2011

Moving beyond class size: The importance of addressing classroom-context on academic achievement

Jean-Jacques Françoisse

MOVING BEYOND CLASS SIZE: THE IMPORTANCE

OF ADDRESSING CLASSROOM-CONTEXT

ON ACADEMIC ACHIEVEMENT

A Dissertation

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

in

Educational Leadership

---

by

Jean-Jacques Françoisse

September 2011

MOVING BEYOND CLASS SIZE: THE IMPORTANCE

OF ADDRESSING CLASSROOM-CONTEXT

ON ACADEMIC ACHIEVEMENT

A Dissertation

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Jean-Jacques Françoisse

September 2011

Approved by:

[redacted]

Marita Mahoney, Co-Chair, Educational Administration

08/20/2011

Date

[redacted]

Matt Riggs, Co-Chair, Psychology

[redacted]

James Hill, Committee Member, Educational Administration

[redacted]

Donna Schnorr, Committee Member, Educational Psychology and Counseling

ABSTRACT

Research has suggested smaller class sizes in early elementary grades lead to improved academic achievement, particularly for minority or low socioeconomic status students. Yet, the impact of class size reduction in middle schools is largely unstudied. Moreover, the mechanisms describing the association between Class Size and Academic Achievement remain elusive. HLM was used to identify these mechanisms, allowing the analysis to be conducted at the classroom and student levels. Classroom-context factors (e.g. Teacher Engagement, Teacher Experience, and Instructional Use of Time) were investigated for moderation effects on the Class Size and Academic Achievement (as measure by standardized test scores in Language Arts and Mathematics) relationship. Both Teacher Engagement and Teacher Experience impacted Language Arts Academic Achievement. Teacher Engagement and Instructional Use of Time (Administrative Tasks) impacted Mathematics Academic Achievement. Additionally, Teacher Engagement moderated the Class Size and Language Arts Academic Achievement relationship depending on English proficiency status. Educational decision makers need to account for the impact of Classroom-context factors beyond Class Size alone. Administrators and policy makers are urged to consider class size reduction along with other alternative ways to raise Academic Achievement rather than an isolate strategy.

ACKNOWLEDGEMENTS

I am most grateful to my dissertation co-chairs, Dr. Mahoney and Dr. Riggs for their guidance, patience, and willingness to share their expertise. I also wish to acknowledge the support of the members of my dissertation committee, Dr. Schnorr and Jim Hill, for providing invaluable suggestions in the areas of engagement and education policies. Dr Joseph. Deem's expertise and guidance in the area of educational assessment have been a source of inspiration.

My thoughts and deepest appreciation also go to my girlfriend Susan for her patience and unwavering support in this project. My children, Caitlin and Jean-Patrick, have also been a constant source of encouragement. I know they will follow the way beyond what I have accomplished.

Lastly, I wish to acknowledge my peers at the California State University, San Bernardino, and my best friend Dr. Allan Aab and his fiancée Arlene.

DEDICATION

To my grand-father, Léon Tassin, one-room schoolhouse headmaster in a small Belgian village, Rognée (Hainaut), who instilled in me perseverance, curiosity, and lifelong love for learning, I will always be grateful.

I also dedicate this work to my parents, Dr. Jacques Francoisse and Marguerite Tassin.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Statement of the Problem

Research has suggested smaller class sizes in primary grades lead to improved Academic Achievement, particularly for minority or low socioeconomic status students. Yet, the mechanisms describing the association between class size and improved academic achievement remain elusive. Student engagement, quality of student-teacher interactions, use of instructional time (group or individualized instruction), and student behavioral changes (e.g., reduced discipline issues, improved attendance) are believed to be key factors leading to improved academic achievement, although these factors impact students differentially. Additionally, these relationships have not been fully explored in middle school grades. This study examined some of these relationships in the context of a large California school district. Five middle schools participated, including two schools which recently received a multi-year grant to implement class size reduction: the Quality Investment Education Act grant (QIEA). While the QIEA contains accountability elements, results of this study will enhance the findings with quantitative observations related to the impact of teacher-student interactions and student demographic backgrounds on achievement.

The Quality Investment Education Act (QIEA) of 2006 was the settlement remedy of a lawsuit filed by the California Teacher Association (CTA) and State Superintendent Jack O'Connell *(California Teachers Association et al. v.*

Schwarzenegger et al., 2006) for California's failure to adequately fund the school revenue limit set by Proposition 98 in fiscal years 2005 and 2006. This remedy provided California's low-performing schools with competitive grants totaling $2.7 billion over a seven-year period until the school year 2013-2014 to implement school improvement strategies, primarily class size reduction, reduction of the counselor-student ratio, high-quality teacher and administrator staff development, and redistribution of experienced teachers. California's lowest performing elementary, secondary and charter schools, ranking in the bottom first and second deciles according to their 2005 Academic Performance Index (API), were invited to apply for the QIEA competitive grant. 448 schools among the applicants were randomly selected on the basis of statutory requirements for geographic locations and grade span characteristics (O'Connell, 2007). As discussed below, the locus and focus of the current research targeted the impact of class size reduction in two middle schools in Southern California that successfully applied for the QEIA competitive grant.

Background of the Research

For well over one hundred years, research has suggested smaller class sizes in the primary grades lead to improved academic achievement as measured by standardized tests, particularly for minority or low socioeconomic status students. However, educational research wrestles with producing coherent explanations of the mechanisms leading to these positive results. At the elementary grade level, student-level factors such as student engagement, demographic characteristics, pro- or anti-learning behaviors and behavioral changes (e.g., reduced discipline issues and improved attendance) have been found to moderate the effects of smaller class size on academic achievement (Finn, Pannozzo, & Achilles, 2003). On the other hand, the literature has also hypothesized that teachers in smaller classes would naturally change their instructional delivery to include higher quality student-teacher interactions. Greater attention to individual needs, student instructional grouping, were also believed to be key factors in improving in academic achievement.

Yet, unlike for student-level factors, researchers disagreed on the effect class size would have on teacher processes such as the assumption that teacher instruction in a smaller setting would naturally foster renewed high-quality, high-content interactions with the students. As evidenced by the emphasis on staff development in the most recent project of class size reduction implementation efforts (Graue, Hatch, Rao, & Oen, 2007; Odden, Picus, Archibald, Goetz, Mangan, & Aportela, 2007), this assumption generates skepticism, and, as a result, remains inconclusive. Further compounding the complexity of these questions is the fact these dynamic mechanisms uncovered in primary grades have not been explored fully at the middle and high school levels.

Research has fallen into three traditions of measuring the effect of class size on student achievement. The first tradition of studies is represented by approximately 100 quasi-studies aimed at establishing a direct relationship between class size and academic achievement. The majority were quasi-experimental while only 14 of these were true experimental designs (Glass & Smith, 1978). Researchers failed to reach a consensus regarding the effect of class size on academic achievement (Glass & Smith, 1978; Graue & Rauscher, 2009; Rockoff, 2009). Extensive reviews of those studies were carried out by Glass and Smith (1979), Glass, Cahen, Smith, and Filby (1982), and Hedges and Stock (1983). One shortcoming the earlier studies remains the relatively small samples sizes. Therefore, generalizability of these studies was limited.

A second tradition studying the effects of class size on student achievement is the econometric work on achievement, an approach measuring the impact of economic input. These studies would typically report data collected at the school level and large-scale samples included multiple regression models involving the control of student characteristics such as previous levels of achievement, socioeconomic status, gender, race, and parent level of education. Reviews of these studies were carried out by Greenwald, Hedges, and Laine (1996), Hanushek (1986, 1989), and Hedges, Laine, and Greenwald (1994). Nye (2004) indicated the best studies took into account data related to individual student prior achievement and socioeconomic status (SES). Studies which controlled for confounding facets such as national school lunch program as sole measure of SES were judged weaker. Secondly, poor independent measurement of class size (most studies averaged the number of students per teacher when some classes are larger than others) in some regression models likely contributed to an overestimation of effect sizes.

To these small-scale experiments and econometric studies is added a third tradition initiated by large-scale randomized experiments such as Project STAR, a study ordered in 1985 by the state of Tennessee. Students in grades K-

3 were randomly assigned to one of three class sizes: small classes (13-17); regular classes (22-26); or, regular classes with a full time aide. Five years after the Project STAR experiment ended the study was extended to investigate the long-term effects of small class sizes on academic achievement. The analysis used in the study was hierarchical linear models. Researchers found the small class effect was larger with minorities than Caucasian students. Unfortunately, this observation was not quantified. Topics for future research were identified as being classroom processes and instruction. However, Project STAR researchers were not able explain the reasons for differences in achievement between groups and the processes leading to increased levels of academic achievement.

Areas for further study in the literature reviewed suggest focusing on other factors favorably impacting academic achievement such as teacher training and the quality of teacher-student interactions, and the type of student grouping during instruction. These elements were of high interest as the present study attempted to shed some light on the mechanisms which allow students in smaller classes to reach a higher level of academic achievement.

Justification of the Study

Justification for this study was two-fold: the use of multi-level statistical analysis of the data; and, participant characteristics.

Most studies in the area of class size reduction carried out before the 1980s used correlational models. Regression statistics were mostly used in the post-1980 years. Recently, researchers have questioned the use of student participants as the main unit of analysis. Indeed, student participants are nested into classrooms, teacher rosters, schools, and even districts. The suggestions that aggregated units of analysis such as the classrooms could influence the relationship between class size and student achievement led to reanalysis of prior results from the late 20th century.

