeling approaches that have strengths, weaknesses, and concerns that future research will continue to build upon. As future discoveries come to light, additional questions are likely to surface as well.

Challenges and concerns with Value Added Models. In each area of educational accountability research there are a variety of concerns that come from different perspectives. Though value added effects have been applied and inferences have

been made regarding schools and teachers, they have not gone without criticisms and expressed concerns. Methodological issues remain to be explored and addressed before these models are accepted with full confidence from researchers and the public. Five of the concerns that Beardsley (2008) listed specific to the Education Value Added Assessment System (EVAAS) should be considered when working with any VAM. The concerns were related to validity, the use of data, the lack of peer review, the handling of missing data, smaller samples regressing to the mean, and the handling of extraneous variables. First, she noted that there were questions concerning the extent to which the value added results were valid, especially the need for content specific assessments to be linked to curriculum and other measures for teacher/school effectiveness. Second, she had concerns about whether the data were used in formative ways that could improve school performance and that many districts failed to use the data in ways to improve their schools because of difficulty interpreting the results. Next, she noted that perhaps one of the most troublesome issues regarding the EVAAS was noted criticisms about the limited outside peer reviews of EVAAS results and proprietary algorithms. Then, she listed concerns about how missing data were related to biasing results and how smaller samples regressed toward the mean. Lastly, were her concerns about extraneous variables that related to the idea that EVAAS failed to include student factors in the model because the EVAAS developers claimed the differences were negligible.

In a special issue of the *Journal of Educational and Behavioral Statistics* on Value Added Modeling, Rubin, Stuart and Zanutto (2004) applauded the efforts of Ballou et al., (2004), McCaffrey et al., (2004), and Tekwe et al. (2004) in their estimation of

value added parameters. However, they did not think their analyses were estimating causal quantities, except under extreme and unrealistic assumptions. Their view was that the estimates of teacher or school effects were not causal, but rather descriptive (Rubin et al., 2004). Without random assignment, causation cannot be inferred. This hinders efforts to infer causation with VAMs because students are rarely, if ever, randomly assigned to teachers and teachers are rarely randomly assigned to schools from a practical standpoint.

Other researchers have voiced additional concerns about VAMs. Raudenbush (2004) addressed the two types of causal effects estimated in school accountability systems, Type A and Type B. A child's outcome would be a function of pre-assigned student characteristics, *S*, and random error, *e*, in addition to two aspects of schools: (1) school context, *C* (e.g., neighborhood) and (2) school practice, *P* (e.g., lecture versus lecture including discussion and hands-on-activities; Raudenbusch & Willms, 1995). Raudenbush (2004) indicated that Type A effects were those of interest to parents choosing a school for their child to attend and Type B effects were of interest to those seeking to hold school personnel accountable for their contributions to student achievement. He described Type A as the difference in a child's potential performance at one school compared to another school where the parents are not as concerned with the context or practice of the schools. He also described Type B as the difference in the child's potential in one school with a particular practice compared to that child's potential outcome in another school with a different practice; here the focus would be the comparison of practice between the two schools. Raudenbusch (2004) reasoned that at best, researchers would be able to estimate Type A effects of interest to parents selecting

schools, but not Type B effects of interest to officials holding schools and teachers accountable for instructional practices because school practices in most systems fail to be defined or observed.

Currently, no one accountability or modeling approach is optimal in all situations and there is still a need for additional studies to further investigate methodological issues and model comparisons. These studies are of significant importance to provide evidence to assist policy makers to choose the best accountability and/or modeling approach for school comparisons. Though numerous problems and concerns exist, VAMs are still considered a viable component of school accountability systems.

Value Added Modeling Characteristics and Modeling Approaches

Within the realm of VAMs, numerous models have similar and differing approaches to investigate school and teacher effects on student academic achievement growth. These models can be distinguished by characteristics such as (1) number of data-points, (2) levels or nesting illustrated by the models, (3) conceptualization of growth, (4) covariate use in the models, (5) school effects as random or fixed, and (6) assumptions about the persistence of school effects. Two of the broad modeling approaches include the Pure Nested Models and the Cross-Classified Random Effects Models. Each of these model types are connected with the purpose of estimating schools' effects on students' achievement scores over time. However, the models differ in various ways depending upon the context investigated and the research questions posed. Some of the models that have been used to investigate value added school effects include the (1) Gain Score

Model, (2) Hybrid Success Model, (3) REACH Model, (4) Layered Effects Model, (5) Cumulative Effects Model, and (6) General Variable Persistence Model. Some of the modeling characteristics, modeling approaches, and specific models can be found in Table 1.

Data points. The number of data points is dependent upon the research/evaluation design. When only one point in time is considered in the study, the design is considered cross-sectional. When two or more points in time are utilized in the study, the design is considered longitudinal. The VAM attributes some of the change over time as a result of experiencing a particular teacher's class or school. Comparisons of change can then be made using two points in time with an infinite number of data collection points for a study. Measurement of change with the Pure Nested Model requires at least two scores (e.g., 4th and 5th grade mathematics scores); however, for the more complex Cross-Classified Random Effects Model, at least three data points are needed when examining growth.

Hierarchical levels. The levels refer to the hierarchical unit of analysis, whether at the student, teacher, school, or district levels. The Pure Nested Model and Cross-Classified Random Effects Model can range in the number of levels considered from two to five, with two or three levels being the most common. Hierarchical Linear Models segment the variance in the data and allow the researcher to examine variance found at each level investigated simultaneously. A two-level model is depicted when students are nested within schools and a three-level model is depicted when students are nested in teachers and teachers are nested within a school.

Growth description. In education, a great deal of focus has been given to how much academic progress students make from year to year in various subjects and it is assumed that progress or growth will take place each year, regardless of the student or school circumstances. Growth, at a minimum, can be described as the difference between two points in time. However, when examining academic growth, the definition can take on a number of different conceptualizations depending upon the research design and models utilized. Growth can be conceptualized as a change between two adjacent scores, the amount of change needed to reach a proficiency target, or the attribution of gains to schools or teacher effects with these effects accumulating in layers from year to year (Doran & Izumi, 2004; McCaffrey et al., 2003; Sanders, Saxton, & Horn, 1997).

Covariate use. Covariates are independent variables that may or may not be used as predictors (e.g., race/ethnicity, SES, AYP) of dependent variables. They are held constant in an analysis to reduce their effects on the outcome of interest used to make comparisons. In many research studies, the covariates may be of interest; however, in others, the covariates may confound or interact with other independent variables, and are therefore excluded. Uncontrolled covariates may lead to incorrect inferences about the relationship between the independent variables and, more specifically, the outcomes may be moderated by one or more covariates, which can also alter interpretations. The choice of whether or not to include covariates will often depend upon the context and research questions to be investigated.

School effects. School effects are the random effects or deflections (whether positive or negative) from a grand average outcome. These school effects are considered

fixed when they assume that the schools are the fixed population to be examined (Tekwe et al., 2004). School effects are considered random when the schools observed are assumed to be a random sample from a larger population of schools (Tekwe et al., 2004). According to McCaffrey et al. (2003), the two methods will tend to yield similar conclusions about variability between schools, but will provide different estimates of individual school effects. Random effects are the natural approach when variance components are of primary interest; however, when the specific intention is to make inferences about a particular set of schools (e.g., in an accountability setting), fixed effects may be preferable (McCaffrey et al., 2003).

Persistence of effects. Persistence describes the degree to which the school effect estimates hold their effect over time. Using a Complete Persistence Model, the persistence of school effects assume that the effects of a previous school on a students' growth (e.g., 4th to 5th grade mathematics growth at School A estimated to be .80) remains undiminished and accumulates year to year. It assumes that, for example, a .80 school effect for 3rd Grade students in School A remains undiminished even when those same students are in 8th Grade at School B. With variable persistence, the previous school effects are estimated each year in future administrations and are allowed to diminish over time rather than assume that the effects accumulate and remain undiminished year after year.

Pure Nested Model. The Pure Nested Models are appropriate for modeling educational data because this type of data is hierarchical by nature. Hierarchical, or nested data, include settings where students are nested within a particular teacher's

classroom, teachers are nested within a particular school, and schools are nested within a particular school district (Raudenbush, & Bryk, 2002; Wiley, 2006). With purely nested situations, it is assumed that the context remains the same at each data point of the study. An example of a Pure Nested Model can be described as measuring students' academic growth in Grades 6 through 8 within one school. Part of the NCLB Act requires states to test students yearly in Grades 3 through 8 for reading and mathematics and, as mentioned earlier, all students must show Adequate Yearly Progress (AYP) toward the goal of 100% proficiency by the year 2014 (Choi, Seltzer, Herman, & Yamashiro, 2007). The Pure Nested Model may be sufficient to address these growth and proficiency issues appropriately when students remain in one context (school).

Three Pure Nested Models that have been utilized in the examination of school effects include the Gain Score Model, the Hybrid Success Model, and the REACH Model. The simpler models have been discussed under a variety of names (e.g., Gain Score Model, Change Score Model, and Covariate Adjustment Model and they are similar in that they measure change as the difference between a current score and previous score (Lockwood et al., 2007; McCaffrey el al., 2004; Tekwe et al., 2004). The Gain Score Model specifies a one-year gain score (current score minus prior score) separately for each year and links student gains to their current-year school's effects (McCaffrey et al., 2003). The Covariate Adjustment Model considers two adjacent years, but is conceptualized slightly differently because it actually regresses the achievement measure for the current year on the previous year; it uses prior scores as covariates in models for current outcomes (Lockwood et al., 2007; McCaffrey et al., 2004).

The Hybrid Success Model includes growth along with proficiency, and the success of a school is a measure of academic growth in the school (Kingsbury & McCall, 2006). The school can be deemed successful depending upon the extent that students are growing "as much" or "more than expected" and growing "toward" or "beyond" proficiency targets (Kingsbury & McCall, 2006). The REACH Model is similar to the Hybrid Success Model in that it focuses on growth and proficiency (Kingsbury & McCall, 2006). At the school level, the percentage of students "at" or "above" the proficient cut point are calculated across all tested grade levels in the school and these calculations are then used to provide an estimated growth rate of the school (Doran & Izumi, 2004). The Pure Nested Models consider the hierarchical nature of the data but are not the most appropriate when the context changes.

Cross-Classified Random Effects Model. Currently, the most complex VAMs are the Cross-Classified Random Effects Models. These models estimate growth trajectories when the context changes during the study period. For example, the lower level units (i.e., students) may occupy more than one higher level unit at level-2 (i.e., schools). These students at level-1 could attend the same elementary school at level-2 together, but attend different middle schools also at level-2. Students would be considered cross-classified, meaning that they are classified in two categories of schools, which are both at level-2. When context changes are ignored and not modeled appropriately, the results may lead to underestimation of the residual errors (Wiley, 2006). When context changes occur within a study, the Cross-Classified Random Effects

Model becomes the best modeling approach because it appropriately models the residuals at each point in time in the study (Raudenbush & Bryk, 2002).

Three of the most commonly referenced Cross-Classified Random Effects Models investigating school effects are the (1) TVAAS or Layered Effects Model of Sanders, Saxton, and Horn (1997), (2) the Cumulative Effects Model (Ponisciak & Bryk, 2005; Raudenbush & Bryk, 2002) and (3) the General Variable Persistence Model of McCaffrey et al. (2004). These three Cross-Classified Random Effects Models can be distinguished by complete or variable persistence, where the persistence would be the school or teacher effects from previous years into future student outcomes. The first two examples of models are examples of the first class of models, which assume complete persistence. The Layered Effects Model of Sanders, Saxton, and Horn (1997) was investigated by Briggs, Weeks & Wiley (2008); McCaffrey et al. (2004); and Tekwe et al. (2004). The Cumulative Effects Model of Raudenbush and Bryk (1993, 2002) was investigated by Briggs et al. (2008), McCaffrey et al. (2004), and Schmitz (2007).

McCaffrey et al. (2004) stated that the Cumulative Effects Model estimates each student's growth across grades by imposing a linear trend line rather than allowing separate means at higher levels at each point in time. They indicated that this model was a multi-grade Gain Score Model, where the mean gain was assumed constant across grades and the variance-covariance matrix for residual error terms from the same student was not diagonal (i.e., gains are not independent across grades). However, the Layered Effects Model, they found, placed no restrictions on the overall grade specific means or the covariance-variance matrix of the repeated test scores from the student. McCaffrey et

al. (2004) indicated that the Cumulative Effects Model and Layered Effects Model use data from all students, even those with partially complete records; this is different from the Gain Score Model or Covariate Adjustment Models, which use only students with two consecutive years of data unless imputation or other missing data methods are applied. They noted that both the Cumulative Effects Model and Layered Effects Model were extensions of Gain Score Modeling.

The second class of Cross-Classified Random Effects Models include the General Variable Persistence Model of McCaffrey et al. (2004) demonstrated by Briggs and Weeks (2007) and Lockwood et al. (2007). The General Variable Persistence Model estimates growth similar to the Layered Effects Model in that it places no restrictions on the covariance-variance matrix of the repeated test scores from a student (McCaffrey et al., 2004). The persistence of effects is not assumed complete as with the Cumulative Effects Model and TVAAS or Layered Effects Models (McCaffrey et al., 2004). Also, the General Variable Persistence Model is limited because the persistence parameters have posed computational challenges when trying to fit the models in HLM, MLWin, R, SAS, or S-Plus software packages, particularly with larger datasets typically found in moderate to large school districts (Lockwood et al., 2007; McCaffrey et al., 2004). Table 1 displays some of the previously mentioned features that may distinguish the different models that have been used in school effects studies. The last column of the table provides references for those seeking more in-depth discussions of the Pure Nested Models and Cross-Classified Random Effects Models.

Table 1
Comparisons of Various Value Added Models and Their Features

Models Types	Data Points	Levels	Growth Description	Covariates Used in References	School Effects	Persistence	School Effects References
I. Pure Nested Models							
Gain Score	2	2	Change Score in Adjacent Years	Varies	Random	Complete	Tekwe et al., 2004; Wang, 2006
Hybrid/PTG	2	2+	Standards Growth	No	Fixed	Complete	Kingsbury & McCall, 2006
REACH/PTG	2+	2+	PAC and ETGR	No	Random	Complete	Doran & Izumi, 2004
II. Cross-Classified Random Effects Models							
Layered Effects	3+	2+	Layered Gains	No	Random	Complete	Briggs et al., 2008; Sanders et al., 1997; Tekwe et al., 2004; Wang, 2006
Cumulative Effects	3+	2+	Linear Gains	No	Random	Complete	Ponisciak & Bryk, 2005; Raudenbush & Bryk, 2002; Schmitz, 2008
General Variable Persistence	3+	2+	Layered Gains	Varies	Both	Variable	McCaffrey et al., 2004; Briggs & Weeks, 2008

Note. Estimated True Growth Rate of School (ETGR); Percentage of students "at" or "above" the proficient cut-point (PAC); Progress Toward a Goal (PTG).

The Gain Score Model and the Layered Effects Model were selected for this study for several reasons. First, these models have been used in previous school effects studies (e.g., Briggs et al., 2008; Sanders et al., 1997; Tekwe et al., 2004; Wang, 2006). Secondly, they allowed comparisons of the random school effects estimated using a simple Gain Score Model. Next, they utilized a more complex hierarchical modeling approach. Lastly, the study was focused on performance and growth instead of proficiency as demonstrated with the Hybrid Success Model and REACH Model. The Cumulative Effects Model was considered for usage in this study in addition to the Gain Score Model and the Layered Effects Model; however, a pilot study conducted to prepare for this study found that the Cumulative Effects Model posed computational challenges due to insufficient memory and convergence problems similar to those found in the studies conducted by Lockwood et al. (2007) and McCaffrey et al. (2004). This study compared the differences in school effect estimates and ranking of schools. A simple Gain Score Model and a complex Layered Effects Model were used with mathematics and reading outcome variables to accomplish these goals.

Though the Gain Score Model and Layered Effects Model were chosen, possible limitations exist when striving to meet educational accountability requirements (AYP). Currently, the Gain Score Model and Layered Effects Model do not have features of monitoring progress to a goal (proficiency) as do the Hybrid Success Model and REACH Model. For Growth Models to be considered for use in AYP school comparisons, they must adhere to the seven bright-line principles. When adhering to the focus of the seven principles, both models can be used or supplemented to address monitoring the proficiency percentages of all students through 2014 and beyond, focusing on grade-level

proficiency versus student or school characteristics. Both models can be used as a means to hold schools and districts accountable for student achievement in reading and mathematics for all student subgroups and measure assessment systems that produce comparable results from grade to grade and year to year. Student participation and achievement can also be incorporated as separate academic indicators in the state accountability system.

The limitations surface again when assessing students' achievement growth in Grades 3 through 8 for reading and mathematics and monitoring student progress as part of the state data system using a Pure Nested Model (i.e., the Gain Score Model). The monitoring of student growth over Grades 3 through 8 would best be accomplished with a Cross-Classified Random Effects Model (i.e., a Layered Effects Model). The Cross-Classified Random Effects Models were developed to accommodate the modeling of growth when the school context changes for instance, advancing from 5th Grade, usually the highest elementary school grade to 6th Grade, usually the lowest middle school grade. When using Pure Nested Models, it is assumed that the school context remains the same, such as having 3rd, 4th, and 5th Grades in one elementary school.

While there is no perfect model for educational accountability today, stakeholders can refine current models to estimate school effectiveness most accurately and ensure that proficiency targets can be met by schools and districts. The progress toward a goal component, which is utilized in the Hybrid Success and REACH Models, can be used with the Gain Score and Layered Effects Models. They can then be utilized by schools, districts, or state officials to address principles related to proficiency, as well as growth. The Gain Score and Layered Effects Models have more strengths than weaknesses for

addressing educational accountability; therefore, they are appropriate for consideration in the current educational accountability discussions and were selected for this study.

The Gain Score Model takes a "difference" score between two adjacent grades for the outcome measure and is utilized in instances where the context remains the same (e.g., gains from Grades 4 to 5). Complete descriptions of the Gain Score Model utilized in this study can be found in the Methods Chapter. Previous studies that have discussed or used the Gain Score Model include Lockwood et al. (2007), Wang (2006), Tekwe et al. (2004), and McCaffrey et al. (2003, 2004).

The Layered Effects Model is most appropriate to model growth when the research context changes from one context to another within a particular study (Raudenbush & Bryk, 2002). The Layered Effects Model has examined school effects in several ways as specified by Briggs and Weeks (2007), Briggs et al. (2008), Sanders et al. (1997), Tekwe et al. (2004), and Wang (2006). The Layered Effects Model has been utilized in numerous studies investigating school and teacher effects. Tekwe et al. (2004) made comparisons of other models with the Layered Effects Model in the investigation of school effects. Ballou, Sanders, and Wright (2004) examined confounding variables with teacher effects. The McCaffrey et al. (2004) study compared other models and focused on examining teacher effects. Briggs et al. (2008) examined the sensitivity of value added modeling to the way an underlying vertical scale was created with school effects being the focus. Wang (2006) compared several models and school rankings with a Monte Carlo simulation study.

The two models used in this study were the Gain Score Model and the Layered Effects Model. The Gain Score Model measures the gain in two adjacent years. The

Layered Effects Model is similar to the Sanders, Saxton, and Horn (1997) TVAAS Model, where school effects are layered year after year, and appropriately estimates the variance when the study context changes. Each of the models described here in the text and in Table 1 demonstrates the differences in complexity of the various modeling options. Depending upon the context, researchers have a variety of model options when modeling school effects on student growth. The next section, Sensitivity of Value Added Effects, builds upon the previous model information and addresses issues of sensitivity of teacher and school effects to different factors (e.g., model specifications, outcome measures).

In summary, the VAM characteristics and modeling approaches vary widely. They differ in the number of data points, growth conceptualization, use of covariates, hierarchical levels, effects as fixed or random, and persistence of effects. Two of the most utilized models in examining school effects are the Gain Score Model and the Layered Effects Model. Each of these models has strengths and weaknesses, and neither of these, nor any other model, have been deemed optimal.

Sensitivity of Value Added Effects

Previous studies have compared value added effect results from varied models and outcomes. These types of studies are called sensitivity analysis studies. Sensitivity analyses are defined as studies that are used to determine whether the results and implications of a study differ when modifications are made to the study (Forrester, Breierova, & Choudhari, 2001). Sensitivity analyses are used to determine how "sensitive" estimates may be to changes in the value of the parameters of the model and

to changes in the structure of the model (Forrester, Breierova, & Choudhari, 2001). Sensitivity analysis studies investigating value added effects have been limited despite their important role in highlighting the different inferences that may occur.

Studies investigating the sensitivity of value added estimates from varied models and outcomes are the focus of the remainder of this literature review. Again, due to the limited amount of sensitivity studies investigating value added effects, this study highlighted the findings and recommendations from teacher effect studies, then, transitioned to findings and recommendations from school effect studies. Studies of teacher effects are different from studies of school effects and may draw varied inferences given the context; however, these studies oftentimes use the same model specifications and outcome measures to estimate value added effects, whether for schools or teachers. Thus, it is also important to compare the similarities and differences in findings from teacher effect studies that compared models and outcomes to the results found from school effect studies, given the current focus on school as well as teacher accountability.

Value added teacher effects. Schmitz (2007) investigated the sensitivity of estimated teachers' effects to different hierarchical linear model parameterizations to determine whether increased model sophistication would lead to substantially different estimated teacher effects. The Schmitz (2007) study was similar to Tekwe et al.'s (2004) in that the simple fixed effect model was used to provide a baseline of teacher effect estimates. The other models Schmitz investigated in 2007 were unadjusted and adjusted two-level Hierarchical Linear Models, unadjusted and adjusted three-level Hierarchical

Linear Models, and unadjusted and adjusted Cumulative Effects Models (a Cross-Classified Random Effects Model).

Schmitz (2007) found that there was little difference in the models' estimation of teachers' effects, besides the adjusted Cumulative Effects Model, which attributed more variability in student gains to teacher effects than the other models. Both the unadjusted and adjusted Cumulative Effects Models outperformed the Gain Score Models; the modeling of growth curves, he found, provided different estimates of teacher effects than those obtained from change scores. He determined there were no systematic differences in consistency of the teacher effect rankings for the lowest and highest 25% of teachers found among the models compared. The outcomes of his study seemed consistent for both extreme groups and better than the average group rankings. Schmitz (2007) then ran additional analyses examining correlations between mathematics and reading from three of the models; these ranged from .442 to .648 and had a mean of .527. As a result, he recommended additional studies modeling different academic subjects separately when estimating teacher effectiveness and utilizing a variety of cohorts to explore the extent to which different contexts impact school and teacher effects. The investigation of outcome measures seemed to be an important area to further explore, along with model specifications, since reading and mathematics outcomes failed to correlate to a high degree in the Schmitz (2007) study.

Lockwood et al. (2007) examined the sensitivity of estimated value added teacher effects to two subscales of a mathematics assessment (Stanford 9) using a Gain Score Model, Covariate Adjustment Model, Complete Persistence Model, and a General Variable Persistence Model with varied degrees of control for a number of student

background characteristics. Scale scores from the Stanford 9 mathematics assessment using only the Problem Solving and Procedures subscales were the outcome measures for the study (Lockwood et al., 2007). The total scale score was available but not used in their study (Harcourt Brace Educational Measurement, 1997; Lockwood et al., 2007). The correlations between the outcome measures from the study ranged from .01 to .46 for years two and three (Lockwood et al., 2007).

Overall, Lockwood et al. (2007) found that the sensitivity of the estimates to models and controls was only slight (i.e., high correlations) in comparison to the sensitivity to the achievement outcomes. Their study indicated that the smallest of any of the correlations related to changing models or controls was .49, which was larger than the largest correlation of .46 between teacher effects from the Procedures and Problem Solving outcomes using any of the combinations of models or controls. Also, across a range of model specifications, their estimates indicated that value added teacher effects were extremely sensitive to the achievement outcome used to create them. They found that not only were value added estimates sensitive to modeling choices, but the outcome measures could be extremely sensitive, possibly rendering inconsistent rankings of schools or teachers. They recommended additional studies to investigate the effects of varying the outcome measure in other contexts, expanding the range of measures to include different test publishers, different item formats, and alternative ways to create sub-scales. Also, they proposed the need to further investigate (a) how findings would be affected by changes in the student or teacher population or outcome measure used and (b) the sensitivity of the estimates to different ways of combining information from test items.

Both the Schmitz (2007) and the Lockwood et al. (2007) sensitivity analysis studies focused on teacher effect estimates with varied model specifications and outcome measures; the discussions and issues were similar in the examination of school effects as well (McCaffrey et al., 2003). There have been a limited number of studies examining the sensitivity of model specifications and outcome measures, whether investigating teacher effects, as in the studies of McCaffrey et al., (2003), Schmitz (2007), and Lockwood et al. (2007), or the investigation of school effects, as in the studies of Briggs and Weeks (2007, 2011), Schmitz and Raymond (2008), and Tekwe et al. (2004). Different models and different outcome measures (whether within the same construct or across subjects) may alter school or teacher effect rankings and should be investigated further for consistency with the findings from the previously mentioned studies. Value added estimates, whether teacher or school, are of interest in facilitating decision making; therefore, continued investigation of decisions and methods that impact these estimates are critical for accuracy and decision making about teachers and schools.

Value added school effects. Tekwe et al.'s study (2004), was one of the earliest value added modeling studies to make comparisons between various models. Their study compared several models, one of which was the Simple Fixed Effects Model, which is a school-specific mean change score minus a district-wide mean change score. Two similar two-level simple change score models were the (1) U_HLMM, with no covariates and a random intercept-only model for one, and (2) the demographic and intake-adjusted change score model, A_HLMM. The last model compared was the multivariate (i.e., TVAAS/Layered Effects Model).

The Tekwe et al. (2004) findings indicated that the global impact of the TVAAS/Layered Effects, compared to the Simple Fixed Effects Model, was small in the study of two years of data. Also, they found consistent disagreement between the TVAAS and the A_HLMM; however, they hypothesized that disagreement would decrease when analyzing three or more points in time. When only two points in time were considered, they recommended the Simple Fixed Effects Model or A_HLMM over TVAAS/Layered Effects. Also, shrinkage by itself and multivariate analysis had little impact on the value added assessment of school performance (Tekwe et al., 2004; Schmitz, 2007). They recommended additional research with these models using similar data and investigations into whether the TVAAS/Layered Effects would produce different results from the Simple Fixed Effects Model when more than two years of data were analyzed. They indicated that the model should also continue to be investigated for methodological issues and in applied situations within different contexts when investigating school and teacher effects.

Tekwe et al. (2004) found that when using the Simple Unadjusted Change Score Model and the Multivariate Layered Effects Model, Pearson correlations ranged from .96 (Grade 4) to .98 (Grade 3) in mathematics. They used the same models for reading and found Pearson correlations ranging from .94 (Grade 4) to .99 (Grade 3). Briggs and Weeks (2011) examined school effects between three model specifications (i.e., Constrained Persistence Model, Alternate Constrained Persistence Model, and a Layered Effects Model). In mathematics, Briggs and Weeks (2011) found that the correlations ranged from .47 (Grade 5) to .93 (Grade 6) and for reading they found correlations ranging from .58 (Grade 5) to .98 (Grade 8).

When reviewing previous sensitivity analysis studies related to value added teacher and school effects, it appears that changes in model specification do not have a significant bearing upon value added results. However, changes in the outcome measure seem to have a significant bearing upon the results when examining teacher effects. This study examined whether a similar pattern would occur with school effects.

Literature Review Summary

Educational accountability is one of the current approaches that is being used to improve schools and the skills of the future workforce in the United States. There are a variety of models that have different purposes, each with strengths and weaknesses. VAMs are one approach that estimate the unique effects schools and teachers have on student achievement growth; however, VAMs are still somewhat controversial due to methodological uncertainties. VAMs cannot be used to make causal inferences about student growth attributed from a particular school or teacher in a specific year, unless random assignment is utilized. While there is no clarity on which models provide the best estimates because the models vary in complexity and purpose, VAMs have characteristics that are more appropriate in certain research contexts.

Sensitivity analysis studies are used to determine the stability of estimates in these models and make steps in determining whether VAMs should be used in educational accountability. Previous research studies with various models have indicated the value added estimates were more sensitive given certain conditions which led to recommendations for future investigations of VAMs addressing a host of methodological issues. Model specifications may not be as important as the outcome measure in teacher

effects, but there have been a limited number of studies that examine the extent to which outcomes impact school effects estimates. Therefore, this study used the Gain Score Model and the Layered Effects Model to examine the sensitivity of school effects to mathematics and reading outcomes, as well as the school ranking when modifications were made to these models and outcomes.

Chapter III: Method

Overview

This chapter discusses the methods used to investigate school effects from outcome measures (mathematics and reading) and model specifications (Gain Score and Layered Effects). It begins with the purpose of the study, the research questions, the research design, a description of the population and sample, and a description of the variables and measures. These are followed by a description of the data collection and the analysis methods used, then proceeds with a description of how missing data and controls were handled. The chapter ends by describing the Gain Score Model and the Layered Effects Model.

Purpose of the Study

The purpose of this study was to examine two Value Added Model specifications to understand similarities and differences in school effect results. Specifically, this study correlated value added school effect estimates, which were derived from two model specifications (Gain Score and Layered Effects) and two outcome measures (mathematics and reading). Next, the school rankings were compared and correlated using the same models and outcome measures (or outcomes). Conducting these analyses (1) demonstrated how stable the value added school effect estimates were when the models and the outcomes were altered, (2) determined whether these models could be used

interchangeably to compare schools, (3) examined whether school effect results were similar to the previous findings for teacher effects, and (4) added to the methodological research literature required for a fuller understanding of the implications of using VAMs for educational accountability.

Research Questions

At the beginning of this study, several specific research questions were posed. The goal was to address issues from previous studies, as well as to explore other areas that had not been addressed in the literature. There were four specific research questions addressed in this study:

> **Research Question 1.** What is the relationship between school effect estimates from the Gain Score Model when mathematics achievement scores are used versus reading achievement scores? What is the relationship between school effect estimates from the Layered Effects Model when mathematics achievement scores are used versus reading achievement scores?
>
> **Research Question 2.** What is the relationship between school effect estimates based on mathematics achievement scores when the Gain Score Model is used versus the Layered Effects Model? What is the relationship between school effect estimates based on reading achievement scores when the Gain Score Model is used versus the Layered Effects Model?

Research Question 3. What is the relationship between school rankings from the Gain Score Model when mathematics achievement scores are used versus reading achievement scores? What is the relationship between school rankings from the Layered Effects Model when mathematics achievement scores are used versus reading achievement scores?

Research Question 4. What is the relationship between school rankings based on mathematics achievement scores when the Gain Score Model is used versus the Layered Effects Model? What is the relationship between school rankings based on reading achievement scores when the Gain Score Model is used versus the Layered Effects Model?

In essence, the first question examined the relationship between school effects from the outcome measures while holding the models constant. The second question examined the relationship between school effects from the models while holding the outcome measures constant. The third question examined the relationship between school rankings from the outcome measures while holding the models constant. Then, the fourth question examined the relationship between school rankings from the models while holding the outcome measures constant.

