University of New Orleans

ScholarWorks@UNO

University of New Orleans Theses and Dissertations

Dissertations and Theses

5-20-2005

A Value-Added Analysis of Teacher Effects on Student Achievement

Ellen Lusco
University of New Orleans

Follow this and additional works at: https://scholarworks.uno.edu/td

Recommended Citation

Lusco, Ellen, "A Value-Added Analysis of Teacher Effects on Student Achievement" (2005). *University of New Orleans Theses and Dissertations*. 265.
https://scholarworks.uno.edu/td/265

This Dissertation is protected by copyright and/or related rights. It has been brought to you by ScholarWorks@UNO with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Dissertation has been accepted for inclusion in University of New Orleans Theses and Dissertations by an authorized administrator of ScholarWorks@UNO. For more information, please contact scholarworks@uno.edu.

A VALUE-ADDED ANALYSIS OF TEACHER EFFECTS ON STUDENT ACHIEVEMENT

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Educational Administration

by

Ellen Ann Lusco

B. A. Nicholls State University, 1996
M. Ed. Nicholls State University, 1998

May, 2005

ACKNOWLEDGEMENTS

I would like to thank Dr. Peggy Kirby for her constant support in my effort to attain a doctoral degree. She made herself available, and I know the road I traveled for this degree would have been much more difficult had it not been for the dedication that Dr. Kirby gives to her students. We worked very closely on this project, and her input was invaluable. She has become more than just a professor; I can safely say that she has become a friend.

Also, a thank you is in order to my two other committee members, Dr. Oescher and Dr. Theodore, for giving invaluable feedback to improve my research design and conclusions. Dr. Oescher pushed me from the beginning to excel, and I am appreciative of his high expectations.

I also would like to thank my mom, Elodie Lusco, just for being my mom and "babysitting" on the long days when I had to work or attend night classes. Thanks Mom for your support.

Finally, a special thank you to our former superintendent, Liz Scurto, for giving me permission to use the data needed for the study and to our current superintendent, Ed Richard, for allowing the continued use of the data; Kathy Tamplain for getting me the IOWA data; and Walt Prejean for creating spreadsheets of student demographic data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The purpose of the present study was to examine the effects of teachers on student achievement using a longitudinal analysis. The analysis was based on the value-added assessment system implemented in Tennessee. In order to conduct this study, data were collected for two sets of students in one school district. The first set of students began Kindergarten in Spring 2000-Spring 2001 and continued to third grade in the Spring 2003-Spring 2004. The second set began Kindergarten in Spring 2001-Spring 2002 and went through second grade in the Spring 2003-Spring 2004 school year.

Using mixed model ANOVAs modeled after Sanders's work, data were examined in thirteen separate analyses using nine independent variables. Year-to-year language arts differences, math differences, and reading differences on the ITBS were the dependent variables.

In all of the year-to-year models, teachers were found to have a profound impact on student achievement. Prior achievement level was found to be another important factor in student achievement, with higher achieving students making consistently higher year-to-year gains than lower-achieving students.

CHAPTER ONE

INTRODUCTION

From early studies such as the Coleman Report (Office of Education, 1966) to the present studies such as value-added assessment (e. g., Sanders, Saxton, & Horn, 1997; Webster & Mendro, 1997), researchers are attempting to find factors that affect student achievement. The earliest studies focused on student and environmental factors involved with student achievement, whereas the current studies focus more on school factors. However, in reference to the classroom, researchers heretofore looked primarily at management, such as safe and orderly climate. Little research focused on teacher quality until recently with the work of William Sanders (e. g. Wright, Horn, & Sanders, 1997). This newer line of research suggests that, of the school factors that promote student achievement, teacher effects have the greatest impact.

Even though the focus of the educational political agenda today is on student achievement (vis a vis test scores) and opportunities for all students to achieve, some children are not reaping the benefits of these opportunities. While current accountability systems typically track the performance of a school, most state accountability systems do not have a system to track individual student gains. In addition, whether or not students are demonstrating gains, the accountability system does not explain the factors that contribute to those gains (or lack thereof). In fact, most accountability models look at student achievement without looking at the longitudinal measures that could uncover teacher effects.

With the implementation of the accountability models, there have been disagreements as to how improvements in student achievement should be accomplished, even within states

(Sanders, 2000). Some argue that educators should be allowed to use their own judgment but should be given more resources to do so while others argue that states must set high standards (Sanders). However, huge variability in effectiveness has been found between schools and districts using the first approach. Furthermore, the use of initiatives and programs has produced mixed results (Sanders).

Student achievement in Louisiana, for example, is measured by a student's performance on standardized tests and the scores from these tests. Okpala, Okpala, and Smith (2001) assert that student achievement is actually "a function of school resources, student ability, student socioeconomic background, and other characteristics" (p. 112). Okpala et al. cite Greenwald, Hedges, and Laine (1996) to define resources as "including (a) school characteristics and facilities and (b) student characteristics such as SES or ability" (p. 112). Greenwald et al. categorize these resources (inputs) into three groups: expenditures, size, and teacher background characteristics. Expenditures include teacher salaries and per-pupil expenditures; size refers to the schools as well as the classes; teacher background includes ability, education, and experience. This study examines the impact of one of these resources – teachers – on student achievement in one Louisiana school district.

School Effectiveness Research

Ellet and Teddlie (2003) divide school effectiveness research in the United States into four distinct phases. Further, they cite the division of three strands defined by Teddlie and Reynolds in 2000. The strands are:

> (1) School effects research, which studies the scientific properties of school effects; (2) effective schools research, which is concerned with the processes of effective schooling; and (3) school

improvement research, which examines the processes whereby schools can be changed for the better (Ellet & Teddlie, p. 112).

With the strands defined, Ellet and Teddlie (2003) discuss the progression of school effectiveness research that began in the 1960's. Even though the phases have been separated into distinct categories, they overlap at some level. They are:

1. Stage One, from the mid-1960s and up until the early 1970s, involved the initial economic driven input-output model.
2. Stage Two, from the early 1970s, saw the beginning of the effective schools studies, which included a wide range of school process variables for study and examined a wider range of school outcomes than Stage One studies.
3. Stage Three, from the late 1970s through the mid-1980s, saw the focus of SER shift towards the incorporation of the effective schools correlates into schools through the generation of various school improvement programs.
4. Stage Four, from the late 1980s to the present day, has involved the introduction of school context factors and of more sophisticated methodologies. (Ellet & Teddlie, p. 112).

Phase One

According to Wenglinsky (2002), the Equality of Educational Opportunity Study, also known as the Coleman Report, was one of the earliest studies to examine factors that influence student achievement. The Coleman Report was a response to the 1964 Civil Rights Act, which called for "a study of inequality of opportunity in education" (Kahlenberg, 2001, p. 1). Data for

the Coleman Report were gathered using questionnaires completed by teachers, principals, superintendents, and students from 4,000 schools in 1965 (Office of Education, 1966). Coleman concluded, unexpectedly, that there were no significant disparities in funding between schools with differing racial make-ups (Kahlenberg). Furthermore, school funding was not closely associated with student achievement. Second, peers were strongly associated with student achievement; that is, it was more advantageous for students to be in school with “middle-class” peers rather than “lower-class” peers. Coleman concluded: “It appears that a pupil’s achievement is strongly related to the educational backgrounds and aspirations of the other students in the school” (Office of Education, p.22). Kahlenberg points out two profound themes. One theme is that other students affect aspirations of students; the other is that, “Numbers matter, in part because the numerical majority in a school sets the tone” (p. 3). In other words, if the majority of peers are from wealthier families, they impact the climate of the school and vice versa. Finally, facilities and curriculum had the least amount of impact on student achievement. However, teacher quality (defined by years of experience, educational background, and ability) was categorized under facilities and curriculum. Even though the result is not found in many of the summary reviews of Coleman’s work, Coleman did address the issue of teacher quality, stating: “The quality of teachers shows a stronger relationship to pupil achievement,” in relation to the other aspects of facilities and curriculum (Office of Education, p.22).

The most prevalent conclusion of the Coleman Report was that family socioeconomic status was the number one predictor of student achievement. When the Coleman Report was analyzed, the findings appeared to determine that schools had little impact on student achievement (Ellet & Teddlie, 2003). Wenglinsky (2002) summarized the findings as follows:

Quantitative research on whether schools matter has generally supported

> the notion that the problems of U. S. education lie outside of the schools. Some research finds that when the social backgrounds of students are taken into account, school characteristics do not seem to influence student outcomes, suggesting that schools do not serve as avenues for upward mobility, but instead reinforce existing social and economic inequalities (p. 2).

Jencks, Smith, Ackland, Bane, Cohen, Gintis, Heyns, and Micholson (1972) determined that there were three main conclusions regarding inequalities in the school: 1) Unequal shares of educational resources are provided for different students and/or groups of students; 2) there is more of an equality in low-cost services than high-cost services; and 3) even if all education were free, it would not equalize educational opportunity due to the premise that not everyone would use these resources.

Jencks et al. (1972) reported that family background was the most prevalent predictor of the extent to which students achieved. The second most prevalent factor was found to be cognitive skills. “The precise effect of cognitive skill is hard to determine however, since we do not know to what extent test scores are a proxy for unmeasured, noncognitive differences between home environments. Race now seems to affect educational attainment almost entirely by affecting test scores and aspirations” (p. 159).

According to Ellet and Teddlie (2003), the Coleman and Jencks works comprised the first phase of school effectiveness research. This phase concentrated largely on economic-driven factors.

Phase Two

According to Fullan (1991) and Ellet and Teddlie (2003), in order to counter the earlier findings, researchers including Brookover (1981), Lezotte, Hathaway, Miller, Passalacqua, and Brookover (1980), and Edmonds (1979) began conducting research on school effectiveness. In order to find evidence for their theories, researchers began studying schools successfully educating students from lower socio-economic backgrounds. These studies began to include behavioral and attitudinal factors as outputs and attempted to explain the processes used in these schools (Ellet & Teddlie).

According to Ellet and Teddlie (2003), the works of Murnane (1975) and Summers and Wolfe (1977) demonstrated that teacher inputs had an impact on student achievement. Using regression models, Murnane found that predicted variance could be determined by school as well as classroom assignments.

The work of Brookover, Beady, Flood, Schweitzer, and Wisenbaker (1979) focused on the social structure using a stratified random sample of elementary schools in order to research student achievement with the hypothesis that the school structure impacts achievement. The researchers examined three areas: school social inputs (student body composition and other personnel inputs), school social structure, and school social climate. According to Brookover et al., findings from their study indicate that school social systems do make a difference in relation to student achievement. For example, the combination of the three variables account for over 85% of the between-school variance in mathematics achievement scores. Brookover, Beamer, Efthim, Hathaway, Lezotte, Miller, Passalacqua, and Tornatzky (1982) added that schools must get rid of the notion of the bell-shaped curve and that there is no reason that schools servicing

poor and minority children should accept failure. They found that “unfortunately, the social system of most low achieving schools is designed to accept failure” (p. 44).

Good, Biddle, and Brophy (1975) reported that there was not much research to determine whether schools, or teachers for that matter, make a difference in student achievement. From their review of the literature, the authors suggested that “time spent in instruction is positively correlated with school achievement” (p. 26). However, in their review of the research, they criticized the fact that many of the early researchers generally disregarded the impact of teachers. Good et al. further stated that much of the previous research did not track individual student gains; instead, mean gains were used to determine effectiveness. Brophy and Evertson (1981) summarized their Student Attribute Study, which focused on the teacher: expectations, attitudes, and student effects on teachers (i. e., how teachers react toward students). The most profound finding was that teacher expectations have an impact on student achievement.

Phase Three

Edmonds (1979) focused research on correlates of effective schools that have an impact on student achievement. The concern was whether poor students were being adequately challenged and receiving an equitable education (Edmonds; Stoll & Fink, 1992). Edmonds reported that teachers in less effective schools pointed to non-school factors as the reason for students having reading difficulties, saying that, as teachers, they were unable to impact student achievement. On the other hand, teachers in more effective schools felt that they did have an impact on students. The third-generation school effectiveness researchers, led by Edmonds, were able to isolate school effects by studying more and less effective schools within the same socio-economic strata. In this manner, they were able to find leadership, climate, and instructional factors that did impact student achievement, even for poor children.

The works of Edmonds (1979) and Lezotte and Bancroft (1985) began to form a foundation that would not only identify an interest in effective schools, but would also be the impetus for creating an interest in implementing strategies to create effective schools (Ellet & Teddlie, 2003). However, their focus on schools servicing only poorer students led to criticisms of biases because of their lack of sampling schools of varying socio-economic strata.

While previous research focused on equity, Wimpelberg focused on efficiency of schools (Wimpelberg, Teddlie, & Stringfield, 1989). Wimpelberg suggested that cumulative resources and curriculum are factors that lead to improved student achievement (Teddlie & Stringfield, 1993). Other factors include leadership, socioeconomic status of peers, discipline, climate, and teachers within the school. All of these factors have an indirect impact on student achievement, except the teacher, who has a direct impact. Further, socioeconomic as well as other personal/environmental factors are not easily controlled aspects of the schools when addressing student achievement. If educators are held accountable for student achievement, it is imperative to isolate those factors that most influence student achievement and that educators can actually influence.

In addition to the previously cited researchers, Coleman reversed his initial finding on the issue of how much of a difference schools make (Coleman, Kilgore, & Hoffer, 1982; Ravitch, 2000). According to Ravitch, Coleman found that schools with high academic expectations and a common academic curriculum, Catholic schools in particular, promoted higher achievement, regardless of the students' backgrounds.

Phase Four

Due to the interest in equity issues in schools, school context factors were generally disregarded during the third phase of school effectiveness research (Ellet & Teddlie, 2003).

Therefore, the fourth stage of effectiveness research began to use more sophisticated methodology. The research began to explore context variables in schools across a wide range of schools. In fact, with the progression of research over time, the trend is toward integrating studies on school effectiveness with teacher effectiveness (Ellet & Teddlie). The theory is that the combination of the two can help to create educational institutions that are more effective in educating students.

Since research has identified definite school effects, federal and state governments have adopted accountability programs. The premise of accountability legislation is that school context variables, more so than student-level factors, affect student achievement. This fourth phase of school effectiveness research (Ellet & Teddlie, 2003) is the impetus for value-added studies.

A Nation at Risk was released in 1983 as a warning to the public of the state of public education. The warning suggested that the economy and society as a whole would be negatively impacted if improvements were not made in the education of all children (Ravitch, 2000). A reform effort, *America 2000,* was presented by the Bush administration, followed by President Clinton enacting *Goals 2000* in 1994 (Sutton & Soderstrom, 1999). Funds were provided to the states to develop standards and assessments. However, Goals 2000 was soon abolished (Ravitch). The most current reform effort enacted by the federal government is the *No Child Left Behind Act*, 2001 (NCLB; Amrein & Berliner, 2003). The goal of this legislation is to increase student motivation and raise student achievement levels. According to NCLB, the goals will be met by testing students annually in the areas of reading and math, grades three through eight (Amrein & Berliner).

According to Sanders (2000), there are consequences to standardized tests due to the fact that politics enters the picture. These consequences include: lowering the achievement level or either lessening the role of the tests or eliminating the tests. Sanders adds,

> The standards movement, as operationalized currently, embodies the concept of a stairstep approach to curricula and its companion, assessment. "What should fourth graders know and be able to do?" is a question defining specific goals for many states and districts – "these are first grade skills; these are second grade skills, etc." As a working definition, this has led to testing regimes that purport to measure the percentage of students within grades who are at mastery, proficient, basic, non-mastery (or whatever language is dangled beside the test results). Inevitably, when the results are presented, it becomes obvious that differences in results among schools and districts are strongly related to socio-economic measures of the demographics of the student population of a school or district (p. 330).

Additionally, standards should be defined in terms of the amount of academic progress rather than a specific achievement level. This would rid much of the debate because research studies display progress can be estimated without socio-economic and ethnic factors involved (Sanders).

Standardized testing is a popular form of accountability for a variety of reasons. Linn (2000) reports four particular reasons that testing and accountability are politically popular. First, it is inexpensive when compared to alternatives such as new instructional programs and the cost of staff development associated with programs. Second, states and districts can easily mandate testing and assessments. Third, implementation can take place quickly (within a four-year elected term). Last, results of tests are easily reported. Naturally, test results are generally poor when

newly implemented. In this way, policymakers are able to demonstrate the impact that their administration has had on educational achievement. Additionally, Dorn (1998) reported that national testing is popular among the general public as shown in the Gallup poll. Dorn further stated that the uses of these statistical systems as accountability measures are useful to politicians because of their “visible power” (p.2).

While each state is allowed to implement its own accountability model, Louisiana accountability measures emphasize student achievement, and reward or threaten to sanction schools based on School Performance Scores (SPS). The SPS is derived from Norm-Referenced Test (NRT) scores (60%), Criterion-Referenced Test (CRT) scores (30%), and attendance/dropout rates (10%). Louisiana is using the IOWA test of basic skills to obtain NRT scores and the Louisiana Educational Assessment Program (LEAP) for CRT scores. IOWA tests are given to third, fifth, and seventh graders throughout the state. LEAP is administered in the fourth and eighth grades where the stakes are high for students; students must pass the state CRT in order to advance to the next grade. One parish school system opted to also administer the NRT tests to kindergarten, first, and second graders for tracking purposes. As of the 2004-05 school year, this particular parish will no longer test kindergarteners and first graders due to budgetary constraints.

With the use of high stakes testing in fourth grade, the Louisiana accountability model is viewing the curriculum as “stair-steps,” if viewed in terms of Sanders. Sanders (2000) would prefer the curriculum to be viewed as a “ramp.” Each student would “move up the same ramp,” but students would not be required to be in the same place at the same time (p. 330). This type of model could alleviate the problems associated with the accountability system.

Significance of Study

Haycock (1998) has reported that teachers are the most important factor contributing to student achievement. Haycock justifies her claims partially on the basis of the Tennessee Value-Added Assessment System (TVAAS) studies (Sanders, Saxton, & Horn, 1997) as well as the Dallas Value-Added Accountability System (Webster & Mendro, 1997). These studies conclude that teachers make a difference in student achievement and that the cumulative effect of having "good" or "bad" teachers for successive years can be dramatic. Because every child deserves a quality education, it is valuable to determine whether teachers make a difference in student achievement and to what extent. In Louisiana, it is children who are penalized by high-stakes accountability tests. But what role does the school, specifically the classroom teacher, play in children's success or failure on these tests?

The expected outcome of the study is that teachers do have a tremendous impact on the educational achievement of students. The current accountability system threatens to sanction schools that demonstrate poor student achievement through state take-over or loss of funds. The state may be threatening takeovers because of bad teaching but ignores that there can be good teaching even in these low-performing schools. Using students' longitudinal progress, this study can provide evidence to support that contention.

By using a longitudinal analysis of individual student test scores, gains of individual students can be tracked. By aggregating student data to the particular teachers, the effects of the teacher can be determined, which will create a more accurate account of individual teacher effectiveness and the impact on students.

Purpose

The problem schools are facing is that not all children are achieving at the same level. The current accountability system is made for the proverbial one-size-fits-all educational system. The current Louisiana accountability program, in particular the LEAP test, is used to either pass or fail students. This system is not responsive to individual student gains, nor does it account for teacher quality as it affects student achievement. In order to gain factual information regarding student progress, research and accountability programs must report individual student and teacher gains. The purpose of the present study was to investigate teacher effects on student achievement in one Louisiana public school district by tracking student gains by teacher. The value-added model developed by Wright, Horn, and Sanders (1997) was adapted for this study.