Furthermore, middle school grades have largely been ignored in class size reduction research. The core of the studies dedicated to class size and its effects has traditionally targeted the elementary grades; early research posited most gains would be obtained in primary elementary grades 1-3 (Glass et al., 1982; Robinson, 1990; Robinson & Wittebols, 1986). Although not quantified, findings at this grade level suggested class size positively affected academic achievement. Yet, these conclusions cannot be extended to the middle school level, which the proposed study will specifically target. Furthermore, in most studies involving quasi-experiment designs, data collection and methods were carried out well after implementation of the program. This was not the case in the current study. Therefore, it was believed the present study would have high internal validity as the concurrent development of the intervention and the study allowed for better control of extraneous factors such as the context elements, as well as student and teacher-level factors.

CHAPTER TWO

REVIEW OF THE LITERATURE

This section provides background knowledge related to prior research undertaken with the aim of better understanding the effect of smaller class sizes on academic achievement in primary and secondary grades. After a brief overview of early empirical studies prior to the 1980s, the focus will turn to the influential state-mandated experiments and quasi-experiments implemented at the onset of the 1990s state and federal accountability programs. Building on the discovered needs for future research, this review does not intend to address public policy questions such as the cost-effectiveness of small class-size programs. Instead, it focuses on the potential academic benefits of such programs as they are related to increasing academic achievement. Lastly, a theoretical model of the dynamics between class size and academic achievement will be suggested, taking into account variables such as student-factors (e.g., motivation, pro-social behavior, anti-social behavior), teacher-factors (e.g., instructional practices, student interactions), and contextual-factors (e.g., school organization, scheduling, internal governance). Central to the study will be whether smaller classes equally benefit all students. Prior to examining the relationship between class size and achievement, it is necessary to define these terms.

Defining Class Size and Academic Achievement

Presently, the construct of class size encompasses a wide variety of instructional settings ranging from student one-on-one tutoring to internet on-line classes serving several hundred students simultaneously. Likewise, the concept of "small" and "smaller" class size evolved greatly in the course of the 20th century.

Econometric studies, aimed at calculating the effects of economic inputs such as per-pupil expenditure, teacher factors, and school resources on education outcomes such as academic achievement, have used the Pupil Teacher Ratio (PTR) extensively (Hanushek & Rivkin, 1997). Consistently, these econometric studies concluded in their causal education production models that class size had little or no effect on academic achievement. PTR and class-size constructs are pivotal in understanding why researchers have not come to a consensus as to the impact of smaller classes on academic achievement. Empirical and econometric studies used PTR extensively, and posited the effect of smaller classes on academic achievement was negligible at best. On the contrary, findings from quasi-experimental and experimental researchers often concluded class size impacted academic achievement.

While class size denotes the average number of students entrusted in the care of one teacher over the course of one year, PTR is computed at the school level. PTR is the number of students within a Local Educational Authority (LEA) divided by the number of licensed personnel (holders of a permanent or temporary teaching license/credential) servicing the student population (Achilles, n.d.). In 2002, the difference between PTR (an administrative metric used mainly for fiscal purposes) and class size (the grouping of students for the purpose of delivering instruction) in U.S. classrooms varied by as many as 10 students (Achilles, Finn, & Pate-Bain, 2002). That is, given a PTR of 17 students to one teacher in a given building, the actual classroom load may be as large as 27 students for one teacher.

While actual class size may vary during the year or even during the same day, PTRs are usually smaller than actual class size since PTR includes licensed personnel not assigned to one classroom or assigned to smaller classes such as those typically required to service special need students. Although PTR and class size both intend to determine the number of students in any given class, it is likely that PTRs would be considerably lower than the actual class size. In fact, it is only at the classroom level that both metrics may be identical (Achilles, n.d.), assuming students are not pulled out during the day. Aggregations at school and district levels assume equal weighting of class loads and fail to account for actual class size variations (Addonizio & Phelps, 2000). Although temporary decreases in class size are possible, when students are pulled out for specialized programs, PTR often underestimates the number of students present in classrooms. When teachers not assigned to classrooms are taken into account in the PTR, effect sizes which linked class sizes to academic achievement are often underestimated since the actual numbers of student sitting in classrooms are often larger than the research assumptions.

Pupil-to-Teacher Ratios in public schools steadily decreased from 35:1 in 1890, to 28:1 in 1940,and 20:1 in 1970 (Hanushek, 1997; Hanushek & Rivkin, 1997). Hanushek (1986) remarked that during 1950-94, PTR dropped 35%, with most of the decrease recorded between 1960 and 1980. Yet, achievement in Mathematics, Science and Reading as measured by the National Assessment of Educational Progress (NAEP) remained consistently flat over the last three decades of the 20th century (Hanushek, 1998; Johnson, 2002). Projection estimates suggested that by Fall 2017, the PTR, including a growing number of recently created positions in special education at both the elementary and secondary levels, will plummet to 14.5:1(Hussar & Bailey, 2008). The steady decrease in PTR combined with stagnant results in academic performance (Scholastic Aptitude Test 'SAT' scores) has been held by some econometric studies (Hanushek, 1997; Hanushek, Rivkin, & Taylor, 1996) as indubitable proof that reducing class size does not result in increased academic achievement. Furthermore, it was also argued that students in other countries were reaching higher levels of academic achievement than their American counterparts in spite of larger class sizes (Hanushek, 1996). It must be emphasized that studies are carried out within a social context, and subsequent findings are partially the product of the researcher's assumption. In an interview, Hanushek revealed his underlying belief in education policies aiming at improving teacher quality rather than supporting class size reduction, estimated in 2009 at $69 billion per year for nationwide implementation (Graue & Rauscher, 2009).

Although these econometric studies suggested lowering the PTRs did not result in gains in academic achievement, proponents of smaller class sizes point to the changing nature of education as necessitating smaller class sizes for effectiveness and efficiency. Indeed, the growth of specialized areas of instruction such as special education gives the illusion that class sizes have been reduced (Achilles et al., 2002; Biddle & Berliner, 2002; Hedges, Laine, & Greenwald, 1994) by lowering the PTR while class size in mainstream classrooms itself remained consistent or even increased. Furthermore, researchers further contended that Hanushek's (1986) conclusions in his meta-analytical work lacked external validity as the sample groups used were small and not representative of the U.S. population (Biddle & Berliner, 2002; Greenwald, Hedges, & Laine, 1996; Krueger, 2003). Moreover, use of PTR alone to describe class size in examining the impact on academic achievement did not control for other contextual factors, such as school demographics; as such, potentially confounding variables were not accounted for (Biddle & Berliner, 2002, 2003; Hedges & Greenwald, 1996). Although meta-analyses allowed for a synthesis of a large number of studies, there were methodological limitations, including being criticized for giving studies of different methodology equal weight (Krueger, 2003). Results from meta-analytical summaries tended to mask the impact of student- and teacher- and school-related contexts as moderating factors in the relationship between class size and student academic achievement.

When disaggregating the data of the National Assessment of Educational Progress (NAEP), the largest gains in academic achievement regardless of class size were made by minorities, with gains between 0.2 to 0.6 *SD* over the period spanning 1970 to 1990 (Grissmer, Kirby, Berends, & Williamson, 1994). However, Grissmer et al. (1994) pointed out that assignment to smaller class size may not have been random as students demonstrating lower academic performance may have been deliberately assigned to smaller remedial classes. This observation is very relevant as this practice continues to this day. For instance, in a recent California lawsuit *(California Teachers Association et al. v. Schwarzenegger et al., 2006*), the court ruled that only underperforming middle schools in the State were eligible for the small class size remedies provided in the QEIA grant settlement.

The difficulty of defining the concept of small class size has been further compounded by multiple methods of calculating class size ratios and the complexity of school master course schedules. As stated earlier, the number of children in any given classroom is likely to vary in the course of the same day, as students are pulled out for specialized interventions. Although researchers (Ehrenberg, Brewer, Gamoran, & Willms, 2001b; Hedges et al., 1994; Slavin, 1989a) agreed class size is a ratio involving students and instructors, studies have been inconsistent or even silent as to how such ratios are obtained. In the large-scale *Coleman Report* (1966), class size was obtained by dividing the student population within a building by the number of faculty, including noninstructional staff such as librarians who do not instruct classes. Since the primary purpose of the *Coleman Report* was to observe the impact of racial segregation on achievement in American schools, class size was, *ipso facto*, aggregated to other measures of "school facilities/resources" and did not account satisfactorily for the impact of class size on achievement within the larger context of public education. Relying on the available data, from large samples of convenience and questionnaires, the study was unable to isolate the impact of class size and achievement.

Furthermore, other factors such as nonassigned teaching staff, pullout of students for differentiated instruction or small group workshops taking place at various times of the day also introduced complications in calculating PTRs. Class size itself includes considerable variations (e.g., allotted time, student characteristics, instructional methods, grade levels, subject areas), which, if left unaccounted for, may result in an underestimation of the true relationship between academic achievement and class size (Ehrenberg, Brewer, Gamoran, & Willms, 2001a). Clearly class size and PTRs differ in that the latter does not account for the actual schooling context in which students are learning and there is no agreement among researchers on a standardized method of calculating such ratios.

Therefore, the researcher must be explicit when defining class size and PTR. Adcock suggested a working definition of class size as "the total number of students enrolled on the last school day of the year divided by the derived school number of core teachers employed on the last of the school year of [a given] school" (1999, April, p. 9). Such a statistic of class size considers only teachers assigned to academic subjects: English/Language Arts, Social Science/History, Mathematics and Science. To add to the terminology confusion of PTR and class size (CS) comes the new term Class Size Reduction (CSR). Unlike PTR and CS, CSR is a matter of public policy assuming that smaller groups result in higher quality of instruction through improving teacher-student interactions and ultimately in higher academic performance (Graue & Rauscher, 2009).