Overview of the Research Design

The approach used to answer the research questions was a non-experimental correlational longitudinal design. This enabled comparisons of the value added school

effect estimates and school rankings obtained between two model specifications and two outcome measures. The school effects were investigated for the outcome measures (mathematics and reading) and the model specifications (Gain Score and Layered Effects Models). Analysis procedures utilized existing Florida Comprehensive Assessment Test (FCAT) mathematics and reading achievement test scores from elementary and middle school students in Grades 3 through 5 and Grades 6 through 8 from a moderately large and rural school district in Florida.

This study was non-experimental, meaning that no variables were manipulated in the study, such as randomly assigning students or teachers to the schools. It was longitudinal, meaning that it followed students from year to year, allowing the estimation of academic achievement gains from year to year, within a particular school. This sample of students not only contained students who remained throughout the study for all six years, but also students who were enrolled one to five years as well. Using both models and outcomes, school effect estimates were generated, which were used in the Pearson product moment correlations. These correlations were used to compare the school effects between models and outcomes. Afterwards, schools were ranked using these school effect estimates and the school rankings were then used in the Spearman rank correlations as with the school effect estimates.

Population and Sample

Florida has 18.8 million residents and is the 4th largest state in population behind California with 37.3 million, Texas with 25.2 million, and New York with 19.4 million residents (United States Census Bureau [USCB], 2010). The USCB

(2012) reported that 21.3% of Florida's population was under the age of 18 in 2010. Toppo (2006), from USA Today, reported that Florida had nine of the 50 largest school districts with three being in the top 10 largest districts. Those three largest districts were Miami-Dade as fourth, Broward as fifth, and Hillsborough as the 10th in the United States. The percentages of residents for the state by race were estimated in 2010 as: White, 75%; Black, 16%; Native American, 0.4%; Asian, 2.4%; Native Hawaiian, 0.1%; Multiracial, 2.5%; Hispanic, 22.5% and White non-Hispanic, 57.9 % (USCB, 2012).

Each school and district throughout the state of Florida is given a school grade, ranging from the best grade, *A,* to the worst, *F*. School grades utilize a point system where schools obtain points for students who score "high" on the FCAT and make annual learning gains. The particular district used in this study had approximately 464,697 residents in 2010, of which 21.2% were under the age of 18 in 2010 (USCB, 2012). The racial make-up was estimated in 2010 as: White, 88.2%; Black, 4.5%; Native American, 0.4%; Asian, 2.1%; Native Hawaiian (value greater than zero but less than half unit of measure shown); Multiracial, 2.2%; Hispanic, 11.7%; and White non-Hispanic, 80.1% (USCB, 2012). A 2007-2008 superintendent report indicated that the district was among the largest 100 school districts within the nation and one of the largest school districts in the state of Florida. Additionally, approximately 86% of the residents 25 and older were high school graduates and approximately 20% had bachelor's degrees or higher. The median family income for the county was $44,228 in 2006-2010, which was slightly less than the state median of $47,661 and the United States' median of $51,914. The graduation rate for the high

schools was 85.5%, which was slightly lower than the state rate of 88.5%. The dropout rate was lower, at 1% compared to the state rate of 1.9% (Florida Department of Education [FDOE], 2011).

The school district had 45 elementary, 15 middle, 13 high, 12 combination, and 20 alternative, charter, and juvenile justice schools that served approximately 57,000 students for at least one point in time. This study included a pilot study to examine school effects from another Value Added Model (VAM), in addition to the Layered Effects Model used in this study. The data provided for the study included six files, one for each year of the study (i.e., 2005-2010). The number of cases per file ranged from 33,470 to 43,013, which included Grades 3 through 12, as well as those in adult education. There was an average of 39,991cases per file, with an average of 3,333 cases per grade.

Convergence problems were experienced in the pilot study using the full dataset with Grades 3 through 5 and Grades 6 through 8 for the district, due to insufficient memory. It was also discovered that estimation problems resulted when the sample sizes were small (i.e., below 50). To avoid these prior challenges, this study examined students in one grade for each year (e.g., 3rd graders in 2005, 4th graders in 2006, etc. ending with 8th graders in 2010). This sample of students was not consistent each year and was inclusive of students who remained in the district throughout the study period, as well as students who joined or left at various times during the study.

During the study period, the district had continual increases in their student population, which prompted the addition of several new schools over the years. With the

increases, the district also had a highly transient student population, which may be one of the causes for students' missing data found within the dataset used in this study. Both situations likely had an impact on frequency counts for each grade level and impacted the school effect estimates and rankings. There were 7,899 unique students who were examined from the 3rd Grade in 2005 through the 8th Grade in 2010, enrolled in the district. Each of these students had at least one test score in reading or mathematics; however, nearly 25% of the students were missing 10 or 11 of the 12 possible scores between reading and mathematics over the six years of the study.

There were 18 schools removed from the analysis due to fewer than 50 students enrolled for more than one of three years in a particular school. These were the charter, alternative, and juvenile justice schools that usually had lower student populations than the traditional elementary and middle schools. Removing these schools excluded 114 students from elementary schools and 129 students from middle schools, due to their school enrollments; 160 students who had duplicate records were also removed. There were a total of 7,496 students who were observed as enrolled in the district as 3rd graders in 2005 through 8th graders in 2010 and these students attended 40 regular elementary and 15 regular middle schools. On average, there were approximately 5,000 students having a reading and/or mathematics score at each grade level. The differences between the total enrolled students and those with actual reading and mathematics scores were attributed to the transience in and out of the district, such as students enrolling at some time during the school year, but leaving before testing took place in the spring at the end of the school year.

Missing data were examined to determine whether the missingness (percentage of missing test scores for a student) correlated, or had a relationship, with the outcome variables. Of the approximately 5,000 students who had mathematics and reading scores, 2,791 had a mathematics score and 2,802 had a reading score for each year of the study. The correlations between the outcome variables and missingness were low and negative and ranged from, -.11 to -.04 in reading and -.16 to -.05 in mathematics. The data were assumed to be missing-at-random, due to the highly transient population of students entering, leaving, and returning on a regular basis over the six-year study with many students having fewer than six test scores in mathematics and reading during the study period. The sample of students included general education students, English Language Learners (ELLs), and Exceptional Student Education (ESE) students who tested in one of the grades in the district during the study period.

The students used for the analyses varied in their demographic characteristics; sample sizes fluctuated due to lack of information provided. Of the student data that were reported, the majority were identified as White and there were fairly equal numbers for gender and free and reduced lunch status. Most students did not have a disability and were classified as non-ESOL (English for Speakers of Other Languages). Specifically, there were 4,888 students with a reported race/ethnic classification: 123 (2.51%) Asian, 259 (5.30%) Black, 857 (17.53%) Hispanic, 21 (0.42%) Native American Indian, 212 (4.33%) Multiracial, and 3,417 (69.91%) White. There were 4,782 students with a reported gender: 2,358 (49.30%) females and 2,424 (50.70%) males. There were 4,834 students with a reported free and reduced price lunch status: 2,196 (45.93%) as eligible and 2,585 (54.07%) as not

eligible. There were 4,720 students with a reported disability status: 3,601 (76.30%) having no disability and 1,119 (23.70%) having some type of disability. There were 6,365 students with a reported Limited English Proficiency status: 5,871 (92.24%) were non-ESOL and 494 (7.76%) in ESOL or having exited an ESOL program during the study period. Overall, the sample used was representative of the district, as a whole, and may be applicable to similar populations throughout the state and/or nation.

Variables and Measures

Previous reliability and validity studies conducted on the FCAT items for the outcome measures used in this study may bring further clarity to the value added results for stakeholders interested in controls and psychometric properties of the tests. Students in Florida test each year with the FCAT in Grades 3 through 10. Those tested included ELL and ESE students enrolled in the tested grade levels. Administration accommodations were provided for eligible ELL and ESE students taking the regular exam (FCAT, 2008). Florida uses the Item Response Theory (IRT) to score and equate FCAT results from year to year (Orr, 2007).

The two outcome measures used for the analysis were the FCAT scores for reading and mathematics for Grades 3 through 8. According to the FCAT (2008), *Understanding FCAT Reports*, the FCAT was designed to align with criterion-referenced standards, but also includes norm-referenced standards to measure student performance in Grades 3 through 10. It reports that the reading portion of the FCAT is designed to measure achievement in applying reading processes to construct

meaning from both informational and literary texts. The mathematics portion of the FCAT measures achievement in five areas: (1) number sense, concepts, and operations; (2) measurement; (3) geometry and spatial sense; (4) algebraic thinking; and (5) data analysis and probability. Both portions of the exam have multiple-choice, gridded-response, short answer, and extended response items (FCAT, 2008). According to FCAT (2008), the progress students make from year to year is tracked by the Developmental Scale Score (DSS), which is a type of scale score used to determine annual progress of students from grade to grade. FCAT (2008) reported that the DSS for both reading and mathematics ranged from 86 to 3,008 across Grades 3 through 10 and gains in these scores could be calculated by subtracting a previous year's DSS from a current year's DSS (e.g., 2008 DSS - 2007 DSS = DSS Gain). This number may be large (for students moving from a low Achievement Level-1 score to a low Achievement Level-2 score) or small (for a student that maintains a high score in Achievement Level-4), indicating that the DSS Gain can be understood best when also considering the achievement level for the two scores (FCAT, 2008). This study only utilized DSS in evaluating student progress across years and in the estimation of the school effects and rankings.

Between 2001 and 2003, reliability of the FCAT scores were examined for mathematics and reading. The internal consistency of the scores from the exams for Grades 3 through 8 ranged from .87 to .94 for both the Classical Cronbach Alpha and IRT Marginal reliability statistics (FDOE, 2004). Similar results were found in the 2007 FDOE Assessment and Accountability Book for 2001-2006. Content, Criterion, and Construct were three types of validity evidence examined for the FCAT. Content

validity evidence was collected from item reviews of educators, item specifications, pilot tests, bias reviews, field tests, and the process of equating one test to another base test to match content coverage and test statistics (FDOE, 2004). Criterion validity was evidenced by correlations on the criterion referenced portion of the Sunshine State Standards (SSS) with scores on the norm-reference portion (Stanford 9) and the correlations in Grade 3 through Grade 10, for reading and mathematics, ranged from .76 to .85 (FDOE, 2004). Construct validity was evidenced by confirmatory and exploratory factor analysis, as well as convergent and discriminant analyses (FDOE, 2004).

In the state of Florida, AYP measurements targeted the performance and participation of various subgroups based on race or ethnicity, socioeconomic status, disability, and ELL. There have been varied opinions about whether to include covariates in VAMs (Ballou, Sanders, & Wright, 2004; McCaffrey et al., 2003). The decision about whether to include controls is dependent upon the research questions and the findings of previous studies. None of the aforementioned subgroup predictors were used with either of the models in this study, which followed the approach of Ballou, Sanders and Wright (2004) to exclude covariates with VAMs, particularly with the TVAAS or Layered Effects Model. The covariates were omitted to estimate the parameters in a similar fashion demonstrated in the research by Sanders. This is because the Growth Model pilot program implemented by Secretary Spellings in 2005 required that grade level proficiency be the standard of achievement and not student or school characteristics (Ballou, Sanders, & Wright, 2004).

Data Collection Procedures

A proposal for this study was developed and submitted to the Director of Research and Evaluation of the district whose data were used. Once permission was granted and data were obtained, a more detailed study proposal was submitted to the Institutional Review Board (IRB) under an Exemption Certification Status. The original data from the district were obtained from the Florida Department of Education by the school district and it was assumed that the integrity of the data obtained was adequate for district analysis, reporting, and for addressing the research questions for this study. The data used for this study were supplied by the district on a compact disc in 2010. The files were SPSS files that were imported into SAS 9.2 and merged by student identification number. Once permission to conduct the research was obtained from the IRB, the data were cleaned for the data analysis.

Data Analysis

Various data analyses took place in investigating the research questions posed in this study. Univariate descriptive analyses, multivariate analyses, bivariate analyses, and missing data were all addressed. Data screening procedures and assumption checks followed those recommended and described by Raudenbush and Bryk (2002). Before analysis took place, variable names in the data were checked for consistency and modified, as needed, to eliminate confusion in later analyses. The analyses were conducted in two phases: (1) preliminary analyses that encompassed univariate statistics and (2) the primary (multivariate and bivariate) analyses, which entailed the specification of proposed models, the estimation of model parameters, and the ranking of schools

based on the school effects. The univariate, multivariate, and bivariate analyses were all conducted using SAS 9.2.

Univariate analyses. The first step in the analysis began with examination of the univariate frequency distributions for the mathematics and reading outcome variables, which provided a descriptive picture of the data. The preliminary data that were examined included the average reading and mathematics scores per grade, the number of students testing at each point in time, the number of schools observed for each point in time, and an examination of the missing frequencies and percentages.

Multivariate analyses. After examining frequencies, percentages, missing data, and decisions regarding controls, the next step was to conduct the primary analyses for this study. The models used were fit to the data, which generated fixed effects, variance components, fit indices, and random effects. The data were then screened for violations of assumptions discussed by Raudenbush and Bryk (2002).

Two key assumptions examined with multilevel models included determining whether the residuals were normally distributed and whether homoscedacity of residuals was observed. These assumptions were examined for the level-1 residuals. Normality was examined using box-and-whisker plots, skewness, and kurtosis values of the level-1 residuals. Homoscedacity was examined using plots of the level-1 residuals against the predicted average values for the outcome measures. A random subset of participants was also used to examine the level-1 residuals and verify the influence diagnostics. In addition to examining the assumptions, the Intraclass Correlation Coefficient (ICC) was calculated using the “between” and “within” school variance components. The ICC measures the proportion of variance in the outcome variable found between groups. Here,

it was the proportion of variance in the outcome variable found between schools (Raudenbush & Bryk, 2002). Even though the ICC was low, the parameters were still estimated using multilevel modeling procedures to determine the amount of clustering that occurs within schools.

This study estimated school effect parameters from two different Value Added Models: the Gain Score Model and the Layered Effects Model. The Gain Score Model utilizes a two-level model, where students are nested within schools and their scores are the difference in previous and current year scores. The Layered Effects Model also utilizes a two-level model, with students at level-1 nested within schools at level-2; however, it measures growth over multiple years and across research contexts. For the Layered Effects Model, students are nested in schools and cross-classified at level-2 according to their school of attendance (i.e., elementary and middle school). These structures allow for variance to be properly modeled when students cross boundaries, such as attending one elementary school and then advancing to middle school. The teacher level was omitted from this analysis to maintain a limited focus on examining school effects. Though the teacher level of effect was omitted, a portion of the student and school variance could be explained with its inclusion. When the Gain Score and Layered Effect models were utilized, random school effects were generated.

The random school effects are deviations from the grand average and can be described as the estimated effect that a school has on a student's academic performance or growth during a specific time period. It was assumed that all students received an average (expected) schooling experience regardless of the school the student attended. In reality, some schools may seem to have a greater effect than do others on students'

academic performance and growth. In comparing school effects, schools are often described as providing an "average," "above average," or "below average" schooling experience to students as a result of attending the school.

There were two main steps involved in obtaining the answer for each of the four research questions for this study. The first step used to answer Questions 1 and 2, was to estimate the school effects for each school using two outcome measures (mathematics and reading) and two model specifications (Gain Score Model and Layered Effects Model). The first step used to answer Questions 3 and 4, was to take the estimated school effects from the outcome measures and model specifications and rank each school from largest to smallest. School effects were estimated for each school from Grades 3 through 5 and Grades 6 through 8. The "3rd Grade effect" is described as the school selection effect and is typically not used in making the school comparisons. These pave the way for the second step to answer the four research questions addressed in this study.

Bivariate analyses. The second step used to answer the research questions was to take the school effects and calculate the Pearson product moment correlation between the outcome measures from both model specifications. Question 1 was then answered by calculating the correlations between the school effects from mathematics and reading while holding the Gain Score Model and Layered Effects Model constant. Similarly, Question 2 was then answered by calculating the correlations between the school effects of the Gain Score Model and Layered Effects Model, while holding mathematics and reading constant. Lastly, Questions 3 and 4 were then answered by calculating the Spearman rank correlations for the school rankings in the same manner as Question 1 and 2. In each instance, "high" and "positive" correlations were an indication of similar

school effect estimates or school rankings between the model specifications and the outcome measures.

In summary, this Data Analysis section described the broad analyses executed to answer the research questions addressed in this study. The preliminary univariate analyses were utilized to describe the data through means, frequencies, and outliers. The primary multivariate and bivariate analyses were instrumental in answering the research questions.

Model Specification and Estimation

Prior to conducting analyses and presenting results for the Gain Score Model and the Layered Effects Model, it is important to clearly understand the model specifications and methods for estimating parameters. The school effects used for the correlations and rankings were estimated from the Gain Score Model at four points in time, (i.e., Grades 4, 5, 7, and 8) and the Layered Effects Model at six points in time (i.e., Grades 3, 4, 5, 6, 7, and 8). These estimates were snapshots of a particular school's effect upon their enrolled students at a particular grade.

Gain Score Model specification. The Gain Score Model is described by Equations 1.1 through 1.7 and the model was utilized in calculating the gain in each student's test scores between two adjacent years (e.g., 3rd Grade to 4th Grade). Those gains were then aggregated to generate an average gain score for each school. Next, average gains for each school were aggregated to generate a grand average gain or fixed effect, which is the average gain in DSS for students for all schools throughout the district in the elementary and middle school grades in reading and mathematics. The

random school effect indicates how many gain-points "above," "below," or "equal with" the grand mean a particular school had between two adjacent years. The gain in two adjacent test scores for a student is a function of the grand average or fixed effect for all schools at a particular grade, plus the random school effect of the student's school on the fixed effect for all schools at a particular grade, plus the student's residual test score in their school at a particular grade.

Level 1 (Student):

$$d_{ijg} = \beta_{0jg} + e_{ijg} \tag{1.1}$$

Level 2 (School):

$$\beta_{0jg} = \gamma_{0g} + u_{0jg} \tag{1.2}$$

The combined model is found in equation 1.3:

$$d_{ijg} = \gamma_{0g} + u_{0jg} + e_{ijg} \tag{1.3}$$

where

$d_{ijg} = \mathrm{Y}_{ijg2} - \mathrm{Y}_{ijg1}$ is the difference in a student's test scores between two adjacent grades,

Y_{ijg} is the DSS for the *i*th student in the *j*th school, at grade g,

β_{0jg} is the average gain between the DSS of students, in the *j*th school at grade g,

γ_{0g} is the grand average gain in DSS between all schools, at grade *g*,

u_{0jg} is the random school effect of the *jth* school on the grand mean, at grade *g,* and

e_{ijg} is the residual test score for *ith* student in the *jth* school at grade *g*.

The e_{ijg} and u_{0j} are independent and randomly distributed above and below the mean of 0: $e_{ijg} \overset{iid}{\sim} N(0,\sigma^2)$, $u_{0j} \overset{iid}{\sim} N(0,\tau_{00})$.

The variance-covariance G-matrix for the random school effects, u_{0jg}, has a block diagonal structure with an identical block for each school and the elements in each block are the variances of the intercepts, τ_{00} (Wang, 2006). Similarly, the level-1 variance-covariance R-matrix (i.e. σ^2 or sigma matrix) for the student residuals, e_{ijg}, also has a diagonal structure with an identical block for each student, *i* (Wang, 2006).

Equations 1.4 through 1.7 demonstrate the gain in test scores across the elementary and middle school grades, Grades 3 through 5 and Grades 6 through 8.

<u>Elementary School:</u>

$$d_{ij4-3} = \gamma_{04} + u_{0j4} + e_{ij4} \tag{1.4}$$

$$d_{ij5-4} = \gamma_{05} + u_{0j5} + e_{ij5} \tag{1.5}$$

<u>Middle School:</u>

$$d_{ij7-6} = \gamma_{07} + u_{0j7} + e_{ij7} \tag{1.6}$$

$$d_{ij8-7} = \gamma_{08} + u_{0j8} + e_{ij8} \tag{1.7}$$

Gain Score Model estimation. The model parameters were estimated through Restricted Maximum Likelihood (REML) using the Newton-Raphson algorithm. The Gain Score Model estimates the gain in test scores for each student, aggregates these gains as an average for each school, and then aggregates each school average to obtain a grand average gain. The school average and grand average gains are the fixed effects. The deflection of each school from the grand average gain is the random school effect and it assumes that the research context remains the same for both points in time. A separate model was run to estimate school effects for each adjacent year.

Layered Effects Model specification. The Layered Effects Model is the most appropriate model for growth when the research context changes from one context to another within a particular study (Raudenbush & Bryk, 2002). The Layered Effects Model found in Equations 1.8 through 1.14 estimates parameters from the grand average DSS, or fixed effect, based on all students in all schools at a particular grade. The school averages are aggregations of students' DSS at each grade. Each school average is then subtracted from the grand average, providing the random school effects, which are deviations of each individual school from the grand average at a particular grade. The random school effects accumulate over time and are assumed not to diminish over time, but remain indefinitely. When using the Layered Effects Model, a student's DSS is a function of the average score for all students in all schools at a certain grade, or fixed effect, plus the sum of all random school effects for each grade, plus the random residual test score for the student in their school at a particular grade.

The general model is specified in equation 1.8 as:

$$Y_{tig} = \mu_g + \sum_{g=1}^{g} u_{gj} + e_{gij} \qquad (1.8)$$

where

Y_{tig} is the score for the *i*th student in the *j*th school, at grade *g*,

μ_g is the average DSS between all students in all schools at grade *g,*

u_{gj} is the random school effect of the *jth* school on the average DSS at grade *g,*

and

e_{gij} is the residual test score for the *ith* student in the *jth* school at grade *g*.

The e_{gij} and u_{gj} are randomly distributed above and below a mean of 0:

$\text{Var}(e_{gij}) = \sigma^2$, $\text{Var}(u_{gj}) = \tau_{00}^2$.

At the student level, the variance-covariance matrix of the e_{ijg} is unstructured, allowing for different variances at each point in time and possibly a nonzero and non-constant correlation of scores from different years or grades (McCaffrey et al., 2004). The Layered Effects Model also allows the variance of school effects, u_{gj}, to vary across grades (McCaffrey et al., 2004). Equations 1.9 through 1.14 demonstrate the layering of school effects across the elementary and middle school grades, Grades 3 through 5 and Grades 6 through 8.

Elementary School:

$$Y_{3ij} = \mu_3 + u_{3j} + e_{3ij} \tag{1.9}$$

$$Y_{4ij} = \mu_4 + u_{3j} + u_{4j} + e_{4ij} \tag{1.10}$$

$$Y_{5ij} = \mu_5 + u_{3j} + u_{4j} + u_{5j} + e_{5ij} \tag{1.11}$$

Middle School:

$$Y_{6ij} = \mu_6 + u_{3j} + u_{4j} + u_{5j} + u_{6j} + e_{6ij} \tag{1.12}$$

$$Y_{7ij} = \mu_7 + u_{3j} + u_{4j} + u_{5j} + u_{6j} + u_{7j} + e_{7ij} \tag{1.13}$$

$$Y_{8ij} = \mu_8 + u_{3j} + u_{4j} + u_{5j} + u_{6j} + u_{7j} + u_{8j} + e_{8ij} \tag{1.14}$$

Before estimating the random school effects with the Layered Effects Model, a Z-matrix was developed separately and integrated with the original data set. The Z-Matrix is an *m x q* incidence matrix that allows for the inclusion of random effects in mixed effects models. This matrix is needed specifically for the Layered Effects Model to capture the random school effects of each school, *j*, at grade *g*. The sets of zg _ j variables are a set of dummy variables, where "1" indicates that student *i* was in school *j* during grade *g*. An illustration of a Z-matrix data set is found in Table 2 which describes two students in two of three school options with each student having six observations. In the example, both students attended the same elementary school, but attended different middle schools.

Table 2

Hypothetical Data Set Illustrating the Z-Matrix Needed for the Layered Effects Models

Student	School	Grade	Score	z4_1	z4_40	z4_50	z5_1	z5_40	z5_50	z6_1	z6_40	z6_50	z7_1	z7_40	z7_50	z8_1	z8_40	z8_50
1	1	3	1400	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	4	1500	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	5	1600	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	40	6	1630	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	40	7	1700	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
1	40	8	1800	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0
2	1	3	1650	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	4	1800	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	5	1840	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	50	6	1900	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2	50	7	2300	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
2	50	8	2450	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1

Layered Effects Model estimation. The model parameters were estimated through REML using the Newton-Raphson algorithm. The estimates of school effects were random versus fixed and were estimated with empirical Bayes estimation methods. The parameters for the school effects are generated differently between the Layered Effects Model and the Gain Score Model. The Layered Effects Model estimates school effects all at one time and generates school effect estimates for each school at each point in time. The SAS code used to fit and estimate the Gain Score and Layered Effects Models can be found in Appendix A. There were 40 elementary schools denoted as schools "1-40" and 15 middle schools denoted by schools "41-55". The cross-classification of students is found in Appendix B, Table B1.

Both the Gain Score Model and the Layered Effects Model estimate random school effects as deviations of individual schools from the grand average for all schools. However, one difference is that the Gain Score Model requires multiple Gain Score Models per adjacent years to estimate these effects, whereas, the Layered Effects Model estimates each year's effects using one model. The coding found in Appendix A of this study provides additional insight into how these parameters were estimated using SAS 9.2.This concludes the Model Specification and Estimation section of this study, which described in detail the models used in estimating the random school effects.

In summary, this chapter described the methods used to estimate the school effects from the model specifications and outcome measures. It included the statement of the purpose of the study, the research questions, the research design

approach, the description of the population and sample utilized, variables and measures, model specifications, and model estimations.

Chapter IV: Results

Overview

This chapter is segmented into two sections that describe the preliminary and primary findings from the data analysis using the Gain Score and Layered Effects Models. The preliminary results include descriptive analyses such as frequencies, means, skewness and kurtosis for mathematics and reading DSS, and an analysis of missing data. The primary results include a brief description of the models, followed by the fixed effects, variance components, and ICCs. For each model, there was an examination of assumptions and residuals. Next, the random school effect estimates, together with the standard errors derived from the Gain Score Model and Layered Effects Model for mathematics and reading outcome measures are provided in tables. For the random school effect estimates, Pearson product moment correlations were calculated from the school effects derived using the two outcome measures and two model specifications. Spearman rank correlation coefficients were calculated from the school rankings in a similar fashion as the school effects that were derived from the outcome measures and model specifications.

Preliminary Results

The preliminary results from the univariate analyses of mathematics and reading

outcome measures include: frequency counts for students and schools used in the calculations, means and standard deviations, skewness, kurtosis, outliers, extreme scores, and missing data. These results provide the descriptive portion about the sample of students and schools utilized in the study. The mobility of students in and out of the district may have some bearing upon the results.

Tables 3 and 4 present the descriptive statistics for the mathematics and reading outcome measures at each grade. Skewness is a measure of symmetry for a distribution of values and kurtosis is a measure of the flatness of the distribution (Cody & Smith, 2006). Outliers were found at each grade level with extreme scores in Grades 5, 7, and 8 for mathematics and Grades 7 and 8 for reading. Outliers were values that extended beyond the box-plot whiskers, but were between three inter-quartile ranges of the box boundaries (Cody & Smith, 2006). The extreme scores were values that extended beyond the three inter-quartile ranges (Cody & Smith, 2006). Values that were outliers or extreme were identified by a "square" in the box-plots. All of the scores were in the range of acceptable scores and therefore the scores remained for all analyses. The box-plots for mathematics and reading are in Figures 1 and 2.

Table 3 presents descriptive summary statistics for mathematics in the elementary and middle school grades. At both levels, scores increased each year with slightly greater increases in the elementary grades. The average score in mathematics was 1471 at the elementary level (i.e., average for 3rd, 4th, and 5th Grades) and 1795 at the middle school level (i.e., average for 6th, 7th, and 8th Grades). The skewness values for mathematics indicated "moderate" to "high" negative skew and the kurtosis values indicated "slight"

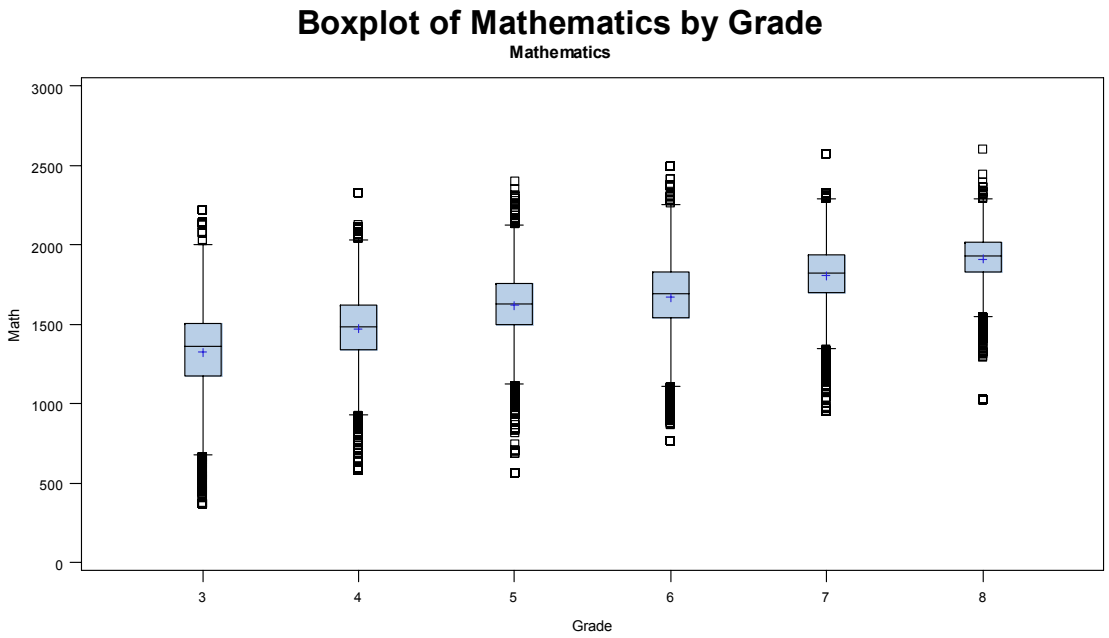

Figure 1. Box-Plot of Mathematics by Grade

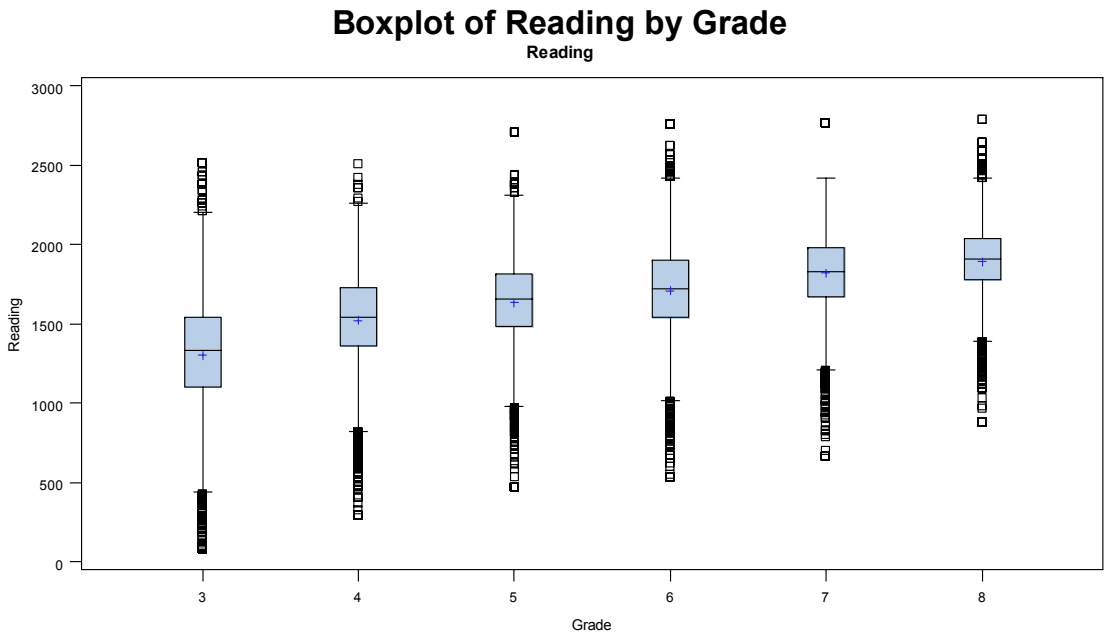

Figure 2. Box-Plot of Reading by Grade

to “high” levels of peaks around the mean. Outliers were found at each grade level and extreme scores were found in Grades 5, 7, and 8. All of the scores were in the range of acceptable scores and were therefore retained for all analyses.