Research Question

The present study focused on one main research question: What effects do teachers have on student achievement? Therefore, the present study used, as a model, studies conducted by Wright, Horn, and Sanders (1997), namely the TVAAS, using controllable and uncontrollable variables. Controllable variables are defined as the variables that can be maintained or changed within the school. Uncontrollable variables are defined as the personal/environmental factors that the school cannot possibly change. By using each student as his/her own control, the TVAAS allows "estimation of school system, school, and teacher effects free of the socioeconomic confoundings that historically have rendered unfair any attempt to compare districts and schools based on the inappropriate comparison of group means" (Sanders et al., 1997; p. 138).

Overview of Methodology

The present study was an adaptation of the TVAAS used in Tennessee. TVAAS uses standardized tests as a way of measuring students' gains. It uses multiple variables and a

complicated formula to find whether or not teachers have an impact on student achievement (Haycock, 1998).

TVAAS uses a variety of factors to determine teacher effects. Student standard scale scores from a given test are figured into the analysis, along with the teacher, class size, and heterogeneity of the class. Standard scale scores are expected to increase as the student progresses through school. Therefore, it is an appropriate scale to use in longitudinal research. Furthermore, Sanders used the norm-referenced portion of the TCAP (Tennessee Comprehensive Assessment Program) due to the fact that it is based on the average level students should approach rather than the criterion-referenced portion which shows what a student should learn. After the information is in the database, a longitudinal analysis is conducted on the data using a mixed methodology approach.

In order to track students, the scores from the TCAP, along with the students' previous teachers, and the school the child attended are part of the students' record for up to five years. In order to have a complete record of each student, data from each year are merged with the prior year's data into one database (Sanders et al., 1997). Because of the tracking of the data, the student serves as his/her own control.

Sanders et al. (1997) use a mixed-methodology approach (built on Henderson's mixed model equations) to longitudinally analyze several factors. Because of the use of the student as his/her own control, a more precise determination of the actual factor(s) that contribute to the student's gain can be identified. Factors that cannot be controlled, such as parents' educational acquisition, race, socioeconomic levels, etc., are partitioned out of the analysis without affecting the results of the study.

This study attempted to adapt the Sanders model for use in one school district in Louisiana. Demographic data on individual students included race, sex, age, special/regular educational instruction, and lunch status. Lunch status is determined by whether the student receives free/reduced lunch or paid lunch. Lunch status is used as a proxy for socioeconomic status, a known educational risk factor. These variables were included for descriptive purposes only; they were not used in the TVAAS model.

School data included whether the school is a “Title I” school. The purpose of determining whether a school is a Title I school is because these schools have a large population of at-risk students as well as highly diverse ethnic and racial populations. Title I schools receive specific funds from the federal government in order to target educational resources or programs for at-risk students. Again, the percentage of at-risk students is determined by the percentage of students receiving free/reduced lunch in a particular school.

The analysis included individual standard scores obtained from the IOWA test for students in grades kindergarten through third. Wright et al.’s (1997) research used student data beginning with the third grade. However, the purpose of these scores being used in the present study is due to the fact that the IOWA test is given each consecutive year in these grades. Even though grades kindergarten through second are not required to be tested by the state of Louisiana, the selected parish requires this testing. In Louisiana, fourth graders are required to take the LEAP test. Therefore, it is not possible to create a valid comparison starting with third grade. The only current way to track an accurate account of student gains is to use consecutive years of standardized tests with the same type of scale.

Potential Limitations and Delimitations

The original study performed in Tennessee obtained a pool of over 65,000 students due to a large database of information. Additionally, information was accessible from more than one district. The present study was delimited to a much smaller sample of students and only scores from one school district. The sample might be considered a convenience sample due to the fact that the researcher is employed by the school system under study; however, the district administration has been trying to find a way to track student progress over time and account for variation in students' longitudinal performance. The Tennessee Value-Added study provides a means to answer these questions. Since only one school district was used, some of the variables used in the original study were not available for the present study, such as data from various school systems in Louisiana.

One limitation is that the data will be collected on K-3 students. Even though that sample is necessary for this research because of the consecutive years the test is given, it will be difficult to generalize the results. Not all districts in Louisiana have implemented testing students in grades K-2. However, the IOWA tests will be changed to iLEAP tests (the format of the iLEAP will match that of the LEAP) within the next year. At that time, a similar CRT will be used in all grade levels which would allow this research method to be repeated using the CRT.

Future studies should be conducted on a larger scale to provide further evidence of effects of teachers and what actually makes a teacher effective. If the data support the hypothesis, it is essential to school systems to find what more effective teachers do to provide success to their students and to guarantee every student has access to effective teachers.

CHAPTER 2

REVIEW OF LITERATURE

Factors that affect student achievement have been studied for more than 45 years. Early studies focused on home/environmental factors affecting student achievement; more recent studies have focused on school factors. The purpose of the present study was to research effects of teachers on student achievement over time. The guiding question was: What impact do teachers have on student achievement over time?

In order to examine the guiding question, this section will begin with a brief review of the literature on school accountability. Next, an in-depth review of two of the primary value-added models will be discussed: the Dallas Value-Added Accountability System and the Tennessee Value-Added Assessment System. Finally, a summary of the literature addressing effective teaching will be cited.

School Accountability

Beginning in the 1990s, federal government reforms were enacted to improve schools. The first was *Goals 2000* in 1994; the current federal reform effort is the 2001 *No Child Left Behind Act*, which makes high-stakes testing more prevalent than ever before (Amrein & Berliner, 2003).

Nearly all school systems have mandated high-stakes testing. The purpose is reportedly to hold schools accountable for student achievement. Often students are held "most" accountable through "high-stakes" tests which serve as gatekeepers of grade-level promotion. Some states have elevated their use of testing data, tracking individual student gains based on aggregated

scores by teacher. These studies, which reveal the effects of individual teachers on student achievement, are known as “value-added” studies. There are two prevalent value-added assessment systems – the Tennessee Value-Added Assessment System (TVAAS) and the Dallas Value-Added Accountability System (DVAAS). These longitudinal analyses of student data are the basis of the present study. Both studies are detailed in this section.

Value-Added Studies

Some of the accountability models currently being used only show whether students know more or less than their peers who took the test in previous year(s) (Holloway, 2000). On the other hand, the value-added models measure what individual students have learned. Furthermore, standardized testing generally places students within a distribution of a normal student population. This is unfair because measuring in these terms will not account for external factors such as socioeconomic status. The Education Improvement Act (EIA) of 1992 contained provisions to ensure fairness. The provisions included having at least three years of student data and schools, systems, and teachers must be assessed on variables outside the TVAAS (Sanders & Horn, 1998). Because value-added studies, such as TVAAS, utilize achievement data as inputs in a longitudinal analysis, external influences are controlled (Holloway). This accountability model tracks academic progress, holding educators accountable for controllable factors (Sanders, 2000). For example, a teacher will not have control over the previous achievement level of a student entering his/her class but will have primary control over the progress made in his/her class (Sanders).

Sanders (2004) cautions that all value-added models are not equivalent in nature. Some of the approaches should be avoided. However, the one thing all value-added models have in common is that they use longitudinal data.

One example of the value of value-added studies involved a middle school in California that was always ranked academically low (Holloway, 2000). Once the value-added approach was used, the school was able to compare final performance of students to their incoming performance. In reality, the school was considered better than a more affluent school in the district for raising scores. “The real difference between the two schools was the students’ incoming academic level – a factor directly related to socioeconomic conditions,” (Holloway, p. 85).

Dallas Value-Added Studies

Investments in research began in the late 1960s and 1970s in the Dallas Independent School District (DISD) in Texas (Cunningham, 1997). In 1984, DISD used multiple regression analyses to track longitudinal student growth using norm-referenced tests. The purpose was to determine whether schools exceeded the expected growth of the students (Stronge & Tucker, 2000). In the 1990s, a state accountability system began to form, being strengthened by the work of William Webster in Dallas (Cunningham). In 1996, Classroom Effectiveness Indices (CEI) were added, using a combination of multiple regression and hierarchical linear modeling (HLM) to control for student and school variables in order to ensure fairness and equitability (Stronge & Tucker).

The main focus of the Dallas accountability system is growth. Growth includes academic growth of students, teachers’ instructional growth, and principals’ instructional leadership growth (Stronge & Tucker, 2000). Effectiveness is determined by the expected growth of the student as compared to the actual growth. Teacher effectiveness is determined by the amount of actual growth as compared to expected growth of students in the teacher’s class(es); school

effectiveness is determined by the actual growth as compared to the expected growth of students within the school (Stronge & Tucker).

The DVAAS began in 1992 as a fair and equitable accountability system to determine effectiveness of schools (Webster & Mendro, 1997). DVAAS has since expanded to identify effective teachers and help with teacher evaluation methods for the DISD. There are several factors used to assess effective schools in the Dallas system. These variables are: CRTs, NRTs, student attendance rates, dropout rates, retention rates of students, student enrollment (honors course, accelerated courses, and advanced diploma plans), graduation rates, and the percentage of students taking college entrance exams – with test results being the pervasive measure (Webster & Mendro; Stronge & Tucker, 2000). To ensure fairness of the system, data analysis is only based on students who are continuously enrolled in a school. Furthermore, at the very least, 95% of the student body must be tested to avoid a school attempting to display higher achievement by avoiding testing of certain students. Along with the longevity of student enrollment safeguard component in the analysis, there is a safeguard against student absences (Stronge & Tucker).

In the DVAAS, a two-stage model of data analysis is required. The first stage of analysis uses multiple regression in order to control effects of the fairness variable. Fairness variables are defined as student differences such as ethnicity, gender socioeconomic status, prior achievement effects, and language proficiency. In order to control school-level variables (mobility, crowding, percentage minority, and socioeconomic status), prior achievement, and attendance effects, a two-level hierarchical linear model (HLM) is employed. Finally, a simple multiple regression model is incorporated using two years of data for each variable by school. In this model, the school is the unit of analysis (Webster & Mendro, 1997).

As determined by the Board of Education and the Commission for Educational Excellence, sorting the information gathered within a particular school can identify effective teachers. Teacher Effectiveness Indices (TEI) were implemented in 1994-95 at the elementary and middle school levels (Webster & Mendro, 1997). School Effectiveness Index (SEI) data were used in order to prepare the TEI. Teachers who taught core courses relevant to NRT and CRT testing data (for example, reading or social studies) were matched with the students for whom the teacher had given grades. SEI data were then computed for each individual teacher. Interpretation was simplified by standardizing the effectiveness data with a mean of 50 and standard deviation of ten.

Webster and Mendro (1997) address the fact that there are technical issues with the TEI component. First, it would be logical to assume that a three-level HLM rather than two-level would be more efficient. However, if a three-level analysis were used, some teachers and schools would have to be dropped from the analysis. Therefore, the two-level HLM is still being used. Second, even though assumed in the model, residuals from the HLM do not have like means. Continued research is being performed to find the best way to standardize the residuals. The third and final technical issue is the sample size. Teachers with six or more students are used for data analysis. However, the size is insufficient if the data are to be used for teacher evaluation purposes. There are some ways to address this issue – the most prominent being the use of more than one year of data.

To conclude their explanation of the DVAAS, Webster and Mendro (1997) address some future considerations. First, issues of changes in testing must be addressed. Second, cheating can occur. However, there is a computerized cheating analysis program. The result of cheating would disqualify a school from receiving rewards while the perpetrator(s) would be dealt with through

the personnel office. A third issue for future analysis would be that of goals. Even though goals are previously set, these goals are not empirically based. Minimum standards of performance should be developed with empirical evidence. Finally, criterion-referenced objectives that are meaningful must be established.

Stronge and Tucker (2000) address six advantages and four disadvantages of using the DVAAS model. The advantages include: 1) The model focuses on fairness and continuous improvement; 2) Because the focus is on improvement, individual differences are identified and addressed; 3) The model demonstrates the advances in technology to analyze data from test scores; 4) The model demonstrates a simple measure to determine student progress and teacher influences; 5) The focus is on student growth and "not absolute achievement"; and 6) It is a "proactive" measure of analysis. According to Stronge and Tucker, the disadvantages are: 1) The model does not account for student attitudes; 2) There must be a commitment of human resources as well as financial resources for such a sophisticated system of data analysis; 3) Many types of testing instruments are used many times each year; and 4) Data can ultimately be improperly interpreted and misused.

Tennessee Value-Added Studies

The second of the two major value-added assessment systems is the Tennessee Value-Added Assessment System (TVAAS). It is premised on the fact that many factors influence the rate of student learning. Some factors are controllable within the education realm and others are not. It is necessary to separate the controllable factors from other influences in the students' lives.

The TVAAS is similar to the DVAAS in many respects (Stronge & Tucker, 2000). Both models are based on student growth/gains and both are based on student outcomes. However, the

TVAAS compares the individual student to himself/herself, whereas the DVAAS compares the students to other students with similar characteristics such as ethnic background and gender. The TVAAS is the first system of its type adopted by the entire state and is the basis for the present study.

Annual testing of students in grades two through eight was implemented in 1990 using the Tennessee Comprehensive Assessment Program (TCAP). Students are tested in five subject areas: math, science, reading, language, and social studies. The scores in each area are input into a system in order to measure student growth as revealed through the TCAP (Stronge & Tucker, 2000).

William Sanders is credited with the research that paved the way for the creation of the TVAAS. Sanders' work at the University of Tennessee Value-Added Research and Assessment Center, focusing on the additive effects of teachers, led to the implementation of TVAAS (Stronge & Tucker, 2000). According to Stronge and Tucker, data suggest that students who are enrolled in the classroom of three high-performing teachers in a row score, on average, in the 96^{th} percentile of math assessments. On the opposing side, students assigned to low-performing teachers three years in a row, average in the 44^{th} percentile of math assessments. One can conclude that teachers have both additive and cumulative effects on student achievement (Stronge & Tucker), especially in math (Sanders, 2004). However, Sanders states that there is no significant evidence of a compensatory effect. Further, the teacher's effect on a student can still be measured two years later, particularly in math. To help students affected by poor teachers, an intervention would be needed (Sanders)

TVAAS uses standardized tests as a way of measuring students' gains. TVAAS is a method of using multiple variables and a complicated formula to find whether or not teachers

have an impact on student achievement (Haycock, 1998). TVAAS uses a variety of factors to determine teacher effects. Demographic variables known to be associated with student achievement such as ethnicity, socioeconomic status and special education status are not necessary to use in the model because the TVAAS uses student gain scores rather than group means to determine achievement. The student data from a given test are figured into the analysis along with the teacher, class size, and heterogeneity of the class. These factors are known to be related to student achievement. Sanders et al. (1997) point out that any other variables highly correlated with curricular outcomes could be used as well. Sanders (2000) argues that in order to use testing data, scales must be correlated with the curriculum.

The students' standard scale scores are used rather than percentiles to give a more accurate account of the students' abilities. Standard scale scores are expected to increase as the student progresses through school. According to Sanders and Horn (1998), scale scores are used "to model their learning patterns" (p.249). Therefore, it is a more appropriate scale to use. Raw test score averages should not be used for reporting due to the fact that these scores are confounded with socioeconomic as well as other uncontrollable factors (Sanders, 2000). In order to make this point, a prime example would be a school serving a poor population could be making good progress, but the average test scores could be lower than the district average. This result would leave the impression that the school is ineffective (Sanders). Furthermore, norm-referenced portions of the TCAP are used due to the fact that this test demonstrates the average level students should approach rather than the criterion-referenced tests, which show what a student should learn. Bratton, Horn, and Wright (1996) additionally address the controversy over norm-referenced/criterion referenced tests, stating that there is no issue with TVAAS.

TVAAS requires the use of a test that correlates with the curriculum as well as one with questions of varying difficulty because of the various levels of achievers in the class (Bratton et al., 1996). Also, when “TVAAS got started, the norm-referenced portion of the TCAP achievement tests best fit the necessary criteria” (p. 24). Furthermore, the problems with using the NRT can be avoided because of the use of a mixed-methodology approach as well as the use of the longitudinal database of students. Braxton et al. compare the two tests in the following table:

Table 1
Bratton et al. Comparison of NRT and CRT in Value-Added Studies

Norm-referenced tests	Criterion-referenced tests
One's scores is compared (referenced) to scores of a peer group which may be local, e.g., one's own school or state, or national.	One's score stands alone, indicating a level of mastery of objectives (criteria) on which the test was based.
NRT's are timed.	CRT's may or may not be timed.
Questions vary as to difficulty, ranging from a few very easy questions to a majority of "grade level" questions to a few very difficult questions.	Questions have a much narrower range of difficulty than NRT's, the vast majority being "on (or below) grade level."
An **average** student is expected to correctly answer only about 60% of the questions. There should be a "reasonable" match between the NRT and the curriculum taught.	An **average** student is expected to correctly answer 100% of the questions. Since a near-perfect match exists between the CRT and the curriculum taught, objectives not mastered by a given student should be retaught and retested.
Several types of scores may be derived from the number of questions answered correctly, but all show how a given student ranks in relation to his/her peers. Score types are: percentiles, stanines, normal curve equivalents, scale scores, and grade equivalent scores.	Scores may be reported as a simple number or percent of questions answered correctly. Objectives or domains may be reported separately as mastery, partial mastery, or non-mastery.
There is no such thing as passing or failing a norm-referenced test.	A pass/fail cut-off score may be set for a criterion-referenced test, as is the case with the TCAP competency test (70%) in language arts and mathematics.

(Bratton et al., p. 25)

TVAAS was at the center of comprehensive educational reform in Tennessee in 1992 (Ceperly & Reel, 1997). It compares the achievement of each student to his/her prior year's

achievement rather than using national norms. By using this type of analysis, the gain of the student is attributed (statistically) to the teacher, school, and district.

> Specifically, the legislation defined TVAAS as a statistical system… which uses measures of student learning to enable the estimation of teacher, school, and school statistical distribution… to account for differences in prior student attainment, such that the impact which teacher, school and school district have on the educational progress of students may be estimated on a student attainment constant basis (Ceperly & Reel, p. 135).

All students from the second through eighth grade currently take the TCAP yearly. The scale scores from these tests are used as data for the TVAAS (Sanders, Saxton, & Horn, 1997). Sanders (2004) states that Tennessee has the largest longitudinally merged database, using the simple gain (Year2 – Year1) as the dependent variable. Furthermore, when analyzing academic progress rather than achievement, the teacher becomes the most important factor in student progress. This fact makes other variables appear to be trivial.

In order to track students, the scores from the TCAP, along with past teachers of the student, and the school the child attended are part of the student's record for up to five years. In order to have a complete record of each student, data from each year are merged with the prior data into one database (Sanders et al., 1997). Because of the tracking of the data, the student serves as his/her own control.

Sanders et al. (1997) use a mixed-methodology approach (built on Henderson's mixed model equations) to longitudinally analyze several factors. Because of the use of the student as his/her own control, a more precise determination of the actual factor(s) that contribute to student

gains can be identified. Factors that cannot be controlled, such as parents' educational acquisition, race, socioeconomic levels, etc., are partitioned out of the analysis without affecting the results of the study.

In summary, the following statistical claims are made concerning the TVAAS Process:

- Individual student records are incorporated into the model as statistical blocks, thereby partitioning exogenous influences on test performance.
- All available data are used, and no imputation for missing data is required.
- Longitudinal analysis over years improves the efficiency of the estimates of the model parameters.
- Additionally, repeated measures across subjects provide similar benefits.
- Educational influences on gain can be estimated from a model that uses scores, not gains.
- Shrinkage estimates of teacher effects provide protection against fortuitous misclassification of individuals.
- The "layered model" improves the efficiency of the estimate of teacher effects.

(Sanders et al., p. 144).