Academic Achievement

The construct of academic achievement in the present study refers to individual norm- and criterion-referenced standardized measures commonly used in K-12 grades. These are administered mostly at the state level (e.g., Iowa Test of Basic Skills [ITBS], California Standards Test [CST], National Assessment of Educational Progress [NAEP], or Stanford Achievement Test [SAT]). Academic achievement differs from academic attainment in that data measuring academic achievement are collected at regular intervals for the purpose of measuring progress. Academic attainment, on the other hand, denotes reaching educational goals or milestones which enhance societal status, such as graduation from an educational institution, or moving up the socioeconomic ladder. Research traditionally reported disaggregated academic achievement results in one or more of the four core subjects (Mathematics, Language Arts, Social Studies, and Science) for groups of students being observed, while other studies, particularly meta-analyses (i.e., Glass & Smith, 1979), combined the academic achievement into a composite for lack of more specific data. Although one could conceive other methods of measuring schooling outcome, such as authentic assessment, standardized testing is more readily available and allowed for a common assessment tool. By and large, commonly reported standardized test results are readily available at the state, district, and school levels.

Historical Context of Class Size Research

As early as the turn of the 20th century, class size and its effects on academic achievement elicited the interest of educational researchers. Ironically, it was the effect of increasing, not reducing, class size that was the topic of the day as school officials were struggling with increased enrollment in grades K-12, the growing cost of educating pupils, and the slow pace of school facilities construction during World War I (Rockoff, 2009).

Prior to World War II, beside anecdotal accounts and empirical observations, 45 studies on class size using primary data were published. While half of these studies involved field experiments, eight used matched pairs (e.g., students in smaller class are matched with similar students attending a larger class) and 13 used correlations of observable data. All 24 field studies but two concluded "average achievement (or achievement growth) was not significantly reduced in larger classes, and in many instances the students in larger classes outperformed their small class counterparts" (Rockoff, 2009, p. 5). At that time, focus was on elementary education, and more sparingly on secondary education (Glass et al., 1982). The early conclusions, that class size was not associated with academic achievement, may have resulted from use of developing methodologies in educational research. On the other hand, the differences in the class sizes being compared (20-35 student and 35-50 student classroom configurations) may have been too large to impact academic achievement regardless of the rigor of the analysis, as findings in later studies were to conclude (Graue, Rauscher, & Sherfinski, 2009; Molnar, Smith, Zahorik, Palmer, Halbach, & Ehrle, 1999; Nye, Hedges, & Konstantopoulos, 2000).

From the 1900s to 1920s, research on the impact of class size on student academic achievement, which were correlational for the most part, demonstrated minimal experimental control (Glass et al., 1982; Rockoff, 2009). By the early 1930s, most of the research efforts related to class size went dormant until interest resurfaced in the 1960s when academic achievement was correlated with school resources (Glass et al., 1982). Experimental and quasi-experimental research greatly expanded in the late 1970s and early 1980s, with the growing unease across the nation that public education was failing (Gardner, Larsen, Baker, & Campbell, 1983).

Although the main body of research in the area of class size and academic achievement focused on increasingly smaller class sizes, comparing classes comprised of between 15 and 35 students, studies prior to the 1970s defined as small classes what would be considered large by today's standards. For instance, while Rice (1902) compared the effectiveness of classes ranging from under 40 students, 40 to 49 students, and 50 students and over, later studies carried out in the 1980s focused on much smaller class sizes, typically of 15 to 22 students versus 23 to 35 students (Molnar et al., 1999; Nye et al., 2000;

Shapson, Wright, Eason, & Fitzgerald, 1980). The first meta-analyses on class size conducted by Glass and Smith (1979) and Glass et al. (1982), included comparisons of classes of 25 students or more with one-on-one tutoring (class size of one). Slavin (1986) pointed out that combining studies involving one-on-one tutoring with more conventional class sizes severely undermined the external validity of Glass' findings. As most educational policies (Burch, 2007; Grissmer, 1999) adopted by individual States in the 1990s involved class size reductions to 25 students at the most, this review will focus on reporting the literature related to this size of classroom, thereby ignoring very small class sizes, such as one-on-one tutoring.

After WWII, two public reports sparked a renewed interest in school reforms and class size research: *A Nation at Risk* (Gardner et al., 1983); and, the *Coleman Report* (Coleman et al., 1966). In the wake of the successful launch of Sputnik by the Soviet Union in 1957, the supremacy of the United States was no longer taken for granted at home; this crisis of confidence culminated 20 years later with the publication of a *Nation at Risk* (Gardner et al., 1983) which pointed at the decline of Scholastic Achievement Test (SAT) scores from 1960s to the 1980s and the lack of international competitiveness of the American educational system. At the state level, Boards of Education closely monitored large programs of class size reduction launched statewide in Tennessee and Wisconsin as the concept of smaller class gained popularity among parents and teachers alike; similar actions controlling class size was seen as an easy mandate for public education entities to implement (Addonizio & Phelps, 2000).

Moreover, opinions in the 1960s were divided as one wondered whether the expected increase in academic achievement realized through the implementation of smaller class size would justify the additional spending of public monies (Graue & Rauscher, 2009; Hanushek & Rivkin, 1997; Rockoff, 2009). The large-scale "state of education" research by Coleman (1966) attributed differences in academic achievement among students to family environment, defined as the number of books available in the home or the socioeconomic status of the unit, and downplayed the role of schooling context, including class size, in academic achievement. The use of archival data at the district and school levels, such as in the Coleman study, presented a distorted picture of class size at the level of the classroom. The lack of distinction between PTR and actual class size likely contributed to Coleman's conclusion that class size was ineffective as a means to improve academic achievement across the nation.

In a commissioned paper design to enlighten public policy in education, the *Coleman Report* (1966), used standardized test scores and questionnaires from teachers and principals of more than 150,000 students in grades one to 12. Coleman et al. (1966) reported class size was a negligible factor in academic achievement on standardized norm-referenced tests in verbal abilities and Mathematics: "Some facilities measures, such as the pupil/teacher ratio in instruction, are not included [in the report] because they showed a consistent lack of relation to achievement among all groups under all conditions" (p. 312).

Disregarding the possible impact of class size on student academic achievement, Coleman concluded the socioeconomic background of the student, the social composition of the student body and the characteristics of the surrounding community were key factors which explained differences in academic achievement among students. This remark continued to be echoed decades later in other studies based on archival data extracted from large archival databases, such as most of current econometric studies. For instance, after reviewing some 400 studies on class size and academic achievement and matching educational inputs with schooling outputs, defined as educational resources and outcomes, Hanushek (1997; Hanushek & Rivkin, 1997) concluded "there is not a strong or consistent relationship between student performance and school resources, at least after variations in family inputs are taken into account" (p. 141).

However, the *Coleman Report* did not distinctly analyze class size as a potential contributing factor; instead class size was combined with other factors such as textbook and library availability under the composite umbrella factor "school facilities/resources." It must be emphasized that the *Coleman Report* defined class size by dividing the student enrolment by the number of school employees (with and without a teaching license) within a building, a potential source of error causing an underestimation of the true relationship between class size and academic achievement. Similar to other econometric studies carried out since (Hanushek, 1998; Rivkin, Hanushek, & Kain, 2005; Wossmann & West, 2006), teacher salaries and other per-pupil expenditures (administration) used as proxy variables for actual class size may mask the true impact of class size on student academic achievement.

Rather than focusing on an unmoderated causal relationship between class size and academic achievement, it would be of greater interest to determine: (a) the marginal gains obtained in small classes over time through time series analysis; and, (b) whether students with different characteristics respond to smaller class instruction in the same fashion (Ehrenberg et al., 2001b). Perhaps, the most compelling objections to the conclusions made in the *Coleman Report* stemmed from its analysis of education at a given point in time. Nevertheless, the same report brought to light other possible confounding factors in the relationship between class size and academic achievement, such as the value of the resources allotted to the schools, the characteristics of instruction including teacher and class size, the characteristics of the school (such as culture), and the characteristics of the community.

This debate over the effectiveness of smaller classes illustrated the divergent and sometimes contradicting interests between government officials and students' families when attempting to answer the question of the economic value of education and the cost benefit of smaller class sizes (Mitchell & Mitchell, 2003). Clearly, research findings and historical contexts cannot be separated. Therefore, research conclusions and findings must be evaluated in the light of the societal issues of the day, the level of sophistication of social research tools, and the political forces at work.

Summary Research Syntheses

In an effort to develop the first comprehensive meta-analysis on the relationship between class size and academic achievement, Glass and Smith (1979) retrieved published empirical class size studies and dissertations since the turn of 1900s, finding over 300 experimental and quasi-experimental studies with usable quantitative data. Glass and Smith (1978, 1979) further focused on 77 experimental studies describing 725 paired comparisons-combinations of student class size broadly categorized in four types (less than 16 students, 17 to 23 students, 24 to 34 students, and over 35 students). In their meta-analysis, Glass and Smith looked at the academic achievement test results of nearly 900,000 students over a 70 year span in a dozen countries.

Glass and Smith (1978, 1979) first approximated the relationship between class size and academic achievement using the model Δ_{S-L}, based on standardized achievement mean differences between pairs of smaller (S) and larger (L) classes divided by the within group standard deviation. Next, rather than creating a matrix with rows and columns representing the class size and the intersecting cell the values of Δ_{S-L}, Glass and Smith used the regression model: $\Delta_{S-L} = \beta_0 + \beta_1 S + \beta_2 S^2 + \beta_3 S^2 + \beta_3 (L-S) + \varepsilon$ to aggregate the findings. Since interpreting the model in terms of class size and academic achievement involved at least three or more dimensions, Glass and Smith imposed a consistency condition on all Δ_{S-L}'s to derive a single curve from the complex regression surface. Imposing arbitrarily the mean *z*-score achievement of 0 to the class-size of 30, the final model was represented by a single regression curve for academic achievement onto class size.