Table 3

Descriptive Statistics for Students in Mathematics

Grade	*n*	*M*	*SD*	Skewness	Kurtosis	Outliers/Extreme
			Elementary Schools (*n* = 40)			
3	4748	1324	282.3	-0.7	1.3	Yes/No
4	4475	1471	238.7	-0.6	2.0	Yes/No
5	4749	1619	213.3	-0.7	2.6	Yes/Yes
			Middle Schools (*n* = 15)			
6	4874	1668	253.9	-0.8	1.8	Yes/No
7	4890	1805	211.6	-0.8	2.5	Yes/Yes
8	4936	1911	160.3	-1.1	4.0	Yes/Yes

Note. *n* indicates the number of students used in the calculations; *M* is the average score; *SD* is the standard deviation; skewness and kurtosis indicate the shape; outliers are values that extended beyond the box-plot whiskers but were between three inter-quartile ranges of the box boundaries; extreme score indicates scores that extend beyond the three inter-quartile ranges.

Table 4 presents descriptive summary statistics for reading in the elementary school and middle school grades. At both levels, mean scores increased each year with greater increases in the elementary grades. The average score in reading was 1485 at the elementary level (i.e., average for 3rd, 4th, and 5th Grades) and 1807 at the middle school level (i.e., average for 6th, 7th, and 8th Grades). The skewness and kurtosis values for reading indicated a fairly normal distribution with “slight” to “moderate” negative skewness and “slight” to “moderate” positive kurtosis, respectively. Outliers were found at each grade and extreme scores were found in Grades 7 and 8.

Table 4

Descriptive Statistics for Students in Reading

Grade	n	*M*	SD	Skewness	Kurtosis	Outliers/Extreme
			Elementary Schools (*n* = 40)			
3	4760	1301	379.6	-0.6	1.2	Yes/No
4	4484	1522	305.2	-1.0	2.6	Yes/No
5	4731	1633	292.4	-0.7	2.4	Yes/No
			Middle Schools (*n* = 15)			
6	4875	1708	306.2	-0.4	1.7	Yes/No
7	4895	1820	262.2	-0.3	2.3	Yes/Yes
8	4928	1894	215.9	-0.5	1.6	Yes/Yes

Note. n indicates the number of students used in the calculations; *M* is the average score; *SD* is the standard deviation; skewness and kurtosis indicate the shape; outliers are values that extended beyond the box-plot whiskers but were between three inter-quartile ranges of the box boundaries; extreme score indicates scores that extend beyond the three inter-quartile ranges.

Table 5 provides a listing of the number of students with test scores in each subject and grade for elementary schools, as well as the overall mean and standard deviation for a particular grade and subject. Below the grand average score, identical information is provided for each school by subject and grade. The numbers of students per school and subject vary by grade, with roughly 117 students per school per grade in elementary schools. The *NA* in the tables indicates that no information was available for the schools; these were schools that were under construction during that timeframe.

Table 6 displays the Pearson product moment correlations between mathematics and reading at the elementary grades. The correlations were moderately high and positive with values ranging from .62 to .82 over three years. Figures C1-C6 in Appendix C provide a visual display of the information found in Table 6.

Table 7 provides a listing of the number of students with test scores in each subject and grade for the middle schools, as well as the overall mean and standard deviation for a particular grade and subject. Below the overall average, the same information is provided for each school by subject and grade. As with the elementary

schools, the number of students vary by school, subject and grade. There were approximately 327 students per school, subject, and grade.

Table 8 displays the Pearson product moment correlations between reading and mathematics in the middle school grades. The correlations were moderately high and positive with values ranging between .67 and .84 over three years. Scatter plots between mathematics and reading are found in Appendix C.

Tables 3 through 8 provide the descriptions of the outcome measures used in the sample for this study. The number of mathematics and reading scores varied from grade to grade. Nearly one-third of the students were missing from one to eleven of the twelve possible test scores (e.g., total of 6 mathematics and 6 reading scores).

Missing Data

The sample had 7,496 students who attended 40 elementary and 15 middle schools; these were unique students who appeared in the dataset at some point during the study. At each grade, there were nearly 5,000 students with mathematics and/or reading schools; these were unique students who appeared in the dataset at some point during the study. At each grade, there were nearly 5,000 students with mathematics and/or reading scores. There were a total of 2,761 students with reading and mathematics scores for each grade. The remaining frequencies and percentages are for the number of DSS that were missing for students in mathematics and reading combined. Nearly 25% of the students had 10 or 11 missing scores (e.g., total of 6 mathematics and 6 reading scores). These were likely to be students who transferred into the district and tested at one or two points in time throughout the study period. Tables D1 through D5, found in Appendix D,

Table 5

Descriptive Statistics in Mathematics and Reading for Grades 3, 4, and 5

School	Subject	Grade 3			Grade 4			Grade 5		
		n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Elementary	Mathematics	4748	1324.0	282.3	4475	1471.0	238.7	4749	1619.0	213.3
Schools	Reading	4760	1301.0	379.6	4484	1522.0	305.2	4731	1633.0	292.4
1	Mathematics	81	1196.3	265.0	65	1365.7	193.2	65	1513.1	165.5
	Reading	81	1126.2	277.1	65	1457.4	188.2	65	1464.5	224.7
2	Mathematics	209	1419.8	226.8	110	1541.4	196.9	116	1682.6	183.6
	Reading	209	1449.5	320.8	110	1609.1	275.3	116	1709.7	272.2
3	Mathematics	146	1414.6	280.9	145	1488.5	234.3	150	1663.1	180.3
	Reading	147	1373.1	361.3	145	1563.5	280.0	150	1686.5	277.0
4	Mathematics	130	1249.5	280.5	120	1428.9	233.7	131	1576.6	192.8
	Reading	131	1226.2	325.1	126	1477.1	297.8	131	1597.6	266.5
5	Mathematics	129	1148.9	325.4	110	1419.5	228.6	99	1622.0	222.3
	Reading	129	1210.1	388.9	109	1490.8	282.5	101	1594.3	272.7
6	Mathematics	146	1247.5	258.7	121	1478.9	235.3	120	1641.1	194.3
	Reading	146	1273.7	356.6	121	1541.6	263.3	120	1634.7	266.1
7	Mathematics	166	1270.2	258.2	122	1381.3	245.5	116	1580.5	194.5
	Reading	167	1186.5	428.4	125	1365.4	322.5	114	1545.0	310.2

Table 5 (Continued)

Descriptive Statistics in Mathematics and Reading for Grades 3, 4, and 5

School	Subject	Grade 3			Grade 4			Grade 5		
		n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Elementary	Mathematics	4748	1324.0	282.3	4475	1471.0	238.7	4749	1619.0	213.3
Schools	Reading	4760	1301.0	379.6	4484	1522.0	305.2	4731	1633.0	292.4
8	Mathematics	141	1219.7	272.9	126	1402.4	249.8	121	1574.9	222.1
	Reading	141	1210.2	374.2	126	1400.6	297.4	121	1517.5	315.3
9	Mathematics		*NA*		83	1533.9	177.8	111	1653.6	155.8
	Reading				82	1613.4	281.0	111	1706.6	219.4
10	Mathematics		*NA*		82	1427.5	239.8	87	1519.3	232.9
	Reading				82	1448.3	299.3	86	1579.5	267.6
11	Mathematics		*NA*		82	1469.6	217.7	89	1658.1	171.7
	Reading				82	1560.6	237.3	90	1679.8	223.1
12	Mathematics	159	1311.5	281.3	149	1410.8	270.3	155	1564.7	197.5
	Reading	159	1254.4	434.7	151	1499.6	322.4	156	1574.3	314.9
13	Mathematics	151	1371.3	254.8	140	1486.6	200.6	148	1620.2	193.4
	Reading	151	1309.1	377.0	140	1516.1	279.0	147	1616.1	280.3
14	Mathematics	148	1439.2	229.0	88	1529.8	239.2	116	1617.7	221.6
	Reading	148	1428.3	333.1	86	1606.8	295.8	115	1652.8	304.6

Table 5 (Continued)

Descriptive Statistics in Mathematics and Reading for Grades 3, 4, and 5

School	Subject	Grade 3			Grade 4			Grade 5		
		n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Elementary	Mathematics	4748	1324.0	282.3	4475	1471.0	238.7	4749	1619.0	213.3
Schools	Reading	4760	1301.0	379.6	4484	1522.0	305.2	4731	1633.0	292.4
15	Mathematics	164	1344.3	316.2	141	1522.9	230.9	149	1658.3	222.5
	Reading	164	1309.8	407.9	141	1559.2	300.6	149	1683.3	323.2
16	Mathematics	113	1337.5	272.7	118	1489.0	204.8	118	1617.8	175.7
	Reading	113	1305.4	376.1	118	1526.5	299.3	116	1675.9	210.8
17	Mathematics	102	1264.4	322.4	85	1438.7	327.9	74	1569.0	295.3
	Reading	102	1283.5	391.9	86	1484.2	317.8	74	1551.1	354.2
18	Mathematics	144	1320.1	249.0	128	1404.0	220.9	137	1555.9	188.1
	Reading	144	1229.7	406.3	132	1449.2	316.7	135	1536.7	316.3
19	Mathematics	107	1366.8	266.7	109	1524.6	166.1	111	1662.9	163.9
	Reading	109	1327.8	310.6	109	1544.0	214.6	111	1661.6	214.3
20	Mathematics	76	1209.5	286.6	54	1450.3	203.5	51	1618.9	199.6
	Reading	76	1144.0	329.5	54	1450.1	246.9	52	1572.1	214.2
21	Mathematics	127	1253.8	282.1	94	1416.6	237.9	90	1621.7	160.1
	Reading	128	1207.3	399.1	93	1466.3	271.1	88	1575.8	246.0

Table 5 (Continued)

Descriptive Statistics in Mathematics and Reading for Grades 3, 4, and 5

School	Subject	Grade 3			Grade 4			Grade 5		
		n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Elementary	Mathematics	4748	1324.0	282.3	4475	1471.0	238.7	4749	1619.0	213.3
Schools	Reading	4760	1301.0	379.6	4484	1522.0	305.2	4731	1633.0	292.4
22	Mathematics	161	1363.4	259.8	104	1442.2	273.6	106	1584.4	200.4
	Reading	161	1294.4	349.1	106	1451.6	380.4	106	1570.2	294.6
23	Mathematics	132	1370.5	253.9	132	1462.6	239.5	134	1587.8	210.0
	Reading	132	1381.8	344.8	131	1536.5	288.8	134	1618.9	349.7
24	Mathematics	109	1409.2	251.3	102	1537.7	234.6	108	1626.3	201.5
	Reading	109	1339.7	442.7	102	1590.6	360.0	107	1703.2	262.3
25	Mathematics	137	1354.4	241.0	125	1488.4	245.1	149	1613.1	224.0
	Reading	136	1344.9	332.2	125	1526.4	283.4	147	1618.6	271.4
26	Mathematics	198	1338.9	289.5	108	1422.9	244.0	113	1623.3	228.5
	Reading	198	1297.8	375.0	108	1508.4	318.5	113	1658.7	304.8
27	Mathematics	116	1199.0	264.6	102	1388.2	228.5	94	1546.5	225.7
	Reading	116	1186.8	369.6	102	1447.6	252.6	95	1599.5	277.8
28	Mathematics	131	1234.2	265.4	104	1421.5	192.8	116	1537.7	233.0
	Reading	131	1204.6	336.7	103	1454.3	279.5	116	1543.7	329.5

Table 5 (Continued)

Descriptive Statistics in Mathematics and Reading for Grades 3, 4, and 5

School	Subject	Grade 3			Grade 4			Grade 5		
		n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Elementary	Mathematics	4748	1324.0	282.3	4475	1471.0	238.7	4749	1619.0	213.3
Schools	Reading	4760	1301.0	379.6	4484	1522.0	305.2	4731	1633.0	292.4
29	Mathematics	99	1390.6	258.2	110	1458.7	220.8	100	1616.5	188.8
	Reading	99	1360.9	363.9	110	1523.6	330.8	100	1586.6	312.2
30	Mathematics	115	1319.8	268.2	106	1450.7	238.3	126	1635.1	177.2
	Reading	115	1292.0	354.6	106	1525.2	324.1	126	1688.1	295.2
31	Mathematics	101	1420.3	262.6	83	1528.6	255.8	80	1668.0	224.0
	Reading	101	1370.9	351.0	83	1568.1	294.7	80	1679.1	252.2
32	Mathematics	134	1426.3	274.4	122	1577.2	275.7	132	1741.7	236.2
	Reading	135	1449.7	392.1	122	1658.6	330.2	132	1758.3	338.0
33	Mathematics	110	1199.8	274.7	95	1374.3	276.2	93	1502.3	215.5
	Reading	110	1144.4	409.2	95	1410.7	358.8	93	1521.7	240.0
34	Mathematics	129	1223.9	312.6	117	1408.8	235.2	121	1534.6	261.9
	Reading	128	1206.2	368.9	115	1430.0	330.9	121	1502.7	334.6
35	Mathematics	97	1261.4	272.1	85	1434.9	258.5	98	1574.9	241.3
	Reading	103	1230.9	316.6	84	1461.1	352.4	98	1578.3	327.4
36	Mathematics	141	1335.3	341.7	119	1534.8	207.1	133	1662.6	237.0
	Reading	142	1309.4	427.5	121	1596.2	267.6	136	1670.5	266.5

Table 5 (Continued)

Descriptive Statistics in Mathematics and Reading for Grades 3, 4, and 5

School	Subject	Grade 3			Grade 4			Grade 5		
		n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Elementary	Mathematics	4748	1324.0	282.3	4475	1471.0	238.7	4749	1619.0	213.3
Schools	Reading	4760	1301.0	379.6	4484	1522.0	305.2	4731	1633.0	292.4
37	Mathematics	157	1400.7	239.0	151	1549.0	187.9	160	1679.9	167.1
	Reading	157	1424.3	359.0	151	1639.8	229.8	160	1688.7	238.0
38	Mathematics	242	1413.5	255.3	177	1543.1	243.6	244	1649.0	229.6
	Reading	242	1454.1	340.1	176	1589.0	321.4	233	1709.2	295.4
39	Mathematics		*NA*		131	1512.5	212.5	135	1677.5	169.0
	Reading				130	1631.3	249.6	134	1754.4	220.0
40	Mathematics		*NA*		129	1542.9	211.9	153	1703.1	213.3
	Reading				130	1549.1	329.1	152	1729.3	276.0

Note. n is the number of students per subjects and grade, *M* is the average score per grade and subject, *SD* is the deviation of the score from the mean score, *NA* indicates the school was in the process of construction and no information was available.

Table 6

Pearson Product Moment Correlations for Mathematics and Reading Scores in Elementary School

	Mathematics 3rd Grade	Mathematics 4th Grade	Mathematics 5th Grade	Reading 3rd Grade	Reading 4th Grade	Reading 5th Grade
Mathematics 3rd Grade	1.0 4,765					
Mathematics 4th Grade	.78 3,821	1.0 4,476				
Mathematics 5th Grade	.78 3,506	.82 4,051	1.0 4,739			
Reading 3rd Grade	.68 4,762	.62 3,827	.62 3,609	1.0 4,777		
Reading 4th Grade	.62 3,825	.67 4,459	.65 4,059	.76 3,830	1.0 4,483	
Reading 5th Grade	.62 3,604	.65 4,046	.68 4,700	.74 3,608	.77 4,055	1.0 4,721

Note. The values with decimals on the top row are the Pearson product moment correlations (r) and all were statistically significant ($p<.0001$); the second row indicates the number of students in the correlations (n).

Table 7
Descriptive Statistics in Mathematics and Reading for Grades 6, 7, and 8

School	Subject	Grade 6			Grade 7			Grade 8		
		n	*M*	*SD*	n	*M*	*SD*	n	*M*	*SD*
Middle	Mathematics	4874	1668.0	254.0	4890	1805.0	211.6	4936	1911.0	160.3
Schools	Reading	4875	1708.0	306.2	4895	1820.0	262.2	4928	1894.0	216.0
41	Mathematics	465	1731.4	204.3	472	1845.6	173.9	482	1942.4	130.1
	Reading	467	1778.7	278.3	471	1887.9	222.5	483	1946.2	184.3
42	Mathematics	288	1584.1	281.3	272	1756.8	205.6	260	1870.6	169.5
	Reading	288	1617.4	335.4	273	1754.1	261.7	262	1825.6	233.2
43	Mathematics	236	1625.7	265.6	229	1771.3	211.6	227	1859.3	186.9
	Reading	238	1634.2	305.2	229	1756.2	243.0	226	1834.4	228.5
44	Mathematics	187	1629.8	276.1	210	1821.4	195.7	203	1923.3	149.2
	Reading	187	1706.6	318.3	210	1852.5	240.3	203	1916.9	197.6
45	Mathematics	556	1736.2	228.7	556	1861.8	200.7	569	1953.5	143.2
	Reading	556	1795.3	289.6	556	1878.1	275.2	568	1953.7	194.8
46	Mathematics	309	1598.1	241.9	305	1750.6	215.2	309	1874.7	171.1
	Reading	308	1615.7	295.5	304	1753.2	248.7	305	1821.1	213.4
47	Mathematics	417	1734.7	246.4	426	1838.4	212.5	437	1938.2	150.5
	Reading	416	1789.6	326.1	426	1864.8	253.3	437	1937.0	215.7

Table 7 (Continued)

Descriptive Statistics in Mathematics and Reading for Grades 6, 7, and 8

School	Subject	Grade 6			Grade 7			Grade 8		
		n	*M*	*SD*	n	*M*	*SD*	n	*M*	*SD*
Middle	Mathematics	4874	1668.0	254.0	4890	1805.0	211.6	4936	1911.0	160.3
Schools	Reading	4875	1708.0	306.2	4895	1820.0	262.2	4928	1894.0	216.0
48	Mathematics	346	1617.9	264.2	340	1732.8	232.4	344	1878.9	150.5
	Reading	346	1646.4	307.5	340	1763.6	268.9	345	1852.7	196.0
49	Mathematics	169	1619.2	263.6	202	1733.8	221.3	199	1871.5	137.0
	Reading	171	1652.8	328.5	201	1757.5	267.5	197	1869.4	223.7
50	Mathematics	289	1660.6	254.8	298	1773.9	222.9	295	1886.3	163.3
	Reading	287	1678.4	308.5	297	1776.5	275.1	293	1850.0	224.7
51	Mathematics	307	1632.0	243.3	285	1771.4	215.0	287	1875.9	180.5
	Reading	306	1683.4	271.9	289	1778.0	258.6	285	1857.8	213.1
52	Mathematics	330	1652.9	260.7	339	1817.6	189.4	340	1907.7	161.9
	Reading	329	1694.8	298.7	339	1825.8	286.1	340	1898.4	236.8
53	Mathematics	380	1665.0	244.6	365	1825.0	207.8	373	1923.8	153.2
	Reading	381	1719.0	271.9	367	1831.0	251.0	371	1886.7	203.7
54	Mathematics	293	1738.9	237.3	309	1880.5	202.3	327	1974.2	151.7
	Reading	294	1771.6	294.7	312	1888.2	259.4	327	1956.6	210.9

Table 7 (Continued)

Descriptive Statistics in Mathematics and Reading for Grades 6, 7, and 8

School	Subject	Grade 6			Grade 7			Grade 8		
		n	*M*	*SD*	n	*M*	*SD*	n	*M*	*SD*
Middle	Mathematics	4874	1668.0	254.0	4890	1805.0	211.6	4936	1911.0	160.3
Schools	Reading	4875	1708.0	306.2	4895	1820.0	262.2	4928	1894.0	216.0
55	Mathematics	302	1633.0	262.7	282	1768.2	208.2	284	1875.6	170.8
	Reading	301	1664.3	299.1	283	1792.0	235.9	286	1868.6	216.9

Note. *n* is the number of students per subjects and grade, *M* is the average score per grade and subject, and *SD* is the deviation of the score from the mean score.

Table 8

Pearson Product Moment Correlations for Mathematics and Reading Scores in Middle School

	Mathematics 6th Grade	Mathematics 7th Grade	Mathematics 8th Grade	Reading 6th Grade	Reading 7th Grade	Reading 8th Grade
Mathematics 6th Grade	1.0 4,840					
Mathematics 7th Grade	.82 4,283	1.0 4,862				
Mathematics 8th Grade	.81 4,033	.84 4,367	1.0 4,898			
Reading 6th Grade	.72 4,829	.69 4,283	.68 4,034	1.0 4,843		
Reading 7th Grade	.68 4,285	.70 4,851	.67 4,373	.77 4,286	1.0 4,867	
Reading 8th Grade	.68 4,035	.69 4,364	.72 4,868	.77 4,037	1.0 4,370	4,889

Note. The values with decimals on the top row are the Pearson product moment correlations (*r*) and all were statistically significant ($p<.0001$), the second row indicates the number of students in the correlations (*n*).

display the number of students missing one or more test scores in mathematics and reading.

This concludes the Preliminary Results section which provided the descriptive statistics for mathematics and reading outcome measures by grade and school along with bivariate correlations for mathematics and reading in the elementary and middle school grades. The Primary Results section, which follows, will present a brief description of the models and the inferential results from the Gain Score and Layered Effects Models for the FCAT mathematics and reading data.

Primary Results

The Primary Results section details all of the results estimated from the outcome measures and model specifications. These details include the model parameter estimates with their standard errors, 95% confidence intervals (CI) for the fixed effects, variance components, and fit indices. When applicable, these also include the Intraclass Correlation Coefficient (ICC) and R-matrix (i.e., sigma matrix) results and the random school effects with standard errors, followed by their correlations. The section then concludes with the school rankings, followed by their correlations.

The first of the two models used to estimate random school effects for this study was the Gain Score Model. The two outcome measures utilized with this model were the FCAT mathematics and reading DSSs. The Gain Score Model estimates random school effects as simple average gain scores between two adjacent years (e.g., Grade 3 and Grade 4). As found in Equations 1.1 through 1.7, the model uses each student's DSS to calculate a difference score between the two adjacent grades. The adjacent grades were Grades 3

and 4, 4 and 5, 6 and 7, and 7 and 8. Each difference score was then aggregated to generate an average gain score for each school. Each school's average gain score was then aggregated to form a grand average gain score for the school level (i.e., elementary and middle). For example, the grand average gain score from Grade 3 to Grade 4 would take the average gain from each elementary school at those grades, add them together, and divide by the total number of elementary schools. The differences between the grand average gain score and the individual school's average gain scores were the random school effects, which are found in Tables 11 and 12. The random school effects indicate the deviations of each school's average gain score from the grand average gain score between two adjacent years (i.e., Grade 3 and Grade 4).

Primary results for Gain Score Models in mathematics. The Gain Score Model provides the perspective of the average gain from year to year. Estimates from the Gain Score Models found no violations to model assumptions in mathematics; see Appendix E for further details. Table 9 lists the parameter estimates for the mathematics scores, along with the 95% CI for the Gain Score Model. In the table, the fixed effects indicate the average gain that occurred from Grades 3 to 4 and Grades 4 to 5 for the elementary schools examined. The Standard Error (SE) is an estimate of the degree to which the sample represents the population and is used in the calculation of the 95% CI. The 95% CI is an estimate of the range where the population mean is likely to fall within the upper and lower limits. The two variance parameter estimates are the intercept and level-1 variance, which indicate the amount of variation between and within units, respectively. The intercept variance is the variance between schools and the level-1 variance is the variance between students within the schools. Between Grades 3 and 4, the average gain score in

mathematics was estimated to be 112.1 points with a 95% CI ranging from 102.1 to 122.0. Between Grades 4 and 5, the average gain score was 146.9 and the 95% CI ranged from 134.3 to 153.8.

Table 10 lists the average gain score estimate for middle schools between Grades 6 and 7 and between Grades 7 and 8. Between Grades 6 and 7, the average gain score was estimated to be 136 points with a 95% CI ranging from 126.4 to 145.6. From Grade 7 to Grade 8, the average gain score was 104.9 with the 95% CI ranging from 97.1 to 112.7.

Table 9
Elementary Fixed Effects, Variance Components, Standard Error, and 95% CI in Mathematics Using the Gain Score Model

	Gain Score Model (Mathematics Grades 3 to 4)		Gain Score Model (Mathematics Grades 4 to 5)	
Parameter Estimate	Parameter (SE) Estimate	95% CI	Parameter (SE) Estimate	95% CI
Fixed Effects				
Intercept (γ_{00})	112.1 (5.1)	102.1 to 122.0	146.9 (3.5)	134.3 to 153.8
Variance Estimates				
Intercept Variance (τ_{00})	945.3 (236.5)	481.8 to 1408.8	411.6 (108.5)	198.9 to 624.3
Level-1 Variance (σ^2)	26760.0 (354.9)	26064.4 to 27455.6	18220.0 (235.0)	17759.4 to 18681.0
Fit Indices	AIC 148833.1	BIC 148836.5	AIC 152691.9	BIC 152695.3

Note. γ_{00} is the average gain score between all elementary schools for the specified time-period; τ_{00} is the variance between elementary schools; σ^2 is the variation of students' scores within schools.

Table 10
Middle Fixed Effects, Variance Components, Standard Error, and 95% CI in Mathematics Using the Gain Score Model

	Gain Score Model (Mathematics Grades 6 to 7)		Gain Score Model (Mathematics Grades 7 to 8)	
Parameter	Parameter (SE) Estimate	95% CI	Parameter (SE) Estimate	95% CI
Fixed Effects				
Intercept (γ_{00})	136.0 (4.9)	126.4 to 145.6	104.9 (4.0)	97.1 to 112.7
Variance Estimates				
Intercept Variance (τ_{00})	335.1 (135.7)	69.1 to 601.1	225.8 (91.2)	47.1 to 404.6

Table 10 (Continued)
Middle Fixed Effects, Variance Components, Standard Error, and 95% CI in Mathematics Using the Gain Score Model

	Gain Score Model (Mathematics Grades 6 to 7)		Gain Score Model (Mathematics Grades 7 to 8)	
Parameter	Parameter (SE) Estimate	95% CI	Parameter (SE) Estimate	95% CI
Level-1 Variance (σ^2)	19832.0 (246.5)	19348.9 to 20315.1	12533.0 (154.2)	12230.8 to 12835.0
Fit Indices	AIC 165057.1	BIC 165058.5	AIC 162462.1	BIC 162463.6

Note. γ_{00} is the average gain score between all middle schools for the specified time-period; τ_{00} is the variance between middle schools; σ^2 is the variation of students' scores within schools.

Intraclass Correlation Coefficient. Before additional analyses were conducted and inferences made, the ICC was calculated to determine the amount of clustering within level-2 units (i.e., schools). The ICC was calculated to measure and determine the proportion of variance in the outcome variables (i.e., mathematics and reading) existing between schools. These estimates were based on gain scores between two adjacent grades. The ICCs in the elementary grades were .034 in the gain from Grade 3 to Grade 4 and .022 with the gain between Grade 4 and Grade 5 in mathematics indicating that 3.4% and 2.2% of the variance in mathematics was between schools at the elementary grades. The ICCs in the middle school grades were .017 with the gain between Grades 6 and 7 and .018 in the gain from Grades 7 to 8. This indicated that 1.7 % and 1.8 % of the variance in the mathematics gain scores was between schools at the middle school grades. Although the ICCs were low for the "between" schools, the parameters were still estimated using multilevel modeling procedures to address the research questions.

The presentation of the results from the fixed effects, variance components, 95% CIs, and ICC provided a picture of the data and model performance overall. The fixed effects provided the overall average gain score between two adjacent grades for all schools

examined. Next, the random school effects were examined, which was the primary focus of this study. The random school effects are the deviations of each school's average gain score from the grand average or fixed effects. The estimates in the following tables for the Gain Score Model are represented as random school effects.

Random school effects in mathematics. Table 11 displays the elementary school random effect estimates for mathematics using the Gain Score Model. The information in the table on the "Elementary Schools" line indicates the grand average gain or fixed effect estimate between two adjacent grades (e.g., Grade 3 to Grade 4) along with the standard error in parentheses. Below the fixed effect estimates are the school numbers, the numbers of students in each school used to estimate the parameters, and the standard error in parentheses. Each estimate indicates the random school effect of each school, which are deviations from the grand average gain DSS from one grade to the next throughout the district. Positive values indicate the degree to which the estimates were above the average gain in the district from grade to grade and negative values indicate the degree to which the estimates were below the average gain in the district from grade to grade. The estimates can vary significantly when changing from grade to grade and gains may not be consistent across subjects. A few of the school's estimates change from positive to negative, negative to positive, or remain fairly similar from grade to grade.

The number of students used in the estimates ranged from 49 to 135 in Grade 3 to Grade 4 and 45 to 167 in Grade 4 to Grade 5. The school effects between Grades 3 and 4 ranged from -59 to 80 (deviations from the grand average) gain-points and -40 to 41 (deviations from the grand average) gain-points between Grades 4 and 5. These estimates

were the number of gain-points a school had “above” or “below” the grand average or fixed effect. For example, the students transitioning from Grade 3 to Grade 4 in school # 7 had an average gain score in mathematics that was 18 gain-points above the grand average gain of 112. Within the same school between Grades 4 and 5, the average gain score was 28 points above the grand average gain score of 147.