The TVAAS uses three different model equations. One model is used for the school system, another for the school, and another for the teacher (Sanders et al., 1997). The system

model equation is as follows: $\gamma_{iklmn} = \mu_{iklm} + e_{iklmn}$. Each symbol represents a specific component of the model: γ represents the test score, μ the fixed system mean score, e the random deviation. Within these variables, the i represents the school system, k the year, l the grade, m the subject, and n the student (Sanders et al., 1997).

The equation for the second component, the school, is as follows: $\gamma_{ioklmn} = \mu_{ioklm} + e_{ioklmn}$. The "o" represents the school. Therefore, this model adds the individual school into the analysis.

The third equation is that of the actual classroom teacher effect on the student. The teacher model is summarized as follows: $\gamma_{ijklmn} = \mu_{ijklm} = \sum_{p=1}^{N_m k} c_{m(ijkl)p} \text{ X } t_{m(ijkl)p} + e_{ijklmn}$.

The TVAAS Model allows for all students to be included in the analysis, even if their records are not complete (Sander et al., 1997). However, students with complete records can be more heavily weighted. Furthermore, the teacher model allows for such factors as departmentalization and team teaching.

Haycock (1998) examined the DVAAS and TVAAS, finding that these models have helped to tie teacher effects to student achievement. She states that poor and minority children should be taught at the same levels as other students, and, if they are, they will achieve at the same higher levels. Stronge and Tucker (2001) also assert that value-added assessments are superior to other accountability systems in that they allow for the identification of good teaching. They also suggest that the TVAAS is more accurate than the DVAAS. This is due to the fact that the TVAAS measures the student against him- or herself; whereas, the DVAAS compares student growth to other students with similar characteristics.

Because of the earlier analyses, the TVAAS will be used as the model to conduct the present study. However, some modifications must be made due to the smaller sample size and accessible data. The main modification is elimination of the first equation because only one

school system is being considered. In addition to the modifications, there are two aspects of this model that must first be addressed. Just as Stronge and Tucker (2000) addressed advantages and disadvantages of DVAAS, they did the same for TVAAS. Second, Sanders' model was recently eliminated in Tennessee.

Stronge and Tucker (2000) cited five advantages and four disadvantages for the TVAAS. The advantages include: 1) Because several years of data are used, experts consider the TVAAS as "robust, fair, reliable, and valid" statistically; 2) Individual differences are accommodated because the focus is on improvement rather than a fixed standard of achievement; 3) The data provide a relatively simple measure of the teacher's ability to influence student gains; 4) The TCAP is aligned with the curriculum in Tennessee, confirming content validity; and 5) There is a positive correlation between supervisory evaluations and teacher effects as measured by the TVAAS. The disadvantages include: 1) A high degree of computing power is necessary due to the complicated data analyses required; 2) Data can be misused and misinterpreted; 3) Testing students annually involves time, money, and human resources; and 4) The purpose of education is multifaceted, and the TVAAS is limited in its ability to measure the student's ability.

The second concern with the TVAAS is the recent legislation proposed by two Democratic politicians to rid the state of Tennessee of the model (Olson, 2004). According to Olson, the concern does not appear to be with value-added assessment but with the attempt by Sanders to determine and adjust for the equivalency of the TCAP year to year.

Since the score is based on an average of three years, Sanders reportedly adjusts for any inconsistency in the scale (Olson, 2004). In the year 2000, the concern grew when it was found that there was a significant difference between the value-added measure and the raw test score after Sanders made adjustments.

TVAAS looks at whether students make more or less progress than expected. The Education Trust is a proponent of value-added assessment because it is the best way to identify effective teachers (Olson, 2004). Students score fifty percentile points higher, on average, using a 100-point scale if they have an effective teacher. Therefore, doing away with value-added measures would limit the ability to identify effective teachers. The Education Trust would prefer to see a few minor adjustments, such as possibly developing multiple test forms at one time and using more of the same test items on a yearly basis (Olson).

Summary of Literature Review

Much of the school effectiveness research has focused on various school factors that influence student achievement. However, it was not until 1982 when Edmonds "called for the marriage of school and classroom research" (p. 98) that there was an acknowledgement that researchers should examine school and classroom factors (together) affecting student achievement (Wimpelberg, Teddlie, & Stringfield, 1989). Since that time, there are a variety of teacher effectiveness variables within the framework of more elaborate school effectiveness research. "However, the tendency is the typical effective schools study to average out classroom effects at the level of the school (or, more, accurately, at the level of a single grade, extrapolated to the school) hides the effects of teacher decisions on individual children and groups within classrooms," (Wimpelberg et al., p. 99).

To further the exploration into the need for teacher effectiveness research, Ellet and Teddlie (2003) cite several researchers calling for an integration of school and classroom level research. Good (1989) surmised that there is a need to investigate how classroom and school processes function and how these processes can be integrated to create better educational

institutions for students. Reports by Creemers and Reezigt (1996) corroborated the need for educational effectiveness studies with the integration of the two separate fields (Ellet & Teddlie).

To complicate matters further, Stronge (2002) suggests that teaching is a complex task that is not easily defined and there are a variety of ways in which effective teaching can be defined. Effectiveness is sometimes defined according to evaluations by administrators or by using student achievement. Furthermore, teachers have a profound and lasting impact on students in a variety of areas (Stronge). Therefore, it is a complicated task to determine the outcomes that define effectiveness on the part of the teacher as well as the best way to measure these outcomes.

Even though research efforts are focusing on teacher effectiveness, the review of the literature has not provided much empirical evidence as to how teaching matters until the work of Sanders and his colleagues (Ellet & Teddlie, 2003). Even though the knowledge base is growing, the debates continue as to whether or not the literature identifying attributes of effective teachers can be trusted (Stronge, 2002; Wright, Horn, & Sanders, 1997). In addition, debates are arising as to whether teacher evaluation should be tied to student achievement. “These debates aside, few attempts have been made to directly measure the influence of individual teachers on the academic progress of large *populations of students* using measurements available from traditional standardized testing program,” (Wright et al.; p. 57).

CHAPTER 3

METHODOLOGY

Overview of Methodology

The *No Child Left Behind Act* requires standardized testing of students, with each state being allowed to implement its own testing program. The current Louisiana Accountability program tests students annually beginning in third grade. To further track students, one parish begins standardized testing of students in the spring of their kindergarten year. While annual testing of students assists in tracking student growth, the current Louisiana model lacks the sufficient database to determine student growth. Rather, the scores are used for purposes of grading and ranking schools.

Beginning with third grade, the ITBS scores are used as 30 percent of a school's performance score (SPS) while LEAP (taken in fourth and eighth grades) determines 60 percent of the SPS. The remaining ten percent is determined by attendance/dropout rates. Therefore, the current accountability model lacks the most important element of the actual purpose of education – student growth.

In order to gain a better understanding of testing data, a value-added approach to student achievement must be implemented. Therefore, the purpose of the present study was to determine teacher effects on student achievement by aggregating student scores by teacher. In this manner, it can be determined which students are progressing by determining which teachers are providing more progress to students.

Population/Sample

One Louisiana school district was used for the present study. This school system educates approximately 19,200 students in forty schools. Nineteen of these schools service nearly 6,000 students in grades K – 3. Of the student population, 62% are white while 38% are minorities with approximately 59% of students receiving free/reduced lunch.

The population for the present study included all students administered the ITBS in kindergarten in 2001, kindergarten and first grade in 2002, first and second in 2003, and second and third in 2004. The sample included all students for whom three consecutive years of test scores are available. To further identify the sample, the goal was to attain a baseline score in kindergarten. Therefore, students who began school in kindergarten in 2001 were tracked through the year 2004 when they tested in the spring as third graders. With these students, the sample included four years of data. Students who began kindergarten in 2002 were tracked for only three consecutive years, reaching second grade in 2004. The total sample size consisted of approximately 3,000 students.

A limited number of academic subjects were used in the sample. Tracking of the students encompassed the use of scale scores in the areas of language arts and mathematics. Reading was only tracked beginning with first grade due to the fact that there is no such score for kindergarten.

Instrument

Louisiana uses the Iowa Test of Basic Skills (ITBS) to obtain NRT scores. The tests are administered to students in the spring of every school year, testing a variety of academic areas: language, reading, mathematics, science, and social studies. However, kindergartners are not given the reading portion of the ITBS. Additionally, scores are reported in several manners: raw

score, percent correct, grade equivalent, standard score or scale score, percentile rank, stanine, and normal curve equivalent. The scale scores from the ITBS were utilized in the formula.

In order to examine the reliability and validity of the ITBS, test developers used evidence from various stages during test development (Hoover, Dunbar, Frisbie, Oberley, Bray, Naylor, Lewis, Ordman, & Qualls, 2003a). In other words, the scores should remain consistent if the students were tested again. "Data reported in the *ITBS Guide to Research and Development* demonstrate evidence of relatively high score reliability" (Hoover et al., p. 14).

Even though issues arise in relation to assessing young children, Hoover et al. (2003b) cite the fact that these issues are more of a philosophical difference. There are appropriate assessments for students in primary grades. The test developers used the same procedures for grade K – 2 assessment reliability as was used with the other forms of the ITBS, i. e., gathering of evidence was conducted at various stages of development (Hoover et al., 2003b).

Methodology

Data for the present study were obtained in two separate formats. The following data for each child were obtained from the school district on an excel spreadsheet: student identification number, grade level, race, gender, enrollment status (i.e., special or regular education), lunch status, teacher, and school. Obtained in hard copy format were the students' ITBS scores. Even though the data were gathered in these two formats, all of the data were entered in excel format and then converted to SPSS for data analysis.

The following equation, based on Sanders' (Wright, Horn, & Sanders, 1997) model will be used: $Y=M + H + C + H*C + T(H*C) + A +A*H + A*C + A*H*C + A*T*H + E$. The original model also contained system-level information. However, system-level notations have been removed. The variables are as follows:

Y = Student's Gain Score (Year 2 – Year 1 on subtest)

M = Overall Mean Gain

H = Heterogeneity in Achievement (Using each student's mean score of the two years, a class mean achievement level and standard deviation are computed; the classes are then divided into 3 levels based on the standard deviations-- low, moderate, and high, with ½ of classes in the moderate level and ¼ in each of the extreme levels.)

C = Class Size (2 groups = small [</=19 students], large [>/=20 students])

H*C = Heterogeneity by Class Size Interaction

T(H*C) = Teacher (nested within heterogeneity and class size)

A = Achievement Level (mean core total for two years broken into four groups by quartile)

A*H = Achievement by Heterogeneity Interaction

A*C = Achievement by Class Size Interaction

A*H*C = Achievement by Heterogeneity by Class Size Interaction

A*T(H*C) = Achievement by Teacher Interaction

E = Random Error Term

All effects in the model are fixed with the exception of T(H*C), A*T(H*C) and E terms which are random effects.

(Wright et al., p. 58-59)

Data Analysis

After all data were entered and converted into the SPSS program, the analyses were computed using a mixed model analyses of variance. The effect of teacher was of primary importance in the study but class size, heterogeneity of achievement level, and interactions also

are reported. It was expected that the classroom teacher would have a significant effect on student achievement even after other variables were accounted for. Analyses were conducted for each subject area (math, language arts, and reading) and each two-year time span. In all, 13 year-to-year models for the three subject area tests were computed (see Table 2). Using models for all years, the estimated teacher effect over time can be computed. In 30 separate models, Wright, Horn, and Sanders (1997) found significant teacher effects in all cases and teacher effects larger than any other effect in 20 of 30 cases.

Table 2
Value-Added Analyses

Kindergarten year	*K – 1st grade gain*	*1st – 2nd grade gain*	*2nd – 3rd grade gain*
2001	Language Arts Math	Language Arts Reading Math	Language Arts Reading Math
2002	Language Arts Math	Language Arts Reading Math	

Delimitations and Limitations of the Methodology

Sanders' (Wright, Horn, & Sanders,1997; Sanders, 2000) research and models based on his work typically involved populations of at least 65,000 students, two sets of school systems, and 54 schools. The present study consisted of approximately 3,000 subjects from 20 schools in only one school system.

CHAPTER 4

RESULTS

Introduction

Two separate student samples were used in this value added study of student achievement. Data for students beginning Kindergarten in the 2000-2001 school year through the third grade in the 2003-2004 school year were used to longitudinally track average student gains for Sample 1 of the study. Sample 2 consisted of students beginning Kindergarten in the 2001-2002 school year through their second grade year in 2003-2004.

Thirteen separate data sets were used in mixed model ANOVAs. The mean differences in student achievement by subject area were the dependent variables in the analyses. Based on the work of Wright, Horn, and Sanders (1997), the following independent variables were used for analysis of each data set:

- heterogeneity in achievement,
- class size,
- heterogeneity by class size interaction,
- teacher nested within heterogeneity by class size interaction,
- achievement level,
- achievement level by heterogeneity interaction,
- achievement level by class size interaction,
- achievement level by heterogeneity by class size interaction, and
- achievement level by teacher nested within heterogeneity by class size interaction.

Sample 1

Variation of Year to Year Population of Students

When conducting the preliminary analysis of the data, it was determined that the attrition rate of students from year to year was of importance to properly analyze the data. Even though the ANOVA excludes student data when only one year of data is present for a particular student, the attrition rate may be of interest in that it too can influence mean class achievement. For the purpose of this study, the number of students reported as "lost" (i.e., attrition) includes those who transferred out of the parish public school system as well as those who were retained in their previous grade level.

Of the 1,463 students who began Kindergarten in the 2000-01 school year, 1,239 of these students returned to the same school system in their first grade year in 2001-02. This represents a loss of 224 students (15.3%) between the Kindergarten and first grade data. Further, a total of 1,650 students began first grade in the 2001-02 school year, a gain of 411 students (24.9%). Given the attrition and gain rates, only 1,239 student data files were used for the ANOVA for the Spring 2001 to Spring 2002 analysis.

For the Spring 2002 to Spring 2003 analysis, data sets were used for students who began first grade in the 2001-02 school year and continued to second grade in the 2002-03 school year. For this set, a total of 1,476 students were enrolled in second grade. Of these students, 1,306 were continuing from the first grade. This totals a loss of 344 (20.8%) students and a gain of 170 (11.5%) students.

The final set of data for Sample 1 consisted of student data from second graders in the 2002-03 school year continuing on to third grade in 2003-04. In all, 1,498 third grade students were enrolled in the parish at the start of the 2003-04 school year. Of these, 1,355 students were

returning from the previous year. This accounts for a loss of 121 (8.1%) students and a gain of 143 (9.5%). Refer to Table 3 for Sample 1 attrition and gain rates.

Table 3
Attrition and Gain Rates for Sample 1

Grade	Total # of Students	Total # of Returning Students	# Lost	% Lost	# Gained	% Gained
K	1,463					
1	1,650	1,239	224	15.3	411	24.9
2	1,476	1,306	344	20.8	170	11.5
3	1,498	1,355	121	8.1	143	9.5

Variable Calculations

Before analyzing data, new variables were calculated for each of the year to year data sets. Due to the fact that the first set of data consisted of four grade levels, three year to year analyses were conducted. For each of these, achievement level ranges as well as heterogeneity in achievement within each class were computed. The achievement level variable was calculated from the average of the Core Battery achievement scores for each student for each two-year period. The mean of the core totals for both years yields an average achievement score by student. Next, average achievement scores were converted into Achievement Levels by grouping scores into quartiles. Achievement levels ranged from 1, the lowest level of average achievement, to 4, the highest level.

Average achievement ranges

For the first analysis (Spring 2001 – Spring 2002), the mean average achievement level was computed and determined to range between 107 and 190 (Refer to Table 4).

Table 4

Average Achievement Range, Spring 2001-Spring 2002

Lowest Average Achievement	Highest Average Achievement	Achievement Level
107.0	134.5	1
135.0	142.5	2
143.0	150.5	3
>150.5		4

For the second year analysis (Spring 2002 – Spring 2003), the mean average achievement level was computed by class and determined to range between 124 and 210 (Refer to Table 5).

Table 5

Average Achievement Range, Spring 2002-Spring 2003

Lowest Average Achievement	Highest Average Achievement	Achievement Level
124.0	151	1
151.5	161	2
161.5	171	3
>171.0		4

For the third analysis of sample one (Spring 2003 – Spring 2004), the mean average achievement level was computed by class and determined to range between 136 and 228 (Refer to Table 6).

Table 6

Average Achievement Range, Spring 2003-Spring 2004

Lowest Average Achievement	Highest Average Achievement	Achievement Level
136.0	169.5	1
170.0	179.5	2
180.0	191.5	3
>191.5		4

Heterogeneity in achievement level ranges

Heterogeneity in achievement was computed for each class. Using each student's mean score of the two years, a class mean achievement level and standard deviation were computed; the classes were then divided into 3 levels based on the standard deviations-- low, moderate, and high, with ½ of classes in the moderate (medium) level and ¼ in each of the extreme levels. Tables 7 through 9 give the ranges for the standard deviations in each of the three Heterogeneity in Achievement levels.

Table 7
Heterogeneity in Achievement, *Spring 2001-Spring 2002*

Class SD Range	Level	n
>/= 11	High	24
8 – 10.999	Moderate	42
<8	Low	25

Table 8
Heterogeneity in Achievement, *Spring 2002-Spring 2003*

Class SD Range	Level	n
>/= 14	High	19
9.51 – 13.99	Moderate	44
<9.5	Low	20

Table 9
Heterogeneity in Achievement, spring 2003-Spring 2004

Class SD Range	Level	n
>/= 16	High	21
11 – 15.999	Moderate	42
<11	Low	18

Value Added Results

Language arts model for Spring 2001-Spring 2002

A mixed model ANOVA was calculated using nine different variables: class size, heterogeneity in achievement, achievement level, heterogeneity in achievement by achievement level, class size by achievement level, class size by heterogeneity in achievement, class size by heterogeneity in achievement by achievement level, teacher nested within class size by heterogeneity, and teacher by achievement level nested within class size by heterogeneity in achievement to determine their effects on Language Arts gain. Of the nine variables, four were found to have a significant ($p < .05$) effect on Language Arts gains in this model. First, a significant main effect for achievement level was found ($F_{(3,880)} = 42.99$, $p < .05$). Second, a significant effect for heterogeneity in achievement level by achievement level interaction was found ($F_{(6,880)} = 4.10$, $p < .05$). Third, a significant effect for heterogeneity by class size interaction was found ($F_{(2,880)} = 4.23$, $p < .05$). Finally, a significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(85, 880)} = 1.89$, $p < .05$; Refer to Table 10).

Table 10
ANOVA Summary Table for Value Added to Language Arts, Spring 2001-Spring 2002

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	50999.75	323	157.89	2.38	0.00	0.4660	0.4660
Intercept	306791.46	1	306791.46	4619.94	0.00		
Class size	5.03	1	5.03	0.08	0.78	0.0001	0.0000
Heterogeneity in ach	51.15	2	25.58	0.39	0.68	0.0009	0.0005
Achievement level	8563.82	3	2854.61	42.99	0.00	0.1278	0.0783
Heterogeneity in ach * ach level	1632.92	6	272.15	4.10	0.00	0.0272	0.0149
Class size * ach level	471.03	3	157.01	2.36	0.07	0.0080	0.0043
Class size * heterogeneity in ach	561.49	2	280.74	4.23	0.01	0.0095	0.0051
Class size * heterogeneity in ach* ach level	93.36	6	15.56	0.23	0.97	0.0016	0.0009
Teacher (Class size * heterogeneity)	10662.21	85	125.44	1.89	0.00	0.1543	0.0974
Teacher * ach level (Class size * heterogeneity in ach)	12656.86	215	58.87	0.89	0.86	0.1780	0.1157
Error	58437.20	880	66.41				
Total	698305.00	1204					
Corrected Total	109436.95	1203					

R Squared = .466 (Adjusted R Squared = .270)

Achievement level was found to be significant in this model. Table 11 gives the means by achievement level. The means table and the Duncan post hoc test (Table 12) show that the higher achieving the student, the more gain there would be in language arts score from kindergarten to first grade.