When compared to larger classes of 40 students, smaller classes of 30, 20, 10 and one students showed standardized differential academic achievement effects of -.05, .05, .26, and .57, respectively. Likewise, when compared to larger classes of 25 students, smaller classes of 20, 15, 10, five, and one student showed standardized differential achievement effects of .04, .13, .26, .41, and .55, respectively. These results included academic achievement scores in Mathematics, Language Arts, and Science, some of which had been combined into composites. Half of these regression analyses involved quasi-experimental or convenience assignment of students to either large or small groups. Translating these *z*-scores into percentile ranks, the gains in the 25 versus 20, 15, 10, five, and one student comparisons are 4, 5, 10, 16, 21 percentile ranks, respectively. 47.2% of the 725 comparisons reported only composite measures of academic achievement.

From the initial 725 paired comparisons of student academic achievement in both smaller and larger groups, 435 (60%) comparisons favored smaller class configurations by showing an increase in academic achievement. Yet, this increase was not quantified as some studies were correlational or lacked empirical support. When focusing on 160 pairs of classes of approximately 18 and 28 students, Glass and Smith (1979) suggested even more distinct differences in achievement: in 111 instances (69%) smaller classes demonstrated a higher level of academic achievement over the larger classes.

Again, this result was not quantified. Regressions analyses based logarithmic models favored smaller classes by nearly one tenth of a standard deviation for the complete set of comparisons (Glass & Smith, 1979). The small effect size for all paired comparisons may be explained by the inclusion of poorly controlled studies in the meta-analysis.

Only 109 of the 725 initial comparisons involved random experimental designs (a total of 14 studies), 81% of which found smaller class sizes led to increased academic achievement as measured by standardized tests or other measures, such as number of promotions to the next grade level. Others types of methodologies reported in the 725 comparisons included: (a) matched: 236 comparisons; (b) repeated measures: 18; and, (c) uncontrolled: 362 comparisons. The last type of methodology involved quasi-experiments which likely weakened conclusive discussion related to the relationship between class size and academic achievement.

Possibly for this reason, Glass (1982) further analyzed the results of the 14 random experimental studies. Glass concluded that an average student taught in a class of 20 students would reach a level of academic achievement higher than that of 60% of students taught in a class of 40 students. At the extreme point of comparison, a student instructed in a class of five students would outperform a student in a class of 40 students by 30 percentile ranks. This study suggested that students in smaller classes achieved at a higher level. Yet, even in the case of experimental comparisons, effect sizes were limited unless the size of the small class dropped below 20 students. Glass and Smith argued in favor of smaller class size: "The major benefits from reduced class size are obtained as size is reduced below 20 pupils" (Glass & Smith, 1978, p. v). Glass and Smith helped move the class size controversy to the center stage in a debate still very much alive to this day.

Two important issues seem to weaken the argument that smaller classes are more effective in increasing academic achievement than larger class sizes. Firstly, the 109 comparisons were aggregated by the Glass and Smith into 30 comparisons. In many instances, the same larger and smaller groups and their performances had been evaluated on the basis of different conditions, such as amount of instruction or subject areas. In other instances, the subject areas measured were combined to create a composite academic achievement score. Secondly, results reported reflected the performance of disparate sizes, such as class of one student compared to a class of 30 students, or a class of 5 students compared to a class of 30 students. Education Research Services (ERS) (1980) claimed the Glass and Smith meta-analysis overemphasized the performance of extremely small instructional settings (one to five students). Hedges and Stock (1983) reanalyzed the Glass meta-analysis and validated findings that class sizes below 20 students were more conducive to promoting academic achievement. Subsequently, the initial analyses by Glass and Smith (1978, 1979), based on earlier theoretical work summarizing psychotherapy studies (Glass, 1976), was further expanded (Glass et al., 1982) to include the implications for educational policy decisions. At the heart of the controversy, is the very concept of practical significance and pragmatic implications of systemic changes towards lowering class sizes. Furthermore, Glass and Smith were criticized for disregarding the cost benefit analysis of implementing a large-scale class reduction policy as most of the academic achievement gains are registered only when moving class size below 15 students (Education Research Services, 1980; Hanushek, 1997, 1998; Rivkin et al., 2005).

Smaller class sizes appeared effective; however, the largest effect sizes were noticed in class sizes of less than 20 students. In their meta-analysis of tutoring classes of nine students or less, Cohen et al. (1982) measured academic achievement and reported effect sizes on 52 studies. The average effect size of these studies (defined as the difference between the means of two groups divided by the standard deviation of the control group) was 0.40 in favor of tutored groups, a gain corresponding to an increase of 16 percentile rank points as compared to the mean percentile rank of 50. Their findings confirmed greater effect sizes (differences of means of both experimental and control groups divided by the standard deviation of the control group) in favor of smaller class sizes. Interestingly, groups tutored by peers (older intermediate elementary students teaching primary students) achieved a greater gain than those entrusted in the teaching of regular teachers (Robinson & Wittebols, 1986). This again suggested the need to further identify contextual variables. Clearly, class size alone does not contribute to greater academic achievement.

Both Glass studies (Glass et al., 1982; Glass & Smith, 1978) supported the opinion largely spread in educational circles that small class sizes were more conducive to student learning. These meta-analyses established the benefit of class sizes of less than 20 students, gave the impetus for statewide experimental class-size reduction, and emphasized the role of teaching processes, such as time-on-task, as underlying explanations for the positive impact of smaller class size on academic achievement.

However, the limited number of experimental analyses retained by Glass et al. (1982) created validity concerns. Slavin (1989a) contended, by limiting the meta-analysis to only 14 experimental studies, the Glass et al. conclusions lost external validity and generalizability at the expense of internal validity. Based on the examination of Glass et al. (1982), it appeared sizeable effects of 0.2 *SD* or greater were observed when comparing groups of 17 students or less to conventional classes of 25 students or more. The greatest effects of class size on academic achievement were found with one-on-one tutoring. Critics of Glass (Hanushek, 1998, 1999) pointed out comparisons between extreme class sizes were of little relevance since these were not reflective of the occurrences in school.

Slavin (1989a) introduced a best evidence synthesis, which combined the elements found in meta-analysis with narrative review. He selected eight random class assignment studies which compared the results of standardized Reading and Mathematics tests in smaller and larger elementary-level classes. For his inclusion criteria, Slavin required studies had to compare larger classes to classes at least 30% smaller with a PTR not to exceed 20:1. The selected studies analyzed smaller class size programs of at least one year in duration, with either random assignment to alternative class sizes, or matching preconditions. Effect sizes were based on the difference between the small class academic achievement mean (experimental group) and the larger class academic achievement mean (control group) divided by posttest standard deviation of the control group. This is the same definition of effect size introduced by Glass and Smith (1978, 1979). On average, the studies in Slavin's analysis compared groups of 27 students to groups of 15 students. Even though these eight studies were well-controlled and documented, the median effect size observed was only +.13 (Slavin, 1989a).

Furthermore, discussions about such small effects as measured by standardized tests in both Mathematics and Language Arts wrongly assumed teacher instructional delivery remained consistent regardless of class size (Slavin, 1989b). The type and quality of interactions between students and teachers, such as explicit direct instruction, had previously been identified as influential factors in the *Coleman Report* (1966). This observation was again echoed by Glass et al. (1982) who noted class size is only one variable impacting effective instruction.

In the wake of the controversy on appropriate use of funding for underachieving schools, the Educational Research Service (ERS) published a report (Porwoll, 1978) on the research on class size citing over 100 studies which suggested small effect sizes (no figures available), most of which used correlational analysis, with some or little control of other potentially confounding constructs such as teacher-, student-, and school-related contexts. Although the ERS research was inconclusive, a subsequent ERS study carried out one decade later corroborated the findings of Glass and Smith (Robinson & Wittebols, 1986) and added an important element to the discussion: Although smaller class sizes appeared to be positively associated with an increase in academic achievement, smaller class sizes alone do not result in increased academic achievement.

Adding to Glass' meta-analysis and Slavin's best evidence synthesis, Robinson (Robinson, 1990; Robinson & Wittebols, 1986) used the related cluster approach to review K-12 research studies conducted between 1950 and 1985, involving class sizes of greater than five students. Robinson aggregated studies in clusters representing important factors influencing class size decisions: subject matters; grade levels; student profiles; instructional practices; and, student behaviors. Results indicated the impact of class size on academic achievement "varies by grade level, pupil characteristics, subject areas, teaching methods, and other learning intervention" (Robinson, 1990, p. 90). The Robinson and Wittebols (1986) meta-analysis unfortunately did not provide any effect sizes but instead classified the studies in three categories as to their stated significant differences: (a) favoring small class sizes; (b) favoring larger class sizes; or, (c) bearing no effect on academic achievement. Robinson concluded the positive effects of class size were consistent in grades K-3, slight in grades 4-8, and imperceptible in grades 9-12. Furthermore, lower SES students were found to benefit most from smaller class sizes. Again, these conclusions did include effect sizes. Nevertheless, Robinson's study suggested the concept of optimal class size without considerations for student characteristics had little relevance in educational research. Smaller class sizes were found to benefit students differently, according to their social contexts, personal background, grade level, and academic subject. Determining optimal class size was described as attempting to determine the quantity of butter needed in a recipe without knowing the nature of the other ingredients (Graue & Rauscher, 2009).

The observation that smaller class size alone does not result in academic achievement corroborates the observations of Coleman (1966) and Glass' second meta-analyses (Glass et al., 1982), which acknowledged class size alone did not account for student differences in on academic achievement. Given this, the focus shifted from a direct relationship between class size and academic achievement to identifying the actual mechanisms which link smaller class size with higher academic achievement.