Table 11

Elementary School Random Effects in Mathematics Using the Gain Score Model

	Mathematics			
	n	3rd to 4th	*n*	4th to 5th
Elementary Schools	3,804	112.5 (5.1)	4,047	146.9 (3.5)
1	61	20.2 (12.0)	63	-11.2 (9.5)
2	91	-27.8 (9.5)	104	4.8 (7.1)
3	118	-41.0 (9.6)	132	30.2 (7.1)
4	99	29.2 (10.1)	108	-18.9 (7.6)
5	99	80.2 (10.3)	87	41.4 (8.0)
6	109	73.8 (9.9)	106	6.8 (7.7)
7	105	17.8 (9.8)	100	28.2 (7.6)
8	105	11.6 (10.1)	99	32.8 (7.8)
9	57	-48.2 (14.1)	80	-24.1 (9.9)
10	72	-7.6 (13.1)	73	-20.3 (10.2)
11	70	-29.7 (13.5)	73	20.4 (10.1)
12	133	-31.7 (9.3)	133	5.4 (7.1)
13	126	-2.5 (9.4)	132	-3.5 (7.2)
14	66	-33.8 (10.8)	81	-22.0 (7.9)
15	121	9.3 (9.4)	135	-25.7 (7.1)
16	107	19.2 (10.0)	111	-10.6 (7.6)
17	70	18.1 (11.6)	61	13.5 (9.0)
18	111	-52.4 (9.9)	116	4.5 (7.5)
19	93	2.8 (10.5)	99	-7.1 (7.9)
20	49	25.2 (13.1)	45	29.0 (10.2)
21	80	5.3 (10.8)	79	23.4 (8.3)
22	92	-58.8 (10.1)	94	-7.1 (7.8)

Table 11 (Continued)

Elementary School Random Effects in Mathematics Using the Gain Score Model

	Mathematics			
	n	3rd to 4th	*n*	4th to 5th
Elementary Schools	3,804	112.5 (5.1)	4,047	146.9 (3.5)
23	113	-6.4 (9.8)	118	-16.3 (7.4)
24	93	7.6 (10.5)	94	-40.1 (8.0)
25	105	10.8 (9.8)	130	-8.1 (7.2)
26	98	-21.6 (9.5)	103	21.4 (7.3)
27	86	21.6 (10.6)	84	-6.8 (8.3)
28	91	-0.6 (10.4)	96	-1.4 (8.0)
29	101	-25.7 (10.3)	96	-4.0 (8.1)
30	93	-11.7 (10.3)	105	10.5 (7.9)
31	73	-16.4 (11.3)	66	-21.8 (8.8)
32	107	-8.7 (9.8)	121	12.2 (7.4)
33	79	20.9(11.0)	82	-10.6 (8.4)
34	100	7.8 (10.2)	104	7.9 (7.7)
35	75	-10.8 (11.1)	81	5.0 (8.6)
36	103	17.6 (9.9)	113	-5.0 (7.5)
37	123	10.9 (9.5)	137	-7.6 (7.0)
38	135	29.7 (8.8)	167	-30.8 (6.4)
39	108	-2.1 (11.4)	120	4.1 (8.5)
40	87	-1.9 (12.3)	119	2.3 (8.6)

Note. The first column indicates each school number, the *n* in the second and fourth columns indicate the total number of students involved in the computation of the gain score in each school (e.g., 3rd to 4th and 4th to 5th), and the third and fifth columns indicate the random school effect estimates with the standard error in parentheses.

Table 12 displays the random school effect estimates in the middle school grades. The number of students used in the school estimates ranged from 167 to 430 between Grade 6 and 7, and 168 to 430 from Grade 7 to Grade 8 in mathematics. The school effects between Grades 6 and 7 ranged from -28 to 31 gain-points, and -18 to 34 gain-points between Grades 7 and 8. These estimates indicate the number of gain-points a school had “above,” “below,” or “equal to” the grand average gain score or fixed effect.

Table 12

Middle School Random Effects in Mathematics Using the Gain Score Model

	Mathematics			
	n	6th to 7th	*n*	7th to 8th
Middle Schools	4,320	135.9 (4.9)	4,411	104.9 (4.0)
41	429	-27.1 (6.1)	444	-10.3 (4.9)
42	242	22.2 (6.8)	230	-0.3 (5.5)
43	211	6.3 (7.0)	208	-15.9 (5.7)
44	172	30.6 (7.3)	193	-5.3 (5.8)
45	490	-7.3 (5.9)	505	-11.2 (4.8)
46	267	17.0 (6.6)	259	16.4 (5.4)
47	372	-28.3 (6.2)	393	-4.6 (5.0)
48	312	-19.7 (6.4)	304	34.0 (5.2)
49	167	-4.1 (7.4)	181	22.1 (5.9)
50	252	-13.4 (6.7)	263	7.6 (5.4)
51	259	-4.9 (6.7)	252	-0.6 (5.4)
52	298	13.0 (6.5)	300	-8.1 (5.2)
53	337	14.2 (6.3)	341	-8.1 (5.1)
54	257	4.8 (6.7)	279	-17.8 (5.3)
55	255	-3.1 (6.7)	259	2.1 (5.4)

Note. The first column indicates each school number, the *n* in the second and fourth columns indicate the total number of students involved in the computation of the gain score in each school (e.g., 6th to 7th and 7th to 8th), and the third and fifth columns indicate the random school effect estimates with the standard error in parentheses.

The Gain Score Model mathematics results section began with the fixed effects and variance components including the 95% CI. The ICC examined the degree of clustering within schools. After the ICC section, the random school effects were listed indicating the estimated effect each school had on their students in specified grades.

Primary results for Gain Score Models in reading. Estimates from the Gain Score Models found no violations to model assumptions in reading, see Appendix E for further details. Table 13 provides the fixed effects, variance estimates, and 95% CI for reading in the elementary school grades. Between Grades 3 and 4, the average gain score

for all schools was estimated to be 166.6 points with a 95% CI ranging from 156.6 to 176.6. Between Grades 4 and 5, the average gain score was 109.2, and the CI ranged from 101.0 to 117.4.

Table 13
Elementary Fixed Effects and Variance Components, Standard Error, and 95% CI in Reading Using the Gain Score Model

	Gain Score Model (Reading Grades 3 to 4)		Gain Score Model (Reading Grades 4 to 5)	
Parameter	Parameter (SE) Estimate	95% CI	Parameter (SE) Estimate	95% CI
Fixed Effects				
Intercept (γ_{00})	166.6 (5.1)	156.6 to 176.6	109.2 (4.2)	101.0 to 117.4
Variance Estimates				
Intercept Variance (τ_{00})	872.4 (242.1)	397.9 to 1346.9	545.9 (158.2)	235.8 to 856.0
Level-1 Variance (σ^2)	48880.0 (647.4)	47611.6 to 50148.9	39873.0 (513.9)	38865.8 to 40880.2
Fit Indices	AIC 156080.9	BIC 156084.2	AIC 162326.5	BIC 162329.9

Note. γ_{00} is the average gain score between elementary schools for the specified time-period; τ_{00} is the variance between elementary schools; σ^2 is the variation of students' scores within schools.

Table 14 provides the fixed effects, variance components, and 95% CI for reading in the middle school grades. Between Grades 6 and 7, the average gain score for middle schools was estimated to be 113.1 points with a 95% CI ranging from 104.9 to 121.3. From Grade 7 to Grade 8, the average gain score for middle schools was estimated to be 70.5 points with a 95% CI ranging from 64.6 to 76.4.

Table 14
Middle Fixed Effects, Variance Components, Standard Error, and 95% CI in Reading Using the Gain Score Model

	Gain Score Model (Reading Grades 6 to 7)		Gain Score Model (Reading Grades 7 to 8)	
Parameter	Parameter (SE) Estimate	95% CI	Parameter (SE) Estimate	95% CI
Fixed Effects				
Intercept (γ_{00})	113.1 (4.2)	104.9 to 121.3	70.5 (3.0)	64.6 to 76.4

Table 14 (Continued)
Middle Fixed Effects, Variance Components, Standard Error, and 95% CI in Reading Using the Gain Score Model

	Gain Score Model (Reading Grades 6 to 7)		Gain Score Model (Reading Grades 7 to 8)	
Parameter	Parameter (SE) Estimate	95% CI	Parameter (SE) Estimate	95% CI
Variance Estimates				
Intercept Variance (τ_{00})	223.3(99.4)	28.5 to 418.1	104.7 (52.6)	1.6 to 207.8
Level-1 Variance (σ^2)	37494.0 (465.9)	35676.0 to 38407.1	26909.0 (330.9)	26260.4 to 27557.6
Fit Indices	AIC 173377.9	BIC 173379.4	AIC 172671.4	BIC 172672.8

Note. γ_{00} is the average gain score between middle schools for the specified time-period; τ_{00} is the variance between middle schools; σ^2 is the variation of students' scores within schools.

Intraclass Correlation Coefficient. In the elementary grades, the ICCs were .017 in the gain from Grade 3 to Grade 4 and .006 between Grade 4 and Grade 5 in reading indicating that 1.7% and 0.6%, respectively, of the variance in reading was between schools at the elementary grades. In the middle school grades, the ICC was .006 in the gain between Grade 6 and Grade 7, and .004 between Grade 7 and Grade 8, which indicated 0.4% and 0.6% of the variance in the reading gain scores was between schools at the middle grades. Although the ICC was low, the parameters were still estimated using multilevel modeling procedures to address the research questions. The model assumptions and the ICCs were examined, then, the random school effect estimates were generated from the models. These results are presented in Table 15.

Random school effects in reading. Table 15 displays the elementary school random effect estimates for reading using the Gain Score Model. The number of students used in the reading estimates ranged from 49 to 135 in Grades 3 to Grade 4, and 46 to 163 in Grades 4 to Grade 5. The school effects in reading between Grades 3 and 4 ranged from -74 to 80 gain-points (deviations from the grand average) and -60 to 44 gain-points (deviations from the grand average) between Grade 4 and 5. These estimates were the

number of gain-points a school had above or below the grand average or fixed effect. When added to the average effect shown in the first row (i.e., 166.6), the school effects provide estimates of the gains of the different schools.

Table 15

Elementary School Random Effects in Reading Using the Gain Score Model

	Reading			
	n	3rd to 4th	*n*	4th to 5th
Elementary Schools	3,814	166.6 (5.1)	4,052	109.2 (4.2)
1	61	80.4 (14.8)	63	-60.2 (12.9)
2	91	-25.0 (11.6)	104	10.7 (9.8)
3	120	-17.3 (11.7)	131	23.3 (9.8)
4	102	25.5 (12.4)	111	8.0 (10.5)
5	98	2.7 (12.7)	88	-6.4 (11.1)
6	109	21.0 (12.1)	106	-6.6 (10.6)
7	106	-15.0 (12.0)	102	43.6 (10.5)
8	105	-22.7 (12.4)	98	11.4 (10.8)
9	57	5.3 (17.1)	80	-4.6 (13.5)
10	72	-27.9 (16.0)	73	5.6 (13.9)
11	70	-18.1 (16.4)	73	0.9 (13.7)
12	134	-3.0 (11.4)	135	-25.5 (9.8)
13	126	2.0 (11.5)	132	0.4 (10.0)
14	65	-25.4 (13.3)	81	-0.5 (10.9)
15	121	-4.0 (11.5)	135	4.7 (9.7)
16	107	12.3 (12.3)	109	17.5 (10.6)
17	71	22.6 (14.2)	62	-37.3 (12.4)
18	112	-8.6 (12.1)	118	4.5 (10.2)
19	95	-18.7 (12.8)	99	2.2 (10.9)
20	49	8.2 (16.0)	46	4.5 (13.7)
21	78	15.3 (13.3)	78	-3.0 (11.5)
22	93	-73.8 (12.3)	97	16.7 (10.8)
23	113	-3.0 (12.0)	118	-4.5 (10.3)
24	93	42.3 (12.8)	93	1.4 (11.1)
25	105	0.2 (11.9)	128	5.3 (10.1)
26	98	50.3 (11.7)	103	-2.9 (10.1)
27	86	28.3 (13.1)	85	14.5 (11.4)
28	90	-14.0 (12.8)	96	11.0 (11.0)

Table 15 (Continued)

Elementary School Random Effects in Reading Using the Gain Score Model

	Reading			
	n	3rd to 4th	*n*	4th to 5th
Elementary Schools	3,814	166.6 (5.1)	4,052	109.2 (4.2)
29	101	-3.4 (12.7)	96	-53.4 (11.2)
30	93	22.0 (12.7)	105	24.6 (10.9)
31	73	-22.7 (13.9)	66	-6.0 (12.1)
32	107	-21.6 (12.1)	121	-9.7 (10.2)
33	79	34.9 (13.5)	82	11.8 (11.5)
34	99	-21.5 (12.6)	101	-3.4 (10.7)
35	76	-17.6 (13.5)	81	15.3 (11.8)
36	105	2.9 (12.1)	116	-7.9 (10.3)
37	123	7.3(11.6)	137	-42.8 (9.6)
38	135	-0.1 (10.6)	163	0.3 (8.9)
39	108	15.0 (14.0)	119	-1.7 (11.8)
40	88	-35.2 (15.0)	121	38.1 (11.7)

Note. The first column indicates each school number; *n* indicates the total number of students involved in the computation of the gain score in each elementary school, 3rd to 4th and 4th to 5th are the random school effect estimates with the standard error in parentheses.

Table 16 displays the middle school random effect estimates. The number of students used in the school estimates ranged from 168 to 490 in the gain between Grades 6 and 7, as well as Grades 7 and Grade 8 for reading. The school effects between Grades 6 and 7 ranged from -25 to 26 gain points, and -15 to 15 gain-points between Grades 7 and 8.

Table 16

Middle School Random Effects in Reading Using the Gain Score Model

	Reading			
	n	6th to 7th	*n*	7th to 8th
Middle Schools	4,287	113.1 (4.2)	4,370	70.5 (3.0)
41	430	-10.8 (6.3)	430	-11.3 (4.8)

Table 16 (Continued)

Middle School Random Effects in Reading Using the Gain Score Model

		Reading		
	n	6th to 7th	*n*	7th to 8th
Middle Schools	4,287	113.1 (4.2)	4,370	70.5 (3.0)
42	243	4.2 (7.3)	243	-4.6 (5.8)
43	212	-1.3 (7.6)	212	9.0 (5.9)
44	172	5.3 (8.1)	172	-7.2 (6.0)
45	490	-20.7 (6.1)	490	0.0 (4.7)
46	265	26.2 (7.2)	265	-4.3 (5.6)
47	371	-24.5 (6.5)	371	-1.0 (5.0)
48	312	12.7 (6.8)	312	8.3 (5.3)
49	168	5.1 (8.1)	168	15.0 (6.2)
50	249	4.2 (7.3)	249	0.0 (5.6)
51	261	-16.2 (7.2)	261	8.9 (5.6)
52	297	7.6 (6.9)	297	3.9 (5.4)
53	338	-4.4 (6.7)	338	-15.2 (5.2)
54	260	-0.5 (7.2)	260	-10.8 (5.5)
55	254	13.0 (7.2)	254	9.2 (5.6)

Note. The first column indicates each school number; *n* indicates the total number of students involved in the computation of the gain score in each school; 6th to 7th and 7th to 8th are the random school effect estimates with the standard error in parentheses.

The Primary results for Gain Score Models in the reading section began with the fixed effects and variance components including the 95% CI. The degree of clustering between schools was then examined using the ICC. Finally, the random school effects, from the Gain Score Model were listed, indicating the deviation of each schools' average gain in DSS from the grand average gain between specified grades.

Primary results for Layered Effects Models in mathematics. The Layered Effects Model found in Equations 1.8 through 1.14 estimates the average DSS, which accounts for past performance of each student in a particular school at a particular grade. The grand average score, or fixed effect, is an aggregation of each school's adjusted

average DSS. The adjusted average DSS for each school was compared to the grand average to get the random school effects. These effects were the difference between the grand average score and the individual school adjusted average DSS at a particular grade. The Layered Effects Models estimate random school effects as empirical Bayes estimates.

The Layered Effects Model provides the average score for each year. Estimates from the Layered Effects Model found no violations to model assumptions in mathematics, see Appendix E for further details. Table 17 lists the elementary parameter estimates from the mathematics scores along with the 95% CI from the Layered Effects Model. The fixed effects indicate the average performance of students in Grades 3, 4, and 5 for all elementary schools examined. The standard error (SE) is the precision of the sample estimate in representing the population and is used in the calculation of the 95% CI. The intercept variance is the variance between schools and the level-1 variance indicates the variance between students' scores within schools (see Table 19).

Table 17 lists the fixed effect or average scores, variance components, and 95% CI in mathematics for the elementary schools by grade. The average score estimates were 1315.1 in Grade 3, 1425.3 in Grade 4, and 1591.9 in Grade 5. The intercept variance was 1686.9 at Grade 3, 1084.1 at Grade 4, and 394.0 at Grade 5 which indicated the variance between schools at the specified grades.

Table 17

Elementary Fixed Effects, Variance Components, Standard Error, and 95% CI in Mathematics Using the Layered Effects Model

	Layered Effects Model (Mathematics)	
Parameter	Parameter (SE) Estimate	95% CI
Fixed Effects (Intercept)		
Grade 3 (γ_{00})	1315.1 (7.3)	1300.9 to 1329.3

Table 17 (Continued)

Elementary Fixed Effects, Variance Components, Standard Error, and 95% CI in Mathematics Using the Layered Effects Model

	Layered Effects Model (Mathematics)	
Parameter	Parameter (SE) Estimate	95% CI
Fixed Effects (Intercept)		
Grade 4 (γ_{10})	1425.3 (8.3)	1430.7 to 1457.9
Grade 5 (γ_{20})	1591.9 (8.7)	1574.8 to 1609.0
Variance Estimates (Intercept Variance)		
Grade 3 (τ_{00})	1686.9 (465.7)	774.1 to 2599.7
Grade 4 (τ_{10})	1084.1 (301.0)	494.1 to 1674.1
Grade 5 (τ_{20})	394.0 (129.2)	140.8 to 647.2
Fit Indices	AIC	BIC
	364934.5	364880.5

Note. γ_{00}, γ_{10}, and γ_{20} are the average scores for elementary schools during the specified time-period; τ_{00}, τ_{10}, and τ_{20} are the variance between elementary schools.

Table 18 lists the fixed effects or average scores, variance components, and 95% CI in mathematics for middle schools by grade. The average score estimate was 1631.0 in Grade 6, 1773.2 in Grade 7, and 1885.4 in Grade 8. The intercept variance was 1331.2 at Grade 6, 293.6 at Grade 7, and 189.4 at Grade 8.

Table 18

Middle Fixed Effects, Variance Components, Standard Error, and 95% CI in Mathematics Using the Layered Effects Model

	Layered Effects Model (Mathematics)	
Parameter	Parameter (SE) Estimate	95% CI
Fixed Effects (Intercept)		
Grade 6 (γ_{30})	1631.0 (10.0)	1611.4 to 1650.6
Grade 7 (γ_{40})	1773.2 (10.8)	1752.0 to 1794.4
Grade 8 (γ_{50})	1885.4 (11.2)	1863.5 to 1907.4
Variance Estimates (Intercept Variance)		
Grade 6 (τ_{30})	1331.2 (533.8)	285.0 to 2377.5
Grade 7 (τ_{40})	293.6 (134.8)	29.4 to 557.8
Grade 8 (τ_{50})	189.4 (86.7)	19.5 to 359.3
Fit Indices	AIC	BIC
	364934.5	364880.5

Note. γ_{30}, γ_{40}, and γ_{50} are the average score for middle schools during the specified time-period; τ_{30}, τ_{40}, and τ_{50} are the variance between middle schools.

The Layered Effects Model does not provide one residual or level-1 variance value as found with the Gain Score Model, but it does provide the 6 x 6 unstructured R-Matrix (i.e., it is 6 x 6 because data from Grades 3 through 8 were used), which has the variance and covariance parameters estimated by maximum likelihood (see Table 19 for an example of one student's R-matrix). The variance estimates are along the diagonal in the table, which indicates the variance of students' scores within schools at specified grades; the information is provided for the elementary and middle school grades. The variances range from 26510 to 75744 within the elementary and middle school grades. The off-diagonal estimates are the covariance estimates between two different grades.

Table 19

Estimated R Matrix for ID 220436 in Mathematics

Grade	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
			Level-1 (σ^2)			
Grade 3	75774					
Grade 4	51920	58162				
Grade 5	46388	43244	47118			
Grade 6	52569	49058	46432	66491		
Grade 7	43070	40139	37600	45743	45606	
Grade 8	33098	30522	28826	34244	29409	26510

Note. σ^2 is the variation of students' scores within schools; they are found in the diagonal of the table. The covariance estimates are found in the off-diagonal of the table.

Random school effects in mathematics. Table 20 displays the elementary school random effect estimates in mathematics using the Layered Effects Model. The estimates indicate the number of points "above," "below," or "equal" to the average score for a particular subject and grade throughout the district.

Table 20

Elementary School Random Effects in Mathematics Using the Layered Effects Model

	Mathematics					
	n	3rd	*n*	4th	*n*	5th
Elementary Schools	4,748	1315.1 (7.3)	4,474	1441.6 (8.3)	4,749	1591.9 (8.8)
1	81	-3.2 (18.3)	65	9.0 (17.5)	65	-5.8 (13.0)
2	209	30.5 (13.3)	110	-22.7 (14.0)	116	-1.3 (11.1)
3	146	52.7 (14.9)	145	-33.2 (14.1)	150	21.8 (10.3)
4	130	-41.3 (15.5)	120	15.5 (14.9)	131	-14.6 (11.0)
5	129	-58.3 (15.8)	110	76.6 (15.1)	99	33.9 (11.5)
6	146	-59.7 (14.9)	121	61.0 (14.5)	120	10.4 (11.0)
7	166	-42.4 (14.2)	122	10.4 (14.4)	116	29.4 (11.2)
8	141	-53.0 (15.4)	126	19.8 (14.8)	121	25.4 (11.1)
9		NA	83	27.9 (14.5)	111	-12.8 (12.3)
10		NA	82	-50.0 (14.0)	87	-23.5 (12.4)
11		NA	82	-14.1 (14.4)	89	20.7 (12.3)
12	159	12.9 (14.7)	149	-36.4 (13.9)	155	4.2 (10.3)
13	151	28.6 (15.1)	140	-3.8 (14.1)	148	-5.6 (10.5)
14	148	84.4 (14.6)	88	-40.6 (15.4)	116	-21.1 (11.9)
15	164	21.2 (14.2)	141	9.9 (13.9)	149	-19.6 (10.3)
16	113	-36.2 (16.3)	118	9.1 (14.9)	118	-11.2 (11.0)
17	102	-52.8 (16.7)	85	19.1 (16.5)	74	8.2 (12.6)
18	144	28.6 (15.1)	128	-53.5 (14.5)	137	-0.6 (10.7)
19	107	7.3 (16.2)	109	3.4 (15.6)	111	-0.7 (11.3)
20	76	-6.4 (19.0)	54	30.0 (18.2)	51	18.6 (13.8)
21	127	-29.9 (15.1)	94	10.8 (16.0)	90	27.5 (11.9)
22	161	42.8 (14.4)	104	-54.9 (14.7)	106	-10.6 (11.5)
23	132	1.7 (15.4)	132	-4.5 (14.6)	134	-15.3 (10.7)
24	109	32.4 (16.4)	102	-0.5 (15.4)	108	-34.5 (11.5)
25	137	5.5 (15.1)	125	7.0 (14.5)	149	-7.7 (10.6)
26	198	3.0 (13.4)	108	-21.4 (14.0)	113	15.7 (11.2)
27	116	-49.5 (16.4)	102	13.7 (15.5)	94	-6.1 (11.8)
28	131	-26.9 (15.6)	104	-0.8 (15.1)	116	-6.4 (11.5)
29	99	43.8 (17.0)	110	-30.9 (15.7)	100	-0.0 (11.4)
30	115	-2.8 (15.9)	106	-15.2 (15.4)	126	9.7 (11.3)
31	101	70.1 (16.5)	83	-18.6 (16.2)	80	-12.0 (12.4)
32	134	53.4 (15.4)	122	-3.2 (14.5)	132	14.5 (10.8)
33	110	-51.7 (16.4)	95	10.3 (15.9)	93	-10.9 (11.9)
34	129	-52.4 (15.4)	117	-0.4 (14.9)	121	-3.4 (11.1)
35	97	1.4 (16.6)	85	1.6 (16.4)	98	7.5 (12.2)
36	141	5.6 (15.1)	119	24.0 (14.5)	133	-2.4 (10.9)

Table 20 (Continued)

Elementary School Random Effects in Mathematics Using the Layered Effects Model

	Mathematics					
	n	3rd	*n*	4th	*n*	5th
Elementary Schools	4,748	1315.1 (7.3)	4,474	1441.6 (8.3)	4,749	1591.9 (8.8)
37	157	26.5 (14.9)	151	6.4 (13.9)	160	-4.2 (10.2)
38	242	29.4 (12.7)	177	12.3 (12.8)	244	-32.8 (9.5)
39		NA	131	5.1 (12.7)	135	6.6 (10.7)
40		NA	129	27.9 (13.2)	153	7.2 (10.9)

Note. The first column indicates each school number; *n* indicates the total number of students involved in the computation of the average score in the elementary schools. The 3rd, 4th, and 5th columns indicate the random school effect estimates with the standard error in parentheses.

Table 21 displays the random effects for the middle schools.

Table 21

Middle School Random Effects in Mathematics Using the Layered Effects Model

	Mathematics					
(SID)	n	6th	n	7th	n	8th
Middle Schools	4,874	1631.0 (10.0)	4,890	1773.2 (10.8)	4,936	1885.4 (11.2)
41	465	27.5 (12.0)	472	-16.2 (7.4)	482	-3.2 (5.9)
42	288	-38.9 (12.5)	272	16.8 (8.5)	260	-3.0 (6.8)
43	236	-9.4 (13.4)	229	-1.9 (8.8)	227	-17.8 (7.1)
44	187	-15.4 (13.7)	210	28.9 (9.3)	203	-1.4 (7.2)
45	556	52.9 (11.8)	556	-5.7 (7.1)	569	-7.1 (5.7)
46	309	-17.6 (12.6)	305	7.5 (8.3)	309	10.0 (6.7)
47	417	45.9 (12.0)	426	-24.9 (7.6)	437	-3.3 (6.1)
48	346	-28.3 (12.6)	340	-19.2 (8.0)	344	29.6 (6.4)
49	169	-25.5 (14.0)	202	-8.9 (9.4)	199	14.2 (7.4)
50	289	-16.8 (12.5)	298	-5.4 (8.4)	295	5.9 (6.7)
51	307	-21.9 (12.4)	285	-3.2 (8.3)	287	0.6 (6.7)
52	330	13.7 (12.5)	339	9.7 (8.1)	340	-9.7(6.4)
53	380	4.1 (12.0)	365	14.3 (7.8)	373	-4.1 (6.2)
54	293	63.2 (13.1)	309	7.7 (8.4)	327	-13.1 (6.6)
55	302	-33.7 (12.5)	282	0.5 (8.3)	284	2.4 (6.7)

Note. The first column indicates each school number; *n* indicates the total number of students involved in the computation of the average score in each middle school. The 6th, 7th, and 8th columns indicate the random school effect estimates with the standard error in parentheses.

Primary results for Layered Effects Models in reading. Estimates from the Layered Effects Model found no violations to model assumptions in reading; see Appendix E for further details. The Layered Effects Model found in equations 1.8 through 1.14 estimates the average DSS accounting for past performance of each student in a particular school at a particular grade. The grand average score, or fixed effect, is an aggregation of each school's adjusted average DSS. The adjusted average DSS for each school is compared to the grand average to get the random school effect. These effects are the differences between the grand average score and the individual school adjusted average DSS at a particular grade.

Table 22 lists the elementary school fixed effect or average scores for reading by grade along with the variance components and their 95% CI. The average score estimate was 1295.3 in Grade 3, 1482.5 in Grade 4, and 1594.5 in Grade 5. The intercept variance was 1443.9 in Grade 3, 591.1 in Grade 4, and 264.6 in Grade 5 which was the variance between schools at the specified grades.

Table 22

Elementary Fixed Effects, Variance Components, Standard Error, and 95% CI in Reading Using the Layered Effects Model

	Layered Effects Model (Reading)	
Parameter	Parameter (SE) Estimate	95% CI
Fixed Effects (Intercept)		
Grade 3 (γ_{00})	1295.3 (7.6)	1280.4 to 1310.2
Grade 4 (γ_{10})	1482.5 (7.6)	1467.6 to 1497.4
Grade 5 (γ_{20})	1594.5 (7.9)	1579.4 to 1610.0
Variance Estimates (Intercept Variance)		
Grade 3 (τ_{00})	1443.9 (437.0)	587.4 to 2300.4
Grade 4 (τ_{10})	591.1 (218.4)	428.1 to 1019.2
Grade 5 (τ_{20})	264.6 (139.1)	0 to 537.2

Table 22 (Continued)

Elementary Fixed Effects, Variance Components, Standard Error, and 95% CI in Reading Using the Layered Effects Model

	Layered Effects Model (Reading)	
Fit Indices	AIC	BIC
	383947.7	383893.7

Note. γ_{00}, γ_{10}, and γ_{20} are the average score for all elementary schools during the specified time-period; τ_{00}, τ_{10}, and τ_{20} are the variance between elementary schools.

Table 23 lists the middle school fixed effect, or average scores, for reading by grade along with the variance components and their 95% CI. The average score estimate was 1662.7 in Grade 6, 1781.9 in Grade 7, and 1859.7 in Grade 8. The intercept variance was 1127.9 in Grade 6, 115.7 in Grade 7, and 5.9 in Grade 8 which was the variance between schools at the specified grades.

Table 23
Middle Fixed Effects, Variance Components, Standard Error, and 95% CI in Reading Using the Layered Effects Model

	Layered Effects Model (Reading)	
Parameter	Parameter (SE) Estimate	95% CI
Fixed Effects (Intercept)		
Grade 6 (γ_{30})	1662.7 (9.6)	1643.9 to 1681.5
Grade 7 (γ_{40})	1781.9 (9.8)	1762.7 to 1801.1
Grade 8 (γ_{50})	1859.7 (9.6)	1840.9 to 1878.5
Variance Estimates (Intercept Variance)		
Grade 6 (τ_{30})	1127.9 (466.2)	214.2 to 2041.7
Grade 7 (τ_{40})	115.7 (78.2)	0 to 269.0
Grade 8 (τ_{50})	5.9 (29.9)	0 to 64.5
Fit Indices	AIC	BIC
	383947.7	383893.7

Note. γ_{30}, γ_{40}, and γ_{50} are the average scores for elementary schools during the specified time-period; τ_{30}, τ_{40}, and τ_{50} are the variances between elementary schools.

Table 24 demonstrates one student's variance-covariance R-matrix. The variance estimates are along the diagonal in the table, which indicates the variance between students within schools at specified grades; this is provided for the elementary

and middle school grades. The variances range from 48656 to 140531 within the elementary and middle school grades. The off-diagonal estimates are the covariance estimates between two different grades.