Table 11
Means of Language Arts Differences for Achievement Level

Dependent Variable:
Language Arts difference

Achlevel	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
1	15.82	0.62	14.60	17.04
2	20.30	0.51	19.29	21.30
3	23.65	0.50	22.68	24.63
4	28.47	0.66	27.18	29.75

Based on modified population marginal mean.

Table 12

Duncan Post Hoc Test for Achievement Level

Ach level	n	Subset Mean 1	2	3	4
1	247	15.66			
2	359		20.40		
3	355			23.46	
4	243				29.24

Note: Means within different subsets are statistically different (p < .05).

Heterogeneity in achievement by achievement level was found to be significant in this model. Table 13 describes the mean language arts gain by the four levels of achievement and three levels of heterogeneity in achievement. Lower-achieving students (achievement levels 1 and 2) performed better in more homogeneous classes whereas higher achieving students (levels 3 and 4) performed better in more heterogeneous classes.

Table 13

Means of Language Arts Differences for Heterogeneity in Achievement by Achievement Level

Dependent Variable:
Language Arts difference

Achlevel	Heterogeneity in Ach	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
1	Low	17.92	1.26	15.44	20.40
	Medium	14.49	0.88	12.76	16.22
	High	16.09	1.20	13.73	18.46
2	Low	21.36	0.86	19.67	23.06
	Medium	19.47	0.74	18.01	20.93
	High	20.59	1.14	18.36	22.83
3	Low	20.76	0.90	19.00	22.52
	Medium	24.54	0.73	23.11	25.96
	High	24.73	1.02	22.73	26.73
4	Low	26.37	1.54	23.35	29.39
	Medium	29.25	0.98	27.32	31.17
	High	28.85	1.05	26.79	30.90

Based on modified population marginal mean.

Heterogeneity in achievement by class size was found to be significant in this model. The following table describes the means for each of the three levels of heterogeneity in achievement by the two levels of class size – small or large (Table 14). Students in smaller classes did better

in language arts when the classes were heterogeneous in ability level whereas students in larger classes required more homogeneous ability grouping.

Table 14

Means of Language Arts Differences for Class Size by Heterogeneity in Achievement

Dependent Variable:
Language Arts difference

Class size	Heterogeneity in Ach	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Small	Low	19.73	0.79	18.18	21.29
	Medium	21.99	0.53	20.95	23.03
	High	23.08	0.64	21.82	24.34
Large	Low	23.20	0.78	21.66	24.74
	Medium	21.64	0.66	20.35	22.93
	High	21.52	1.01	19.54	23.50

Based on modified population marginal mean.

Teacher within class size by heterogeneity in achievement was significant in this model. Because of the large number of teachers, ranges of teacher effects were calculated and the lowest and highest levels of achievement for each class size and heterogeneity in achievement are reported (see Table 15). It seems that the teacher has a significant effect on language arts achievement even when heterogeneity of achievement and class size are taken into account. There is a wide range of teacher effects within each class size by heterogeneity group.

Table 15

Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Language Arts Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	13.10 – 27.33
	Med heterogeneity	12.26 – 27.22
	High heterogeneity	15.67 – 28.08
Large class size	Low heterogeneity	14.42 – 26.79
	Med heterogeneity	11.73 – 29.91
	High heterogeneity	17.40 – 26.38

Measures of effect size in ANOVA are measures of the degree of association between an effect (e.g., a main effect, an interaction, a linear contrast) and the dependent variable. They can be thought of as the correlation between an effect and the dependent variable. Eta squared is the proportion of the total variance that is attributed to an effect. It is calculated as the ratio of the effect variance (SS_{effect}) to the total variance (SS_{total}). One of the problems with Eta squared is that the values for an effect are dependent upon the number and magnitude of other effects. The partial Eta squared is often preferred. It is the proportion of the effect plus error variance that is attributable to the effect. The formula differs from the Eta squared formula in that the denominator includes the SS_{effect} plus the SS_{error} rather than the SS_{total} Note that partial Eta squareds are not additive. Nonetheless, the partial Eta squared of .1543 for the *teacher within class size by heterogeneity* effect can be interpreted as that effect explaining 15.43% of the variability in students' improvement from one year to the next (see Figure 1 below).

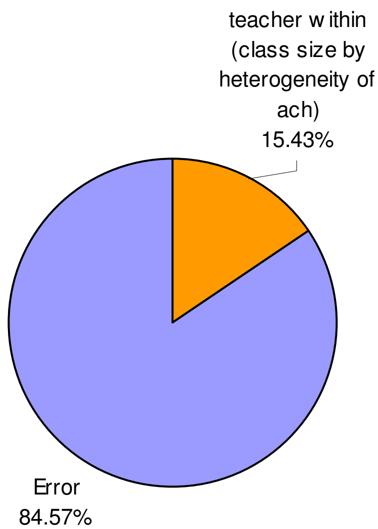

Fig. 1. *Sample 1: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable on language arts improvement from Kindergarten to first grade*

Math model for Spring 2001-Spring 2002

Math gains were determined using the same nine variables established in the Language Arts portion of this model. Of the nine variables, three were found to have a significant effect on Math gains. First, a significant main effect was found for achievement level ($F_{(3,876)} = 42.02$, $p < .05$). Second, a significant effect for heterogeneity by class size interaction was found ($F_{(2,876)} = 3.85$, $p < .05$). Third, a significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(85,876)} = 2.84$, $p < .05$; Refer to Table 16).

Table 16

ANOVA Summary Table for Value Added to Math, Spring 2001-Spring 2002

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	55940.67	323	173.19	2.47	0.00	0.4768	0.4768
Intercept	122315.86	1	122315.86	1745.54	0.00		
Class size	174.57	1	174.57	2.49	0.11	0.0028	0.0015
Heterogeneity in ach	58.45	2	29.23	0.42	0.66	0.0010	0.0005
Achievement level	8833.12	3	2944.37	42.02	0.00	0.1258	0.0753
Heterogeneity in ach * ach level	279.78	6	46.63	0.67	0.68	0.0045	0.0024
Class size * ach level	209.96	3	69.99	1.00	0.39	0.0034	0.0018
Class size * heterogeneity in ach	539.44	2	269.72	3.85	0.02	0.0087	0.0046
Class size * heterogeneity in ach* ach level	157.12	6	26.19	0.37	0.90	0.0026	0.0013
Teacher (Class size * heterogeneity)	16913.31	85	198.98	2.84	0.00	0.2160	0.1442
Teacher * ach level (Class size * heterogeneity in ach)	14115.72	215	65.65	0.94	0.72	0.1870	0.1203
Error	61384.12	876	70.07				
Total	361456.00	1200					
Corrected Total	117324.79	1199					

R Squared = .477 (Adjusted R Squared = .284)

Achievement level was found to be significant for this model. Table 17 gives the means by achievement level and Table 18 gives the Duncan post hoc results. As the achievement level increased so did the mean gain in math score from kindergarten to first grade.

Table 17
Means of Math Differences for Achievement Level

Dependent Variable:
Math difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	7.69	0.64	6.44	8.94
2	12.83	0.52	11.80	13.86
3	16.68	0.51	15.67	17.68
4	19.60	0.68	18.26	20.93

Based on modified population marginal mean.

Table 18
Duncan Post Hoc Test for Achievement Level

Achlevel	n	Subset Mean			
		1	2	3	4
1	247	7.81			
2	357		12.99		
3	355			16.28	
4	241				19.79

Heterogeneity in achievement by class size was found to be significant in this model. The following table describes the means for each of the three levels of heterogeneity in achievement as well as the two levels of class size (Table 19). Students in smaller classes demonstrated more gains when in more heterogeneous classes while students in larger classes demonstrated more gains when in more homogeneous classes.

Table 19
Means of Math Differences for Class Size by Heterogeneity in Achievement

Dependent Variable:
Math difference

Class size	Heterogeneity in Ach	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Small	Low	12.91	0.81	11.31	14.51
	Medium	14.16	0.54	13.09	15.23
	High	15.51	0.66	14.21	16.81
Large	Low	15.89	0.81	14.31	17.47
	Medium	12.85	0.68	11.51	14.19
	High	12.22	1.04	10.18	14.26

Based on modified population marginal mean.

Teacher within class size by heterogeneity in achievement was significant in this model. Because of the large number of teachers, ranges of teacher effects were calculated as well as the lowest and highest levels of achievement for each class size and heterogeneity in achievement are reported. Even when heterogeneity in achievement and class size are taken into account, it appears the teacher again has a significant impact on student achievement. Table 20 describes the wide range of teacher effects within each class size and level of heterogeneity.

Table 20
Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Math Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	2.75 – 21.08
	Med heterogeneity	5.72 – 27.49
	High heterogeneity	5.83 – 23.00
Large class size	Low heterogeneity	7.44 – 23.00
	Med heterogeneity	6.36 – 19.50
	High heterogeneity	9.40 – 16.64

Figure 2 below depicts the proportion of total variance in math gains that is attributable to the teacher within class size by heterogeneity of achievement level for the math model.

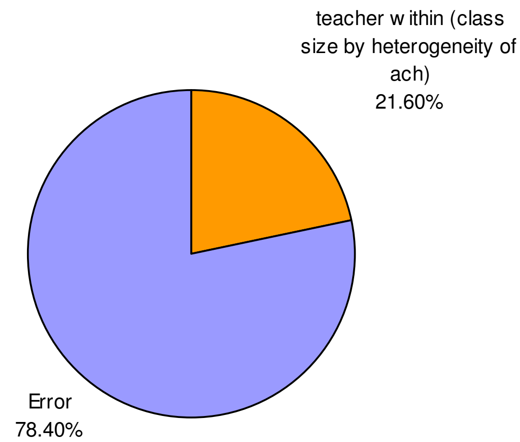

Fig. 2. *Sample 1: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable on math improvement from Kindergarten to first grade*

Language arts model for Spring 2002-Spring 2003

In the second year of the model, a mixed model ANOVA was again calculated to determine the effects of the nine different variables. Of the nine variables, only two were found to have a significant effect on Language Arts gains. A significant main effect for achievement level was found ($F_{(3,970)} = 21.40$, $p < .05$). A second significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(75,970)} = 2.13$, $p < .05$; Refer to Table 21).

Table 21

ANOVA Summary Table for Value Added to Language Arts, Spring 2002-Spring 2003

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	92002.65	295	311.87	2.28	0.00	0.4090	0.4090
Intercept	228081.52	1	228081.52	1663.86	0.00	0.6317	
Class size	470.95	1	470.95	3.44	0.06	0.0035	0.0021
Heterogeneity in ach	796.44	2	398.22	2.91	0.06	0.0060	0.0035
Achievement level	8802.48	3	2934.16	21.40	0.00	0.0621	0.0391
Heterogeneity in ach * ach level	833.28	6	138.88	1.01	0.42	0.0062	0.0037
Class size * ach level	221.83	3	73.94	0.54	0.66	0.0017	0.0010
Class size * heterogeneity in ach	80.69	2	40.35	0.29	0.75	0.0006	0.0004
Class size * heterogeneity in ach* ach level	480.28	6	80.05	0.58	0.74	0.0036	0.0021
Teacher (Class size * heterogeneity)	21930.58	75	292.41	2.13	0.00	0.1416	0.0975
Teacher * ach level (Class size * heterogeneity in ach)	28578.96	197	145.07	1.06	0.29	0.1769	0.1270
Error	132967.59	970	137.08				
Total	776326.00	1266					
Corrected Total	224970.23	1265					

R Squared = .409 (Adjusted R Squared = .229)

Achievement level was found to be significant in this model. Table 22 gives the means by achievement level. Table 23 gives the Duncan post hoc for each level. It is determined by the two tables that the higher the achievement level the more gain in language arts scores from first to second grade.

Table 22
Means of Language Arts Differences for Achievement Level

Dependent Variable:
Language Arts difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	14.76	1.05	12.70	16.81
2	17.37	0.73	15.95	18.79
3	21.54	0.71	20.14	22.94
4	26.67	0.84	25.03	28.32

Based on modified population marginal mean.

Table 23
Duncan Post Hoc Test for Achievement Level

Achlevel	n	Subset Mean			
		1	2	3	4
1	181	13.39			
2	360		16.76		
3	372			22.25	
4	353				27.44

Teacher within class size by heterogeneity of achievement was significant in this model. Once again, it appears that the teacher has a significant effect on language arts achievement even when heterogeneity in achievement and class size are taken into account (see Table 24).

Table 24
Range of Teacher within Class Size by Heterogeneity of Achievement Effects on Language Arts Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	9.25 – 27.93
	Med heterogeneity	12.21 – 33.61
	High heterogeneity	17.48 – 24.25
Large class size	Low heterogeneity	8.67 – 27.36
	Med heterogeneity	11.22 – 30.75
	High heterogeneity	19.13 – 28.98

In this model, 14.16% of the students' achievement can be attributed to teacher within size by heterogeneity (Figure 3).

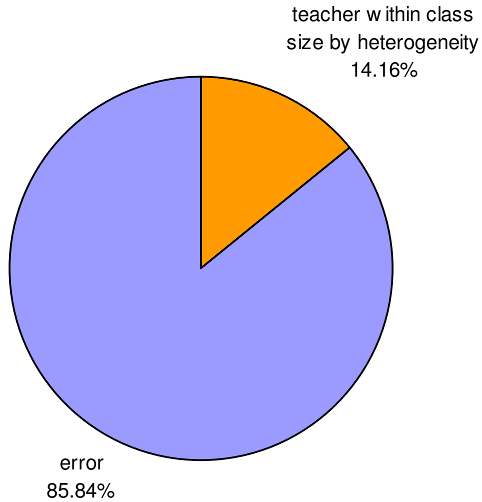

Fig. 3. *Sample 1: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable on language arts improvement from first to second grade*

Math model for Spring 2002-Spring 2003

Of the nine variables calculated, four were found to have a significant effect in this model. First, a significant main effect for class size was found ($F_{(1,969)} = 5.17$, $p < .05$). Second, a significant main effect for heterogeneity in achievement was found ($F_{(2,969)} = 6.21$, $p < .05$). Third, a significant main effect for achievement was found ($F_{(3,969)} = 9.07$, $p < .05$). Finally, a significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(75,969)} = 4.31$, $p < .05$; Refer to Table 25).

Table 25
ANOVA Summary Table for Value Added to Math, Spring 2002-Spring 2003

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	60407.78	294	205.47	2.84	0.00	0.4627	
Intercept	217304.51	1	217304.51	3002.27	0.00	0.7560	
Class size	374.55	1	374.55	5.17	0.02	0.0053	0.0029
Heterogeneity in ach	898.61	2	449.31	6.21	0.00	0.0127	0.0069
Achievement level	1970.47	3	656.82	9.07	0.00	0.0273	0.0151
Heterogeneity in ach * ach level	791.90	6	131.98	1.82	0.09	0.0112	0.0061
Class size * ach level	276.11	3	92.04	1.27	0.28	0.0039	0.0021
Class size * heterogeneity in ach	2.86	2	1.43	0.02	0.98	0.0000	0.0000
Class size * heterogeneity in ach* ach level	186.49	6	31.08	0.43	0.86	0.0027	0.0014
Teacher (Class size * heterogeneity)	23372.76	75	311.64	4.31	0.00	0.2500	0.1790
Teacher * ach level (Class size * heterogeneity in ach)	15240.40	196	77.76	1.07	0.25	0.1785	0.1167
Error	70136.21	969	72.38				
Total	629963.00	1264					
Corrected Total	130543.99	1263					

R Squared = .463 (Adjusted R Squared = .300)

Class size was found to be significant in this model. Table 26 demonstrates that larger class sizes gained significantly more than smaller class sizes in math. This difference was small (1.42).

Table 26
Means of Math Differences for Class Size

Dependent Variable:
Math difference

Class size	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Small	19.10	0.39	18.34	19.86
Large	20.52	0.46	19.62	21.41

Based on modified population marginal mean.

Heterogeneity in achievement level was found to be significant in this model. Classes with a higher level of heterogeneity in achievement showed the most gain in math score from first to second grade as demonstrated by the mean and Duncan pos hoc tables (Tables 27 and 28).

Table 27
Means for Math Differences for Heterogeneity in Achievement

Dependent Variable:
Math difference

Heterogeneity in Ach	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Low	17.77	0.60	16.58	18.95
Medium	19.69	0.39	18.92	20.46
High	21.91	0.68	20.57	23.25

Based on modified population marginal mean.

Table 28
Duncan Post Hoc Test for Heterogeneity in Achievement Level

Heterogeneity in Ach	n	Subset Mean		
		1	2	3
Low	312	17.21		
Medium	751		20.36	
High	201			22.22

Achievement level was found to be significant in this model. Table 29 gives the means by achievement level. The highest achieving students showed the most gain and the lowest achieving students showed the least gain in math score from first to second grade. However, the

two medium levels (Levels 2 and 3) were not significantly different from one another in achievement according to the Duncan's post hoc (Table 30).

Table 29
Means for Math Differences for Achievement Level

Dependent Variable:
Math difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	17.83	0.77	16.32	19.34
2	18.76	0.53	17.73	19.79
3	18.66	0.52	17.64	19.67
4	22.93	0.61	21.74	24.13

Based on modified population marginal mean.

Table 30
Duncan Post Hoc Test for Achievement Level

Achlevel	n	Subset Mean		
		1	2	3
1	179	16.19		
2	359		17.82	
3	372		18.84	
4	354			24.92

Note: Means within different subsets are statistically different ($p < .05$).

Teacher within class size by heterogeneity in achievement was significant in this model. Again, there are a large number of teachers. Due to this fact, ranges of teacher effects were calculated as well as the lowest and highest levels of achievement for each class size and heterogeneity in achievement are reported. There is a wide range of teacher effects within class size by heterogeneity, and it appears that the teacher has a significant effect on math achievement even when nested within class size and heterogeneity (Table 31).

Table 31
Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Math Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	9.53 – 22.88
	Med heterogeneity	7.16 – 32.14
	High heterogeneity	11.10 – 27.83
Large class size	Low heterogeneity	11.07 – 24.60
	Med heterogeneity	12.12 – 30.72
	High heterogeneity	18.06 – 29.71

Figure 4 demonstrates that 25% of the variability in student achievement is attributable to teacher within class size by heterogeneity in achievement.

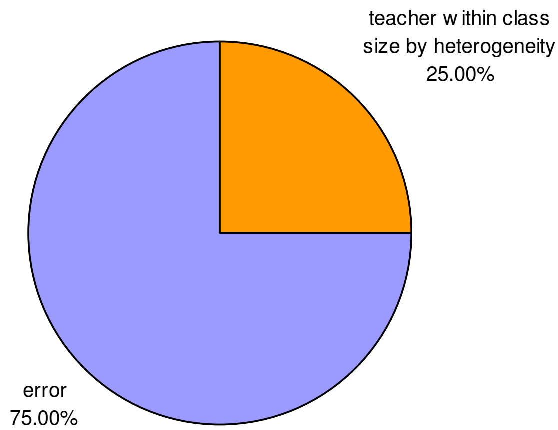

Fig. 4. *Sample 1: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable on math improvement from first to second grade*

Reading model for Spring 2002-Spring 2003

Reading data are collected for students beginning in the first grade. Therefore, this is the first model to include such data. Of the nine variables used in this model, three were found to

have significant effects on Reading gains. First, a significant main effect for achievement level was found ($F_{(3,969)} = 11.72$, $p < .05$). Second, a significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(75,969)} = 1.54$, $p < .05$). Third, a significant effect for teacher by achievement level interaction nested within class size by heterogeneity interaction was found ($F_{(197,969)} = 1.31$, $p < .05$; Refer to Table 32).