Robinson's (1990) research announced a new direction which recognized the complexity of the relationship between academic achievement and class size. The need to control potentially confounding constructs such as student past academic achievement, already emphasized by Glass et al. (1982), became central in most post-1980s class size studies as researchers recognized previous studies carried out on academic achievement and class sizes suffered from poor sampling, methodological flaws, or inadequate design of quasi-experiments (Finn, 2002; Slavin, 1989a). Research methodology was called to be more sophisticated and to account for differential effects on various student groupings (e.g., achievement, ethnicity, English mastery) within different contexts (e.g., school setting, class size, and instructional methods). Meanwhile, it is noteworthy to point out that research on class sizes at secondary or postsecondary levels continues to be limited to this day.

Critics of the Glass and Smith analysis (1979), such as Slavin (1989a), contended shortcomings of some studies selected within the meta-analysis included: short duration (as little as 100 hours of differentiated instruction); compared disproportionate sizes (one-on-one tutoring vs. 25 student class); or, evaluated subjects of nonacademic nature (such as tennis). However, most of the Glass and Smith conclusions were later sustained by subsequent research on large-scale class size reduction projects (Finn, 1998).

In spite of methodological differences, the research syntheses carried out by Glass (Glass et al., 1982; Glass & Smith, 1978, 1979), Slavin (1984, 1986; 1989a), and Robinson and Wittebols (1986), all concluded students enrolled in classes of less than 20 students performed better. Furthermore, smaller class sizes were associated with a significant increase in academic achievement, especially among the primary grades (K-3). Robinson and Wittebols (1986) and Smith, at al. (1982) announced a new research direction, indicating clearly reducing class size alone is not directly related to an increase in academic achievement unless teachers adopt different classroom procedures and instructional methods. Robinson (1990) also concluded that economically disadvantaged students as those who were most likely to benefit from smaller classes. Ten out 15 grade level (K-3, 4-8 and 9-12) studies on class size and academic achievement of low SES or minority students were found to favor smaller classes. These results were also found for low SES or minority students in the middle school/junior high grades in five of these 15 studies (Robinson & Wittebols, 1986). This observation was corroborated by later research (Ehrenberg et al., 2001a; Finn, 1998; Finn & Achilles, 1999; Krueger, 1999; White, 1982). Different demographics and different instructional context could no longer be accounted through district and school level archival data alone.

In summary, a consensus in the research has identified smaller classes (less than 20) as more conducive to producing higher academic achievement. Likewise, contextual factors such as family background (Coleman et al., 1966), student characteristics (Finn et al., 2003; Robinson, 1990; Slavin, 1986), or institutional resources (Gardner et al., 1983; Hanushek, 1997; Rivkin et al., 2005) were thought to impact academic achievement. Yet, to what extent remains at the heart of the controversy.

Large-scale State Experiments

Project Prime Time

First piloted in 1981-82 in a limited-size experiment of class size reduction in primary grades K-2 with student-ratios of 14:1, this five-year project initiated by Indiana Governor Lamar Alexander (future Secretary of Education during the George H. W. Bush presidency) started in earnest in 1984-85 with class size reduction of a PTR of 18:1 in grades K-3. By 2008-09, *Project Prime Time* was in its 25th year of implementation (Indiana Department of Education, 2010).

An early study (McGiverin, Gilman, & Tillitski, 1989) investigated the performance of second grade students at the end of two years of reduced class size instruction (19.1:1) demonstrated greater student academic achievement in Reading and Math measured by standardized tests than their counterparts in large classes averaging 26.4 students. Ten studies yielding 24 comparisons with 1,148 scores in Mathematics and Reading were combined into one analysis. Six studies involving randomly selected *Prime Time* schools were compared to four studies carried out on three schools with regular size classes. A total of 1,940 *Prime Time* student scores on standardized tests (Cognitive Ability Test – CAT, Iowa Test of Basic Skills – ITBS) in Mathematics and Reading were compared to the related performance of 2,027 students from larger classes in these ten studies. The mean differences between groups divided by the two groups pooled standard deviation were averaged within a meta-analysis to yield an effect size of .34 *SD* for all subtests (McGiverin et al., 1989). This analysis suggested *Project Prime Time* students enrolled in smaller class performed better academically. Yet, interestingly, the Indiana Department of Education stated on its Prime Time web page (Indiana Department of Education, 2010), "Lowering class size, alone, will not bring about better teaching and learning." Although the idea that smaller class size positively impacts student achievement is not questioned here, quality instruction and student engagement appear to be emphasized. More research was suggested to measure the impact of these constructs in the relationship between class size and academic achievement.

<u>Project Student Teacher Achievement Ratio Project</u>

From 1985 to 1989, the *Student Teacher Achievement Ratio Project* (STAR), carried out in Tennessee was the first statewide randomized class size reduction experiment of the kind, involving 76 schools, 1,200 teachers and 12,000 K-3 students over four years. Students were randomly assigned to either (S) a small class (typically 13 to 17 students), (R) a nonreduced class (22 to 26 students), or (A) a nonreduced class with a full-time instructional aide. Class sizes were reduced by one-third (seven students) on average (Wossmann & West, 2006). Teacher assignments were also randomized. This configuration was to continue over the four years of the experiment and data were collected from various sources including teacher interviews, academic achievement data, classroom observations, and teacher questionnaires. Students remained in this configuration from kindergarten until completion of grade 3. The following year, all students returned to full-size classes. In grades K-3, students enrolled in small classes consistently performed better than their nonreduced class counterparts on standardized tests (Stanford Achievement Test). After adjusting for nonrandom attrition and transition between groups, Krueger estimated effect sizes on academic achievement, expressed in standard deviations, to be .19 in third grade, .28 in first grade, and .20 in kindergarten (1999). Overall, students were found to perform better in Reading and Mathematics, outperforming their fellow students enrolled in regular classes by an average of .22 *SD*. Translated into percentile ranks, the differential between STAR classroom and non-STAR classroom was about five percentiles in K, 8.6 in first grade, and five to six percentiles in both second and third grade (Krueger, 1999).

Concurrently, researchers heading Project STAR reached similar conclusions. Effect sizes calculated as the mean score for small class (S) minus the mean score for regular class (R) and teacher-aide class (A) configurations [S-(R+A)/2] expressed in standard deviation units after four years. All students were found to benefit from smaller classes. Data collected in grades K-3 indicated higher academic achievement in small class configurations, with attainment ranging from 0.15 to 0.25 standard deviations as compared to larger class configuration performance. However, effect sizes of academic achievement were typically two to three times larger for minority students than for Caucasian students (Finn, 1998; Finn & Achilles, 1999). Follow-up data collected in subsequent years, from grades 4 to 8, suggested achievement gains were maintained (Finn et al., 2003). The design of the study was strengthened by the within-school implementation of the three configurations (S, R, and A) which allowed for better control of potentially confounding variables such as school setting (urban, suburban, rural), the socioeconomic status of the students, per-pupil expenditures, and gender of the students. All differences in academic achievement between groups favored small class sizes of about 18 students versus the larger class size configurations (e.g., 24 students with or without a teacher assistant). Gender and school settings were not found to interact with class size to result in higher academic achievement. As documented in STAR teachers' logs, the benefits of reduced class sizes extended beyond academic achievement. Teachers reported: (a) fewer class interruptions; (b) increased time-on-task; (c) faster resolution of potential discipline problem; (d) faster teacher feedback to students; (e) greater individualization of instruction; and, (6) a greater social integration on the part of the students, resulting in positive pro-social behaviors (e.g., collaboration, peer help, activity participation, and student engagement level) (Pate-Bain & Achilles, 1986).

In contrast, nonexperimental researchers using education production (econometric) models noted student attrition, cross-contamination of control and experimental groups (occurring when parents pressured the school administration for their child to be moved from larger to smaller class configurations), nonrandom assignment of teachers (administrator selection), and possible Hawthorne effects as potentially undermining the experimental sturdiness of STAR (Hanushek, 1999; Krueger, 1999; Rivkin et al., 2005). Isolating cohorts of students who remained in the program for four years (48% of the kindergartners initially enrolled), Hanushek calculated the performance of both control and experimental groups to be much lower than the estimates calculated by STAR program evaluators. For instance, while third-grade students in small classes performed 0.22 *z*-score above the nonreduced classes, the gap between reduced and nonreduced cohorts after four years was only 0.14. Similarly, in Mathematics, the gap between yearly samples and 4-year cohort for the same grade decreased from 0.18 *SD* to 0.10 *SD*. The treatment effect was mitigated by student mobility and possibly student SES since students with lower SES demonstrated higher mobility. This does not imply class size should not be considered. The evidence indicated class size reduction affects students differently (Finn & Achilles, 1999). In support of these views, Nye et al. (2004) remarked that public policies should target urban schools with high poverty student populations. In conclusion, most of the evidence in favor of class size reduction revealed smaller classes benefitted students differently according to individual student circumstances.

Based on this evidence, the federal government actively promoted class size reduction, citing STAR as a *prima facie* case in favor of expanding the small class size concept across the nation (United States. Congress Senate. Committee on Health Education Labor and Pensions, 1999). Although types of educational reforms, such as staff development, are effective in raising the level of academic achievement, public policies across most states promoted smaller class sizes under the pressure of public opinion, teacher unions, and parent groups (Grissmer, 1999).

Until the end of the millennium, the class size debate sharply divided proponents and opponents of smaller class sizes as local governments considered additional expenditures with the aim of reducing the inequalities Coleman first reported as strongly associated to socioeconomic status and race. The interest in class size reduction as a tool to improve student academic achievement culminated in 1998 with a U.S Department of Education and the Office of Educational Research and Improvement commissioned report (Finn, 1998). This report purported to be an overview of the previous two decades (late 1970s to late 1990s) of research on class size reduction, with the goal of providing evidence to guide and prioritize national educational policies, and clarify questions related to academic effects, cost-benefit analysis of small class sizes, and implications for practice and student behavior.