Table 24

Estimated R Matrix for ID 220436 in Reading

Grade	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
			Level-1 (σ^2)			
Grade 3	140531					
Grade 4	93576	99688				
Grade 5	88158	76351	93339			
Grade 6	89202	78122	76307	99419		
Grade 7	73389	64162	63764	67269	72901	
Grade 8	58579	51552	50265	54953	47169	48656

Note. σ^2 is the variation of students' scores within schools; they are found in the diagonal of the table. The covariance estimates are found in the off-diagonal of the table

Random school effects in reading. Table 25 displays the random elementary school effect estimates in reading using the Layered Effects Model. The estimates indicate the number of points "above," "below," or, "equal to" the average score for a particular subject and grade throughout the district.

Table 25

Elementary School Random Effects in Reading Using the Layered Effects Model

	Reading					
	n	Grade 3	*n*	Grade 4	*n*	Grade 5
Elementary Schools	4,760	1295.3 (7.6)	4,483	1482.5 (7.6)	4,731	1594.5 (7.9)
1	81	-5.4 (20.5)	65	27.4 (17.4)	65	-21.2(13.5)
2	209	32.8 (15.2)	110	-19.3 (14.8)	116	-1.3 (12.3)
3	147	36.3 (17.0)	145	-0.9 (14.8)	150	12.7 (11.8)
4	131	-29.9 (17.6)	126	10.7 (15.4)	131	2.0 (12.1)
5	129	-1.3 (18.0)	109	2.3 (15.7)	101	-3.4 (12.6)
6	146	.4 (17.1)	121	11.2 (15.2)	120	-1.5 (12.2)
7	167	-43.0 (16.2)	125	-11.1 (15.1)	114	18.8 (12.3)
8	141	-38.7 (17.6)	126	-19.2 (15.5)	121	6.7 (12.3)

Table 25 (Continued)

Elementary School Random Effects in Reading Using the Layered Effects Model

	Reading					
	n	Grade 3	*n*	Grade 4	*n*	Grade 5
Elementary Schools	4,760	1295.3 (7.6)	4,483	1482.5 (7.6)	4,731	1594.5 (7.9)
9		NA	82	37.1 (14.7)	111	-1.0 (12.8)
10		NA	82	-49.3 (14.4)	86	-2.9 (13.0)
11		NA	82	11.2 (14.7)	90	6.5 (12.9)
12	159	-4.5 (17.0)	151	-6.0 (14.7)	156	-15.4 (11.7)
13	151	6.1 (17.3)	140	5.8 (14.9)	147	0.7 (11.8)
14	148	54.2 (16.7)	86	-19.4 (15.9)	115	-2.5 (12.7)
15	164	6.9 (16.3)	141	7.9 (14.7)	149	3.6 (11.7)
16	113	-34.9 (18.5)	118	5.5 (15.5)	116	10.0 (12.2)
17	102	-20.0 (18.9)	86	-6.1 (16.7)	74	-17.7 (13.3)
18	144	-40.8 (17.3)	132	-15.1 (15.2)	135	-7.0 (12.0)
19	109	5.1 (18.2)	109	-5.5 (15.9)	111	3.5 (12.5)
20	76	-12.3 (21.3)	54	10.8 (17.9)	52	5.8 (13.9)
21	128	-40.1 (17.1)	93	11.4 (16.4)	88	0.2 (12.9)
22	161	1.9 (16.5)	106	-50.1 (15.4)	106	-0.9 (12.5)
23	132	16.7 (17.6)	131	-2.6 (15.3)	134	-5.6 (12.0)
24	109	-3.9 (18.7)	102	31.3 (15.9)	107	6.5 (12.6)
25	136	5.9 (17.3)	125	-0.6 (15.1)	147	1.5 (11.9)
26	198	-26.2 (15.5)	108	21.9 (14.8)	113	3.6 (12.3)
27	116	-21.8 (18.7)	102	11.8 (16.0)	95	4.9 (12.8)
28	131	-33.8 (17.9)	103	-6.9 (15.7)	116	3.9 (12.5)
29	99	34.2 (19.2)	110	-18.5 (16.1)	100	-24.4 (12.6)
30	115	-16.9 (18.0)	106	8.0 (15.9)	126	14.4 (12.4)
31	101	44.8 (18.7)	83	-6.9 (16.5)	80	-5.4 (13.2)
32	135	65.2 (17.7)	122	-5.7 (15.2)	132	-3.0 (12.0)
33	110	-59.4 (18.7)	95	15.4 (16.3)	93	7.4 (12.9)
34	128	-62.1 (17.6)	115	-28.4 (15.6)	121	-9.3 (12.3)
35	103	-6.9 (18.5)	84	-5.9 (16.5)	98	7.8 (13.0)
36	142	17.1 (17.3)	121	14.9 (15.1)	136	-4.8 (12.1)
37	157	33.1 (17.1)	151	2.8 (14.7)	160	-23.1 (11.6)
38	242	50.9 (14.6)	176	-7.9 (24.5)	233	-4.0 (11.0)
39		NA	130	37.6 (13.1)	134	6.7 (11.9)
40		NA	130	6.0 (13.6)	152	27.3 (11.9)

Note. The first column indicates each school number; *n* indicates the total number of students involved in the computation of the average score in each elementary school. The columns for Grade 3, 4, and 5 indicate the random school effect estimates with the standard error in parentheses.

Table 26 displays the school effects in the middle grades.

Table 26

Middle School Random Effects in Reading Using the Layered Effects Model

	Reading					
	N	6th	*n*	7th	*n*	8th
Middle Schools	4,875	1662.7 (9.6)	4,897	1781.9 (9.8)	4,928	1859.7 (9.6)
41	467	36.0 (12.2)	471	-1.8 (6.8)	483	-0.6 (2.3)
42	288	-28.5 (12.7)	273	-3.9 (7.6)	262	-0.7 (2.4)
43	238	-38.1 (13.6)	229	0.1 (7.8)	226	0.3 (2.4)
44	187	13.8 (13.9)	210	4.5 (8.1)	203	-0.2 (2.4)
45	556	62.0 (11.9)	556	-10.8 (6.6)	568	0.3 (2.3)
46	308	-41.3 (12.8)	304	8.8 (7.5)	305	-0.4 (2.4)
47	416	52.7 (12.3)	426	-15.2 (7.0)	437	-0.4 (2.3)
48	346	-29.3 (12.8)	340	5.4 (7.2)	345	0.8 (2.3)
49	171	-12.5 (28.5)	201	6.3 (8.2)	197	1.1 (2.4)
50	287	-24.5 (25.4)	297	1.1 (7.6)	293	-0.3 (2.4)
51	306	-15.7 (12.6)	289	-7.9 (7.5)	285	0.5 (2.4)
52	329	14.3 (12.7)	339	4.7 (7.3)	340	0.4 (2.3)
53	381	-2.4 (12.2)	367	-4.4 (7.1)	371	-1.1 (2.3)
54	294	50.2 (13.3)	312	2.0 (7.6)	327	-0.5 (2.3)
55	301	-34.6 (12.7)	283	11.1 (7.5)	286	0.8 (2.4)

Note. The first column indicates each school number; *n* indicates the total number of students involved in the computation of the average score in each middle school. The columns for Grades 6, 7, and 8 indicate the random school effect estimates with the standard error in parentheses.

Figure 3 provides a visual display of all of the elementary school random effect estimates presented in Tables 11, 15, 20, and 25. Figure 4 provides a visual display of all of the middle school random effect estimates presented in Tables 12, 16, 21, and 26. The fixed effect is the estimated grand average gain score between two adjacent grades for the Gain Score Model and the grand average score at a particular grade for the Layered Effect Model; both of which are denoted by the straight line extending from the "0." The numbers and figures in the graphs denote the deflection of the school's average gain

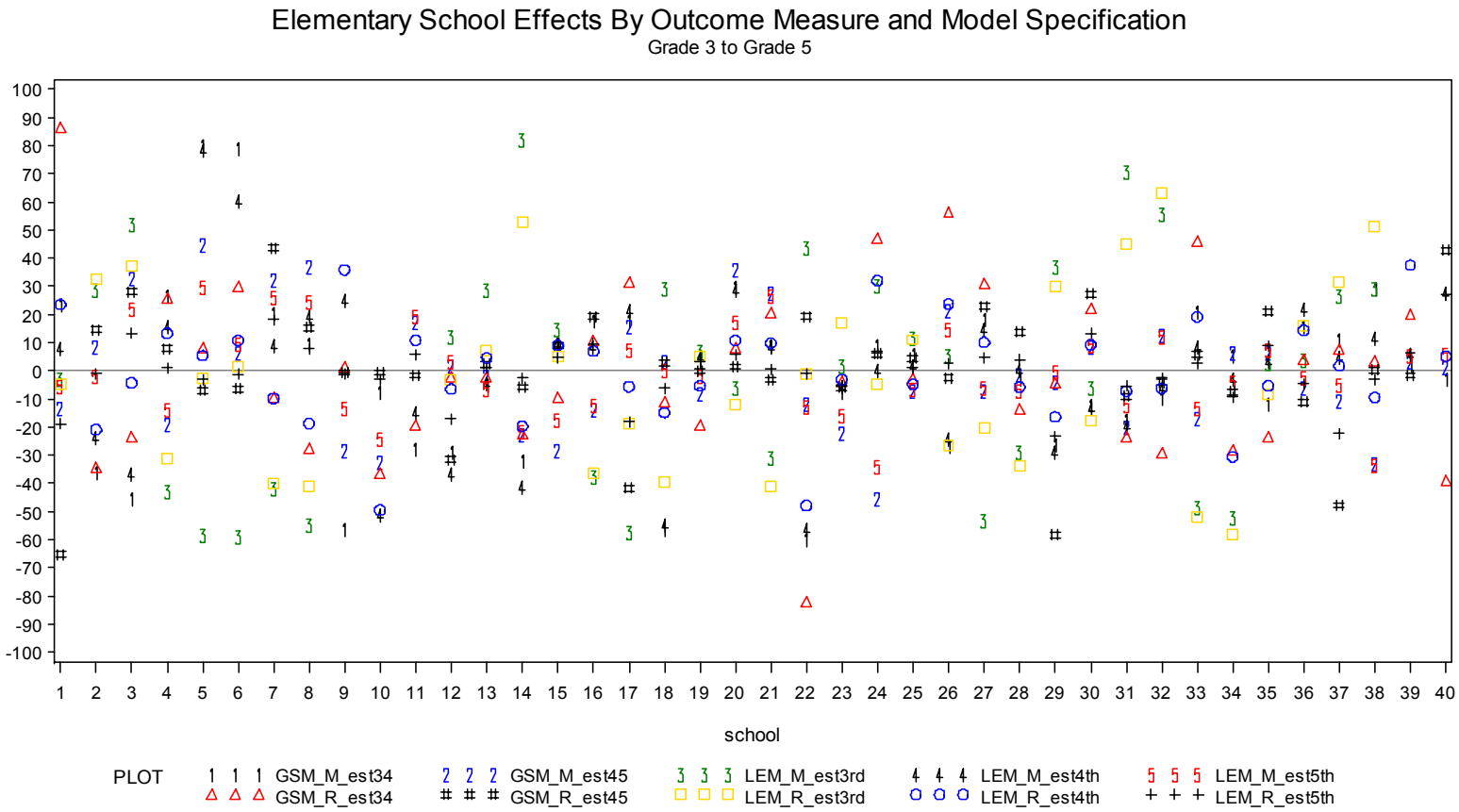

Figure 3. Elementary School Random Effects by Outcome Measure and Model Specification. The horizontal axis indicates the elementary school numbers. The vertical axis indicates the deflection of each school's gains score and average score from the grand average gain score and average score. The grand average gain score and grand average score for all elementary schools is denoted by "0," which is the the average gain and average score for all elementary schools for the specified grade(s). The denotations for the legend were M for Mathematics, R for Reading, GSM for Gain Score Model, LEM for Layered Effects Model, and "est" is the estimate for the specified grade(s).

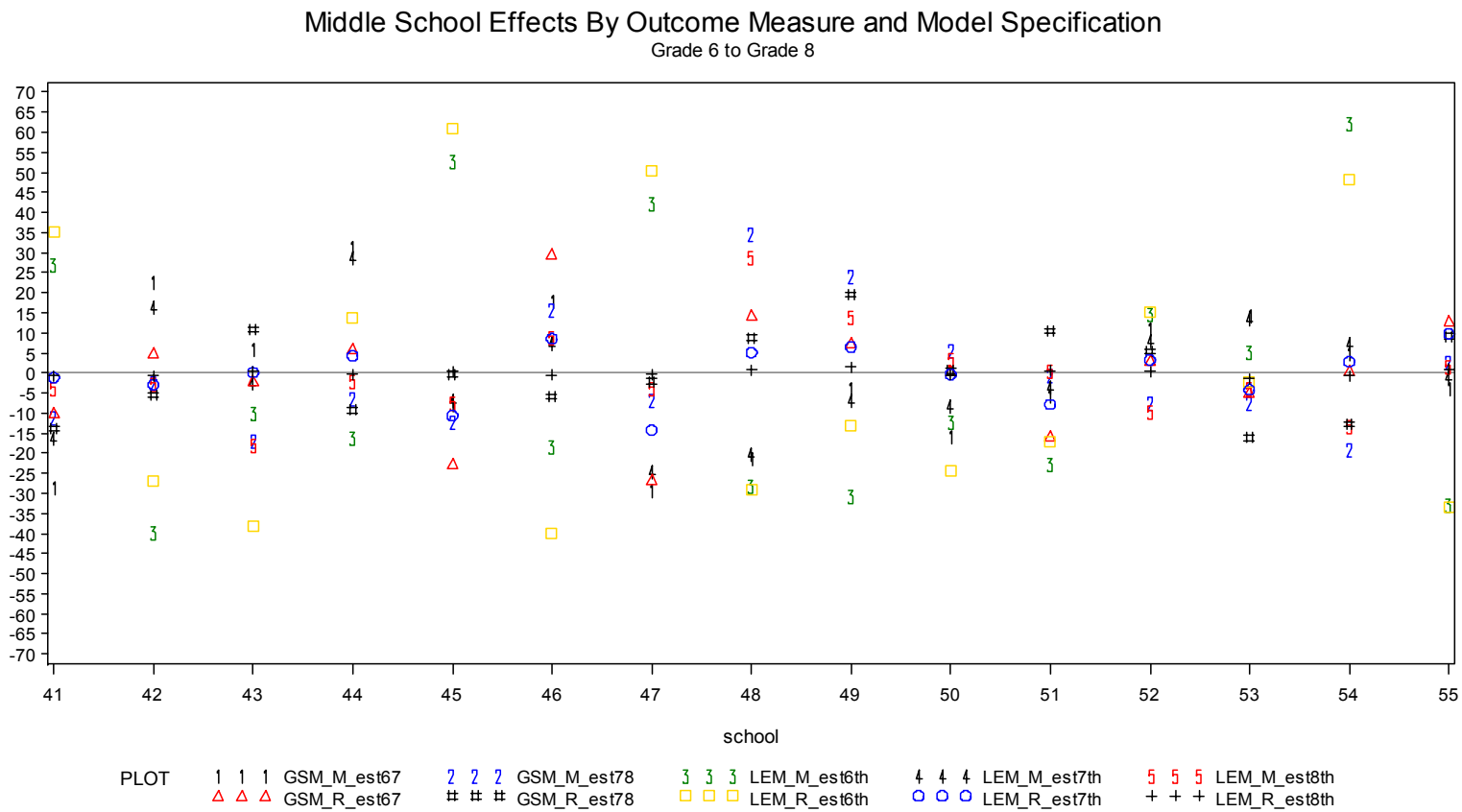

Figure 4. Middle School Random Effects by Outcome Measure and Model Specification. The horizontal axis indicates the middle school numbers. The vertical axis indicates the deflection of each school's gains score and average score from the grand average gain score and average score. The grand average gain score and grand average score for all middle schools is denoted by "0," which is the the average gain and average score for all middle schools for the specified grade(s). The denotations for the legend were M for Mathematics, R for Reading, GSM for Gain Score Model, LEM for Layered Effects Model, and "est" was the estimate for the specified grade(s).

score from the grand average gain score or the grand average score for the specified grade(s).

This concludes the estimates for the fixed effects, variance components, and random school effects along with figures for the Gain Score and Layered Effects Models.

School effects correlations. This section addresses the research questions for this study by taking the random school effect estimates for the mathematics and reading outcome measures, and examining the relationship between the two sets of estimates. Pearson product moment correlations were derived from the estimates between the outcome measures from both models and then between model specifications from both outcome measures. Correlations that are in "**bold**" indicate the primary correlations examined with the other correlations provided for informational purposes only. The correlational findings were organized into the categories of "low," "moderate," or "high."

Gain Score Model (outcome measures). Question 1 sought to determine the relationship between mathematics and reading estimates. Table 27 lists the elementary school effect correlations using the Gain Score Model between mathematics and reading estimates. The correlation was .43 between Grades 3 and 4 and .15 between Grades 4 and 5.

Table 28 lists the middle school effects correlations using the Gain Score Model between mathematics and reading. The correlations between the mathematics and reading school effects using the Gain Score Model were .48 between Grades 6 and 7 and Grades 7 and 8.

Layered Effects Model (outcome measures). Table 29 lists the school effect correlations using the Layered Effects Model between the mathematics and reading

outcome measures in the elementary grades. The correlations were .77 at Grade 3, .50 at Grade 4, and .21 at Grade 5.

Table 27
Pearson Product Moment Correlations Using the Gain Score Model between Mathematics and Reading for the Elementary School Grades

	Mathematics Grades 3 to 4	Mathematics Grades 4 to 5	Reading Grades 3 to 4	Reading Grades 4 to 5
	Elementary Schools (n = 40)			
Mathematics Grades 3 to 4	1.0			
Mathematics Grades 4 to 5	.13	1.0		
Reading Grades 3 to 4	**.43**	-.09	1.0	
Reading Grades 4 to 5	-.07	**.15**	-.40	1.0

Note. The values in the table are the Pearson product moment correlations (r).

Table 30 lists the school effect correlations using the Layered Effects Model between the mathematics and reading outcomes in the middle school grades. The correlations were .97 at Grade 6, .59 at Grade 7, and .37 at Grade 8.

Mathematics (model specifications). Question 2 sought to examine the relationship between Gain Score Model and Layered Effects Model estimates using the mathematics and reading outcome measures. These correlations compared school effects from the Gain Score Model between two adjacent elementary grades (e.g., Grade 3 and Grade 4) to the school effects of the latter estimate of the adjacent elementary grades using the Layered Effects Model. In Table 31 the elementary mathematics school effect correlations were .84 at Grade 4 and .96 at Grade 5.

Table 32 displays the correlations between the school effects using mathematics between the Gain Score and the Layered Effects Models. The correlations were .95 at Grade 7 and .82 at Grade 8.

Reading (model specifications). Table 33 provides the elementary school reading effect correlations. These correlations were .74 at Grade 4 and .92 at Grade 5. Table 34 lists the middle school correlations between estimates from the Gain Score and Layered Effects Model. The correlations were .83 at Grade 7 and .91 at Grade 8.

This concludes the School Effects Correlation section which provided the results used to address research Questions 1 and 2. These results were the Pearson product moment correlations between school effects from the outcome measure and model specification estimates. The School Ranking section, next, will provide the results to answer research Questions 3 and 4. The school effect estimates in the previous section were used to rank each school from highest to lowest, with the lowest number indicating the highest ranking. The rankings were then correlated using the Spearman rank correlation coefficient, which is the appropriate method when correlating ranks and the results are comparable with the Pearson product moment correlations.

The correlation results are the Spearman rank correlation coefficients between the school rankings from the outcome measure and model specification estimates. Question 3 was similar to Question 1 in examining the relationship between estimates from the school effects between the outcome measures using both model specifications, but differed because it used the school rankings in place of the actual numerical value for the school effects. Similarly, Question 4 was similar to Question 2 in that it examined the relationship between estimates from the school effects between model specifications using both outcome measures, but differed in that it used the school rankings in place of

Table 28

Pearson Product Moment Correlations Using the Gain Score Model between Mathematics and Reading for the Middle School Grades

	Mathematics Grades 6 to 7	Mathematics Grades 7 to 8	Reading Grades 6 to 7	Reading Grades 7 to 8
		Middle Schools (n = 15)		
Mathematics Grades 6 to7	1.0			
Mathematics Grades 7 to 8	-.19	1.0		
Reading Grades 6 to 7	**.48**	.53	1.0	
Reading Grades 7 to 8	-.23	**.48**	.14	1.0

Note. The values in the table are the Pearson product moment correlations (r).

Table 29

Pearson Product Moment Correlations Using the Layered Effects Model between Mathematics and Reading for the Elementary School Grades

	Mathematics Grade 3	Mathematics Grade 4	Mathematics Grade 5	Reading Grade 3	Reading Grade 4	Reading Grade 5
			Elementary Schools (n = 40)			
Mathematics Grade 3	1.0					
Mathematics Grade 4	-.69	1.0				
Mathematics Grade 5	-.37	.28	1.0			
Reading Grade 3	**. 77**	-.19	-.25	1.0		

Table 29 (Continued)

Pearson Product Moment Correlations Using the Layered Effects Model between Mathematics and Reading for the Elementary School Grades

	Mathematics Grade 3	Mathematics Grade 4	Mathematics Grade 5	Reading Grade 3	Reading Grade 4	Reading Grade 5
Reading Grade 4	-.23	**.50**	.07	-.14	1.0	
Reading Grade 5	-.19	.14	**.21**	-.31	.18	1.0

Note. The values in the table are the Pearson product moment correlations (*r*).

Table 30

Pearson Product Moment Correlations Using the Layered Effects Model between Mathematics and Reading for the Middle School Grades

	Mathematics Grade 6	Mathematics Grade 7	Mathematics Grade 8	Reading Grade 6	Reading Grade 7	Reading Grade 8
			Elementary Schools (*n* = 15)			
Mathematics Grade 6	1.0					
Mathematics Grade 7	-.12	1.0				
Mathematics Grade 8	-.71	.29	1.0			
Reading Grade 6	**. 97**	-.17	-.68	1.0		
Reading Grade 7	-.31	**.59**	.49	-.37	1.0	
Reading Grade 8	-.51	-.26	**.37**	-.43	.15	1.0

Note. The values in the table are the Pearson product moment correlations (*r*).

Table 31

Pearson Product Moment Correlations Using Mathematics between the Gain Score and Layered Effects Models in the Elementary School Grades

	GSM Grades 3 to 4	GSM Grades 4 to 5	LEM Grade 4	LEM Grade 5
		Elementary Schools (n = 40)		
GSM Grades 3 to 4	1.0			
GSM Grades 4 to 5	.13	1.0		
LEM Grade 4	**.84**	.21	1.0	
LEM Grade 5	.13	**.96**	.28	1.0

Note. The values in the table are the Pearson product moment correlations (r). GSM indicates the Gain Score Model and LEM indicates the Layered Effects Model.

Table 32

Pearson Product Moment Correlations Using Mathematics between the Gain Score and Layered Effects Models in the Middle School Grades

	GSM Grades 6 to 7	GSM Grades 7 to 8	LEM Grade 7	LEM Grade 8
		Middle Schools (n = 15)		
GSM Grades 6 to 7	1.0			
GSM Grades 7 to 8	-.18	1.0		
LEM Grade 7	**.95**	-.25	1.0	
LEM Grade 8	.31	**.82**	.29	1.0

Note. The values in the table are the Pearson product moment correlations (r). GSM indicates the Gain Score Model and LEM indicates the Layered Effects Model.

Table 33

Pearson Product Moment Correlations Using Reading between the Gain Score and Layered Effects Models in the Middle School Grades

	GSM Grades 3 to 4	GSM Grades 4 to 5	LEM Grade 4	LEM Grade 5
		Elementary Schools (n = 40)		
GSM Grades 3 to 4	1.0			
GSM Grades 4 to 5	-.40	1.0		
LEM Grade 4	**.74**	-.12	1.0	
LEM Grade 5	-.18	**.92**	.18	1.0

Note. The values in the table are the Pearson product moment correlations (r). GSM indicates the Gain Score Model and LEM indicates the Layered Effects Model.

Table 34

Pearson Product Moment Correlations Using Reading between the Gain Score and Layered Effects Models in the Middle School Grades

	GSM Grades 6 to 7	GSM Grades 7 to 8	LEM Grade 7	LEM Grade 8
		Middle Schools (n = 15)		
GSM Grades 6 to 7	1.0			
GSM Grades 7 to 8	.14	1.0		
LEM Grade 7	**.83**	-.07	1.0	
LEM Grade 8	.30	**.91**	.15	1.0

Note. The values in the table are the Pearson product moment correlations (r). GSM indicates the Gain Score Model and LEM indicates the Layered Effects Model.

the actual numerical value school effects. None of the schools had the same ranking using any outcome measure or model specification.

School rankings. This section provides the information for the school rankings and Spearman rank correlation coefficients between the school rankings using the Gain Score Model. The random school effects from Tables 11 and 12 were used to rank each school in order from 1-40 in the elementary grades and 41-55 in the middle school grades. Lower ranking numbers for schools (i.e., 1, 2, 3, etc.) indicated higher school effects for the specified grades. The rankings for the schools in mathematics and reading were then correlated as done previously with the school effects in Questions 1 and 2.

Gain Score Model (mathematics). Table 35 lists the elementary school numbers, the number of students in each school, and the school rankings. The school's ranking on the gain scores between Grades 3 and 4, and Grades 4 and 5 were based upon the school effects from Tables 11 and 12. Schools with larger school effects have higher school rankings and are represented by lower ranking numbers. For instance, school #7 ranked 11th from Grade 3 to Grade 4 and ranked 5th from Grade 4 to Grade 5, compared to school #8 that ranked 13th from Grade 3 to Grade 4, and 2nd from Grade 4 to Grade 5. The rankings were similar in some instances, but quite different for the majority when comparing across grade (i.e., 3rd to 4th and 4th to 5th Grades).

Table 35

Elementary School Rankings in Mathematics Using the Gain Score Model

	Mathematics			
	n	3rd to 4th	*n*	4th to 5th
Elementary Schools	3,757	112.1 (5.1)	4,060	147.3 (3.5)
1	61	8	63	31
2	91	33	104	16
3	118	37	132	3

Table 35 (Continued)

Elementary School Rankings in Mathematics Using the Gain Score Model

	Mathematics			
	n	3rd to 4th	*n*	4th to 5th
Elementary Schools	3,757	112.1 (5.1)	4,060	147.3 (3.5)
4	99	4	108	33
5	99	1	87	1
6	109	2	106	13
7	105	11	100	5
8	105	13	99	2
9	57	38	80	37
10	72	26	73	34
11	70	34	73	8
12	133	35	133	14
13	126	24	132	21
14	66	36	81	36
15	121	16	135	38
16	107	9	111	30
17	70	10	61	9
18	111	39	116	17
19	93	20	99	26
20	49	5	45	4
21	80	19	79	6
22	92	40	94	25
23	113	25	118	32
24	93	18	94	40
25	105	15	130	28
26	98	31	103	7
27	86	6	84	24
28	91	21	96	20
29	101	32	96	22
30	93	29	105	11
31	73	30	66	35
32	107	27	121	10
33	79	7	82	29
34	100	17	104	12
35	75	28	81	15
36	106	12	113	23
37	123	14	137	27
38	135	3	167	39

Table 35

Elementary School Rankings in Mathematics Using the Gain Score Model

	Mathematics			
	n	3rd to 4th	*n*	4th to 5th
Elementary Schools	3,757	112.1 (5.1)	4,060	147.3 (3.5)
39	108	23	120	18
40	87	22	119	19

Note. The *n* indicates the number of students involved in the computation with consecutive scores between the respective grades; the numbers below the consecutive grades (i.e., 3rd to 4th) indicate the ranking of the school based on the school effect estimates.

Table 36 displays the middle school rankings, which can be interpreted as the ranking of each school based upon the school effect the school has on student performance and/or growth.

Table 36

Middle School Ranking in Mathematics Using the Gain Score Model

	Mathematics			
	n	6th to 7th	*n*	7th to 8th
Middle Schools	4,325	135.9 (4.9)	4,328	104.9 (4.0)
41	429	14	444	12
42	242	2	230	6
43	211	6	208	14
44	172	1	193	9
45	490	11	505	13
46	267	3	259	3
47	372	15	393	8
48	312	13	304	1
49	167	9	181	2
50	252	12	263	4
51	259	10	252	7
52	298	5	300	10
53	337	4	341	11
54	257	7	279	15
55	255	8	259	5

Note. The *n* indicates the number of students involved in the computation with consecutive scores between the respective grades; the number below the consecutive grades (i.e., 6th to 7th) indicates the ranking of the school based on the school effect estimates.

Gain Score Model (reading). Table 37 lists the elementary school numbers, the number of students in each school who had two consecutive scores, and the school rankings in reading. The school's ranking between Grades 3 and 4, and Grades 4 and 5 were based upon the school effects from Tables 11 and 12. Schools with larger school effects have higher school rankings which are represented by lower ranking numbers.

Table 37

Elementary School Rankings in Reading Using the Gain Score Model

	Reading			
	n	3rd to 4th	*n*	4th to 5th
Elementary Schools	3,838	166.6 (5.2)	4,071	109.2 (4.2)
1	61	1	63	40
2	91	36	104	12
3	120	28	131	4
4	102	6	111	13
5	98	17	88	32
6	109	9	106	33
7	106	27	102	1
8	105	34	98	10
9	57	15	80	30
10	72	38	73	14
11	70	30	73	21
12	134	22	135	36
13	126	18	132	22
14	65	37	81	24
15	121	24	135	16
16	107	12	109	5
17	71	7	62	37
18	112	25	118	18
19	95	31	99	19
20	49	13	46	17
21	78	10	78	27

Table 37 (Continued)

Elementary School Rankings in Reading Using the Gain Score Model

	Reading			
	n	3rd to 4th	*n*	4th to 5th
Elementary Schools	3,838	166.6 (5.2)	4,071	109.2 (4.2)
22	93	40	97	6
23	113	21	118	29
24	93	3	93	20
25	105	19	128	15
26	98	2	103	26
27	86	5	85	8
28	90	26	96	11
29	101	23	96	39
30	93	8	105	3
31	73	35	66	31
32	107	33	121	35
33	79	4	82	9
34	99	32	101	28
35	76	29	81	7
36	105	16	116	34
37	123	14	137	38
38	135	20	163	23
39	108	11	119	25
40	88	39	121	2

Note. The *n* indicates the number of students involved in the computation with consecutive scores between the respective grades; the number below the consecutive grades (i.e., 4th to 5th) indicates the ranking of the school based on the school effect estimates.

Table 38 displays the middle school rankings, which can be interpreted as the ranking of each school based upon the school effect the school has on student performance and/or growth. The rankings were similar in some instances, but quite different for the majority when comparing across grades (i.e., 6th to 7th and 7th to 8th Grades).