Achievement level was found to be significant in this model. Table 33 gives the means by achievement level, and Table 34 gives the Duncan's post hoc results. Higher-achieving students showed the most gains in reading score from first to second grade.

Table 32

ANOVA Summary Table for Value Added to Reading, Spring 2002-Spring 2003

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	50220.46	295	170.24	2.08	0.00	0.3873	0.3873
Intercept	119442.96	1	119442.96	1456.93	0.00		
Class size	248.05	1	248.05	3.03	0.08	0.0031	0.0019
Heterogeneity in ach	128.05	2	64.03	0.78	0.46	0.0016	0.0010
Achievement level	2883.33	3	961.11	11.72	0.00	0.0350	0.0222
Heterogeneity in ach * ach level	422.47	6	70.41	0.86	0.52	0.0053	0.0033
Class size * ach level	135.53	3	45.18	0.55	0.65	0.0017	0.0010
Class size * heterogeneity in ach	3.54	2	1.77	0.02	0.98	0.0000	0.0000
Class size * heterogeneity in ach* ach level	536.80	6	89.47	1.09	0.37	0.0067	0.0041
Teacher (Class size * heterogeneity)	9465.67	75	126.21	1.54	0.00	0.1065	0.0730
Teacher * ach level (Class size * heterogeneity in ach)	21186.28	197	107.54	1.31	0.01	0.2105	0.1634
Error	79441.43	969	81.98				
Total	434245.00	1265					
Corrected Total	129661.88	1264					

R Squared = .387 (Adjusted R Squared = .201)

Table 33

Means of Reading Differences for Achievement Level

Dependent Variable:
Reading difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	9.79	0.81	8.19	11.38
2	14.46	0.56	13.36	15.56
3	16.28	0.55	15.19	17.36
4	18.93	0.65	17.66	20.20

Based on modified population marginal mean.

Table 34
Duncan Post Hoc Test for Achievement Level

Achlevel	n	Subset Mean 1	2	3	4
1	180	9.16			
2	360		14.03		
3	372			15.93	
4	353				19.84

Teacher within class size by heterogeneity in achievement was significant in this model. Even when class size and heterogeneity in achievement are taken into account, it appears that teachers still statistically impact student achievement. Table 35 demonstrates the wide variability in teacher effects within class size and heterogeneity.

Table 35
Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Reading Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	3.42 – 19.19
	Med heterogeneity	7.85 – 23.83
	High heterogeneity	9.15 – 18.90
Large class size	Low heterogeneity	10.32 – 19.66
	Med heterogeneity	11.48 – 26.26
	High heterogeneity	12.48 – 19.35

Figure 5 depicts the proportion of variance in reading gains that is attributed to teacher within class size by heterogeneity.

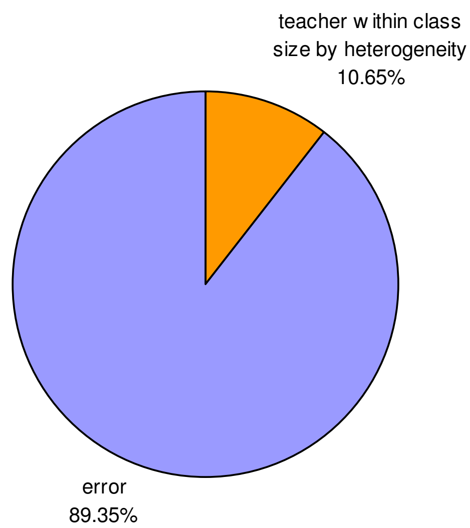

Fig. 5. *Sample 1: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable on reading improvement from first to second grade*

In this model, achievement level by teacher nested within class size by heterogeneity in achievement is significant. Table 36 demonstrates the ranges in teacher effect with the other three variables taken into account. Ranges in teacher effects vary widely even when calculated with achievement level within class size by heterogeneity interaction.

Table 36

Achievement Level by Teacher within Class Size by Heterogeneity in Achievement Effects on Reading Differences

Achlevel	Heterogeneity in Achievement	Class Size	Range of Teacher Effects
1	Low	Small	-1.50 – 35.00
		Large	5.00 – 15.33
	Medium	Small	-12.00 – 28.00
		Large	.40 – 19.00
	High	Small	-1.00 – 17.67
		Large	-2.00 – 20.00
2	Low	Small	7.50 – 25.00
		Large	12.27 – 21.00
	Medium	Small	5.00 – 26.00
		Large	6.67 – 24.00
	High	Small	-1.00 – 30.50
		Large	13.83 – 23.40
3	Low	Small	-4.75 – 21.50
		Large	7.50 – 18.86
	Medium	Small	5.50 – 30.00
		Large	6.63 – 37.00
	High	Small	10.40 – 18.00
		Large	10.00 – 20.75
4	Low	Small	-10.00 – 30.00
		Large	12.00 – 22.42
	Medium	Small	0.00 – 45.00
		Large	15.00 – 28.75
	High	Small	9.00 – 25.83
		Large	17.33 – 24.00

Language arts model for Spring 2003-Spring 2004

The Spring 2003-Spring 2004 model is the third and final longitudinal data set for the first sample in this study. Of the nine variables, four were found to have a significant effect on Language Arts gains in this model. First, a significant main effect for achievement level was found ($F_{(3,1019)} = 13.36$, $p < .05$). Second, a significant effect for class size by heterogeneity in achievement interaction was found ($F_{(2,1019)} = 3.41$, $p < .05$). Third, a significant effect for the three-way interaction of class size by heterogeneity in achievement by achievement level was

found ($F_{(6,1019)} = 2.14$, $p = .05$). Finally, a significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(75,1019)} = 2.55$, $p < .05$; Refer to Table 37).

Table 37
ANOVA Summary Table for Value Added to Language Arts, Spring 2003-Spring 2004

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	94673.53	294	322.02	1.79	0.00	0.3403	0.3403
Intercept	246499.96	1	246499.96	1368.57	0.00		
Class size	163.32	1	163.32	0.91	0.34	0.0009	0.0006
Heterogeneity in ach	768.69	2	384.35	2.13	0.12	0.0042	0.0028
Achievement level	7216.65	3	2405.55	13.36	0.00	0.0378	0.0259
Heterogeneity in ach * ach level	1675.27	6	279.21	1.55	0.16	0.0090	0.0060
Class size * ach level	635.51	3	211.84	1.18	0.32	0.0035	0.0023
Class size * heterogeneity in ach	1227.89	2	613.94	3.41	0.03	0.0066	0.0044
Class size * heterogeneity in ach* ach level	2313.42	6	385.57	2.14	0.05	0.0124	0.0083
Teacher (Class size * heterogeneity)	32969.14	75	439.59	2.44	0.00	0.1523	0.1185
Teacher * ach level (Class size * heterogeneity in ach)	39850.32	196	203.32	1.13	0.13	0.1784	0.1432
Error	183537.10	1019	180.11				
Total	739208.00	1314					
Corrected Total	278210.63	1313					

R Squared = .340 (Adjusted R Squared = .150)

Achievement level was found to be significant in this model. Table 38 gives the means by achievement level. The lowest achieving students (Level 1) demonstrated the lowest amount of gains in Language Arts while the highest achieving students (Level 4) demonstrated the highest gains. However, the two medium levels of achievers (Levels 2 and 3) were not significantly different from one another as determined through the Duncan post hoc test (Table 39).

Table 38

Means of Language Arts Differences for Achievement Level

Dependent Variable:
Language Arts Difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	14.72	1.04	12.67	16.77
2	18.56	0.82	16.94	20.17
3	19.11	0.87	17.40	20.81
4	22.89	0.96	21.00	24.79

Based on modified population marginal mean.

Table 39

Duncan Post Hoc Test for Achievement Level

Ach level	n	Subset Mean		
		1	2	3
1	288	14.41		
2	345		17.67	
3	350		19.65	
4	331			22.62

Heterogeneity in achievement by class size was found to be significant in this model. The following table describes the means for each of the three levels of heterogeneity in achievement as well as the two levels of class size (Table 40). Students in both small and large class sizes gained the most when in more heterogeneous classes. However, students in small classes with a low level of heterogeneity gained more than students in small class sizes with moderate heterogeneity. On the other hand, students in large classes with moderate levels of heterogeneity gained more than those in large class sizes with low levels of heterogeneity in achievement (Table 40).

Table 40

Means of Language Arts Differences for Class Size by Heterogeneity in Achievement

Dependent Variable:
Language Arts Difference

Class size	Heterogeneity in ach	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Small	Low	18.59	1.18	16.26	20.91
	Medium	17.23	0.95	15.38	19.09
	High	19.02	1.03	17.01	21.03
Large	Low	17.90	1.63	14.71	21.09
	Medium	20.23	0.86	18.54	21.92
	High	21.56	1.86	17.90	25.21

Based on modified population marginal mean.

The three-way interaction of class size by heterogeneity in achievement by achievement level was found to be significant in this model. The following table describes the means for each of the three levels of heterogeneity, the two levels of class size, and the four levels of achievement (Table 41). In every achievement level, students in large class sizes gained the most. In the two medium levels of achievement, students in large class sizes with high heterogeneity gained the most. In the lowest achievement level, the students in the moderately heterogeneous large classes gained the most. In the highest achievement level, students in more homogeneously grouped large classes gained the most. In the lowest two achievement levels, large class sizes with low levels of student heterogeneity gained the least. In Levels 3 and 4, students in small classes gained the least. Further, in Level 3, small, moderately heterogeneous classes showed the least gain, and in Level 4, small, low heterogeneously grouped classes gained the least. Therefore, it can be determined that students may generally gain more if grouped heterogeneously except in cases of high achieving students who to be less affected by within-class achievement mix.

Table 41

Means of Language Arts Differences for Class Size by Heterogeneity in Achievement by Achievement Level

Dependent Variable:
Language Arts Difference

Achlevel	Class size	Heterogeneity in ach	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
1	Small	Low	16.02	2.90	10.34	21.70
		Medium	12.07	1.99	8.16	15.98
		High	14.45	2.01	10.50	18.41
	Large	Low	10.29	2.51	5.37	15.20
		Medium	18.19	2.16	13.95	22.42
		High	11.88	4.37	3.30	20.47
2	Small	Low	16.37	1.80	12.84	19.89
		Medium	17.67	1.51	14.70	20.64
		High	20.55	2.14	16.36	24.75
	Large	Low	14.51	3.08	8.47	20.56
		Medium	19.49	1.57	16.42	22.57
		High	23.06	4.22	14.78	31.35
3	Small	Low	23.02	2.20	18.70	27.34
		Medium	15.35	1.80	11.82	18.88
		High	19.67	2.13	15.49	23.84
	Large	Low	17.20	3.22	10.89	23.51
		Medium	19.95	1.44	17.12	22.78
		High	25.76	3.51	18.88	32.65
4	Small	Low	19.26	3.04	13.31	25.22
		Medium	23.59	2.20	19.27	27.91
		High	21.50	1.92	17.72	25.28
	Large	Low	25.99	2.71	20.67	31.31
		Medium	23.29	1.73	19.90	26.68
		High	25.52	2.51	20.58	30.45

Based on modified population marginal mean.

Teacher within class size by heterogeneity in achievement was significant in this model. Ranges of teacher effects were calculated and the lowest and highest levels of achievement for each class size and heterogeneity in achievement are reported. Table 42 provides evidence that teacher effects vary greatly and that teachers impact student achievement even when class size and heterogeneity in achievement are taken into account.

Table 42

Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Language Arts Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	9.02 – 26.33
	Med heterogeneity	2.17 – 27.27
	High heterogeneity	9.49 – 32.63
Large class size	Low heterogeneity	6.75 – 23.49
	Med heterogeneity	9.28 – 30.64
	High heterogeneity	17.71 – 25.65

The proportion of the total variance that is attributable to heterogeneity within class size by heterogeneity is depicted below (Figure 6).

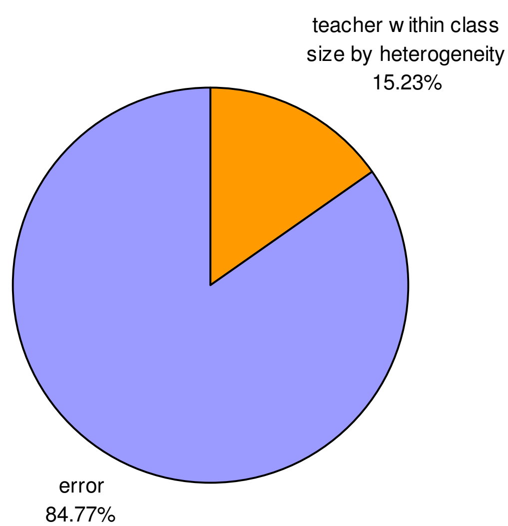

Fig. 6. *Sample 1: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable on language arts improvement from second to third grade*

Math model for Spring 2003-Spring 2004

Of the nine variables used to determine impact on student gains, three were found to be significant in the math model for Spring 2003-2004. First, a significant main effect for achievement level was found ($F_{(3,1015)} = 12.38$, $p < .05$). Second, a significant main effect for class size x heterogeneity interaction was found ($F_{(2,1015)} = 3.60$, $p < .05$). Third, a significant main effect for teacher nested within class size x heterogeneity interaction was found ($F_{(75,1015)} = 2.06$, $p < .05$; Table 43).

Table 43
ANOVA Summary Table for Value Added to Math, Spring 2003-Spring 2004

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	55142.96	293	188.20	1.76	0.00	0.3373	0.3373
Intercept	151314.32	1	151314.32	1417.79	0.00		
Class size	20.23	1	20.23	0.19	0.66	0.0002	0.0001
Heterogeneity in ach	384.91	2	192.45	1.80	0.17	0.0035	0.0024
Achievement level	3964.84	3	1321.61	12.38	0.00	0.0353	0.0243
Heterogeneity in ach * ach level	1198.03	6	199.67	1.87	0.08	0.0109	0.0073
Class size * ach level	70.36	3	23.45	0.22	0.88	0.0006	0.0004
Class size * heterogeneity in ach	769.31	2	384.66	3.60	0.03	0.0071	0.0047
Class size * heterogeneity in ach* ach level	510.30	6	85.05	0.80	0.57	0.0047	0.0031
Teacher (Class size * heterogeneity)	16467.66	75	219.57	2.06	0.00	0.1320	0.1007
Teacher * ach level (Class size * heterogeneity in ach)	19701.37	195	101.03	0.95	0.68	0.1539	0.1205
Error	108326.39	1015	106.73				
Total	465998.00	1309					
Corrected Total	163469.35	1308					

R Squared = .337 (Adjusted R Squared = .146)

Achievement level was found to be significant in this model. Table 44 gives the means by achievement level. The Duncan post hoc test is reported in Table 45. As demonstrated in the following two tables, students gain more as their respective levels of average achievement increases in math score from first to second grade.

Table 44
Means of Math Differences for Achievement Level

Dependent Variable:
Math difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	11.83	0.81	10.25	13.42
2	13.86	0.63	12.62	15.10
3	16.16	0.66	14.86	17.46
4	18.46	0.74	17.00	19.92

Based on modified population marginal mean.

Table 45
Duncan Post Hoc Test for Achievement Level

Ach level	n	Subset Mean			
		1	2	3	4
1	287	10.44			
2	344		13.86		
3	348			16.69	
4	330				19.18

Heterogeneity in achievement by class size was found to be significant in this model. The following table describes the means for each of the three levels of heterogeneity in achievement as well as the two levels of class size (Table 46). Students in large classes with a medium level of heterogeneity in achievement level showed the most gain while students in large classes with low heterogeneity showed the least gain in math score. Small classes did better with low heterogeneity.

Table 46
Means of Math Differences for Class Size by Heterogeneity in Achievement Level

Dependent Variable:
Math difference

Class size	Heterogeneity in ach	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Small	Low	15.42	0.91	13.63	17.21
	Medium	14.48	0.73	13.05	15.90
	High	14.07	0.80	12.51	15.63
Large	Low	14.06	1.25	11.60	16.51
	Medium	16.63	0.66	15.33	17.93
	High	15.32	1.43	12.50	18.13

Based on modified population marginal mean.

Teacher within class size by heterogeneity in achievement was significant in this model. Lowest and highest levels of achievement for each class size and heterogeneity of achievement and the range of teacher effects for each level are reported. Teacher effects vary widely as demonstrated in Table 47.

Table 47
Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Math Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	2.96 – 22.67
	Med heterogeneity	5.78 – 22.21
	High heterogeneity	6.98 – 20.23
Large class size	Low heterogeneity	10.38 – 16.64
	Med heterogeneity	8.25 – 25.52
	High heterogeneity	8.96 – 21.11

Figure 7 gives a pictorial view of the variability in students' improvement in math from one year to the next.

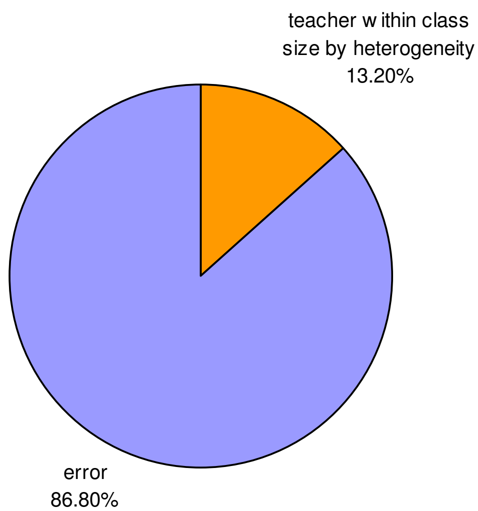

Fig. 7. *Sample 1: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable on math improvement from second to third grade*

Reading model for Spring 2003-Spring 2004

Of the nine variables utilized to determine effects on student achievement, only two were found to have a significant effect on Reading gains in this model. First, a significant main effect for achievement level was found ($F_{(3,1020)} = 3.63$, $p < .05$). Second, a significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(75,1020)} = 1.70$, $p < .05$; Refer to Table 48).

Table 48
ANOVA Summary Table for Value Added to Reading, Spring 2003-Spring 2004

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	38693.00	294	131.61	1.32	0.00	0.2752	0.2752
Intercept	114167.92	1	114167.92	1142.73	0.00		
Class size	146.18	1	146.18	1.46	0.23	0.0014	0.0010
Heterogeneity in ach	73.52	2	36.76	0.37	0.69	0.0007	0.0005
Achievement level	1087.40	3	362.47	3.63	0.01	0.0106	0.0077
Heterogeneity in ach * ach level	167.41	6	27.90	0.28	0.95	0.0016	0.0012
Class size * ach level	107.66	3	35.89	0.36	0.78	0.0011	0.0008
Class size * heterogeneity in ach	93.91	2	46.95	0.47	0.63	0.0009	0.0007
Class size * heterogeneity in ach* ach level	296.66	6	49.44	0.49	0.81	0.0029	0.0021
Teacher (Class size * heterogeneity)	12734.19	75	169.79	1.70	0.00	0.1111	0.0906
Teacher * ach level (Class size * heterogeneity in ach)	19164.45	196	97.78	0.98	0.57	0.1583	0.1363
Error	101905.79	1020	99.91				
Total	370861.00	1315					
Corrected Total	140598.79	1314					

R Squared = .275 (Adjusted R Squared = .066)

Achievement level was found to be significant in this model. Table 49 gives the means by achievement level. The Duncan post hoc test (Table 50) shows that there is no significant difference between Levels 1 and 2 or between Levels 3 and 4. However, there is a significant difference in achievement between the first two levels (1 and 2) and the second two levels (3 and 4). As with all previous models, higher achieving students gain more.