Project Wisconsin's Academic Achievement Guarantee

Building on the knowledge gained from the Tennessee experiment, *Wisconsin's Academic Achievement Guarantee* (SAGE) was launched as a five-year intervention program targeting low SES students in primary grades K-3.

Initiated in the 1996-97 school year, the program design included four components: (a) class size reduction to meet a teacher-student ratio of 1to15 (including arrangements such as two teachers for 30 students); (b) extended school day; (c) implementation of "rigorous" curricula; and, (d) staff development and a system of professional accountability. Thirty schools from 21 school districts meeting the criteria of 50% low SES students (based on free school lunch participation) began the program. K-1 grades were targeted the first year, and grades two and three were added in subsequent years. 14 schools with nonreduced class sizes (typically 22 to 24 students) in seven districts which participated in the SAGE program were deemed comparable based on family income, achievement in reading, ethnicity, and K-3 enrollment. These provided control data in this quasi-experiment. The intent of the researchers was to maintain classroom cohorts intact across the five years of the program. However, after the first year of implementation parents of students receiving instruction in nonreduced classrooms began to pressure school officials, requested their child to be transferred to smaller class size settings. Such switches from control to experimental subgroups contaminated the results of the study, which ultimately showed no greater gains for students with lower SES (Mosteller, 1995). Anecdotal records by the experimental group teachers suggested students demonstrated fewer instances of disruptive behavior, an increased desire to participate, and a more appreciative attitude towards others (Mosteller, 1995). Teachers further indicated potential discipline problems could be handled in a timely manner, and that academic learning time, including reteaching and instructional differentiation could be blended within lesson delivery. Towards the end of Project SAGE, under pressure from middle class parents (who did not meet the low SES requirement) to the state legislature, small class sizes (and presumably similar benefits) were extended to nondisadvantaged students. This move was qualified as readily available "insurance" (Graue & Rauscher, 2009, p. 11) in a more-is-better mindset. Again, the ethical researcher should question whether limited resources should be spent on equality for the sake of reaching equity.

California Class-Size Reduction

In 1996, following the successes of Project STAR and SAGE, the California legislature provided schools with over $1 billion to reduce class size. Unlike the other programs, CSR was not experimental and affected a staggering 1.6 million students at a projected cost of $1.5 billion per year (Bohrnstedt & Stecher, 1999), effectively reducing average PTR in grades K-3 classrooms from 28.6 students to no more than 20 students per teacher. By the 1998-99 school year, 98.5% of all eligible LEAs had embraced this voluntary program, servicing 92% of K-3 students enrolled in California schools (Bohrnstedt & Stecher, 1999). However, some school districts, such as Modesto Elementary (18,000 student Average Daily Attendance) and other small LEAs chose not to participate as their class sizes were already around 25 students (Illig, 1997). Whether it was believed that this size was small enough to be of academic benefit or the district was unwilling to accept the terms of the class size reduction grant is unclear.

At the end of its first year of implementation, approximately 18,400 additional teachers were hired, a figure that would increase a year later to 23,500 (Bohrnstedt & Stecher, 1999). The following school year 1997-98, the Governor's Budget suggested expanding CSR to fourth grade. The State Legislative Analyst's Office (Schwartz & Warren, 1997) recommended against the initiative, citing several obstacles impeding current and even future efforts of school reform through CSR in California, namely a shortage of qualified teachers and a lack of suitable facilities.

The rapid implementation across four levels, grades K-3, departed from the models followed in Tennessee (STAR) and Wisconsin (SAGE) in that CSR was introduced in three grade levels the very first year of class size reduction implementation in California, a move widely regarded as counterproductive (Achilles et al., 2002). Although the initial per-pupil funding of $600 was later raised to approximately $800, the CSR program was severely underfunded from the start as compared to the $2,000 per pupil additional funding of Project SAGE (Biddle & Berliner, 2002). California CSR also presented considerable challenges as compared to STAR. First, whereas in Tennessee where large classes had been reduced from classes of 22-26 students down to smaller classes of 13-17, California's overcrowded classrooms in the same primary grades averaged 33 students prior to CSR. California students were also more diverse than their Tennessee counterparts, with a larger population of English Learners and greater ethnic diversity. Furthermore, unlike California, Tennessee had space to accommodate class downsizing (Bohrnstedt & Stecher, 1999).

Due to these implementation characteristics, CSR in California had unintended effects upon poor and nonEnglish speaking students; the very students it sought to help. Overcrowded urban schools catering to lower SES students experienced the greatest difficulty in attracting qualified teachers and providing adequate facilities (Stecher, Bohrnstedt, Kirst, McRobbie, & Williams, 2001). For example, the California Legislative Analyst's Office reported in the first year of CSR implementation that over 90% of teachers in more affluent districts were credential holders versus approximately 75% of the teachers in urban, low SES districts (Schwartz & Warren, 1997). As a result, schools servicing students with minority and low SES profiles were perhaps the last ones to benefit from full implementation.

Experiment and Quasi-Experiment Research Summary

The first generation in class size research investigated whether or not class size improved academic achievement. By and large, research established that class size reduction positively impacted achievement. Yet, effect sizes estimated between 0.10 and 0.20 overall should not be compared to an absolute zero. On the contrary, some researchers argued that compared to other interventions yielding equal or superior results, class size reduction remained less effective compared to other interventions. For instance, Hattie (2005) ranked class size among 46 factors impacting academic achievement based on over 4,000 effects sizes derived from over 500 meta-analyses summarizing approximately 300,000 studies of factors linked to student academic achievement. The top ten influences (effect sizes; number of studies) were determined to be: feedback (0.8; 13,209); direct instruction (0.8;1,925); prior achievement (0.80; 619); lack of disruptive students (0.79; 1,511); quality of teaching (0.67; 808); phonological awareness (0.66; 429); early intervention (0.64; 30,275); peer assessment (0.63; 308); challenging goals (0.59; 959); and, self-assessment (0.56; 521). The average mean effect size for all 46 factors effecting student achievement was 0.40. Class size fell below this average, with an overall effect size of 0.13 (2,559 studies), which aligned with the 0.10 to 0.20 average of effects sizes found in major studies (see Table 2.1).

The small effect sizes for class size reduction found in all studies combined with some wide variations (SAGE, in particular) created new questioning and avenues for research. A second generation of class size studies was initiated with the goals of uncovering the mechanisms linking small class sizes and higher academic achievement, prompted by the acknowledgement that class size reduction alone may be a necessary but nonsufficient condition towards improving academic achievement.

Why then would class size and its modest effects be chosen over other intervention types as the primary instrumental policy of school reform in the late 20th century? Perhaps, the answer lies in what Graue and Rauscher (2009, p. 12) described as the "perfect storm." Indeed, class size was not a hard sell to parents, teachers, and politicians. It also coincided with a time of increased

Table 2.1

Effects Sizes of Landmark Meta-Analyses, Experiments and Quasi-Experiments

Author	Project	Studies	Class size	Compa-risons	Effect -size[a]	Subject
Glass & Smith (1980)	Meta-analysis	59	15-25	371	0.24	Composite
Slavin (1989b)	Meta-analysis	8	15-25	20	0.13	All subjects
McGiverin et al. (1989)	PRIME TIME	10	19.1-26.4	1	0.34	All subjects
Finn & Achilles (1999)	STAR	1	15-23	1	0.15–0.27	All subjects
Molnar et al. (1999)	SAGE	1	14-25	1	0.16	Reading
Molnar et al. (1999)	SAGE	1	14-25	1	0.20	Language
Molnar et al. (1999)	SAGE	1	14-25	1	0.25	Mathematics
Bohrnste t & Stecher (1999)	CSR California	1	20-30	1	0.05-0.10	All subjects

Note. Table as cited in Bohrnstedt & Stecher (1999), Finn & Achilles (1999), Glass & Smith (1978), Grissmer (1999), Hattie (2005), Molnar et al. (1999), Slavin (1989b).

[a] Effect sizes types are not defined but are believed to be Glass's Δ, and Hedge's d, prevalent at the time.

accountability and positive economic growth allowing additional resources to be injected into education. An additional example further illustrated the political context: Frank Mosteller (an evaluator of the Tennessee STAR Project) when interviewed by Graue and Rauscher (2009) indicated that California Governor

Pete Wilson had a bad experience with the powerful lobby of the California Teachers Association. Subsequently, he consented to spending additional monies in class size reduction rather than placing the monies in the general funds, a move that may have meant salary increases for California teachers. Class size, though expensive and less cost-effective than other school reform, was chosen as public policy for its political appeal to all stakeholders, from parents to teachers and politicians.

Contextual Factors and Academic Achievement

For decades, researchers suspected that direct causal models failed to adequately represent the complexity of the relationship between class size and academic achievement. Therefore, the next generation of research on class size was compelled to look inside the black box between predictors and outcomes. In the last decade, a consensus emerged in the educational community that studies had to look beyond simple direct relationship and unpack the complexity of indirect relationships. Research now focused on potential moderating factors in the model associating class size and academic achievement.

As most studies concurred that class size did impact academic achievement at least to some degree (Glass et al., 1982; Graue & Rauscher, 2009; Robinson & Wittebols, 1986; Slavin, 1986), especially in the primary K-3 grades, with minority students (Biddle & Berliner, 2003; Finn & Achilles, 1999), and with lasting effects (Finn, Gerber, & Boyd-Zaharias, 2005; Krueger & Whitmore, 2001; Nye, Hedges, & Konstantopoulos, 2001) but without significantly reducing the achievement gap (Konstantopoulos, 2008), it was evident that class size reduction affected students differently regardless of identical reduction in class size. As a result, the next wave of research tackled the mechanisms linking the constructs of smaller class sizes and academic achievement.