Table 38

Middle School Rankings in Reading Using the Gain Score Model

	Reading			
	n	6th to 7th	*n*	7th to 8th
Middle Schools	4,325	113.0 (4.3)	4,328	104.9 (4.0)
41	430	12	445	14
42	243	7	243	11
43	212	10	212	3
44	172	5	172	12
45	490	14	490	7
46	265	1	265	10
47	371	15	371	9
48	312	3	312	5
49	168	6	168	1
50	249	8	249	8
51	261	13	261	4
52	297	4	297	6
53	338	11	338	15
54	260	9	260	13
55	254	2	254	2

Note. The *n* indicates the number of students involved in the computation with consecutive scores between the respective grades; the number below the consecutive grades (i.e., 7h to 8th) indicates the ranking of the school based on the school effect estimates.

Layered Effects Model (mathematics). Table 39 displays the elementary school rankings using the Layered Effects Model in mathematics. These indicate the ranking of the school in a particular grade compared to other schools in the district at the same level.

Table 39

Elementary School Rankings in Mathematics Using the Layered Effects Model

	Mathematics					
	n	3rd	*n*	4th	*n*	5th
Elementary Schools	4,748	1315.1 (7.3)	4,474	1441.6 (8.3)	4,749	1591.9 (8.7)
1	81	22	65	17	65	25
2	209	8	110	33	116	20

Table 39 (Continued)

Elementary School Rankings in Mathematics Using the Layered Effects Model

	Mathematics					
	n	3rd	*n*	4th	*n*	5th
Elementary Schools	4,748	1315.1 (7.3)	4,474	1441.6 (8.3)	4,749	1591.9 (8.7)
3	146	4	145	35	150	5
4	130	27	120	9	131	34
5	129	34	110	1	99	1
6	146	35	121	2	120	10
7	166	28	122	13	116	2
8	141	33	126	7	121	4
9		NA	83	4	111	33
10		NA	82	38	87	38
11		NA	82	29	89	6
12	159	14	149	36	155	16
13	151	10	140	27	148	24
14	148	1	88	37	116	37
15	164	13	141	15	149	36
16	113	26	118	16	118	31
17	102	32	85	8	74	12
18	144	11	128	39	137	17
19	107	15	109	21	111	19
20	76	23	54	3	51	7
21	127	25	94	12	90	3
22	161	6	104	40	106	29
23	132	19	132	28	134	35
24	109	7	102	24	108	40
25	137	17	125	18	149	28
26	198	18	108	32	113	8
27	116	29	102	10	94	26
28	131	24	104	25	116	27
29	99	5	110	34	100	18
30	115	21	106	30	126	11
31	101	2	83	31	80	32
32	134	3	122	26	132	9
33	110	30	95	14	93	30
34	129	31	117	23	121	22
35	97	20	85	22	98	13
36	141	16	119	6	133	21

Table 39

Elementary School Rankings in Mathematics Using the Layered Effects Model

	Mathematics					
	n	3rd	*n*	4th	*n*	5th
Elementary Schools	4,748	1315.1 (7.3)	4,474	1441.6 (8.3)	4,749	1591.9 (8.7)
37	157	12	151	19	160	23
38	242	9	177	11	244	39
39		NA	131	20	135	15
40		NA	129	5	153	14

Note. The *n* indicates the number of students involved in the computation with scores for the respective grades in each school; the numbers below the consecutive grades (i.e., 3rd, 4h and 5th) indicate the ranking of the school based on the school effect estimates.

Table 40 displays the middle school rankings using the Layered Effects Model in mathematics. These indicate the ranking of the school in a particular grade compared to other schools in the district at the same level.

Table 40

Middle School Rankings in Mathematics Using the Layered Effects Model

	Mathematics					
	n	6th	*n*	7th	*n*	8th
Middle Schools	4,874	1631.0 (10.0)	4,890	1773.2 (10.8)	4,936	1885.4 (11.2)
41	465	4	472	13	482	9
42	288	15	272	2	260	8
43	236	7	229	8	227	15
44	187	8	210	1	203	7
45	556	2	556	11	569	12
46	309	10	305	6	309	3
47	417	3	426	15	437	10
48	346	13	340	14	344	1
49	169	12	202	12	199	2
50	289	9	298	10	295	4
51	307	11	285	9	287	6
52	330	5	339	4	340	13
53	380	6	365	3	373	11

Table 40 (Continued)

Middle School Rankings in Mathematics Using the Layered Effects Model

	Mathematics					
	n	6th	*n*	7th	*n*	8th
Middle Schools	4,874	1631.0 (10.0)	4,890	1773.2 (10.8)	4,936	1885.4 (11.2)
54	293	1	309	5	327	14
55	302	14	282	7	284	5

Note. The *n* indicates the number of students involved in the computation with scores for the respective grades in each school; the numbers below the consecutive grades (i.e., 6th, 7th, and 8th) indicate the ranking of the school based on the school effect estimates.

Layered Effects Model (reading). Table 41 displays the elementary school rankings using the Layered Effects Model in reading.

Table 41

Elementary School Rankings in Reading Using the Layered Effects Model

	Reading					
	n	Grade 3	*n*	Grade 4	*n*	Grade 5
Elementary Schools	4,760	1295.3 (7.6)	4,483	1482.5 (7.6)	4,731	1594.5 (7.9)
1	81	20	65	4	65	38
2	209	8	110	36	116	24
3	147	5	145	22	150	4
4	131	27	126	13	131	18
5	129	17	109	20	101	29
6	146	16	121	11	120	25
7	167	33	125	32	114	2
8	141	30	126	35	121	8
9		NA	82	2	111	23
10		NA	82	39	86	27
11		NA	82	10	90	10
12	159	19	151	27	156	36
13	151	12	140	17	147	20
14	148	2	86	37	115	26
15	164	11	141	15	149	15
16	113	29	118	18	116	5
17	102	24	86	28	74	37

Table 41 (Continued)

Elementary School Rankings in Reading Using the Layered Effects Model

	Reading					
	n	Grade 3	*n*	Grade 4	*n*	Grade 5
Elementary Schools	4,760	1295.3 (7.6)	4,483	1482.5 (7.6)	4,731	1594.5 (7.9)
18	144	32	132	33	135	34
19	109	14	109	24	111	17
20	76	22	54	12	52	12
21	128	31	93	9	88	21
22	161	15	106	40	106	22
23	132	10	131	23	134	33
24	109	18	102	3	107	11
25	136	13	125	21	147	19
26	198	26	108	5	113	16
27	116	25	102	8	95	13
28	131	28	103	30	116	14
29	99	6	110	34	100	40
30	115	23	106	14	126	3
31	101	4	83	29	80	32
32	135	1	122	25	132	28
33	110	34	95	6	93	7
34	128	35	115	38	121	35
35	103	21	84	26	98	6
36	142	9	121	7	136	31
37	157	7	151	19	160	39
38	242	3	176	31	233	30
39		NA	130	1	134	9
40		NA	130	16	152	1

Note. The n indicates the number of students involved in the computation with scores for the respective grades in each elementary school; the numbers below the consecutive grades (i.e., 3rd, 4th, and 5th) indicate the ranking of each elementary school based on the school effect estimates.

Table 42 displays the middle school rankings using the Layered Effects Model in reading. These indicate the ranking of a school in a particular grade compared to other schools in the district at the same level.

Table 42

Middle School Rankings in Reading Using the Layered Effects Model

	Reading					
	n	6th	*n*	7th	*n*	8th
Middle Schools	4,875	1662.7 (9.6)	4,897	1781.9 (9.7)	4,928	1859.7 (9.6)
41	467	4	471	10	483	13
42	288	11	273	11	262	14
43	238	14	229	9	226	6
44	187	6	210	6	203	8
45	556	1	556	14	568	7
46	308	15	304	2	305	11
47	416	2	426	15	437	10
48	346	12	340	4	345	3
49	171	8	201	3	197	1
50	287	10	297	8	293	9
51	306	9	289	13	285	4
52	329	5	339	5	340	5
53	381	7	367	12	371	15
54	294	3	312	7	327	12
55	301	13	283	1	286	2

Note. The *n* indicates the number of students involved in the computation with scores for the respective grades in each middle school; the numbers below the consecutive grades (i.e., 6th, 7th, and 8th) indicate the ranking of the school based on the school effect estimates.

This concludes the School Rankings section which provided a narrative description of the random school rankings for mathematics and reading using the Gain Score Model and Layered Effects Model in the elementary and middle school grades. The School Ranking Correlations section, next, describes the correlations between the school rankings, which were the primary values used in answering research Questions 3 and 4 of this study. Questions 3 and 4 examined the relationship between school rankings based on the outcome measures (i.e., mathematics and reading) and model specifications (i.e., Gain Score Model and Layered Effects Model). These rankings were derived from the school effect estimates and were correlated using the Spearman rank correlation

coefficient. These correlations were estimated in the same fashion as with the Pearson product moment correlations used with the school effects.

School ranking correlations.

Gain Score Model (outcome measures). Table 43 lists the Spearman rank correlation coefficients between mathematics and reading school rankings using the Gain Score Model for the elementary grades. The elementary school correlation was .51 between Grades 3 and 4 and .07 between Grades 4 and 5. Table 44 lists the Spearman rank correlation coefficients between the mathematics and reading school rankings in the middle school grades. The middle school correlations were .47 between Grades 6 and 7 and .41 between Grades 7 and 8.

Layered Effects Model (outcome measures). Table 45 lists the Spearman rank correlation coefficients between the mathematics and reading school rankings in the elementary school grades using the Layered Effects Model. The correlations were .74 at Grade 3, .46 at Grade 4, and .22 at Grade 5. Table 46 lists the Spearman rank correlation coefficients between the mathematics and reading school rankings in the middle school grades using the Layered Effects Model. The correlations between mathematics and reading were .78 at Grade 6, .17 at Grade 7, and .38 at Grade 8.

Mathematics (model specifications). Table 47 lists the school ranking correlations in the elementary grades using mathematics between the Gain Score Model and the Layered Effects Model. The correlations were .79 at Grade 4 and .96 at Grade 5. Table 48 lists the correlations between the school rankings in the middle school grades between the Gain Score Model and the Layered Effects Model using mathematics.

Table 43

Spearman Rank Correlation Coefficients Using the Gain Score Model between Mathematics and Reading in the Elementary School Grades

	Mathematics Grades 3 to 4	Mathematics Grades 4 to 5	Reading Grades 3 to4	Reading Grades 4 to 5
		Elementary Schools (n = 40)		
Mathematics Grades 3 to 4	1.0			
Mathematics Grades 4 to 5	-.01	1.0		
Reading Grades 3 to 4	**.51**	-.05	1.0	
Reading Grades 4 to 5	-.04	**.07**	-.24	1.0

Note. The values in the table are the *Spearman rank correlation coefficients* (r_s).

Table 44

Spearman Rank Correlation Coefficients Using the Gain Score Model between Mathematics and Reading in the Middle School Grades

	Mathematics Grades 6 to 7	Mathematics Grades 7 to 8	Reading Grades 6 to 7	Reading Grades 7 to 8
		Middle Schools (n = 15)		
Mathematics Grades 6 to7	1.0			
Mathematics Grades 7 to 8	-.10	1.0		
Reading Grades 6 to 7	**.47**	.57	1.0	
Reading Grades 7 to 8	-.23	**.41**	.23	1.0

Note. The values in the table are the *Spearman rank correlation coefficients* (r_s).

Table 45

Spearman Rank Correlation Coefficients Using the Layered Effects Model between Mathematics and Reading in the Elementary School Grades

	Mathematics Grade 3	Mathematics Grade 4	Mathematics Grade 5	Reading Grade 3	Reading Grade 4	Reading Grade 5
			Middle Schools (n = 15)			
Mathematics Grade 3	1.0					
Mathematics Grade 4	-.72	1.0				
Mathematics Grade 5	-.35	.18	1.0			
Reading Grade 3	**.74**	-.28	-.26	1.0		
Reading Grade 4	-.27	**.46**	.06	-.17	1.0	
Reading Grade 5	-.22	.17	**.22**	-.37	.29	1.0

Note. The values in the table are the *Spearman rank correlation coefficients* (r_s).

Table 46

Spearman Rank Correlation Coefficients Using the Layered Effects Model between Mathematics and Reading in the Middle School Grades

	Mathematics Grade 6	Mathematics Grade 7	Mathematics Grade 8	Reading Grade 6	Reading Grade 7	Reading Grade 8
			Middle Schools (n = 15)			
Mathematics Grade 6	1.0					
Mathematics Grade 7	-.11	1.0				
Mathematics Grade 8	-.73	-.28	1.0			
Reading Grade 6	**.78**	-.16	-.49	1.0		
Reading Grade 7	-.46	**.17**	.50	-.54	1.0	
Reading Grade 8	-.38	-.34	**.38**	-.24	.45	1.0

Note. The values in the table are the *Spearman rank correlation coefficients* (r_s).

Table 47

Spearman Rank Correlation Coefficients Using Mathematics between the Gain Score Model and the Layered Effects Model in the Elementary School Grades

	GSM Grades 3 to 4	GSM Grades 4 to 5	LEM Grade 4	LEM Grade 5
		Elementary Schools (n = 40)		
GSM Grades 3 to 4	1.0			
GSM Grades 4 to 5	-.01	1.0		
LEM Grade 4	**.79**	.10	1.0	
LEM Grade 5	.01	**.96**	.18	1.0

Note. The values in the table are the *Spearman rank correlation coefficients* (r_s). GSM indicates the Gain Score Model and LEM indicates the Layered Effects Model.

Table 48

Spearman Rank Correlation Coefficients Using Mathematics between the Gain Score Model and the Layered Effects Model in the Middle School Grades

	GSM Grades 6 to 7	GSM Grades 7 to 8	LEM Grade 7	LEM Grade 8
		Middle Schools (n = 15)		
GSM Grades 6 to 7	1.0			
GSM Grades 7 to 8	-.10	1.0		
LEM Grade 7	**.94**	-.24	1.0	
LEM Grade 8	-.16	**.94**	-.28	1.0

Note. The values in the table are the *Spearman rank correlation coefficients* (r_s). GSM indicates the Gain Score Model and LEM indicates the Layered Effects Model.

Table 49

Spearman Rank Correlation CoefficientsUsing Reading between the Gain Score Model and the Layered Effects Model in the Elementary School Grades

	GSM Grades 3 to 4	GSM Grades 4 to 5	LEM Grade 4	LEM Grade 5
		Elementary Schools (n = 40)		
GSM Grades 3 to 4	1.0			
GSM Grades 4 to 5	-.24	1.0		
LEM Grade 4	**.76**	-.13	1.0	
LEM Grade 5	.02	**.83**	.29	1.0

Note. The values in the table are the *Spearman rank correlation coefficients* (r_s). GSM indicates the Gain Score Model and LEM indicates the Layered Effects Model.

Table 50

Spearman Rank Correlation CoefficientsUsing Reading between the Gain Score Model and the Layered Effects Model in the Middle School Grades

	GSM Grades 6 to 7	GSM Grades 7 to 8	LEM Grade 7	LEM Grade 8
		Middle Schools (n = 15)		
GSM Grades 6 to 7	1.0			
GSM Grades 7 to 8	.23	1.0		
LEM Grade 7	**.93**	.34	1.0	
LEM Grade 8	.32	**.92**	.45	1.0

Note. The values in the table are the *Spearman rank correlation coefficients* (r_s). GSM indicates the Gain Score Model and LEM indicates the Layered Effects Model.

The correlations were both .94 at Grade 7 and Grade 8.

Reading (model specifications). Table 49 lists the school ranking correlations in the elementary grades using reading between the Gain Score and the Layered Effects Model. The correlations were .76 at Grade 4 and .83 at Grade 5. Table 50 lists the school ranking correlations in the middle grades using reading between the Gain Score and the Layered Effects Model. The correlations were .93 at Grade 7 and .92 at Grade 8.

Summary

Table 51 provides a summary of the school effects and school ranking correlation coefficients for the outcome measures and model specifications. When using the mathematics and reading correlations from the Gain Score Model, correlations for school effects and school rankings were low, when advancing from Grade 4 to Grade 5 and moderate in the other grades. When using the Layered Effects Model, the correlations were low at Grade 5 for the school effects and the school rankings, and at Grade 7 for the school rankings. In the other grades, correlations were moderate to high. When using the Gain Score Model and Layered Effects Model correlations from mathematics, both the school effects and school rankings were high in each grade. When using reading, the correlations were also large for each of the grades.

This concludes the Primary Results section of this chapter which displayed the findings from all of the analyses that were used to answer the research questions. Table 51 summarizes these results. Chapter V will discuss the findings, limitations of the study, and directions for future research investigating school effects and school rankings, then the chapter will conclude with closing remarks.

Table 51

Summary of Pearson Product Moment Correlations (r) and Spearman Rank Correlation Coefficient (r_s)

Mathematics-Reading relationship by Model Specification				Gain Score Model-Layered Effects Model relationship by Outcome Measure		
		r	r_s		*r*	r_s
Gain Score Model				Mathematics		
Grades	3 to 4	.43	.51	Grade 4	.84	.79
Grades	4 to 5	.15	.07	Grade 5	.96	.96
Grades	6 to 7	.48	.47	Grade 7	.95	.94
Grades	7 to 8	.48	.41	Grade 8	.82	.94
Layered Effects Model				Reading		
Grade	3	.77	.74			
Grade	4	.50	.46	Grade 4	.74	.76
Grade	5	.21	.22	Grade 5	.92	.83
Grade	6	.97	.78			
Grade	7	.59	.17	Grade 7	.83	.93
Grade	8	.37	.38	Grade 8	.91	.92

Note. The *Pearson product moment correlations* from the school effects are denoted by *r* and the *Spearman rank correlation coefficients* from the school rankings are denoted by *r*.

Chapter V: Discussion

Overview

This chapter provides a discussion of the results from this study. It begins with a review of the methods used and, next, lists the research questions, limitations of the study, implications for the field, and directions for future research, before concluding with closing remarks. The purpose of this study was to examine two Value Added Model specifications to understand similarities and differences in school effect results. Specifically, this study correlated value added school effect estimates, which were derived from two model specifications and two outcome measures. Next, the school rankings were compared and correlated using the same models and outcome measures. This study used a non-experimental, correlational, longitudinal design to examine existing student achievement data from a sample of approximately 5,000 students in Grades 3 through 5 and Grades 6 through 8 from a moderately large and rural school district in Florida. These students attended 40 elementary and 15 middle schools and tested at least one time in the district between the 3rd Grade in 2005 and the 8th Grade in 2010. The population was highly transient, with some students remaining in the district during the entire period of this study and others transferring in and out of the district.

The two outcome measures used to estimate school by grade effects were students' Florida Comprehensive Assessment Test (FCAT) scores in mathematics and reading. The two model specifications used to estimate school by grade effects were the

Gain Score and Layered Effects Models. The school effect estimates from the two outcomes and models were used in the calculation of the Pearson product moment correlation coefficients. Next, each school was ranked based upon the school effect estimates, then the school rankings were used in calculating the Spearman rank correlation coefficients. Correlational findings were organized into the categories of "low" (high sensitivity), "moderate," or "high" (low sensitivity). This study built upon previous studies examining school effects, by examining four research questions:

Research questions.

Research Question 1. What is the relationship between school effect estimates from the Gain Score Model when mathematics achievement scores are used versus reading achievement scores? What is the relationship between school effect estimates from the Layered Effects Model when mathematics achievement scores are used versus reading achievement scores?

Research Question 2. What is the relationship between school effect estimates based on mathematics achievement scores when the Gain Score Model is used versus the Layered Effects Model? What is the relationship between school effect estimates based on reading achievement scores when the Gain Score Model is used versus the Layered Effects Model?

Research Question 3. What is the relationship between school rankings from the Gain Score Model when mathematics achievement scores are used versus reading

achievement scores? What is the relationship between school rankings from the Layered Effects Model when mathematics achievement scores are used versus reading achievement scores?

Research Question 4. What is the relationship between school rankings based on mathematics achievement scores when the Gain Score Model is used versus the Layered Effects Model? What is the relationship between school rankings based on reading achievement scores when the Gain Score Model is used versus the Layered Effects Model?

School Effect Relationship by Outcome Measures

Answering Question 1 required an examination of the relationship between the school effect estimates from the mathematics and reading outcome measures, first using the Gain Score Model, then the Layered Effects Model. When using the Gain Score Model, the Pearson product moment correlations between the school effects from mathematics and the school effects from reading ranged from .15 (Grades 4 to 5) to .48 (Grades 6 to 7 and Grades 7 to 8). These low to moderate results indicate that the choice of the outcome measure may substantially alter conclusions about which schools are most and least effective, when using the Gain Score Model.

When using the Layered Effects Model, school effects from the mathematics outcome measures were correlated with the school effects from the reading outcome measures. The Pearson product moment correlations ranged from .21 (Grade 5) to .97 (Grade 6). Both the elementary and middle school grades' correlations were high

(Pearson product moment correlations for Grades 3 and 6 were .77 and .97, respectively) and were moderately low in the latter years using the Layered Effects Model (Pearson product moment correlations for Grades 5 and 8 were .21 and .37, respectively). The low correlations in the higher grade of elementary schools (Grade 5) and middle schools (Grade 8) suggest that for these years, the choice of outcome measure is likely to alter conclusions about which schools are most and least effective when using the Layered Effects Model to estimate school effects.

Schmitz (2007) examined the sensitivity of effect estimates to the outcome measures for teacher effects, not school effects. The results, however, were similar to this study's results in that the effect estimates were sensitive to which outcome measures (mathematics and reading achievement scores) were used. He also examined the correlation between teacher effects based on mathematics and reading in the elementary grades using three model specifications: (1) a Simple Fixed Effects Model, (2) a Conditional two-level Random Effects Model, and (3) a Conditional Cumulative Effects Model. He found that the results from the Cumulative Effects and the Layered Effects Models were somewhat similar, both estimating the random effects; his correlations ranged from .44 to .65 and had a mean of .53. Though the effects and models used differed between the Schmitz (2007) study and this present study, both found moderate correlations.

The results of this study examined school effects and indicated that different outcome measures will likely provide different value added effects. In situations where awards or sanctions are given based on these effects, the results for a school may be different depending upon the outcome measure examined. Caution should be considered

when attempting to rank schools based on value added effects when varied outcome measures are examined.

School Effect Relationship by Model Specification

Answering Question 2 required an examination of the relationship between school effect estimates from the Gain Score Model and school effect estimates from the Layered Effects Model, first for the mathematics outcome measure, then for the reading outcome measure. When using the mathematics outcome measure, the Pearson product moment correlations between the school effects from the Gain Score Model and the school effects from the Layered Effects Model ranged from .82 (Grade 8) to .96 (Grade 5). The high correlations may indicate that the gains in the school effect estimates (Gain Score Model) correspond closely with the average school effect estimates (Layered Effects Model) when examining mathematics. Therefore, the choice of model specification is not likely to substantially alter conclusions about which schools are most and least effective when mathematics is used as the outcome measure.

Similarly to the mathematics outcome measure, high correlations were found for the reading outcome measure, but to a lesser degree. These Pearson product moment correlations ranged from .74 (Grade 4) to .92 (Grade 5). The high correlations in the reading outcome measure may indicate that the gains in the school effect estimates from the Gain Score Model correspond closely with the average school effect estimates from the Layered Effects Model. Therefore, conclusions about which schools are most and least effective when examining these model specifications will likely be similar within the same outcome measure (e.g., mathematics or reading).

Tekwe et al. (2004) also found high correlations using mathematics outcome measures in their study, which compared school effects using four models (i.e., Simple Fixed Effects Model, Simple Unadjusted Change Score, Demographic and Intake Adjusted Change Score, and a Multivariate Layered Effect Model). They found that when using the Simple Unadjusted Change Score Model and the Multivariate Layered Effects Model, the Pearson product moment correlations ranged from .96 (Grade 4) to .98 (Grade 3) in mathematics. Briggs and Weeks (2011) examined school effects using three model specifications (i.e., Constrained Persistence Model, Alternate Constrained Persistence Model, and a Layered Effects Model) and found that the correlations ranged from .47 (Grade 5) to .93 (Grade 6) in mathematics. When Tekwe et al. (2004) examined reading outcome measures using the Simple Unadjusted Change Score Model and the Multivariate Layered Effects Model, they found that the correlations ranged from .94 (Grade 4) to .99 (Grade 3) and Briggs and Weeks (2011) found that the reading correlations ranged from .58 (Grade 5) to .98 (Grade 8).

This current study and the previous research studies compared varied model specifications and found moderate (.47) to high (.99) correlations when using mathematics and reading outcome measures. In high-stakes situations, the differences between the model estimates may be too large to determine whether either estimate would be sufficient in making decisions; however, in lower stakes situations, the results may suffice to make decisions regardless of the model. This current study's and previous research studies' results demonstrated that school effect estimates were more sensitive to outcome measures than they were to model specifications; therefore, arguments for using

complex models over simpler models (e.g., Layered Effects Model versus the Gain Score Model) may be questionable.

School Ranking Relationship by Outcome Measures

Answering Question 3 required using the school effect estimates from mathematics and reading outcome measures for each model and ranking each school. The rankings were then used to calculate the Spearman rank correlation coefficients, which examined the relationship between school rankings based on the mathematics outcome measure and the reading outcome measure using first, the Gain Score Model, then the Layered Effects Model. Results from the school rankings followed a similar pattern as the school effects. Since the school effect correlations were low to moderate, one would expect fairly similar correlation coefficients when correlating the school rankings, particularly when using the same outcome measures and model specifications. The coefficients were as expected. When using the Gain Score Model, the Spearman rank correlation coefficients between the school rankings based on the mathematics and reading outcome measures ranged from .07 (Grades 4 to 5) to .51 (Grades 3 to 4).

When using the Layered Effects Model, the Spearman rank correlation coefficients between the school rankings based on mathematics and reading outcome measures ranged from .17 (Grade 7) to .78 (Grade 6). In the first elementary school grade level (Grade 3) and the first middle school grade level (Grade 6), school ranking correlations were moderately high, indicating that schools' rankings in the mathematics outcome measure tended to correspond fairly well with their rankings in the reading outcome measure. This relationship tended to decrease gradually in Grades 4 and 5. In

the middle school grades, the opposite pattern emerged where these school rankings were low in Grade 7, but emerged to be moderately low in Grade 8.

School Ranking Relationship by Model Specification

Answering Question 4 required an examination of the relationship between school rankings based on the Gain Score and the school rankings based on the Layered Effects Model, first for the mathematics outcome measures, then for the reading outcome measures. When using the mathematics outcome measure, the Spearman rank correlation coefficient between the school rankings from the Gain Score Model and the school rankings from the Layered Effects Model ranged from .79 (Grade 4) to .96 (Grade 5). When using the reading outcome measure, the Spearman rank correlation coefficient between the school rankings from the Gain Score Model and the school rankings from the Layered Effects Model ranged from .76 (Grade 4) to .93 (Grade 7). This seemed to indicate that schools with "high" school rankings based upon gain scores, also had "high" school rankings based upon the average school performance in the mathematics and reading outcome measures.

The results in this study also demonstrated that the value added school effect estimates were more sensitive to outcome measures than they were to the model specifications. This pattern emerged when examining school effects and school rankings using the Gain Score and Layered Effects Models with mathematics and reading outcome measures. This seemed to be the pattern in previous research studies as well (Briggs & Weeks, 2011; Schmitz, 2007; Tekwe et al., 2004). Estimates of school effects will likely be similar, regardless of the model specification; however, the school effects would also

likely be more sensitive to outcome measures. Therefore, determining which school is more or less effective depends more upon the academic subjects (e.g., outcome measures) examined rather than the models used to generate the school effect estimates.

Limitations of the Study

This study was limited in three ways that may restrict generalization to different populations. First, the study was limited by a majority-transient student sample that progressed through elementary and middle school grades. These students included those who entered the district after 2005 and/or left the district sometime between 2005 and 2010. These were students who had one or more test scores missing, indicating that the student did not test in the district, tested in another district, or moved from another state. The degree to which this transience was more frequent in one or more race/ethnic groups was not examined. Only about 40% of the students in the sample had mathematics and reading scores for the six years examined. The effects of the school on student performance and growth may be more difficult to estimate, given the challenge of continual adjustments in school populations. Examining the performance and growth of all students for the six years may allow for more accurate estimates of the schools effects on student academic performance and growth.

A second limitation of this study was the utilization of models with only two levels. The levels examined were the students at level-1 and the schools at level-2, omitting the teacher level. According to Briggs and Weeks (2011), when only school effects are included without teacher effects, the school effects are likely to represent an aggregation of teacher effects of student achievement, but they are also likely to capture

the influence of administrative leadership and policies that might fall under the "school climate" heading. In addition, the extent to which estimated school effects are biased when teacher effects have been omitted is still unclear and there have been no studies examining both teacher and school effects in one value added model (Briggs & Weeks, 2011). Anytime aggregation occurs, information is lost and, in this case, the degree of variance between teachers within schools is lost in the school effects. Due to limited research, the degree to which including the teacher level impacts the school effects is unclear.

A third limitation of the study was the utilization of only two model specifications. This study utilized the Gain Score Model and the more complex Layered Effects Model to obtain estimates and rankings of the schools. There were a variety of models that could be used to estimate value added school effects and rankings as indicated in Table 1. The findings of this study implied that school effects were not as sensitive to model specifications as they were to outcome measures. However, the inclusion of covariates may provide another perspective of the schools' effect upon students' academic performance and growth. Depending upon the research goals, some models may be better suited than others.

Directions for Future Research

Many of the studies examining value added effects use end-of-the-year state achievement exams (i.e., summative assessments) as the outcome measure of interest. As found with this study, different outcome measures can provide different perspectives of school effectiveness. Future studies may consider other content area outcomes measured

by end of semester exams (e.g., science) and the use of composite scores (e.g., mathematics combined with science and/or reading, etc.) to determine to what extent if any, the school effects differ from using other measures or combinations. Though summative assessments have been the predominant outcome measure in the previous research, the use of formative assessments in the estimation of value added school effects should be further investigated. Schools likely vary in their effectiveness at different points in time during the same school year and even across years due to a variety of factors (e.g., principal and teacher turnover). These types of results may be useful for diagnostic purposes to support more targeted school improvement efforts.

The findings from this study seemed to indicate that the model specification has less impact on school effects and school rankings than do the outcome measures; however, some models may be better than others depending upon the context. Of the two models used in this study, the Gain Score Model was easiest to implement and obtain estimates. The Z-matrix utilized in the Layered Effects Model added a degree of complexity that may or may not be warranted given the findings of this current study and previous studies. The Gain Score and Layered Effects Models for this current study demonstrated similar results related to school effects and school rankings. The fairly homogenous population may have had some influence on these findings.

Future research may consider examining school effects from these models using more diverse populations. Reardon and Raudenbush (2008) noted the most significant finding of their study was the importance of modeling heterogeneity in effects. Researchers may investigate further if schools' effects are homogenous for all populations (e.g., SES, race, gender); meaning, do the models provide school effect

estimates that are the same for all subgroups within a school or are school effects differential based upon the subgroups? The ability for schools and teachers to have the same effect on all students regardless of their race/ethnicity, gender, SES, or other characteristics seems unlikely, but is still questionable until further investigated.