Table 49
Means of Reading Differences for Achievement Level

Dependent Variable:
Reading difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	10.73	0.78	9.19	12.26
2	11.93	0.61	10.73	13.13
3	13.91	0.65	12.64	15.18
4	14.98	0.72	13.57	16.39

Based on modified population marginal mean.

Table 50
Duncan Post Hoc Test for Achievement Level

Ach level	n	Subset Mean	
		1	2
1	288	11.40	
2	345	11.83	
3	350		14.00
4	332		15.48

Teacher within class size by heterogeneity in achievement was significant in this model. As with all of the previous models, the range of teacher effects varies widely within class size by heterogeneity levels (Table 51). It appears that no matter what the class size or heterogeneity level, teachers have a significant effect on student achievement.

Table 51
Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Reading Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	5.56 – 19.39
	Med heterogeneity	2.28 – 17.63
	High heterogeneity	8.30 – 20.19
Large class size	Low heterogeneity	10.67 – 16.84
	Med heterogeneity	6.15 – 20.90
	High heterogeneity	9.69 – 17.95

Just over 11% of the variance in student achievement can be attributed to teacher within class size by heterogeneity (Figure 8).

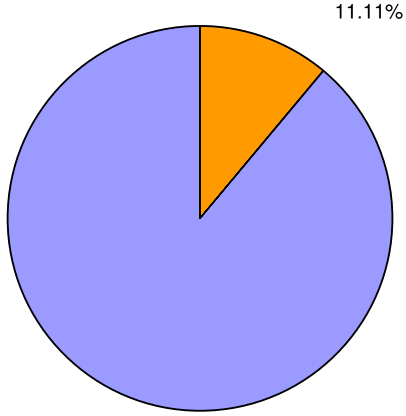

Fig. 8: *Sample 1: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable on reading improvement from second to third grade*

Sample 2

Variation of Year to Year Population of Students

Of the 1,217 students who began Kindergarten in the 2001-02 school year, 1,203 of these students returned to the same school system in their first grade year in 2002-03. This represents a loss of 214 (15.1%) between the Kindergarten and first grade. Furthermore, a total of 1,538 students began first grade in the 2002-03 school year, a gain of 335 (21.8%). Given the attrition and gain rates, only 1,203 student data files were used for the ANOVA for the Spring 2002 to Spring 2003 analysis.

For the Spring 2003 to Spring 2004 analysis, data sets were used for students who began first grade in the 2002-03 school year and continued to second grade in the 2003-04 school year.

For this set, a total of 1,401 students were enrolled in second grade. Of these students, a total of 1,268 were continuing from the first grade. This totals a loss of 270 (17.6%) students and a gain of 133 (9.5%) students. Refer to Table 52 for Sample 2 attrition and gain rates.

Table 52
Attrition and Gain Rates for Sample 2

Grade	Total # of Students	Total # of Returning Students	# Lost	% Lost	# Gained	% Gained
K	1,417					
1	1,538	1,203	214	15.1	335	21.8
2	1,401	1,268	270	17.6	133	9.5

Variable Calculations

Average achievement ranges

For the first analysis (Spring 2002 – Spring 2003), the mean average achievement level was computed and determined to range between 106 and 180 (Refer to Table 53).

Table 53
Average Achievement Range, Spring 2002-Spring 2003

Lowest Average Achievement	Highest Average Achievement	Achievement Level
106.0	136.0	1
136.5	144.0	2
144.5	151.5	3
>151.5		4

For the second year-to-year analysis (Spring 2003 – Spring 2004), the mean average achievement level was computed and determined to range between 126 and 201 (Refer to Table 54).

Table 54
Average Achievement Range, Spring 2003-Spring 2004

Lowest Average Achievement	Highest Average Achievement	Achievement Level
126.0	152.5	1
153.0	162.5	2
163.0	171.5	3
>171.5		4

Heterogeneity in achievement level ranges

For the second sample of data, heterogeneity in achievement was computed for each class. A class mean achievement level and standard deviation were computed using each student's mean composite score for the two years. The classes were then divided into three levels. Tables 55 and 56 gives the ranges for the standard deviations in each of the three Heterogeneity in Achievement levels.

Table 55
Heterogeneity in Achievement, Spring 2002-Spring 2003

Class SD Range	Level	n
>/= 10	High	26
7.6 – 9.9	Moderate	43
<7.6	Low	19

Table 56
Heterogeneity in Achievement, Spring 2003-Spring 2004

Class SD Range	Level	n
>11.99	High	22
9.29 - <12	Moderate	34
<9.3	Low	20

Value Added Results

Language Arts Model for Spring 2002-Spring 2003

A mixed model ANOVA was calculated using nine different variables (established in Sample 1) to determine their effects on student achievement gains. Of the nine variables, four were found to be significant in this model. First, a significant main effect for heterogeneity in achievement was found ($F_{(2,873)} = 3.13$, $p < .05$). Second, a significant main effect for achievement level was found ($F_{(3,873)} = 11.88$, $p < .05$). Third, a significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(80,873)} = 2.62$, $p < .05$). Finally, a significant effect for achievement level by teacher interaction within class size by heterogeneity interaction was found ($F_{(208,873)} = 1.21$, $p < .05$; Refer to Table 57).

Heterogeneity in achievement was found to be significant in this model. Table 58 gives the means by heterogeneity in achievement level. However, the Duncan post hoc test (Table 59) shows that there is no statistical significance between the three levels of heterogeneity in language arts gain.

Table 57

ANOVA Summary Table for Value Added to Language Arts, Spring 2002-Spring 2003

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	68515.44	311	220.31	2.26	0.00	0.4456	0.4456
Intercept	321659.08	1	321659.08	3293.56	0.00		
Class size	242.50	1	242.50	2.48	0.12	0.0028	0.0016
Heterogeneity in ach	611.39	2	305.70	3.13	0.04	0.0071	0.0040
Achievement level	3479.43	3	1159.81	11.88	0.00	0.0392	0.0226
Heterogeneity in ach * ach level	774.21	6	129.03	1.32	0.24	0.0090	0.0050
Class size * ach level	83.09	3	27.70	0.28	0.84	0.0010	0.0005
Class size * heterogeneity in ach	89.72	2	44.86	0.46	0.63	0.0011	0.0006
Class size * heterogeneity in ach* ach level	412.86	6	68.81	0.70	0.65	0.0048	0.0027
Teacher (Class size * heterogeneity)	20440.73	80	255.51	2.62	0.00	0.1934	0.1329
Teacher * ach level (Class size * heterogeneity in ach)	24543.89	208	118.00	1.21	0.04	0.2235	0.1596
Error	85259.80	873	97.66				
Total	779261.00	1185					
Corrected Total	153775.24	1184					

R Squared = .446 (Adjusted R Squared = .248)

Table 58

Means of Language Arts Differences for Heterogeneity in Achievement

Dependent Variable:
Language Arts difference

Heterogeneity in ach	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Low	22.46	0.78	20.93	23.98
Medium	21.85	0.52	20.82	22.88
High	23.78	0.62	22.56	25.01

Based on modified population marginal mean.

Table 59
Duncan Post Hoc Test for Heterogeneity in Achievement

Heterogeneity in ach	n	Subset Mean
		1
Medium	556	22.55
Low	276	22.99
High	353	23.63

Achievement level was found to be significant in this model. Table 60 gives the means by achievement level, and Table 61 gives the results of the Duncan post hoc test. Both of these tables demonstrate that the higher the achievement level of the student the more gain there is in language arts score from kindergarten to first grade.

Table 60
Means of Language Arts Differences for Achievement Level

Dependent Variable:
Language Arts difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	17.65	0.82	16.04	19.26
2	21.67	0.57	20.55	22.79
3	23.52	0.66	22.22	24.81
4	27.72	0.83	26.09	29.34

Based on modified population marginal mean.

Table 61
Duncan Post Hoc Test for Achievement Level

Ach level	n	Subset Mean			
		1	2	3	4
1	239	17.78			
2	378		21.24		
3	329			24.23	
4	239				29.19

Teacher within class size by heterogeneity in achievement is significant in this model. Again, there is a wide range of teacher effects within each class size by heterogeneity group. It

appears that the teacher has a significant impact on language arts achievement even when the other two variables are included (Table 62).

Table 62
Range of Teacher within Class Size by Heterogeneity of Achievement Effects on Language Arts

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	14.86 – 28.39
	Med heterogeneity	11.54 – 37.78
	High heterogeneity	12.44 – 28.92
Large class size	Low heterogeneity	11.86 – 34.57
	Med heterogeneity	10.93 – 28.73
	High heterogeneity	13.99 – 33.73

Figure 9 demonstrates the proportion of the total variance in language arts gains that is attributable to teacher within class size by heterogeneity.

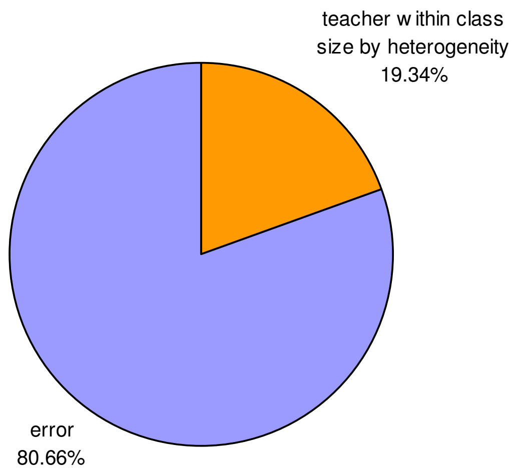

Fig. 9: *Sample 2: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable on language arts improvement from Kindergarten to first grade*

Achievement Level by Teacher nested within class size by heterogeneity of achievement in this model. Table 63 displays the large variation in teacher effects. Teachers impact student

achievement even when achievement level, class size, and heterogeneity are factored into the model.

Table 63
Range of Achievement Level by (Teacher within Class Size by Heterogeneity of Achievement) Effects on Language Arts

Achlevel	Heterogeneity in Achievement	Class Size	Range of Teacher Effects
1	Low	Small	8.75 – 29.00
		Large	6.00 – 35.00
	Medium	Small	-3.00 – 35.00
		Large	0.00 – 27.00
	High	Small	12.50 – 31.00
		Large	10.00 – 25.00
2	Low	Small	13.63 – 27.50
		Large	12.43 – 38.50
	Medium	Small	8.00 – 45.00
		Large	11.00 – 30.20
	High	Small	11.75 – 44.67
		Large	14.20 – 28.50
3	Low	Small	15.25 – 34.00
		Large	12.20 – 36.75
	Medium	Small	2.00 – 36.00
		Large	5.50 – 34.20
	High	Small	12.00 – 41.33
		Large	13.75 – 33.00
4	Low	Small	14.40 – 32.50
		Large	-6.00 – 38.50
	Medium	Small	5.00 – 45.00
		Large	20.50 – 38.67
	High	Small	10.00 – 41.00
		Large	10.00 – 46.50

Math model for Spring 2002-Spring 2003

Of the nine variables used in this model, four were found to have a significant effect on Math gains. First, a significant main effect for achievement level was found ($F_{(3,873)} = 8.06$, $p < .05$). Second, a significant effect for heterogeneity by class size interaction was found ($F_{(2,873)} = 6.28$, $p < .05$). Third, a significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(80,873)} = 2.12$, $p < .05$). Last, a significant effect for achievement level by

teacher interaction nested within class size by heterogeneity interaction was found ($F_{(206,873)}$ = 1.31, $p < .05$; Refer to Table 64).

Table 64
ANOVA Summary Table for Value Added to Math, Spring 2002-Spring 2003

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	36551.76	309	118.29	2.15	0.00	0.4322	0.4322
Intercept	172141.19	1	172141.19	3129.51	0.00		
Class size	70.40	1	70.40	1.28	0.26	0.0015	0.0008
Heterogeneity in ach	149.72	2	74.86	1.36	0.26	0.0031	0.0018
Achievement level	1329.80	3	443.27	8.06	0.00	0.0269	0.0157
Heterogeneity in ach * ach level	332.74	6	55.46	1.01	0.42	0.0069	0.0039
Class size * ach level	40.48	3	13.49	0.25	0.86	0.0008	0.0005
Class size * heterogeneity in ach	690.41	2	345.20	6.28	0.00	0.0142	0.0082
Class size * heterogeneity in ach* ach level	219.15	6	36.52	0.66	0.68	0.0045	0.0026
Teacher (Class size * heterogeneity)	9307.05	80	116.34	2.12	0.00	0.1623	0.1100
Teacher * ach level (Class size * heterogeneity in ach)	14873.69	206	72.20	1.31	0.01	0.2365	0.1759
Error	48020.12	873	55.01				
Total	426051.00	1183					
Corrected Total	84571.88	1182					

R Squared = .432 (Adjusted R Squared = .231)

Achievement level was found to be significant in this model. Table 65 gives the means by achievement level and Table 66 reports the results of the Duncan post hoc test. The lowest achieving students show the least gain and the highest achieving students show in the highest gain in math score from kindergarten to first grade. However, the two middle levels (Levels 2 and 3) did not demonstrate significant differences in gains from one another.

Table 65
Means of Math Differences for Achievement Level

Dependent Variable:
Math difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	13.58	0.62	12.37	14.79
2	16.69	0.43	15.85	17.53
3	17.51	0.49	16.54	18.48
4	18.83	0.62	17.61	20.05

Based on modified population marginal mean.

Table 66
Duncan Post Hoc Test for Achievement Level

Ach level	n	Subset Mean		
		1	2	3
1	238	13.37		
2	378		16.97	
3	329		17.50	
4	238			19.93

Heterogeneity in achievement by class size was found to be significant for this model. The following table describes the means for each of the three levels of heterogeneity in achievement as well as the two levels of class size (Table 67). It appears that students in smaller class sizes gain more when placed in highly heterogeneous groups while students in larger class sizes gain more when placed in more homogeneous groups.

Table 67
Means of Math Differences for Class Size by Heterogeneity in Achievement

Dependent Variable:
Math difference

Class size	Heterogeneity in ach	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Small	Low	16.53	0.75	15.06	18.01
	Medium	15.51	0.49	14.55	16.48
	High	17.61	0.56	16.52	18.71
Large	Low	19.06	0.92	17.26	20.87
	Medium	17.41	0.62	16.20	18.63
	High	15.51	0.85	13.85	17.17

Based on modified population marginal mean.

Teacher within class size by heterogeneity in achievement is significant in this model. Table 68 demonstrates that teachers affect student gains no matter what the class size or heterogeneity in achievement level.

Table 68
Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Math Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	11.75 – 20.51
	Med heterogeneity	7.50 – 23.47
	High heterogeneity	13.87 – 23.61
Large class size	Low heterogeneity	13.18 – 24.68
	Med heterogeneity	13.64 – 22.08
	High heterogeneity	10.98 – 21.44

Figure 10 gives a pictorial view of the proportion of variance in math gain attributed to teacher within class size by heterogeneity.

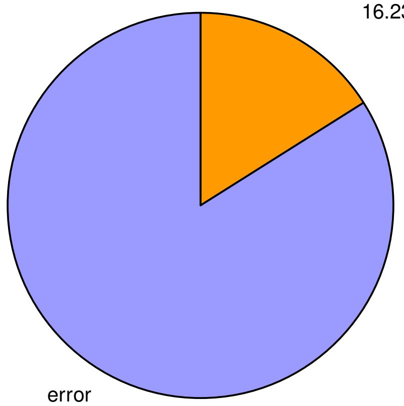

Fig. 10: *Sample 2: Effect size (partial Eta squared) for the* **"***teacher within class size by heterogeneity" variable on math improvement from Kindergarten to first grade*

Achievement level by teacher nested within class size by heterogeneity in achievement was significant in this model. Table 69 breaks down the range of teacher effects by class size by heterogeneity in achievement by achievement level. It appears that teachers impact student achievement even when the other variables are calculated into the model. There is a wide range of variability in all levels of this model.

Table 69

Range of Achievement Level by (Teacher within Class Size by Heterogeneity in Achievement) Effects on Math Differences

Achlevel	Heterogeneity in Achievement	Class Size	Range of Teacher Effects
1	Low	Small	10.00 – 18.44
		Large	12.20 – 20.00
	Medium	Small	4.00 – 20.50
		Large	6.25 – 20.00
	High	Small	4.17 – 22.00
		Large	7.00 – 14.50
2	Low	Small	10.86 – 26.00
		Large	11.33 – 28.29
	Medium	Small	8.60 – 27.67
		Large	11.00 – 21.50
	High	Small	2.33 – 23.20
		Large	10.50 – 23.50
3	Low	Small	11.22 – 26.00
		Large	15.20 – 23.75
	Medium	Small	-2.25 – 29.67
		Large	10.00 – 25.50
	High	Small	11.00 – 29.00
		Large	10.00 – 27.00
4	Low	Small	9.00 – 21.00
		Large	14.00 – 26.67
	Medium	Small	2.00 – 36.00
		Large	11.00 – 26.25
	High	Small	10.00 – 29.00
		Large	7.00 – 26.83

Language arts model for Spring 2003-Spring 2004

Spring 2003-Spring 2004 for Sample 2 is the last of the five data sets utilized in this study. The Language Arts model begins the second set of analyses for Sample 2. Again, nine variables were used. Of the nine, only two were found to have a significant effect on Language Arts gain in this model. First, a significant main effect for achievement level was found ($F_{(3,952)}$ = 51.39, $p < .05$). Second, a significant effect for teacher nested within class size by heterogeneity was found ($F_{(70,952)}$ = 2.37, $p < .05$; Refer to Table 70).

Table 70
ANOVA Summary Table for Value Added to Language Arts, Spring 2003-Spring 2004

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	90442.20	281	321.86	2.47	0.00	0.4218	0.4218
Intercept	253902.83	1	253902.83	1950.07	0.00		
Class size	199.67	1	199.67	1.53	0.22	0.0016	0.0009
Heterogeneity in ach	27.92	2	13.96	0.11	0.90	0.0002	0.0001
Achievement level	20073.31	3	6691.10	51.39	0.00	0.1394	0.0936
Heterogeneity in ach * ach level	119.80	6	19.97	0.15	0.99	0.0010	0.0006
Class size * ach level	217.44	3	72.48	0.56	0.64	0.0018	0.0010
Class size * heterogeneity in ach	108.62	2	54.31	0.42	0.66	0.0009	0.0005
Class size * heterogeneity in ach* ach level	1185.98	6	197.66	1.52	0.17	0.0095	0.0055
Teacher (Class size * heterogeneity)	21630.68	70	309.01	2.37	0.00	0.1486	0.1009
Teacher * ach level (Class size * heterogeneity in ach)	23136.78	188	123.07	0.95	0.68	0.1573	0.1079
Error	123952.10	952	130.20				
Total	747931.00	1234					
Corrected Total	214394.31	1233					

R Squared = .422 (Adjusted R Squared = .251)

Achievement level was found to be significant in this model. Table 71 gives the means by achievement level, and Table 72 shows the results of the Duncan post hoc test. Higher-achieving students showed the most gain in language arts score from first to second grade. It can be assumed from this model that the higher the achievement level of the student, the more gain there would be in achievement.

Table 71
Means of Language Arts Differences for Achievement Level

Dependent Variable:
Language Arts difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	11.38	1.00	9.41	13.35
2	17.21	0.73	15.78	18.64
3	22.11	0.71	20.73	23.50
4	28.27	0.79	26.72	29.82

Based on modified population marginal mean.