Researchers were also divided as to the effect of class size reduction on teacher-, student-, and school-contexts. Some insisted the attitudes and dispositions of the students were responsible for structural changes since teachers do not fundamentally change their practices from larger to smaller classes (Betts & Shkolnik, 1999; Clotfelter, Ladd, & Vigdor, 2006; Hanushek, 1971; Mitchell & Mitchell, 1999; Shkolnik, 1997). On the other hand, another school of thought argued smaller class sizes caused teachers to change their instructional delivery, modify their interactions with the students, or increase cooperative learning opportunities (Blatchford, 2005; Evertson & Burry, 1989; Zahorik, Halbach, Ehrle, & Molnar, 2003). In this debate, it is important to recognize that, while class size reduction created the opportunity for changing student-teacher interactions, maximizing the instructional potential of smaller groups relied on teacher expertise and school leadership. For instance, Rice's (1999) regression model at the classroom level predicted instructors in smaller high school Math and Science classes were spending less time on noninstructional tasks and devoted more individual attention to their students.

Research next focused on what constituted best practices within smaller class configurations. Instructional orientation (e.g., explicit step-by-step instruction, scaffolding, and frequent/immediate feedback on performance), management style (e.g., clear rules and procedures, seamless transitions between activities, logical sequencing of activities, reward system), and individualization focus (e.g., students articulating their thought in a dialectic communication with the teacher) were three traits identified as most effective teaching practices in the elementary grades (Zahorik et al., 2003).

Reducing class size was found particularly beneficial for lower-performing students in Mathematics. Biddle and Berliner (2002) pointed out that young students in primary grades benefitted from smaller classes as the acculturation process into schooling is facilitated. Teachers also reported to enjoy a higher level of job satisfaction (Bourke, 1986; Glass et al., 1982), increased collaboration with the home, and paid more individual attention to their students (Smith, Molnar, & Zahorik, 2003; Zahorik, 1999). The weakness of these conclusions was that these were only collateral findings within studies not directly aimed at uncovering the relationships between class size and academic achievement. Critics of these findings also pointed out the lack of consistency across studies. For instance, Betts and Shkolnick (1999), after collecting data on 2,170 classes of high school Mathematics, noted the teachers did not spend more time preparing for their classes or reviewing additional materials even though these structural changes allowed them to do so. Interestingly, they noted time shifted from whole group instruction to individual help with increased academic time devoted to review. In a similar qualitative study (Blatchford, Baines, Kutnick, & Martin, 2001), interactions between upper-elementary students and their teacher were increased by as much as 50%. However, studies dedicated to unpacking teacher contextual factors might have been affected by the biased opinions of teachers, whose working conditions had improved with reduction in class size (Graue & Rauscher, 2009). For instance, during the four years of Tennessee STAR, 1,000 teachers commented on the numerous ways smaller class size changed instructions including: faster coverage of the material allowed for expanded topics; use of supplemental texts and activities; student engagement with concrete materials; and, individualized instruction, to name a few (Pate-Bain, Achilles, Boyd-Zaharias, & McKenna, 1992). Clearly, teachers associated the better working conditions generated by smaller class size configurations with job satisfaction, and by extension higher productivity.

Years of teaching experience; highest degree conferred; and, professional development are the teacher factors most commonly considered in the research literature. The understanding of moderating factors such as teacher qualifications and student background in the relationship between class size and academic achievement was further enhanced by a national study conducted by the ETS Policy Information Center (Wenglinsky, 1997). This study was somewhat unique as it bridged the gap between econometric studies and quasi-experimental research. The study originated from a school finance approach, attempting to link spending of public funds and the overt goal of schooling: academic achievement. Therefore, it was only nonintentionally that Wenglinsky stumbled on the connection between class size and academic achievement.

The scope of *When Money Matters* (Wenglinsky, 1997), not unlike the *Coleman Report* thirty years earlier, covered the nation; however, with dramatically different conclusions. Using district-level data from three different databases maintained by the National Center for Educational Statistics (NCES), Wenglinsky grouped 10,000 fourth-graders in 203 districts and 10,000 eight-graders in 182 districts according to socioeconomic status. The linking of these databases allowed differentiation between types of spending in a way not previously possible at the time the *Coleman Report* was produced. Furthermore, the *Coleman Report* was unable to consider cost of education variation across states. Indeed, aggregated spending per-pupil-expenditure (PPE) cannot account for the types of expenditures incurred, some of which were positively linked to academic achievement while some were not. Wenglinsky suggested a model (Figure 2.1) which resolutely departed from direct causal class size-academic achievement models found in education production - also known as econometric - studies, (e.g. Coleman et al., 1966; Hanushek, 1998).

Through a series of multivariate regressions, Wenglinsky (1977) concluded increasing school district administration and instructional expenditures to decrease PTRs raised fourth-grader academic achievement in Mathematics as measured by the National Assessment of Educational Progress (NAEP – no data/effect sizes were reported).

The decrease in PTR was believed to decrease behavioral problems among students and set a positive tone to school environment. Administration and instructional PPEs were positively linked to an increase in academic

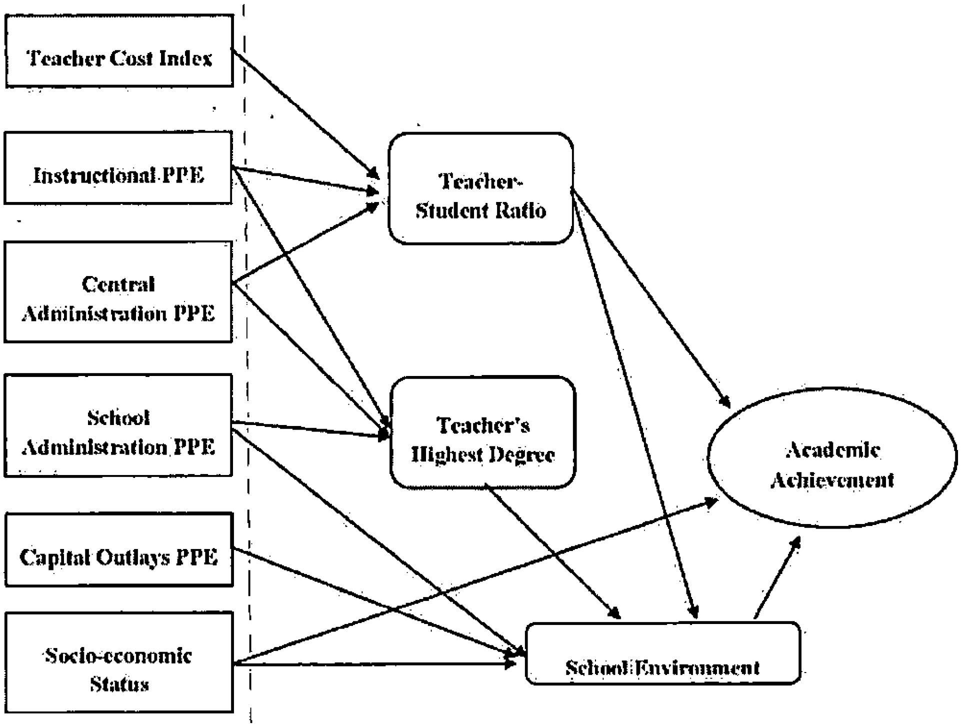

Figure 2.1. Wenglinsky hypothesized paths to achievement. Adapted from Wenglinsky, H. (1977). When money matters: How educational expenditures improve student performance and how they don't. A policy information perspective *Policy Issue Perspectives.* Princeton, NJ: Policy Information Center, Educational Testing Service.

achievement in 8th grade. Interestingly, spending on facilities, school-level administration, and expenditures to recruit highly educated teachers were not found to be directly associated to academic achievement. Wenglinsky concluded, "Because the [previous] studies did not specify measures of school environment, the effect of school spending on achievement as moderated by environment remains unstudied" (Wenglinsky, 1997, p. 21). In the middle/junior high grades academic achievement appeared to be moderated by an increase in social integration attributed to smaller class size. Building a 2 X 2 factorial matrix combining district with above- and below-average socioeconomic status (SES) and districts with above- and below-average teacher cost, Wenglinsky concluded the largest gains in achievement in Mathematics were obtained in districts with below-average SES students and above-average teacher cost. In eighth grade, PTR was linked to a positive school environment (low teacher- and student-absenteeism, respect of property, low class cutting rate, low tardiness rate, teacher control over instruction/course content). Positive school environment, in turn was positively associated with higher achievement in Mathematics. In the light of these findings at the school level, more research is needed to refine these observations at the classroom level, particularly at the junior high/middle school level. This direction for future research partially provides justification for the present study.

Teacher quality is often referred to a combination of licensure status and years of experience. Yet, calculating effects of teacher contextual factors in the relationship between class size and academic achievement has been impeded by the "positive matching" of students and teachers (Clotfelter et al., 2006), exemplified by more affluent, better educated students assigned to classes of more qualified teachers as a result of parental interventions or requested teacher assignments. Teacher characteristics impact both quality instruction and academic achievement, yet in different ways. A regression analysis of class size reduction in third grade calling for a composite of teacher characteristics (e.g., percentage of teacher in their first year of teaching, percentage of teachers in their second year, percentage of teachers not fully licensed, and percentage of student with no graduate education) led Jepsen and Rivkin (2002) to similar conclusion when reviewing California CSR. Jepsen and Rivking along with other researchers (Betts & Shkolnik, 1999; Rivkin et al., 2005; Rockoff, 2009) found little or no evidence that "teacher certification or education was significantly associated to the quality of instruction" as per student achievement metrics (Jepsen & Rivkin, 2002, p. 45). On the other hand, novice teachers were associated with a decrease in Mathematics and Reading achievement of four percentage points (for students exceeding the national median – test unknown), thereby canceling the positive effect possibly created by class size reduction.