The use of value added estimates (i.e., quantitative data) as the sole measure in making decisions about schools' effectiveness can be problematic and may lead to decisions based on inaccurate information. Future research may want to go a step further to examine the qualitative characteristics of schools (e.g., school culture) and whether these characteristics would be sustained or vary over time. Mixed method studies (qualitative and quantitative) examining school effects may provide additional credibility for, and/or against, the use of school effect estimates to evaluate school effectiveness.

Closing Remarks

Educational accountability emphasizes holding states, districts, schools, educators, and students responsible for student academic achievement (Education Week, 2004). Status and Value Added Models are the common types of educational accountability models in place today that are used to measure student academic achievement and evaluate schools. Both of these types of models provide a unique perspective on academic achievement. Status Models provide estimates of the percentage of students who are proficient in a given year. Value Added Models provide estimates of the academic achievement gains that students make in one or more years and attempt to identify the unique contributions of districts, schools, and teachers towards students'

academic achievement. Both models have strengths and weaknesses, but to date there is no one perfect educational accountability model.

Value added modeling has been implemented as a part of several states' and districts' education accountability systems as a means to evaluate the effects of districts, schools, and teachers on students' academic performance and growth. There are arrays of value added model specification and outcome measure choices that may be used in estimating school effects for educational accountability systems. However, the findings of this study and the previous research studies of Briggs and Weeks (2011), Lockwood et al. (2007), Schmitz (2007), and Tekwe et al. (2004) using a variety of VAMs seem to indicate that the value added effect estimates are less sensitive to model specifications than they are to the outcome measures used in generating the estimates. Therefore, determining which school is more or less effective seems to depend more upon academic subjects examined, rather than the models used to generate the school effect estimates.

Value Added Models provide an additional perspective that can be used to support diagnostic conclusions; however, depending upon the outcome measure used, school effect estimates can vary significantly. Given the considerable variability in the outcome measure correlations found in this study and in other studies (e.g., Briggs & Weeks, 2011), these models should not be used to make high-stakes decisions (i.e., school and teacher evaluations). It is likely that the effect schools have on student academic performance is largely dependent upon the context, with the context being a combination of principal leadership, teacher quality, student populations, and parent engagement among other factors. Researchers should continue to investigate the accuracy of the various models and methods used to evaluate schools. These models and methods

should be fair, scientifically based, and should consider context before making final judgments and policy decisions.

References

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value added assessment of teachers. *Journal of Educational and Behavioral Statistics*, *29*(1), 37-65.

Beardsley, A. A. (2008). Methodological concerns about the education Value Added Assessment System. *Educational Researcher*, *37*(2), 65-75.

Briggs, D. C., & Weeks, J. (2007). *The persistence of value added school effects*. Paper presented at the 2008 annual meeting of the American Educational Research Association, Division D, New York, NY. March 27, 2008. Retrieved June 8, 2009, from http://www.colorado.edu/education/faculty/derekbriggs/presentations.html.

Briggs, D.C., Weeks, J., & Wiley, E. (2008). *The sensitivity of value added modeling to the creation of a vertical score scale*. Retrieved June 9, 2009, from http://www.wcer.wisc.edu/news/events/natConf_papers.php.

Briggs, D. C., & Weeks, J. (2011). The persistence of school-level value-added. *Journal of Educational and Behavioral Statistics*, *36*(5), 616-637.

Choi, K., Seltzer, M., Herman, J., & Yamashiro, K. (2007, Fall). Children left behind in AYP and non-AYP schools: Using student progress and the distribution of student gains to validate AYP. *Educational Measurement: Issues and Practice, 26*(3), 21-32.

Cody, R., & Smith, J. (2006). Describing data. In S. Yagan (Ed.), *Applied statistics and the SAS programming language* (5th ed.). Upper Saddle River, NJ: Elsevier Science Publishing Co., Inc.

Doran, H., & Izumi, L. (2004). *Putting education to the test: A value added model for California*. Retrieved September 17, 2009, from http://www.heartland.org/policybot/results/15626/Putting_Education_To_The_Test_A_ValueAdded_**Model**_For_California.html.

Education Week. (2004). *Accountability.* Retrieved June 21, 2012, from http://www.edweek.org/ew/issues/accountability/.

Florida Comprehensive Assessment Test. (2008). *Understanding FCAT Reports 2008,* Tallahassee. State of Florida, Department of State.

Florida Department of Education. (2004). *Assessment and Accountability Briefing Book*. Retrieved August 17, 2009, from http://fcat.fldoe.org/pdf/fcataabb.pdf.

Florida Department of Education. (2007). *Assessment and Accountability Briefing Book*. Retrieved August 17, 2009, from http://www.fldoe.org/asp/.

Florida Department of Education. (2011). *Florida Public High School Graduation Rates 2010-2011*. Retrieved February 23, 2012, from http://www.fldoe.org/eias/eiaspubs/word/GradRate-1011.doc.

Forrester, J., Breierova, L., & Choudhari, M., (2001). *An introduction to sensitivity analysis*. Retrieved July 31, 2012, from http://ocw.mit.edu/courses/sloan-school-of-management/15-988-system-dynamics-self-study-fall-1998-spring-1999/readings/sensitivityanalysis.pdf.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., & Williams, A. (2005). *Policymakers guide to growth models for school accountability: How do accountability models differ*? Retrieved May 1, 2009, from http://www.ccsso.org/publications/details.cfm?publicationid=287.

Harcourt Brace Educational Measurement. (1997). *Stanford achievement test series: Technical data report* (9th ed.). San Antonio: Harcourt Brace and Company.

Hershberg, T. (2005, February). *Value added assessment and systematic reform: A response to the challenge of human capital development*. Paper presented at the Aspen Institute's Congressional Institute on the Challenge of Education Reform, Standards, Accountability, Resources and Policy. Retrieved August 1, 2010, from http://www.cgp.upenn.edu/ctr_pubs.html.

Hershberg, T. (2005). Value added assessment and systematic reform: A response to the challenge of human capital development. *Phi Delta Kappan*, *87*(4), 276-283. Retrieved August 1, 2010, from http://www.cgp.upenn.edu/ctr_pubs.html.

Kingsbury, G. G., & McCall, M. S. (2006). The hybrid success model: Theory and practice. In R.Lissitz (Ed.), *Longitudinal and Value Added Models of student performance*. Maple Grove, MN: JAM Press.

Lockwood, J., Doran, H.C., & McCaffrey, D.F. (2003). Using R for estimating longitudinal student achievement models. *The Newsletter of the R Project*, *3*(3), 17-23.

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V. N., & Martinez, J. F. (2007). The sensitivity of value added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, *44*(1), 47-67.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value added models for teacher accountability*. Retrieved February 22, 2012, from http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf.

McCaffrey, D. F., Lockwood, J., Loretz, D., Louis, T., & Hamilton, L. (2004). Models for value added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*(1), 67-101.

Orr, C. (2007). *Grade 3 FCAT construction and equating*. Retrieved August 17, 2009, from http://fcat.fldoe.org/fcatadvisory/ppt/FCATGrade3DistrictAdvisory.ppt.

Ponisciak, S., & Bryk, A. S. (2005). Value added analysis of the Chicago public schools: An application of hierarchical models. In R. Lissitz (Ed.), *Value Added Models in Education Theory and Applications*. Maple Grove, MN: JAM Press.

Raudenbush, S.W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics*, *18*(4), 321-349.

Raudenbush, S. W. (2004). *What are value added models estimating and what does this imply for statistical practice?* Retrieved July 13, 2012 from: http://www.ssicentral.com/hlm/techdocs/value-added04.pdf.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models*: *Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Willms, J. D. (1995). *The estimation of school effects*. Retrieved July 13, 2012 from: http://www.unb.ca/crisp/pdf/9501.pdf.

Reardon, S. F., & Raudenbush, S. W. (2008). *Assumptions of Value Added Models for estimating school effects*. Paper presented at the National Conference on Value Added Modeling, Madison, Wisconsin on April 22-24, 2008. Retrieved February 22, 2012, from http://www.wcer.wisc.edu/news/events/natConf_papers.php.

Rubin, D., Stuart E., & Zanutto, E. (2004). A potential outcome's view of value added assessment in education. *Journal of Educational and Behavioral Statistics*, *29*, 101-116.

Sanders, W. L., & Horn, S. (1994). The Tennessee Value Added Assessment System: Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299-311.

Sanders, W. L., & Horn, S. (1998). Research findings from the Tennessee Value Added Assessment (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*(3), 247-256.

Sanders, W. L., Saxton, A., & Horn, S. (1997). The Tennessee Value Added Assessment System: A quantitative, outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press.

Schmitz, D. D. (2007). An empirical sensitivity analysis of value added teachers' effect estimates to hierarchical linear model parameterizations (Doctoral dissertation, University of Northern Colorado, 2007). *Dissertations & Theses: Full Text database* (Publication No. AAT 3280271).

Schmitz, D.D., & Raymond, K.J. (2008). *The utility of the cumulative effects model in a statewide analysis of student achievement*. Paper presented at the American Educational Research Association Annual Meeting, New York, NY on March 27, 2008.

Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M., Roth, J., . . .Resnick, M. (2004). An empirical comparison of statistical models for value added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29(1), 11-36.

Toppo, G. (2006, June 20). Big-city schools struggle with graduation rates. *USA Today*. Retrieved August 15, 2009, from http://www.usatoday.com/news/education/2006-06-20-dropout-rates_x.htm.

United States Census Bureau. (2010). *Demographic Profiles: Census 2010*. Retrieved July 22, 2012, from http://www.census.gov/compendia/statab/2012/tables/12s0016.pdf.

United States Census Bureau. (2012). *State and county quick facts: Florida*. Retrieved February 23, 2012, from http://quickfacts.census.gov/qfd/states/12000.html.

United States Department of Education. (n.d.). *Elementary and secondary education: A blueprint for reform*. Retrieved September 18, 2009, from http://www2.ed.gov/policy/elsec/leg/blueprint/publication_pg4.html.

United States Department of Education. (1983). *A nation at risk: The imperative for educational reform*. Retrieved September 18, 2009, from http://www2.ed.gov/pubs/NatAtRisk/index.html.

United States Department of Education. (2002). *Elementary & secondary education: Key policy letters signed by the education secretary or deputy secretary*. Retrieved September 18, 2009, from http://www.ed.gov/policy/elsec/guid/secletter/020724.html.

United States Department of Education. (2009). *Press releases: Secretary Spellings approves additional growth model pilots for 2008-2009 school year*. Retrieved May 1, 2009, from http://www.ed.gov/news/pressreleases/2009/01/01082009a.html.

United States Department of Education. (2009). *Elementary & secondary education: A blueprint for reform, the reauthorization of the Elementary and Secondary Education Act*. Retrieved May 7, 2010, from http://www2.ed.gov/policy/elsec/leg/blueprint/publicationtoc.html.

United States Department of Education. (2010). *Elementary & secondary education: Executive summary*. Retrieved July 27, 2010, from http://www2.ed.gov/nclb/overview/intro/execsumm.html.

Wang, Z. (2006). Comparison of value added models for school ranking and classification: A Monte Carlo study. *Proquest Dissertations and Theses 2006*, *803*(0288). (Publication Number: AAT 3296572).

Wiley, E. W. (2006). *A practitioner's guide to value added assessment: Educational policy studies laboratory research monograph*. Retrieved February 22, 2012, from http://nepc.colorado.edu/publication/a-practitioners-guide-value-added-assessment-educational-policy-studies-laboratory-resea.

Wisconsin Policy Research Institute Inc. (2009, October 12). *Value added testing: Improving state testing and teacher compensation in Wisconsin*. Retrieved October 12, 2009, from http://www.wpri.org/Reports/Volume22/Vol22No4/Vol22No4.html.

Wright, S., Horn, S., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67.

Appendices

Appendix A: SAS Code for Gain Score and Layered Effects Models

Gain Score Model

```
title'Gain Score Model_4th to 5th Math';
proc mixed data=perm.bryce3 noclprint;
class id schlid_es1;
model MGS_4_5=/solution ddfm=bw;
random int/sub=schlid_es1 solution rcorr;
run;
```

Layered Effects Model

```
title'Layered Effects Model';
proc mixed data=perm.bryce3 method=REML scoring=100 convh=10E-4 update
noclprint;
class id;
model math= t0 t1 t2 t3 t4 t5/noint solution;
random z0_1-z0_38/type=toep(1)solution;
random z1_1-z1_38/type=toep(1)solution;
random z2_1-z2_53/type=toep(1)solution;
random z3_39-z3_53/type=toep(1)solution;
random z4_39-z4_53/type=toep(1)solution;
random z5_39-z4_53/type=toep(1)solution;
repeated/Type=un sub=id r rcorr;
run;
```

Appendix B: Cross-Classification of Students in Elementary and Middle School

Table B1

Cross-Classification of Students in Elementary and Middle School

	Middle School															
	NA	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
NA		78	56	21	24	193	87	158	63	36	71	53	104	59	132	42
1	35	1	1	31	1	3	NA	NA	1	NA	NA	1	21	1	NA	NA
2	26	109	1	NA	NA	4	1	1	1	NA	NA	1	1	1	NA	NA
3	49	2	NA	1	2	45	1	71	1	NA	2	NA	4	1	12	1
4	49	NA	NA	NA	7	19	NA	NA	89	1	NA	NA	4	1	3	1
5	43	1	1	58	33	1	1	NA	3	NA	NA	2	5	NA	NA	1
6	49	7	12	NA	NA	1	11	2	1	2	70	6	NA	6	2	2
7	65	3	67	NA	NA	1	2	NA	1	2	2	15	1	5	2	4
8	68	3	1	NA	1	NA	93	4	NA	2	6	3	NA	2	NA	2
9	12	3	NA	NA	NA	3	1	96	3	1	NA	NA	NA	NA	5	1
10	17	NA	3	NA	NA	1	4	2	NA	NA	2	63	NA	5	1	8
11	9	85	NA	NA	NA	NA	4	1	1	NA	1	1	NA	2	NA	NA
12	47	NA	2	5	13	5	2	1	121	1	NA	1	5	1	1	NA
13	64	1	NA	3	34	3	1	1	88	NA	NA	NA	NA	NA	NA	2
14	36	1	1	1	1	1	1	12	2	3	1	1	5	1	87	NA
15	48	12	11	NA	NA	NA	54	NA	NA	2	52	4	NA	3	NA	5
16	22	NA	NA	88	17	2	2	1	2	NA	NA	NA	6	NA	NA	2
17	47	2	33	1	NA	1	2	1	NA	1	20	3	1	4	NA	2
18	58	3	5	NA	1	NA	2	NA	1	7	4	17	NA	8	2	82
19	36	11	20	NA	NA	NA	4	NA	NA	2	48	5	1	7	2	6
20	31	1	NA	43	2	NA	NA	1	NA	1	NA	NA	3	NA	NA	NA
21	39	2	2	1	NA	NA	3	1	NA	1	1	69	1	6	1	9
22	50	3	8	1	NA	1	4	NA	1	1	8	67	1	5	NA	10
23	45	2	NA	3	1	3	4	NA	3	1	NA	NA	103	1	4	NA
24	24	NA	1	10	84	NA	NA	NA	2	NA	NA	4	2	NA	2	NA
25	34	118	2	NA	NA	2	3	2	1	3	4	NA	NA	5	NA	3
26	36	15	3	1	NA	NA	3	2	NA	NA	4	1	NA	84	1	2

Appendix B (Continued)

Table B1 (Continued)

Cross-Classification of Students in Elementary and Middle School

	Middle School															
	NA	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
27	45	1	2	NA	1	1	1	NA	1	61	NA	3	NA	6	5	14
28	51	3	3	2	3	NA	4	1	1	8	NA	3	1	3	NA	81
29	30	1	3	NA	NA	NA	2	2	NA	65	NA	3	NA	8	2	16
30	39	1	5	NA	NA	2	1	NA	NA	3	3	4	NA	99	1	5
31	32	12	2	NA	NA	1	10	2	NA	NA	47	2	1	2	NA	2
32	28	1	NA	1	NA	2	NA	20	1	1	NA	2	3	NA	102	1
33	48	1	3	NA	NA	1	70	2	NA	NA	2	3	2	NA	NA	NA
34	43	3	87	NA	NA	NA	5	1	1	2	5	5	1	6	NA	4
35	39	1	3	1	NA	NA	2	NA	NA	5	1	8	NA	35	1	35
36	41	NA	NA	NA	NA	6	NA	100	3	3	NA	1	6	1	13	1
37	48	NA	1	NA	1	146	NA	NA	1	NA	NA	NA	3	NA	2	1
38	83	NA	1	1	4	94	2	1	9	NA	2	1	117	NA	1	NA
39	20	44	NA	NA	NA	1	2	16	1	1	2	1	NA	61	NA	NA
40	22	1	1	1	1	132	1	1	4	2	NA	NA	4	NA	1	2
	NA	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55

Note. NA indicates no students in this school.

Appendix C: Scatter-Plot of Reading by Mathematics

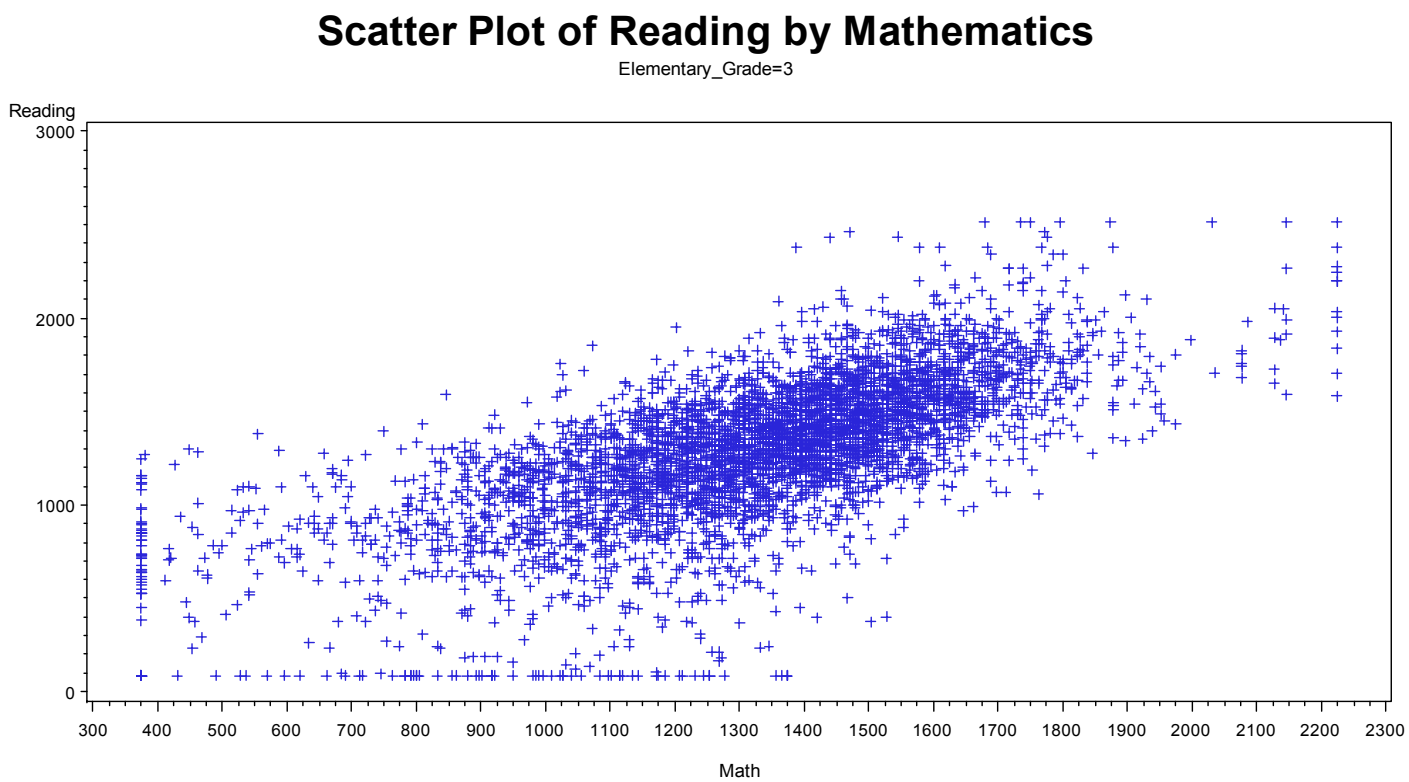

Figure C1. Scatter Plot of Mathematics by Reading for Grade 3.

Appendix C (Continued)

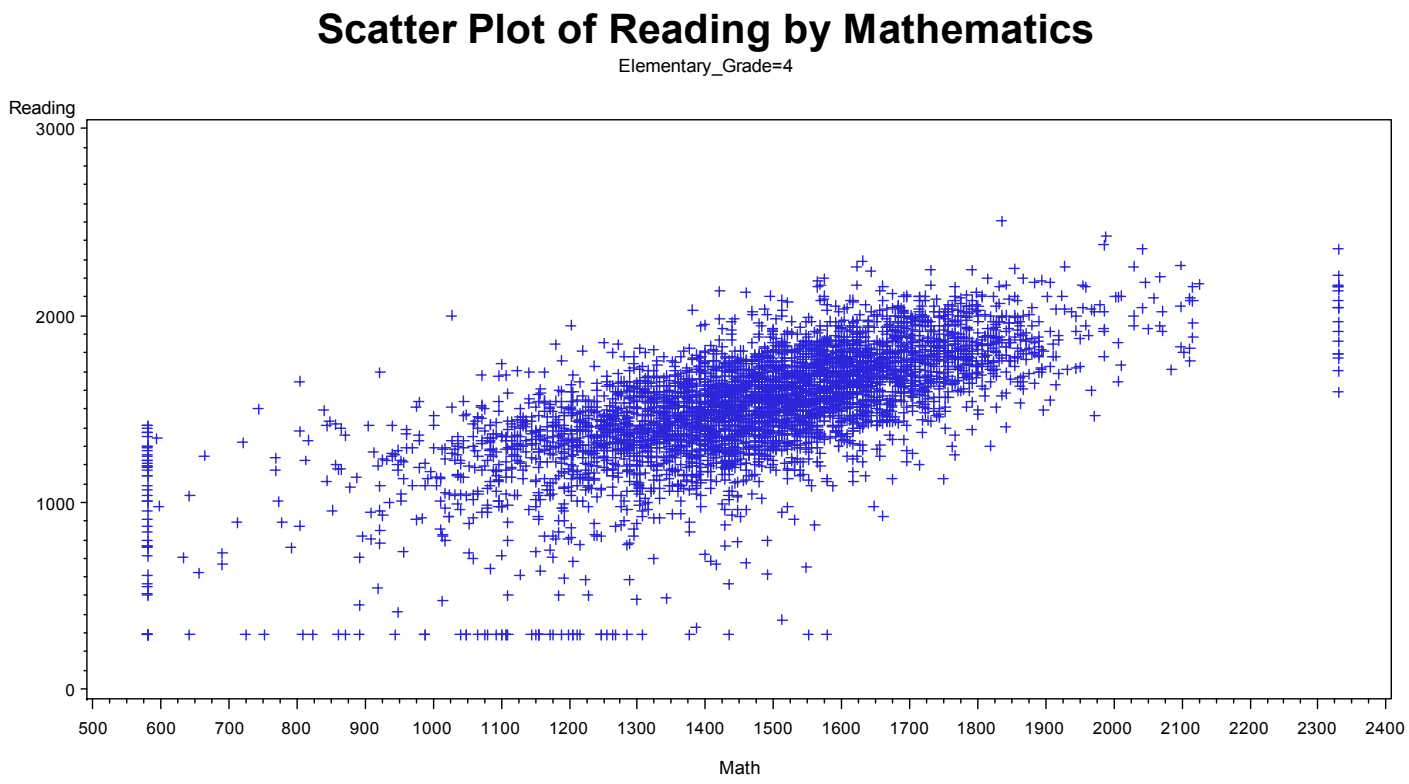

Figure C2. Scatter Plot of Mathematics by Reading for Grade 4.

Appendix C (Continued)

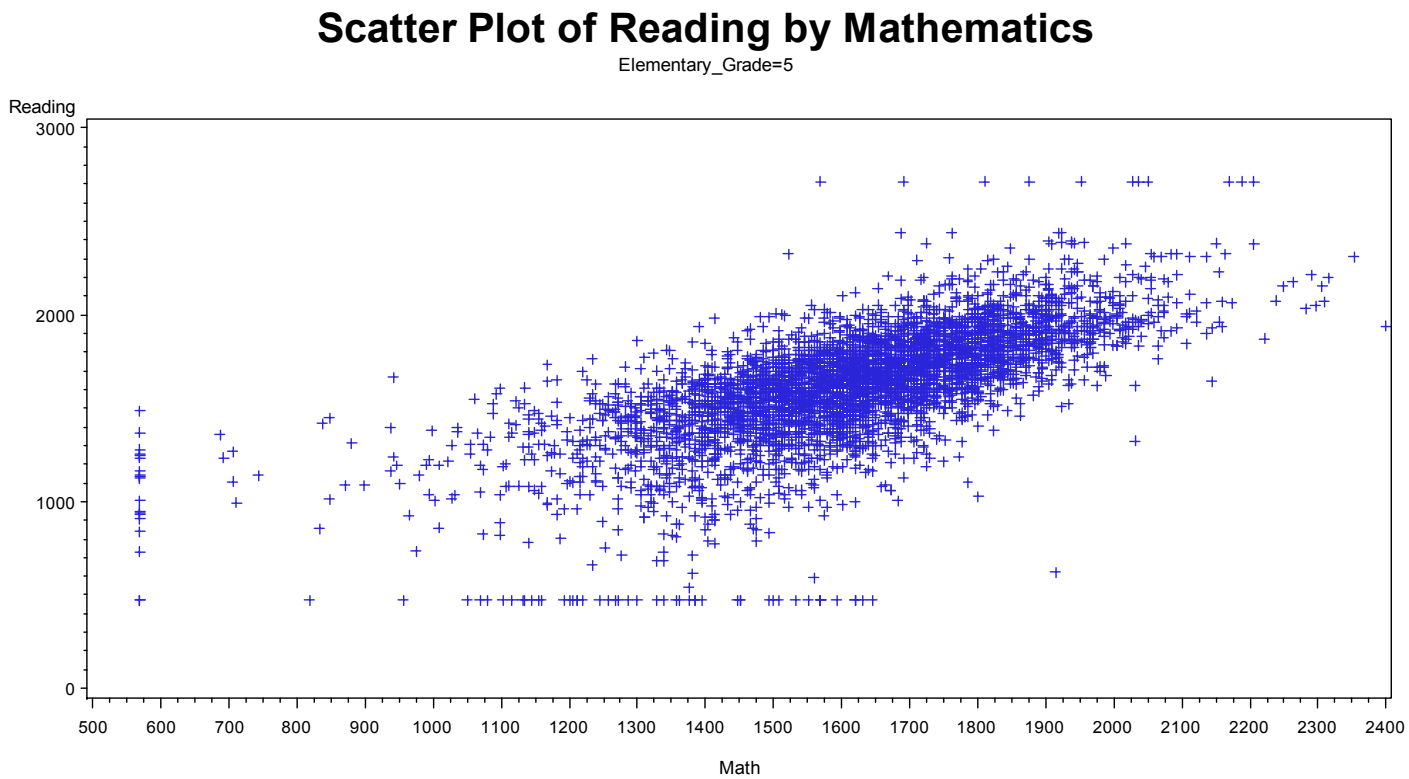

Figure C3. Scatter Plot of Mathematics by Reading for Grade 5.

Appendix C (Continued)

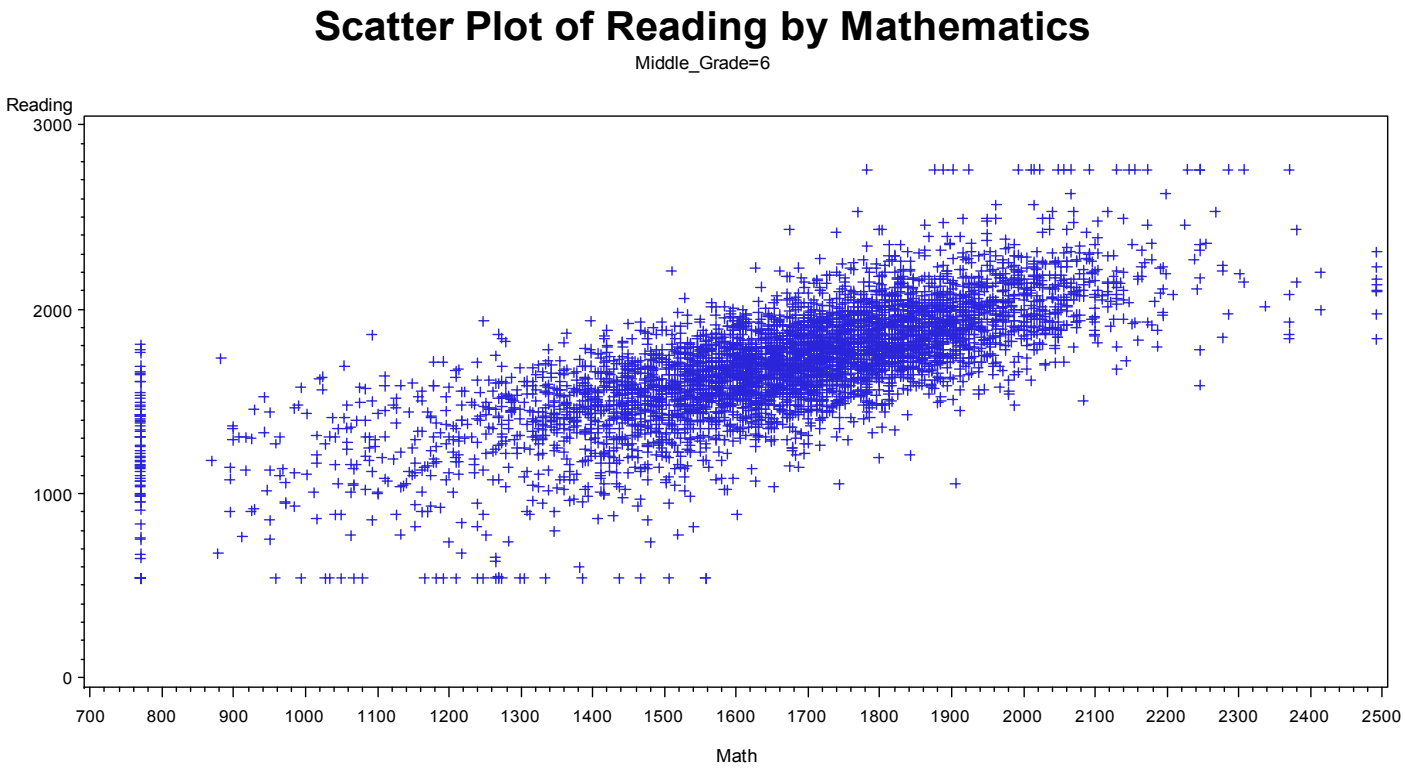

Figure C4. Scatter Plot of Mathematics by Reading for Grade 6.