Table 72
Duncan Post Hoc Test for Achievement Level

Ach level	n	Subset Mean			
		1	2	3	4
1	192	11.53			
2	344		17.37		
3	360			22.43	
4	338				27.80

Teacher within class size by heterogeneity in achievement was significant in this model. Table 73 gives the ranges of teacher effects. The lowest and highest levels of achievement for each class size and heterogeneity in achievement are reported. This model provides evidence that teachers impact student achievement whether students are in small or large classes and whether in low, medium, or high heterogeneity of achievement levels.

Table 73
Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Language Arts Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	12.29 – 38.75
	Med heterogeneity	11.77 – 29.59
	High heterogeneity	10.73 – 28.54
Large class size	Low heterogeneity	11.48 – 29.50
	Med heterogeneity	11.11 – 28.34
	High heterogeneity	13.03 – 25.78

The proportion of the total variance of student achievement attributable to teacher within class size by heterogeneity is depicted below (Figure 11).

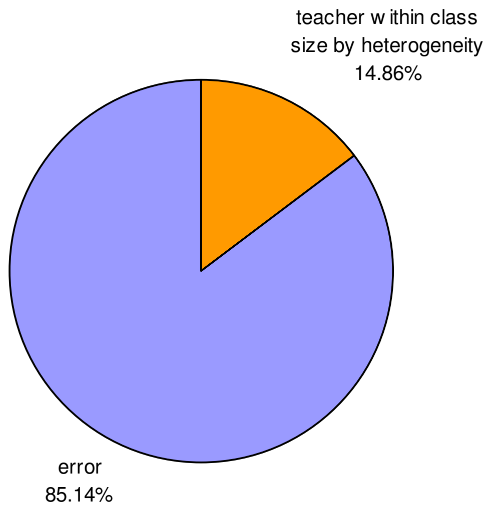

Fig. 11: *Sample 2: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable language arts improvement from first to second grade*

Math model for spring 2003-spring 2004

Of the nine variables, three were found to have a significant effect on Math gains in this model. First, a significant main effect for heterogeneity in achievement was found ($F_{(2,948)} = 3.86$, $p < .05$). Second, a significant main effect for achievement level was found ($F_{(3,948)} = 38.47$, $p < .05$). Third, a significant effect for teacher nested within class size heterogeneity interaction was found ($F_{(70,948)} = 4.60$, $p < .05$; Refer to Table 74).

Table 74
ANOVA Summary Table for Value Added to Math, Spring 2003-Spring 2004

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	55774.12	281	198.48	3.46	0.00	0.5064	0.5064
Intercept	179744.11	1	179744.11	3134.74	0.00		
Class size	6.16	1	6.16	0.11	0.74	0.0001	0.0001
Heterogeneity in ach	442.26	2	221.13	3.86	0.02	0.0081	0.0040
Achievement level	6617.04	3	2205.68	38.47	0.00	0.1085	0.0601
Heterogeneity in ach * ach level	641.50	6	106.92	1.86	0.08	0.0117	0.0058
Class size * ach level	36.88	3	12.29	0.21	0.89	0.0007	0.0003
Class size * heterogeneity in ach	157.47	2	78.74	1.37	0.25	0.0029	0.0014
Class size * heterogeneity in ach* ach level	143.39	6	23.90	0.42	0.87	0.0026	0.0013
Teacher (Class size * heterogeneity)	18482.48	70	264.04	4.60	0.00	0.2537	0.1678
Teacher * ach level (Class size * heterogeneity in ach)	12251.19	188	65.17	1.14	0.12	0.1839	0.1112
Error	54357.68	948	57.34				
Total	467372.00	1230					
Corrected Total	110131.80	1229					

R Squared = .506 (Adjusted R Squared = .360)

Heterogeneity in achievement was found to be significant in this model. Table 75 describes the means for the levels of heterogeneity in achievement. The Duncan post hoc test (Table 76) shows that there is no significant difference between the low and medium levels of heterogeneity, but there is a significant difference between high heterogeneity and low/medium levels. Gains were higher for the high heterogeneity group.

Table 75
Means of Math Differences for Heterogeneity in Achievement

Dependent Variable:
Math difference

Heterogeneity in ach	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	15.07	0.54	14.00	16.13
2	16.57	0.39	15.80	17.35
3	17.91	0.49	16.96	18.87

Based on modified population marginal mean.

Table 76
Duncan Post Hoc Test for Heterogeneity in Achievement Level

Hetero in ach	n	Subset Mean	
		1	2
1	316	15.95	
2	569	16.43	
3	345		19.06

Achievement level was found to be significant in this model. Table 77 gives the means by achievement level. The means table and Duncan post hoc test (Table 78) show that the higher achieving the student, the more gains there would be in math score from second to third grade.

Table 77
Means of Math Differences for Achievement Level

Dependent Variable:
Math difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	13.01	0.67	11.69	14.33
2	14.70	0.48	13.75	15.65
3	16.29	0.47	15.37	17.21
4	22.10	0.52	21.07	23.12

Based on modified population marginal mean.

Table 78
Duncan Post Hoc Test for Achievement Level

Ach level	n	Subset Mean 1	2	3	4
1	190	12.71			
2	343		14.11		
3	359			16.73	
4	338				22.78

Teacher within class size by heterogeneity in achievement was significant in this model. Table 79 gives the ranges of teacher effects. The lowest and highest levels of achievement for each class size and heterogeneity in achievement are reported. From the data provided, it can be assumed that teachers impact students in each class level as well as heterogeneity level.

Table 79
Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Math Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	5.04 – 36.58
	Med heterogeneity	8.51 – 24.50
	High heterogeneity	8.42 – 25.67
Large class size	Low heterogeneity	7.71 – 22.58
	Med heterogeneity	12.05 – 21.69
	High heterogeneity	13.46 – 23.07

The effect size for the teacher within class size by heterogeneity is depicted in Figure 12.

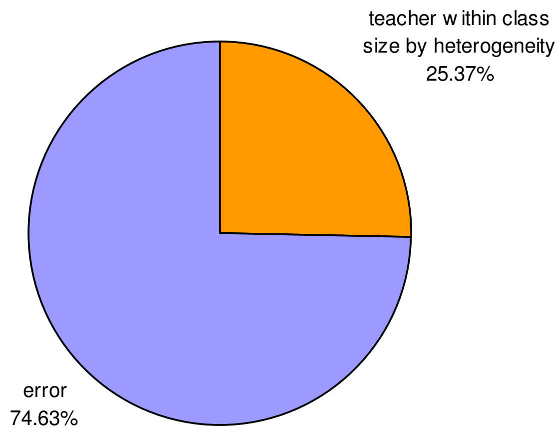

Fig. 12: *Sample 2: Effect size (partial Eta squared) for the "teacher within class size by heterogeneity" variable math improvement from first to second grade*

Reading model for Spring 2003-Spring 2004

The Reading model is the last of the data sets. Of the nine variables, three were found to have significant effects on Reading gains in this model. First, a significant main effect for achievement level was found ($F_{(3,951)} = 34.80$, $p < .05$). Second, a significant effect for class size by achievement level interaction was found ($F_{(3,951)} = 2.57$, $p = .05$). Third, a significant effect for teacher nested within class size by heterogeneity interaction was found ($F_{(70,951)} = 2.07$, $p < .05$; Refer to Table 80).

Table 80
ANOVA Summary Table for Value Added to Reading, Spring 2003-Spring 2004

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Eta Squared
Corrected Model	46890.78	281	166.87	1.91	0.00	0.3603	0.3603
Intercept	195296.23	1	195296.23	2230.80	0.00		
Class size	90.86	1	90.86	1.04	0.31	0.0011	0.0007
Heterogeneity in ach	35.79	2	17.90	0.20	0.82	0.0004	0.0003
Achievement level	9140.94	3	3046.98	34.80	0.00	0.0989	0.0702
Heterogeneity in ach * ach level	790.20	6	131.70	1.50	0.17	0.0094	0.0061
Class size * ach level	675.32	3	225.11	2.57	0.05	0.0080	0.0052
Class size * heterogeneity in ach	143.68	2	71.84	0.82	0.44	0.0017	0.0011
Class size * heterogeneity in ach* ach level	893.05	6	148.84	1.70	0.12	0.0106	0.0069
Teacher (Class size * heterogeneity)	12675.75	70	181.08	2.07	0.00	0.1321	0.0974
Teacher * ach level (Class size * heterogeneity in ach)	16421.63	188	87.35	1.00	0.50	0.1647	0.1262
Error	83255.80	951	87.55				
Total	515333.00	1233					
Corrected Total	130146.59	1232					

R Squared = .360 (Adjusted R Squared = .171)

Achievement level was found to be significant in this model. Table 81 gives the means by achievement level, and the Duncan post hoc test (Table 82) confirms the results of the means tables. Higher-achieving students showed the most gain in reading score from first to second grade.

Table 81
Means of Reading Differences for Achievement Level

Dependent Variable:
Reading difference

Achlevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	12.39	0.82	10.78	14.01
2	15.88	0.60	14.71	17.05
3	18.57	0.58	17.44	19.71
4	22.65	0.65	21.38	23.93

Based on modified population marginal mean.

Table 82
Duncan Post Hoc Test for Achievement Level

Achlevel	n	Subset Mean			
		1	2	3	4
1	192	12.53			
2	344		15.48		
3	359			18.34	
4	338				22.12

Class size by achievement level was found to be significant in this model. In this model, the lowest achievement level (Level 1) showed the most gain in reading score when placed in a small class size. However, the three remaining achievement levels showed slightly more gain in reading score when in larger class sizes (Table 83).

Table 83
Means of Reading Differences for Class Size by Achievement Level

Dependent Variable:
Reading difference

Achlevel	Class size	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	Small	13.65	1.01	11.66	15.64
	Large	10.11	1.41	7.34	12.88
2	Small	15.13	0.74	13.68	16.58
	Large	17.04	1.00	15.07	19.00
3	Small	18.48	0.83	16.86	20.10
	Large	18.72	0.73	17.29	20.15
4	Small	22.63	0.95	20.77	24.49
	Large	22.69	0.80	21.12	24.25

Teacher within class size by heterogeneity in achievement was significant in this model. Table 84 gives the ranges of teacher effects. As with every other model in this study, the effects of teachers are evident even when nested within class size by heterogeneity interaction.

Table 84
Range of Teacher within Class Size by Heterogeneity in Achievement Effects on Reading Differences

Class Size	Heterogeneity in Ach	Range of teacher effects
Small class size	Low heterogeneity	5.04 – 31.17
	Med heterogeneity	13.43 – 20.83
	High heterogeneity	10.58 – 25.05
Large class size	Low heterogeneity	14.98 – 20.37
	Med heterogeneity	12.41 – 26.63
	High heterogeneity	10.70 – 21.28

Just over 13% of the variance in student achievement gain is attributable to teacher within class size by heterogeneity (Figure 13).

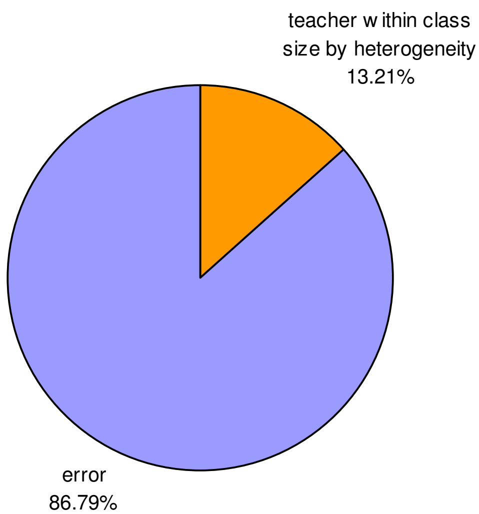

Fig. 13: *Sample 2: Effect size (partial Eta squared) for the* **"***teacher within class size by heterogeneity" variable reading improvement from first to second grade*

Summary

Thirteen analyses were conducted for the nine individual variables and three dependent variables in the five separate sets of data. The following displays the number of significant findings out of the thirteen analyses:

- 13 out of 13 Teacher nested within Class Size x Heterogeneity in Achievement
- 13 out of 13 Achievement Level
- 5 out of 13 Class Size x Heterogeneity in Achievement
- 3 out of 13 Heterogeneity in Achievement

- 3 out of 13 Teacher x Achievement Level nested within Class Size x Heterogeneity in Achievement
- 1 out of 13 Class Size
- 1 out of 13 Heterogeneity in Achievement x Achievement Level
- 1 out of 13 Class Size x Achievement Level
- 1 out of 13 Class Size x Heterogeneity in Achievement x Achievement Level

After analyzing the data, it can be determined that teachers and achievement level are the two most important variables affecting student achievement. All of the data sets showed a wide variance in student gain when using the teacher nested within class size by heterogeneity.

The four levels of achievement were found to have differential impact on student gains. In all cases, the relationship was linear, with higher-achieving students posting higher gains than lower-achieving students.

One other possible emerging pattern was found with heterogeneity in achievement by class size. In three of the five significant analyses, heterogeneously grouped classes fared better when classes were small. Two of these findings were from math classes and one from language arts. One other significant result in a math class found the inverse. Lower heterogeneity was better with small classes. The final significant result from a language arts class found that small classes with neither high nor low heterogeneity in achievement did best. Therefore, no definitive conclusions can be determined in regard to heterogeneity by class size.

Heterogeneity in achievement as a main effect was significant in three analyses. In two of these cases, both in math classes, the higher heterogeneity groups had larger gains. The other case, from a language arts class, showed that the higher and lower heterogeneity groups did better than the moderate heterogeneity group, but the differences were very small.

The one class size main effect showed a very weak advantage for the large class. The one significant achievement by heterogeneity interaction effect (from a language arts class) showed that lower-achieving students gained more in less heterogeneous classes. Similarly, the one significant effect for achievement by class size interaction (from a reading class) showed that lower-achieving students performed best in small classes.

These findings are related to earlier value added research findings. The results will be discussed in detail in Chapter 5.

CHAPTER FIVE

DISCUSSION

Introduction

For the present study, data were collected from one school district in Louisiana to longitudinally track student gains in language arts, math, and reading that could be attributable to teacher effects after accounting for prior achievement, average class achievement, heterogeneity of class achievement, and class size. The data included standard scores for two separate sets of students. The first set of data consisted of students who began Kindergarten in the 2000-2001 school year through third grade in the 2003-2004 school year. The second set of data contained students who began Kindergarten in the 2001-2002 school year through their second grade year in 2003-2004.

Using the data, 13 year-to-year analyses were conducted using mixed model ANOVAs. Nine independent variables, heterogeneity in achievement, class size, heterogeneity by class size interaction, teacher nested within heterogeneity by class size interaction, achievement level, achievement level by heterogeneity interaction, achievement level by class size interaction, achievement level by heterogeneity by class size interaction, and achievement level by teacher nested within heterogeneity by class size interaction, were used to determine their individual and interactive effects on the mean differences in student achievement in three subject areas (language arts, math, and reading).

Limitations of the present study included three areas. First, the population available for the present study was relatively small considering that the work of Wright, Horn, and Sanders

(1997) consisted of more than 65,000 student records. The present study included approximately 3,000 students. Second, Wright et al. used two school systems whereas only one school system was used in this model. Third, in Sanders' model, Wright et al. used scores from students in second grade for the baseline data, while the baseline data for this study was obtained from Kindergarten students.

Conclusions

Teacher Effects

The profound impact of the teacher on student achievement was confirmed by the range in effects in each model when teacher was nested within class size by heterogeneity interaction. The findings of this study remain consistent with those of Wright et al. (1997). The teacher was found in that study to significantly impact student achievement in 30 out of 30 analyses, and was found, in 20 of the 30 analyses, to have a larger effect size than any other factor (Wright et al.). Thomas and Stockton (2003) revealed that students in low socioeconomic classrooms demonstrated significantly less gain on the Norm-referenced Assessment Program in Texas. However, the variance became larger once the teacher was added into the variable because of the influence of the teacher on student achievement. In the present study, the teacher within class size by heterogeneity effect was significant in 13 of 13 analyses. In 10 of 13 analyses, the eta squared for the teacher nested within class size by heterogeneity was higher than that of any other significant F result. In the three cases with higher eta squared, these were also based on teacher effects (achievement by teacher nested within class size by heterogeneity). These findings indicate that there is, in reality, no comparison between the magnitude of teacher effects and the other significant variables on student achievement. Furthermore, the consistency of the findings stands out as quite unusual and leaving no doubt that the results are beyond chance.

Teachers are the most important factor in student achievement when all environmental variables are controlled.

The following Table (85) describes the grade level, percentile ranks, standard scores, and grade equivalents for the ITBS (Hoover et al., 2003b). Due to the fact that Kindergarten was used as the baseline and is the first grade to take the assessment, only the median standard score can be reported.

Table 85
ITBS Standard Scores with Percentile Ranks and Grade Equivalents

Grade	Scale Scores		Percentile Ranks		Grade Equivalents	
K		(130)				
1	142 - 161	(150)	25 - 75	(50)	1.5 - 2.4	(1.8)
2	156 - 182	(168)	25 - 75	(50)	2.1 - 3.6	(2.8)
3	170 - 204	(185)	25 - 75	(50)	2.9 - 5.1	(3.8)

The standard score is used to describe the location of the student on a continuum of scores. Percentile ranks demonstrate the ranking of the student in comparison with other students. In other words, if a student scores in the 75th percentile, that student has scored equal to or higher than 75% of the other students in the comparison. Table 85 demonstrates the range of typical scores for students at each grade level, Kindergarten through third. The number in parentheses is the median score within the range. Finally, grade equivalents describe the location of a student in terms of grade levels and months. For example, 1.5 can be interpreted as the fifth month of first grade (Hoover et al.). Grade equivalents and percentile ranks are based on standard scores.

In this study, ranges in teacher effects within class size by heterogeneity in achievement in reading achievement varied widely from first to second grade in Spring 2002-Spring 2003 (Data from Table 35). Student reading gains in small classes with low heterogeneity in achievement ranged from 3.42 to 19.19. If two students began the year with the median scale score of 150, one would finish the year with a scale score of 153.42 for the least effective teacher

and the other with a scale score of 169.19 for the most effective teacher. Their ending grade level equivalents would be either 2.0 (20th percentile) or 2.9 (51st percentile), depending on the teacher. A second example would have one student ending the year with a 2.2 grade level equivalent (29th percentile) or 3.1 grade level (62nd percentile), again, depending upon the teacher.

Achievement Level

Prior achievement of the student was a significant factor in student academic gain, with higher achieving students consistently gaining more across all models. Wright et al. (1997) found that out of 30 analyses of achievement level, 26 were found to have a significant effect. Furthermore, 10 out of the 30 were found to have the largest effect size. Unlike this study, Wright et al. found no single pattern in which level of prior achievement affected students consistently. The lowest achieving group in the Sanders' study had the most gains 12 times, middle groups eight times, and highest group six times. The least gains occurred 15 times in the highest achieving group, six times in the middle groups, and five times in the lowest group (Wright et al.).

Wright et al. (1997) suggested a concern that a common pattern appears to be that high achieving students made the least gains. Educators often assume that higher achieving students will excel no matter what. Wright et al.believe that many teachers are teaching to the average and below average students, not allowing enrichment for the higher achievers. Conversely, regression to the mean would suggest that higher achieving students would demonstrate the least amount of gain; the opposite was a consistent finding in this study. Several propositions can be raised from these findings.

Proposition 1. Higher quality teachers are often given the highest achieving students. If the best teachers would be assigned to teach the lower achieving students, evidence suggests that the achievement gap could be decreased and even closed (Haycock, 1998). Higher quality teachers tend to higher achieving schools as they gain more experience.

Proposition 2. More resources are often provided to schools with higher achieving students. Gustafson (2002) states that poorer children must be given the opportunity to experience enriching activities in school. Libraries with a staff that demonstrates the excitement of reading should be a primary focus. Teachers need the freedom to expose students to music, art, plays, and field trips. However, the mandates of the standards-based movement hinder the teacher’s ability to provide such experiences (Gustafson). Greenwald et al. (1996) suggest that there is a positive relationship between resources and student achievement. They suggest that an increase in per-pupil expenditure would increase teacher salaries, assisting in retaining more experienced and educated teachers. However, Okpala et al. (2001) determined that expenditures are not correlated to increase educational outcomes for students. Furthermore, Sutton and Soderstrom (1999) indicate that expenditures make no difference to the student because most of the money allocated funnels into personnel costs. Also, Greenwald et al. noted that only small increases in student achievement would be gained by using the expenditures to reduce class size. No clear conclusion may be drawn due to the inconsistencies in the research in this area (Okpala et al.). However, if teachers are considered the number one resource and higher achieving students are placed with more effective teachers, one could assume that putting more money into retaining teachers with higher effectiveness would increase student achievement. After gaining some experience, some teachers attempt to move to districts in close proximity that pay teachers higher salaries.

Proposition 3. The self-fulfilling prophecy may be a factor in that teacher expectations yield lower gains (Purkey & Smith, 1983). Too often, it is assumed that poor and minority children are unable to learn. However, if these students are taught at the same (high) levels as other students, they will be able to achieve at the same (high) levels (Haycock, 1998). "We document the clear relationship between low standards, low-level curriculum, undereducated teachers and poor results" (Haycock, p. 2). Additionally, Johnson, Livingston, Schwartz, and Slate (2001) suggest that all students are capable of learning if they are in a school where the school leader and teachers behave in such a manner as to foster this belief.

Other Findings

Heterogeneity is achievement by class size interaction was significant in five of the analyses in the present study. However, there was no typical pattern in the results. The main effect of heterogeneity in achievement was found to be significant in three year-to-year models. Larger gains in more heterogeneous classes were found twice, and in lower and higher heterogeneity groups once. Based on Sanders' analyses (Wright et al., 1997), two out of thirty of the heterogeneity main effects were found to be significant (one with lower heterogeneity and one with higher heterogeneity), which would be expected to occur by chance.

The main effect of class size was determined to be significant in one analysis in the present study, with a slight advantage for larger class sizes. Wright et al. (1997) found significant main effects of class size in three of the 30 analyses, determining that class size was not significant. The same conclusion can be drawn for class size in the present study. Furthermore, interactions in both the present study and Sanders model displayed a number of significant interactions. However, these effects appear in both studies to be less important than achievement and heterogeneity by achievement. Additionally, Sanders and his colleagues concluded that

intraclassroom heterogeneity, whether a main effect or interaction, does not impact student achievement (Wright et al.). Similarly, the results of this study indicate the same. Achilles, Finn, and Pate-Bain (2002) found class size to be a significant predictor of student achievement; the difference could be the way in which class sizes are broken down. Achilles et al. looked at data involving 13 – 17 students (small classes), 22 – 25 students (regular classes), or classes with teachers and paraprofessionals. The data for class sizes were manipulated if the class involved teachers and paraprofessionals.

Implications

The findings of this study clearly answer the question as to whether teachers impact student achievement. It is evident through all of the analyses that teachers are the most important factor in student achievement. Every data set demonstrated wide variability from teacher to teacher in student achievement no matter what the class size or heterogeneity in achievement level in these classes.

A quality education is necessary for the future success of all students. The implications from past and present research are vast. By using the information gathered from this latest phase of school effectiveness research, practitioners can learn and attempt to implement ways to improve the achievement level of all students.

It is not the purpose of the present study to deny the fact that environmental factors affect student achievement. It is a known fact that many students from the lower socio-economic strata start school at lower levels than their peers from higher socio-economic backgrounds. Okpala et al. (2001) reiterate that there is a direct correlation between family background and academic achievement of students. Furthermore, student characteristics (such as poverty) demonstrate a negative impact on achievement (Darling-Hammond, 2000). However, these are factors that the

school system cannot control. Therefore, schools must find ways to work around these uncontrollable factors, and value-added research is clear in its conclusions that teachers make a difference even when the effects of home background are controlled through use of gain scores.

It may be advantageous for students from lower SES backgrounds to be placed in more heterogeneous classes, enabling them to receive resources provided to students of higher SES status (Caldas & Bankston, 1997). Not only would they receive the added resources, they would also be exposed to a diversity of students from varying backgrounds. Furthermore, since effects of teachers are additive and poor and minority students are more likely to be assigned to the least effective teachers, the effects of these teachers tend to be forced on these students (Mendro, 1998).

All states, including Louisiana, must implement better accountability models that focus on student gains. “We need accountability systems that provide comprehensive, detailed, and accurate information that goes beyond raw numbers and can help transform teaching and learning” (Hershberg, Simon, & Lea-Kruger, 2004; p. 27). The current use of raw test scores creates problematic issues. They are not tracking students longitudinally nor factoring out environmental factors (Hershberg et al.).

The value added models will not improve student achievement. However, the data is valuable, if analyzed correctly. If properly used, value added assessments can be a guide for instructional and professional development. Sparks (1998) indicates that staff development activities are important to the development of teacher expertise. Administrators must bear part of the burden by creating an environment that encourages best practices for teachers (Hershberg et al., 2004).

Variability in student achievement typically varies more within schools rather than across schools in a district (Sparks, 1998). Therefore, value added models are one way to hold both administrators and teachers accountable for student learning (Hershberg et al., 2004). The model may not be able to explain why certain teachers in a particular school are effective, but they can recognize which teachers are more effective (Archer, 1999).

If the variability within schools is large, this brings about yet another implication. Many times programs are implemented in schools to bring about change (Mendro & Bembry, 2000). Program evaluation was often the focus, but the focus has shifted from evaluating programs to evaluating teacher and school effectiveness. “We measured the results of program after program and very few had noticeable positive effects and some had negative effects” (Mendro & Bembry, p. 2). In part, the failure of programs can be because they are often poorly implemented. With the few programs that showed positive effects, their effects often declined over time. The teacher’s role in implementation may be the key factor in the successful sustainability of educational programs. Therefore, schools must provide an equitable opportunity for all students, providing lower achieving students with more effective teachers. Second, teacher assignment patterns should be carefully observed since one teacher cannot compensate for an ineffective teacher in only one year. Finally, staff development should consist of scientifically based research (Mendro & Bembry). Darling-Hammond (2000) states that investments should be made to improve teacher quality as quality may be related to improving student achievement. Also, states differ in the investments made to improve teacher quality even though research indicates that what teachers know and can do directly affects student achievement.

With the implementation of value added assessments, changes in school and teacher evaluations are the new focus (Mendro & Bembry, 2000). Furthermore, since the assessments are

able to identify ineffective teachers, steps can be taken to remediate these teachers (Mendro, 1998). “If the ultimate goal is to improve the academic growth of student populations, one must conclude that improvement of student learning begins with the improvement of relatively ineffective teacher regardless of the student placement strategies deployed within a school” (Wright et al., 1997; p. 66).

Wimpelberg (1986) cites characteristics that are important to effectiveness. One such factor is the expectation level of teachers as well as administrators. If this factor is important, teachers and administrators must have a shared mission for the school that is centered on the children (Wimpelberg). Rice (2003) adds that there are five categories of effective teachers: experience, preparation programs and degrees, certification, coursework, and teacher test scores. Even though Rice lists the categories of effective teacher, she acknowledges the fact that there are gaps in the research as to exactly what makes a quality teacher. Yet another list of characteristics of good teachers is addressed by Sparks (1998). Sparks explains that teachers need a “deep knowledge of content, repertoire of instructional skills, knowledge regarding skills, and attitudes that support high levels of learning” (p. 2).

Future Research

The focal point of this study was the impact of the teacher on student achievement. Due to the positive effects of teachers on student achievement, further research is needed in order to find what makes teachers effective. Rice (2003) states that “researchers, practitioners, policy makers, and the public have been unable to reach a consensus about what specific qualities and characteristics make a good teacher” (p. 3). On the other hand, Haycock (1998) indicates that there are some teacher characteristics that impact student achievement: strong verbal and math skills, deep content knowledge, and teaching skill.

Future studies must also be expanded to the school and district levels. Sanders et al. (1997) detail their research into these contexts. However, only one district was available for the current study, and the school-level model was not implemented as with Sanders' work.

Johnson et al. (2001) cite school leadership as an important variable to school effectiveness. If teachers are nested within schools, then research should be conducted on what makes an effective leader and what correlation leaders have to the impact of the teacher as well as the school.

Finally, some research has suggested that educational programs are not the sole answer to improved student achievement (Mendro & Bembry, 2000). If this is the case, research should be conducted on the use of these programs, particularly the implementation fidelity, and how teacher effectiveness interacts with program effectiveness.

Much is yet to be learned about how to close the persistent achievement gap between poor and minority students. This research and other value added studies on which it was based clearly and profoundly points to one element as most critical in closing the gap: a highly motivated and skilled teacher in every classroom.

REFERENCES

Achilles, C. M., Finn, J. D., & Pate-Bain, H. (2002). Measuring class size: Let me count the ways. *Educational Leadership*, 59(5), 24-26.

Amrein, A. L., & Berliner, D. C. (February, 2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership*, *60*(5), 32-38.

Archer, J. (1999). Sanders 101. *Education Week*. Retrieved April 19, 2003 from http://www.edweek.org/ew/vol-18/34sander.h18.

Bratton, Jr., S. E., Horn, S. P., & Wright, S. P. (1996). *Using and Interpreting Tennessee's Value-Added Assessment System: A Primer for Teachers and Principals*. Tennessee: University of Tennessee.

Brophy, J. E., & Evertson, C. M. (1981). *Student Characteristics and Teaching*. New York: Longman.

Brookover, W. (1981). *Effective secondary schools*. Philadelphia: Research for Better Schools.

Brookover, W., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *School Social Systems and Student Achievement: Schools Can Make a Difference*. New York: J. F. Bergin Publishers, Inc.

Brookover, W., Beamer, L., Efthim, H., Hathaway, D., Lezotte, L., Miller, S., Passalacqua, J., & Tornatzky, L. (1982). *Creating Effective Schools: An In-Service Program for Enhancing School Leaning Climate and Achievement*. Florida: Learning Publications, Inc.

Caldas, S. J., & Bankston III, C. (May/June, 1997). Effect of school population Socioeconomic status on individual academic achievement. *The Journal of Educational Research*, 90, 269-277.

Ceperley, P. E., & Reel, K. (1997). The impetus for the Tennessee value-added accountability system. J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp. 133-136). Thousand Oaks, CA: Corwin Press, Inc.

Coleman, J. S., Kilgore, S., & Hoffer, T. (1982). *Public and Private Schools.* New York: Basic Books.

Creemers, B. P. M., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement*, 7(3), 197-228.

Cunningham, L. L. (1997). In the beginning. J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp. 75-80). Thousand Oaks, CA: Corwin Press, Inc.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of State policy evidence. *Education Policy Analysis Archives*. Retrieved February 24, 2004 from http://epaa/asu.edu/epaa/v8n1.

Dorn, S. (January 1998). The political legacy of school accountability systems. *Education Policy Analysis Archives*, *6*(1), 1-34. Retrieved August 23, 2003 from http://epaa.asu.edu/epaa/v6n1.html

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, *37*(1), 15-27.

Ellet, C. D., & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness, and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17(1), 101-128.

Fullan, M. G. (1991). *The New Meaning of Educational Change*, 2nd Ed. New York: Teachers College Press.

Good, T. L. (1989). *Classroom and School Research: Investments in Enhancing Schools*. Columbia, MO: Center for Research in Social Behavior.

Good, T. L., Biddle, B. J., & Brophy, J. E. (1975). *Teachers Make a Difference*. New York: Holt, Rinehart, and Winston.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effects of school resources on student achievement. *Review of Educational Research*, 66(3). 361-396.

Gustafson, J. P. (Winter, 2002). Missing the mark for low-ses students. *Kappa Delta Pi Record*, 38(2), 60-63.

Haycock, K. (Summer, 1998). Good teaching matters: How well-qualified teachers can close the gap. *The Education Trust*, 3(2).

Hershberg, T., Simon, V. A., & Lea-Kruger, B. (February, 2004). Measuring what matters. *American School Board Journal*, 27-31.

Holloway, J. H. (2000). Research link/A value-added view of pupil performance. *Educational Leadership*, 57(5), 84-85.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Bray, G. B., Naylor, R. J., Lewis, J. C., Ordman, V. L., & Qualls, A. L. (2003a). *The IOWA Tests: Interpretive Guide for School Administrators*. Iowa: Riverside Publishing.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Bray, G. B., Naylor, R. J., Lewis, J. C., Ordman, V. L., & Qualls, A. L. (2003b). *The IOWA Tests: Interpretive Guide for Teachers and Counselors*. Iowa: Riverside Publishing.

Jencks, C., Smith, M., Ackland, H, Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Micholson, S. (1972). *Inequality: A Reassessment of the Effect of Family and Schooling in America*, New York: Basic Books, Inc.

Johnson, J. P., Livingston, M., Schwartz, R. A., & Slate, J. R. (2001). What makes a good elementary school? A critical examination. *The Journal of Educational Research*, *93*(6), 339-348.

Kahlenberg, R. D. (Summer, 2001). Learning from James Coleman. *Public Interest*, 144, p. 54. Retrieved February 29, 2004 from http://seg802.ocs.lsu.edu:2061/DeliveryPrintSave.asp?tb=1&_ug=dbs+eric+sid+E22BFA5…

Lezotte, L. W., & Bancroft, B. (1985). Growing use of effective schools model for school improvement. *Educational Leadership*, 42(3), 23-27.

Lezotte, L., Hathaway, D., Miller, S., Passalacqua, J., Brookover, W. (1980). *School learning climate and student achievement*. East Lansing: Michigan State University, Center for Urban Affairs.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.

Mendro, R. L. (1998). Student achievement and school and teacher accountability. *Journal of Personnel Evaluation in Education*, 12(3), 257-267.

Mendro, R., & Bembry, K. (2000). *School evaluation: A change in perspective*. (ERIC Document No. ED443874).

Murnane, R. J. (1975). *The Impact of School Resources on the Learning of Inner City Children*. Cambridge, MA: Ballinger Publishing Co.

Office of Education (1966). *Equality of Educational Opportunity – Summary*. (ERIC Document No. ED015953).

Olson, L. (2004). Tennessee reconsiders value-added assessment system. *Education Week*, 23(25), 9.

Okpala, C. O., Okpala, A. O., & Smith, F. E. (November/December, 2001). Parental involvement, instructional expenditures, family socioeconomic attributes, and student achievement. *The Journal of Educational Research*, *95*(2), 110-115.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal*, 83(4), 427-452.

Ravitch, D. (2000). *Left Back: A Century of Failed School Reforms*. New York: Simon & Schuster.

Rice, J. K. (August, 2003). Teacher quality: Understanding the effectiveness of teacher attributes. *Economic Policy Institute*. Retrieved June 22, 2004 from http://www.epinet.org/content.cfm/books_teacher_quality_execsum_intro.

Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education,* 14(4), 329-339.

Sanders, W. L. (November, 17, 2004). *Beyond No Child Left Behind. Keynote* address at the annual meeting of the Mid-South Educational Research Association, Gatlinburg, TN.

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

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value-added assessment system: A quantitative, outcomes-based approach to educational assessement. J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp. 137-162). Thousand Oaks, CA: Corwin Press, Inc.

Sparks, D. (April, 1998). Teacher expertise linked to student learning. *National Staff Development Council*, p. 2.

Stoll, L., & Fink, D. (1992). Effecting school change: The Halton approach. *School Effectiveness and School Improvement*, *3*(1), 19-41.

Stronge, J. H. (2002). Qualities of effective teachers. Retrieved October 17, 2004 from http://www.ascd.org/publications/books/102007/intro.html

Stronge, J. H., & Tucker, P. D. (2000). *Teacher Evaluation and Student Achievement.* National Education Association of the United States.

Stronge, J. H., & Tucker, P. D. (2001). Teacher evaluation and student achievement. *From a Consumer Perspective: Briefings on Educational Research from the Education Consumers Consultants Network*. Retrieved March 19, 2004 from http://www.education-consumers.com/briefs/march2001.shtm

Summers, A. A., & Wolfe, B. L. (1977). Do schools make a difference? *American Economic Review*, 67, 639-652.

Sutton, A., & Soderstrom, I. (1999). Predicting elementary and secondary school achievement with school-related and demographic factors. *The Journal of Educational Research,* 92(6), 330-338.

Teddlie, C., & Stringfield, S. (1993). *Schools Make a Difference: Lessons Learned from a 10-Year Study of School Effects*. New York: Teachers College Press.

Thomas, J., & Stockton, C. (2003). *Socioeconomic Status, Race, Gender, & Retention: Impact on Student Achievement*. Retrieved February 24, 2004 from http://www.usco.edu/essays/vol72003/stockton.pdf.

Webster, W. J. & Mendro, R. L. (1997). The Dallas value-added accountability system. J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp. 81-99). Thousand Oaks, CA: Corwin Press, Inc.

Wenglinsky, H. (February, 2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, 10(12), 1-27. Retrieved February 24, 2004 from http://epaa.asu.edu/epaa/10n12

Wimpelberg, R. K. (1986). *Bureaucratic and cultural images in the management of more and less effective schools*. (Eric Document No. ED269898).

Wimpelberg, R. K., Teddlie, C., & Stringfield, S. (1989). Sensitivity to context: The past and future of effective schools research. *Educational Administration Quarterly*, 25(1), 82-107.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57-67.

Appendix

University of New Orleans

Human Subjects Approval

University Committee for the Protection of Human Subjects in Research
University of New Orleans

Form Number: 03JAN05

(please refer to this number in all future correspondence concerning this protocol)

Principal Investigator: Ellen Lusco   Title: Doctoral Student

Faculty Supervisor: Peggy Kirby   *(if PI is a student)*

Department: Educational Administration   College: Education

Project Title: A Value-Added Analysis of Teacher Effects on Student Achievement

Date Reviewed:

Dates of Proposed Project Period   From 12/01/04   to 05/01/05

****approval is for one year from approval date only and may be renewed yearly.***

Note: Consent forms and related materials are to be kept by the PI for a period of three years following the completion of the study.

Approval Status	Date
☐ Full Committee Approval	
☒ Expedited Approval	12-10-04
☐ Continuation	
☐ Rejected	

☐ The protocol will be approved following receipt of satisfactory response(s) to the following question(s) within 15 days:

Committee Signatures:

Laura Scaramella (signature) — Laura Scaramella, Ph.D. (Chair)

Pamela Jenkins, Ph.D.

Anthony Kontos, Ph.D.

Richard B. Speaker, Ph.D.

Gary Talarchek, Ph.D.

Kari Walsh

L. Allen Witt, Ph.D.

VITA

Ellen Lusco received her Bachelor in Education from Nicholls State University in Thibodaux, Louisiana in 1996. While working as a classroom teacher in an elementary school, she went on to pursue a Master of Education in Guidance Counseling in 1998 and a plus 30 in Educational Administration from Nicholls. She continued her education at the University of New Orleans, attaining a Ph. D. in Educational Leadership.

After teaching for six years, Ellen was promoted to Guidance Counselor at a high school. She is currently an administrator in the tenth largest school system in Louisiana and is looking forward to advancement in the area of educational leadership.