Appendix C (Continued)

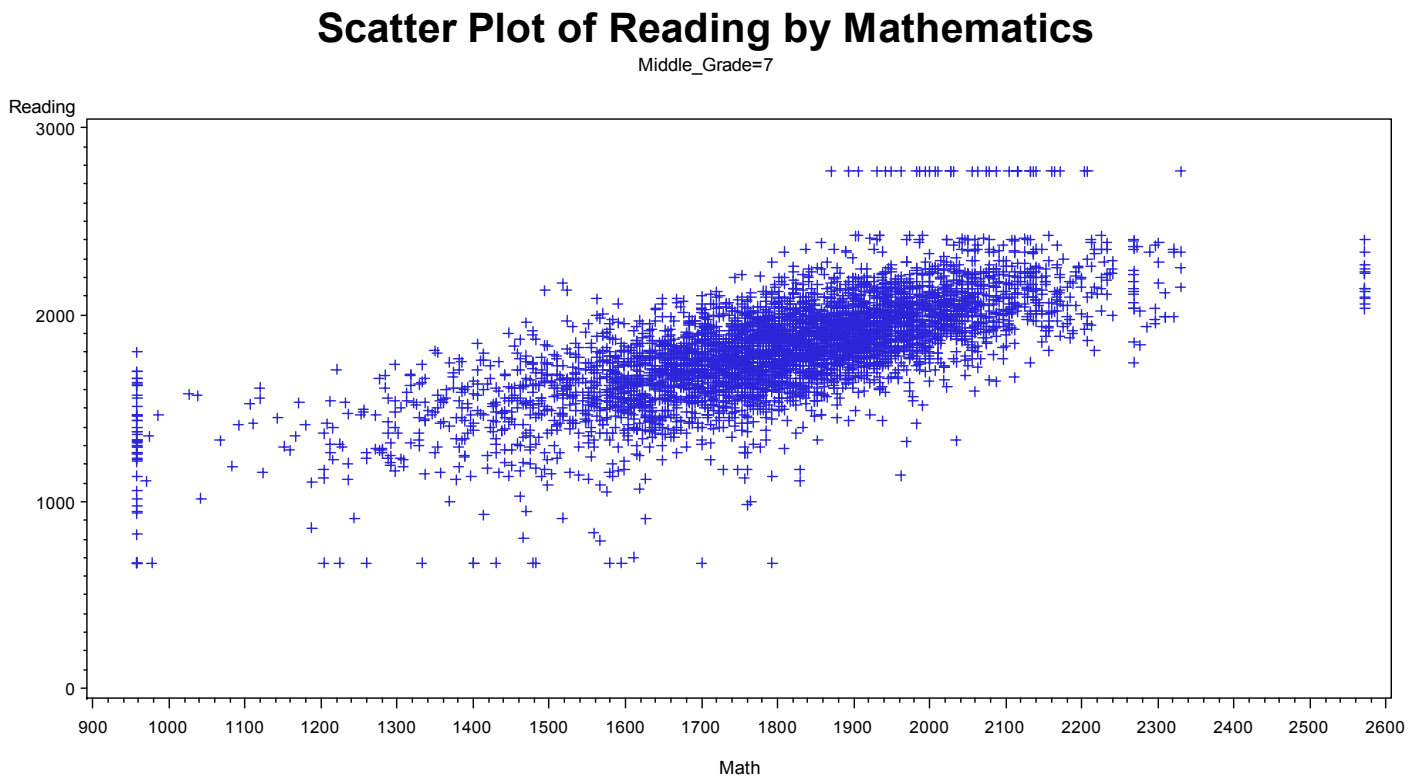

Figure C5. Scatter Plot of Mathematics by Reading for Grade 7.

Appendix C (Continued)

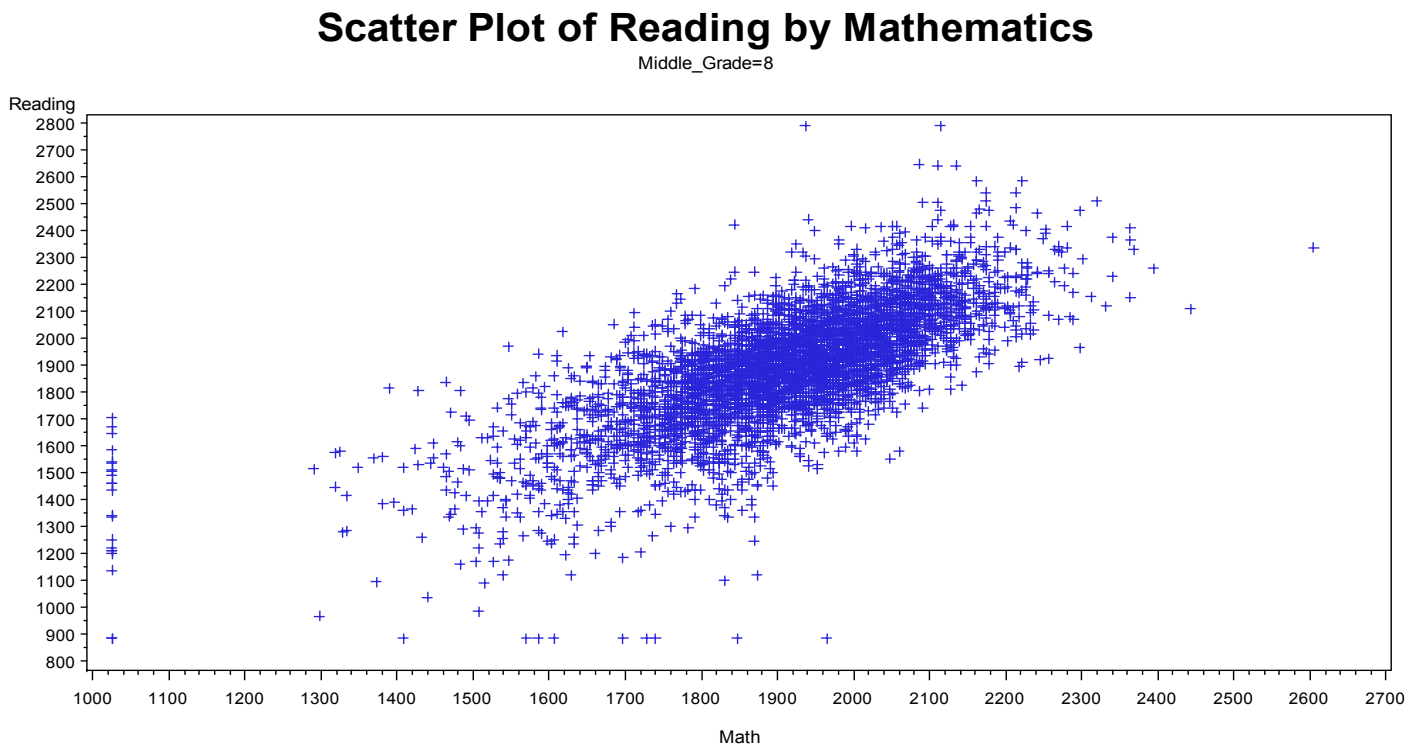

Figure C6. Scatter Plot of Mathematics by Reading for Grade 8.

Appendix D: Missing Data in Mathematics and Reading

Table D1

Number of Students with Missing Data in Mathematics and Reading DSS (N=7,496)

nmiss_all	Frequency (Percentage)	Cumulative Frequency (Percentage)
0	2761 (36.83%)	2761 (36.83 %)
1	66 (.88 %)	2827 (37.71 %)
2	617 (8.22 %)	3443 (45.93 %)
3	24 (0.32 %)	3468 (46.26 %)
4	650 (8.66 %)	4117 (54.92 %)
5	41 (0.54 %)	4158 (55.47 %)
6	772 (10.29 %)	4929 (65.76 %)
7	18 (0.24 %)	4948 (66.00 %)
8	819 (10.92 %)	5766 (76.92 %)
9	19 (0.26 %)	5786 (77.18 %)
10	1675 (22.34%)	7460 (99.52 %)
11	36 (0.48%)	7496 (100.00 %)

Note. nmiss_all indicates the number of missing test scores between mathematics and reading that a student had ranging from zero to eleven. Frequency (percentage) indicates the number and percentage of students missing test scores by nmiss_all. Cumulative frequency (percentage) indicates the cumulative frequencies and percentages of missing test scores going from zero to eleven for mathematics and/or reading.

Appendix D (Continued)

Table D2

Number of Students Missing a Mathematics Score (N=7,496)

nmissm	Frequency (Percentage)	Cumulative Frequency (Percentage)
0	2791 (37.23%)	2791 (37.23%)
1	663 (8.84 %)	3453 (46.07 %)
2	687 (9.16 %)	4140 (55.22 %)
3	805 (10.73 %)	4944 (65.96 %)
4	832(11.09%)	5776 (77.05 %)
5	1701 (22.69 %)	7476 (99.74%)
6	20 (0.26 %)	7496 (100.00 %)

Note. nmissm indicates the number of missing test scores for mathematics that a student had ranging from zero to six. Frequency (percentage) indicates the number and percentage of students missing test scores by nmissm. Cumulative frequency (percentage) indicates the cumulative frequencies and percentages of missing test scores going from zero to six for mathematics.

Table D3

Number of Students Missing a Reading Score (N=7,496)

nmissr	Frequency (Percentage)	Cumulative Frequency (Percentage)
0	2802 (37.38 %)	2802 (37.38 %)
1	650 (8.68 %)	3453 (46.06 %)
2	683 (9.11 %)	4135 (55.16 %)
3	799 (10.65 %)	4934 (65.82 %)
4	842 (11.23 %)	5775 (77.04 %)
5	1705 (22.74 %)	7480 (99.78 %)
6	16 (0.22 %)	7496 (100.00 %)

Note. nmissr indicates the number of missing test scores for reading a student had ranging from zero to six. Frequency (percentage) indicates the number and percentage of students missing test scores by nmissr. Cumulative frequency (percentage) indicates the cumulative frequencies and percentages of missing test scores going from zero to six for reading.

Table D4

Number of Students Missing a Mathematics Score by Grade (N=7,496)

Grade	Frequency (Percent)	Missing (Percent)
Elementary		
3	4748(63.33 %)	2248 (36.67 %)
4	4475 (59.70%)	3021 (40.30 %)
5	4749 (63.33 %)	2747 (36.67 %)
Middle		
6	4874 (65.02 %)	2622 (34.98 %)
7	4890 (65.23 %)	2606 (34.77 %)
8	4396 (58.65 %)	3100 (41.35 %)

Note. Grade indicates the grade levels in elementary and middle school. Frequency (percentage) indicates the number and percentage of students having a test score in mathematics by grade level. Missing (percentage) indicates the number of students missing a score in mathematics by grade.

Table D5

Number of Students Missing a Reading Score by Grade (N=7,496)

Grade	Frequency (Percent)	Missing (Percent)
Elementary		
3	4760 (63.50 %)	2736 (36.50 %)
4	4484 (59.82 %)	3012 (40.18 %)
5	4731 (63.11 %)	2765 (36.89 %)
Middle		
6	4875 (65.03 %)	2621 (34.97 %)
7	4895 (65.30 %)	2601 (34.97 %)
8	4928 (65.74 %)	2568 (34.26 %)

Note. Grade indicates the grade levels in elementary and middle school. Frequency (percentage) indicates the number and percentage of students having a test score in reading by grade level. Missing (percentage) indicates the number of students missing a score in reading by grade.

Appendix E: Violation of Model Assumptions

Violations of model assumptions can lead to misinterpretation of results and faulty inferences regarding the sample and population under investigation. There are a variety of methods that can be utilized to evaluate the integrity of models such as examining fit indices and/or checking the model assumptions. Two key model assumptions that are typically examined include the assumption that level-1 residuals are normally distributed around a mean of "0" and the homoscedacity of residuals or homogeneity of variance. Both assumptions were examined to determine if the data were consistent with the model assumptions. The normality assumption was evaluated by examining box and whisker plots and skewness and kurtosis values.

The level-1 residuals in mathematics for the Gain Score Model were found to be fairly normally distributed around the mean values for each point in time and the skewness and kurtosis values indicated weak to high skewness and highly leptokurtic peakedness. Homoscedacity was examined from plots of the level-1 residuals against the predicted values for the outcome measures. Figure E1 displays the level-1 residuals in mathematics for each grade change period. The level-1 residuals for mathematics indicated that the test scores were fairly homogenous at each grade and across schools with no test score seeming to have an enormous influence on the results based on the influence diagnostics that were also examined.

The test scores in reading for the Gain Score Model were found to be fairly normally distributed around the mean values for each point in time and the skewness and kurtosis values indicated slight to high skewness and high peakedness. Homoscedacity was examined from plots of the level-1 residuals against the predicted values for the outcome measures. The level-1 residuals for reading in Figure E2 indicated that the test scores were fairly homogenous at each grade and across schools with no test score seeming to have a large influence on the results based on the influence diagnostics that were also examined.

The skewness and kurtosis values for mathematics using the Layered Effects Model were moderate to highly negatively skewed and slight to highly leptokurtic, respectively. Outliers were found at each grade level and extreme scores in Grades 5, 7, and 8. All of the scores were in the acceptable range and were utilized for all analyses. Homoscedacity was examined from plots of the level-1 residuals against the predicted values for the outcome measures. Figure E3 displays the level-1 residuals in mathematics for Grades 3 through 8. The level-1 residuals for mathematics indicated that the test score residuals were fairly homogenous at each grade and no test score seemed to have a huge influence on the results based on the influence diagnostics that were also examined.

The skewness and kurtosis values for reading using the Layered Effects Model indicated a fairly normal distribution with slight to moderately negative skewness and slight to moderately positive kurtosis. Outliers were found at each grade and extreme scores were found in Grades 7 and 8. All of the scores were in the acceptable range and were utilized for all analyses. Homoscedacity was examined from plots of the level-1 residuals against the predicted values for the outcome measures. Figure E4 displays the

level-1 residuals in reading for Grades 3 through 8. The level-1 residuals for reading indicated that the test scores were fairly homogenous at each grade and no test score residuals seemed to have a large influence on the results based on the influence diagnostics that were also examine

Appendix E (Continued)

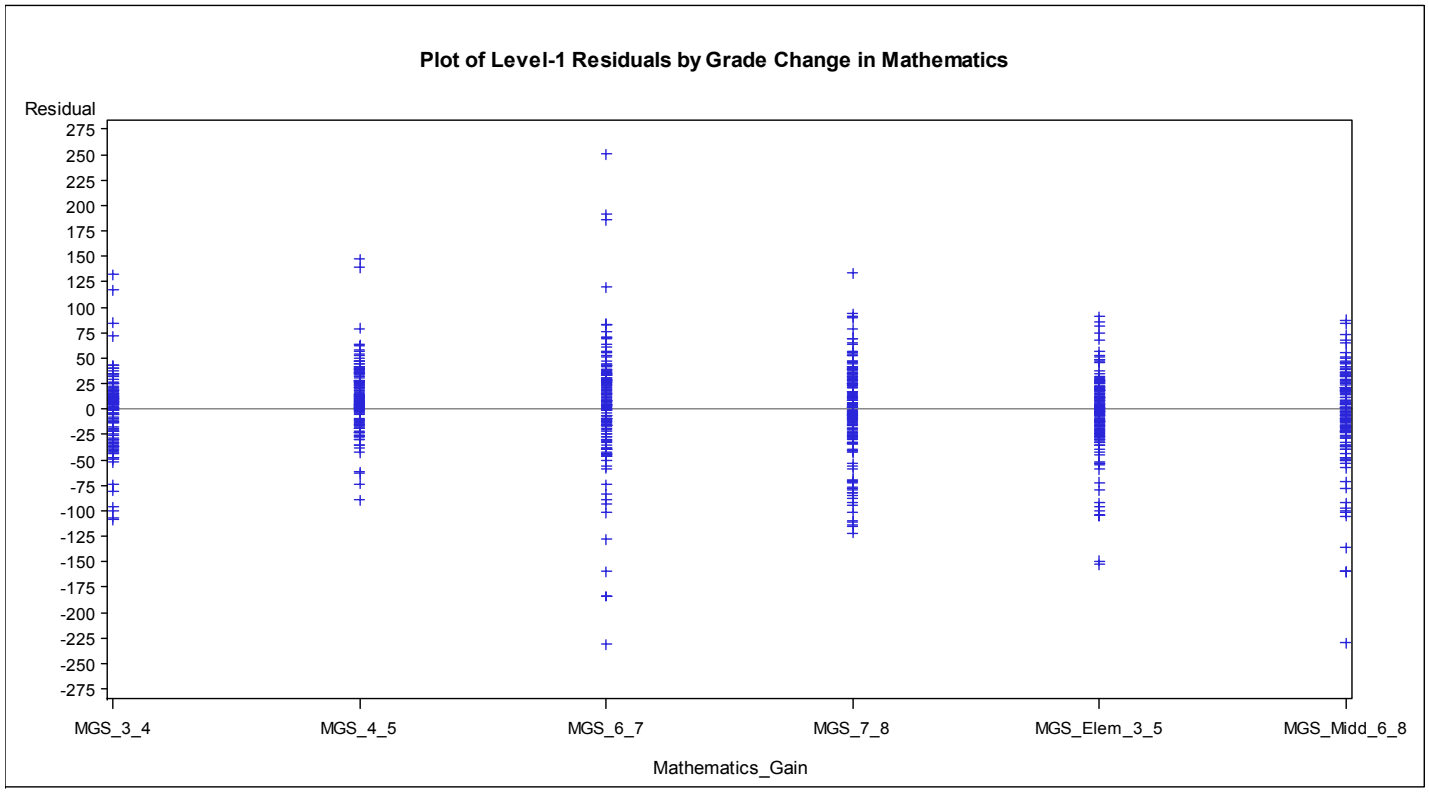

Figure E1. Plot of Level-1 Residuals by Grade Change in Mathematics. MGS indicates mathematics average gain within the indicated grades (e.g., MGS_3_4 indicates the residual average gain between Grade 3 and Grade 4). The grand average gain score for a subset of elementary and middle school students is denoted by "0." The grand average is the average change of the subset of elementary and middle school students between the respective grades. The "+" above and below the "0" indicates the deflection of the respective subset of students from the grand average change score.

Appendix E (Continued)

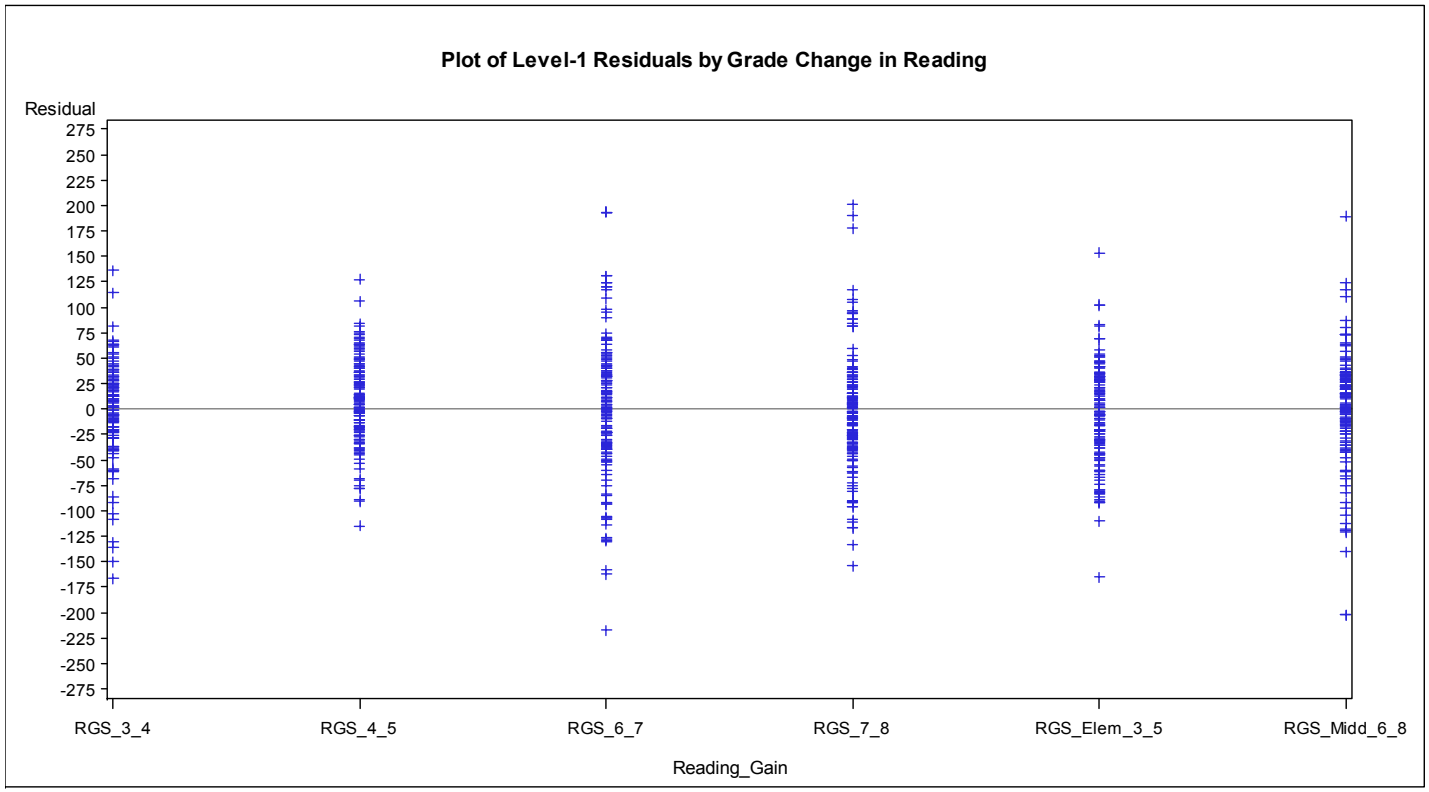

Figure E2. Plot of Level-1 Residuals by Grade Change in Reading. RGS indicates mathematics average gain within the indicated grades (e.g., RGS_3_4 indicates the residual average gain between Grade 3 and Grade 4). The grand average gain score for a subset of elementary and middle school students is denoted by "0." The grand average is the average change of the subset of elementary and middle school students between the respective grades. The "+" above and below the "0" indicates the deflection of the respective subset of students from the grand average change score.

Appendix E (Continued)

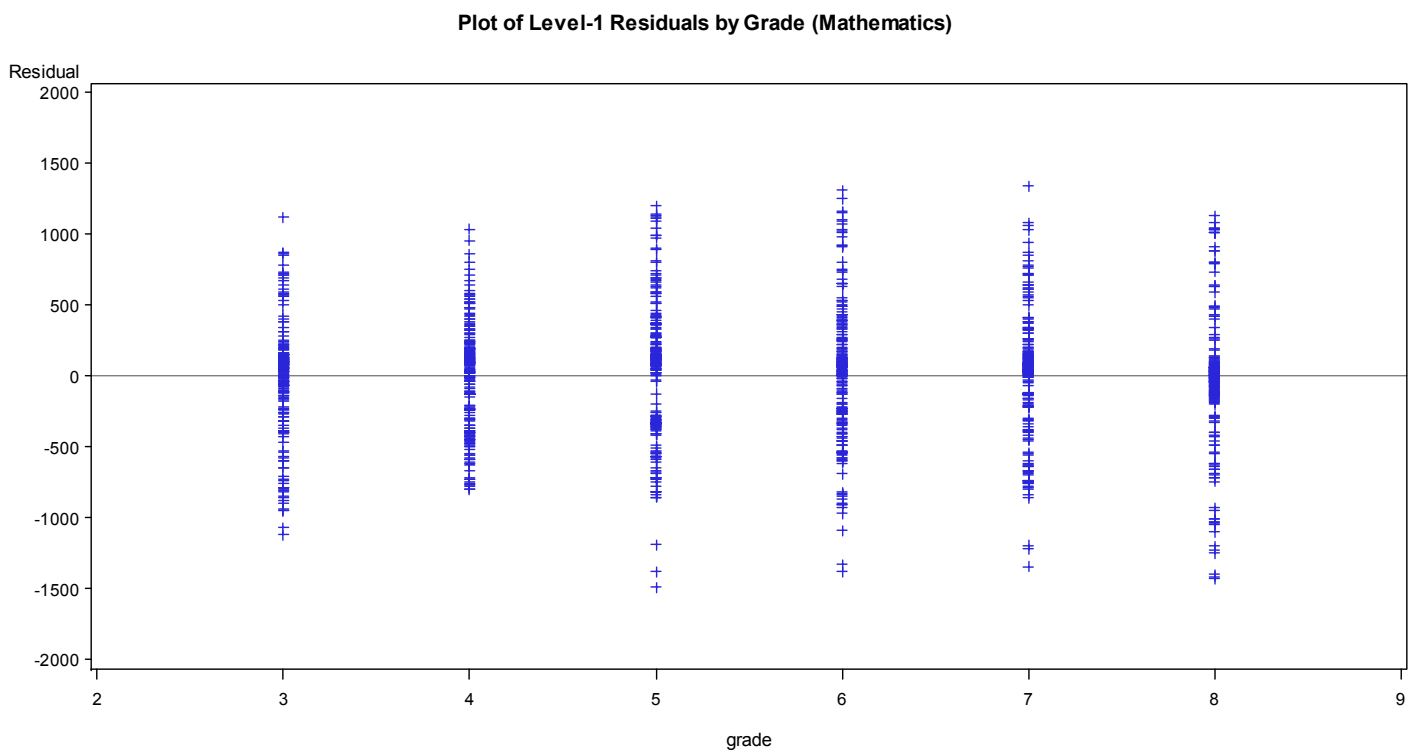

Figure E3. Plot of Level-1 Residuals in Mathematics by Grade. The horizontal axis indicates the grade and the vertical axis is the residual. The "0" line indicates the grand average score in mathematics at the specified grades for a subset of students. The grand average is the average score of the elementary and middle school subset of students at the respective grades. The "+" above and below the "0" indicates the deflection of the respective students from the grand average score.

Appendix E (Continued)

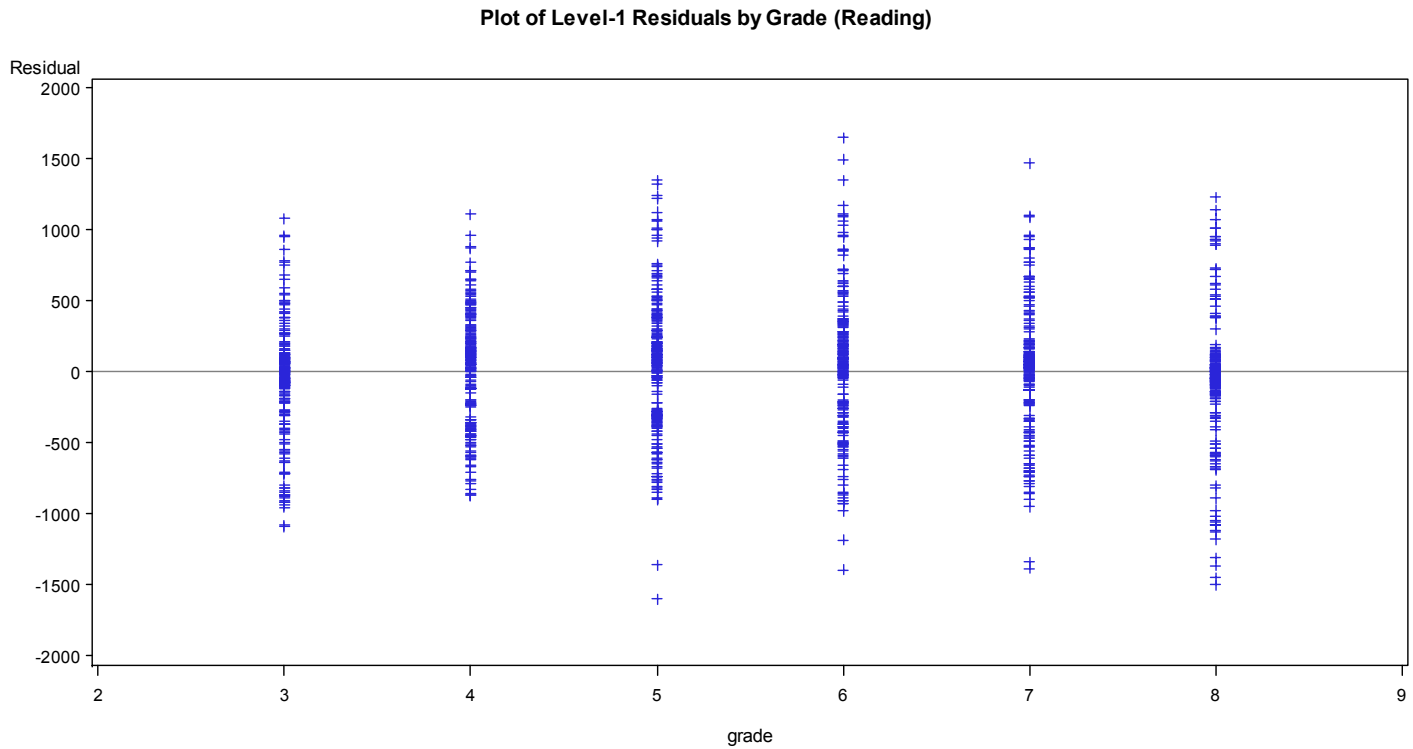

Figure E4. Plot of Level-1 Residuals in Reading by Grade. The horizontal axis indicates the grade and the vertical axis is the residual. The "0" line indicates the grand average score in reading at the specified grades for a subset of students. The grand average is the average score for the subset of students by grade (e.g., 3rd, 4th, 5th, and 6th, 7th, and 8th). The "+" above and below the "0" indicates the deflection of the respective students from the grand average score.

About the Author

Bryce L. Pride received his B.S. in Biology from Austin Peay State University with a minor in Chemistry and a M.Ed. in Curriculum and Instruction from Tennessee State University. He received his Ph.D. in Educational Measurement and Evaluation from the University of South Florida. His research interests include survey design, hierarchical linear modeling, value added modeling, data mining/analytics techniques and applications, and program evaluation. Bryce currently works as the Supervisor of Research and Evaluation in a school district in Florida and is a consultant to not-for-profit and for-profit organizations seeking research and evaluation services.

Formerly, Bryce worked as a teacher, manager of the Tennessee Comprehensive Assessment Program (TCAP), and a manager and evaluator of various programs such as GEAR UP, School Choice, College of Education Mini Grants, EVIA marketing research, and MOSTI programming for alternative science teacher certification. He has delivered numerous academic presentations and his memberships have included the American Educational Research Association (AERA), American Evaluation Association (AEA), Florida Educational Research Association (FERA), and the National Council for Measurement in Education (NCME).