In subsequent work, Wenglinsky (2000) suggested beyond certification and professional development of teachers, instructional practices accounted for the most influential factor in increasing academic achievement as measured by Grade 8 Mathematics NAEP of the 1996 administration. The above combined characteristics (Figure 2.2) were found to have a greater impact on academic achievement.

Instructional practices based on hands-on activities in Science as well as comprehensive summaries of the curricula such as in group reports seemed to favor higher level of thinking skills associated with improved academic achievement both in Science and Mathematics (Wenglinsky, 2000). Instructional practices also impacted other factors believed to indirectly impact academic achievement: time on task, time spent by the teacher on administrative task as well as, time spent addressing student disruptions.

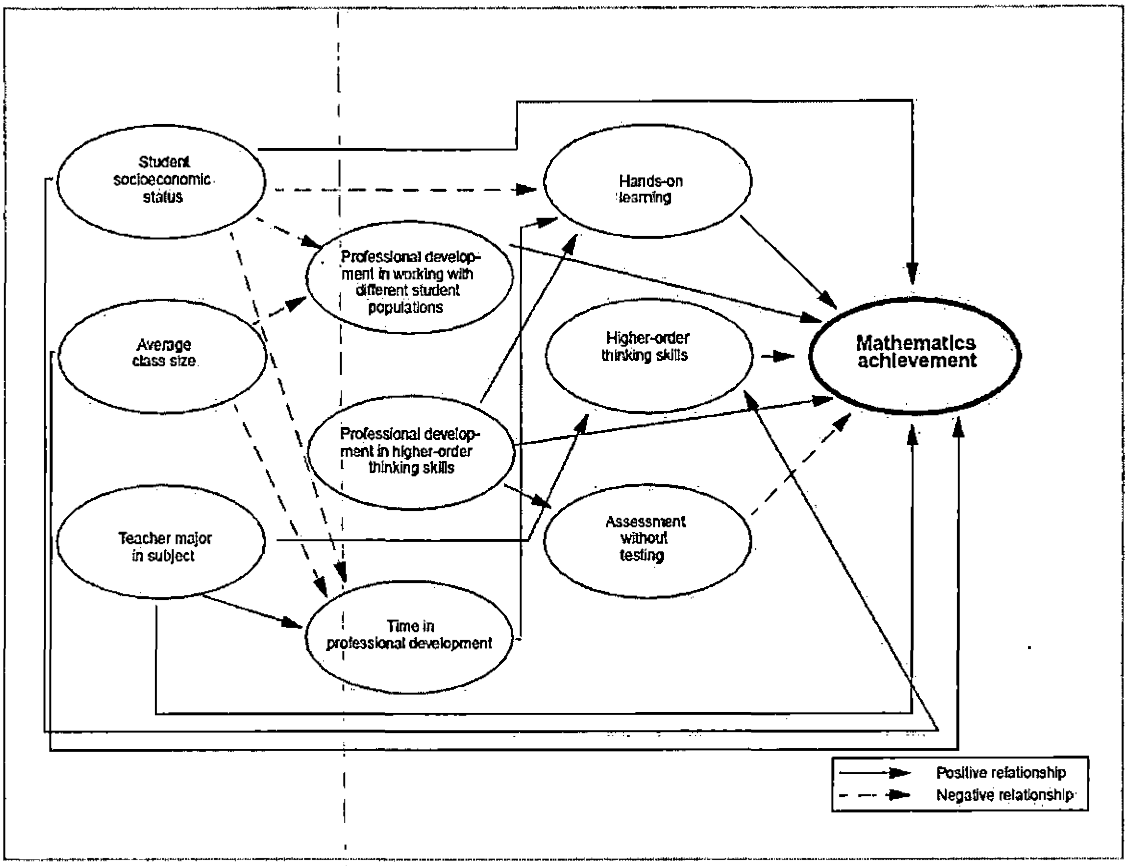

Figure 2.2. Links among teacher inputs, professional development, and student performance in mathematics. Adapted from Wenglinsky, H. (2000). How teaching matters: Bringing the classroom back into discussions of teacher quality *ETS Policy and Research Reports* Policy Information Center, Educational Testing Service.

Due to the complexity of designing teacher context analyses and qualitative observations, on the one hand, and the finding that teachers do not change their methods when class size is reduced (Allington, Stuetzel, & Shake, 1986), a possible causal link between smaller class size, better instruction, and improved academic achievement is far from reaching a consensus in educational research. Allington et al. (1986) observed teachers involved in small reading group programs were found to revert to using worksheets and whole group instruction. Bourke (1986) studied the extent to which a causal relationship between class size and academic achievement in elementary Mathematics was moderated through instructional practices using a hierarchical regression model including three blocks: background factors (students, school, teachers); background factors and class size; and, background factors, class sizes, and teaching practices. Once the multiple regression model established a positive link between smaller class sizes and achievement, the following teaching practices were associated with higher achievement: greater use of groupings in larger classes; whole class instruction in smaller class; greater number of interactions between students and teacher; and, increased time monitoring student work in smaller classes. Interestingly, the first block (including background factors such as teacher experience, previous level of student achievement, and teacher experience) accounted for 29% of the variance explained. When adding class size to the model, 37% of the variance in academic achievement was explained. Finally, the total model including the last block (teaching practices) account for 85% of the variance explained. Clearly, teaching practices more than certification or experience, impact student achievement. Furthermore, it was also suggested that the quality and intensity of teaching is inversely proportional to class size. As larger groups are more likely to be heterogeneous, teachers tend to reach out to all students by adapting their instruction. In so doing, they tend to lower their teaching standards so that students with average abilities may succeed (Schussler, 2009).

Although the possible positive main effects of class size reduction on academic achievement are further compounded by more effective teacher instructional practices, researchers (Betts & Shkolnik, 1999; Finn et al., 2003; Odden et al., 2007) argued students' attitudes and responses were also more likely to moderate any possible causal relationship.

Students in larger classes tended to engage in more peer-to-peer interactions, not only for off-task activities or disruptions, but also for on-task activities (Blatchford, Edmonds, & Martin, 2003). Since teachers were less likely to provide small group instruction, peers tended to obtain clarifications from one another, and the question arose whether small classes did not create a counterproductive situation where students tended to be overly dependent. Perhaps, student engagement is the most commonly cited benefit of smaller classes (Deutsch, 2003; Downer, Rimm-Kaufman, & Pianta, 2007; Finn et al., 2003; Schussler, 2009). Smaller classes appeared to increase motivation based on cohesiveness between instructor and students; to a lesser extent, similar benefits were observed laterally in peer-to-peer relations (Bolander, 1973). One possible explanation lies in that teachers in smaller classes are more likely to convey positive academic support and the belief that all students can succeed (Schussler, 2009). The analysis revealed that class size substantially explained variations in individual and intragroup vertical (student-teacher) motivation level, and, to a lesser extent, it also explained variations in intragroup lateral (student-to-student) motivation levels. Students tended to be less distracted in smaller classes and exhibit less nonparticipatory or disruptive behaviors (Finn & Achilles, 1999; Smith et al., 2003). These pro-social and anti-social student behaviors were further conceptualized. Finn et al. (2003) proposed four mechanisms to explain the impact of small classes on student academic engagement: diffusion of responsibility; social loafing; group cohesiveness; and, psychological sense of community. These factors are related to a sense of belonging. As class size increases, students perceived their collaborative roles in the class as being of less importance, and their sense of responsibility towards the group decreases accordingly (social loafing).

Along with teacher techniques, student behavioral changes related to class size are central factors to understanding the association between class size and academic achievement. Today, researchers set out to better understand the unique characteristics of one-on-one tutoring (Bloom, 1984) with the hope of replicating beneficial practices in the context of larger classes. As one-on-one tutoring was associated with gains of approximately two sigmas (standard deviation, i.e., a 40 percentile gain) on standardized test scores, the central question needs to focus on determining the most influential contextual factors. This set a new direction for research, which prompted the reanalysis of some of the large experimental programs such as STAR and SAGE. Research methodologies departed from the strict quantitative approach applied in econometric studies (Coleman et al., 1966; Hanushek, 1998, 1999) to include qualitative elements such as case studies, classroom observations, and student-teacher-questionnaires (Blatchford, Russell, Bassett, Brown, & Martin, 2007). How the teacher-, school-, and student-contexts moderate the relationship between smaller classes and academic achievement has seldom been the object of research at the middle and high school levels. In light of the potential benefits for at-risk students, the study of class size and academic achievement at the middle and high school levels is urgently needed as research is very limited.

Using a dataset form the Longitudinal Studies of American Youth, students in 100 middle and high schools were followed over a five year period starting in 1987, Shkolnik (1997) hypothesized that most studies on class size and academic achievement suggested little or no effects as the classroom student average level of ability was uncontrolled. She concluded controlling class ability was necessary as high achievers seemed to be placed in larger classes, while students of lower ability may be placed in smaller classes.

Research prior to the 1990s was largely focused on establishing a direct causal effect between class size and Academic Achievement. Effects sizes were estimated between .10 and .20 *SD* overall. As statistical tools became more sophisticated, researchers attempted to understand the mechanisms of this relationship. Classroom context (teachers and instruction factors) were studied as moderators, while striving for better control of student variables. The present study followed this tradition.

Hypothesized Model

As previously suggested, the relationship between smaller class size and academic achievement as measured by standardized assessment in English/Language Arts and Mathematics is moderated by classroom context factors (see Figure 2.3). A moderated relationship was favored instead of a mediated relationship as the current study focuses on the impact of classroom-context factors and class size on Academic Achievement. In a moderation model, an interaction of two factors impact the outcome, while at the same time impacting that outcome each individually (Baron & Kenny, 1986).

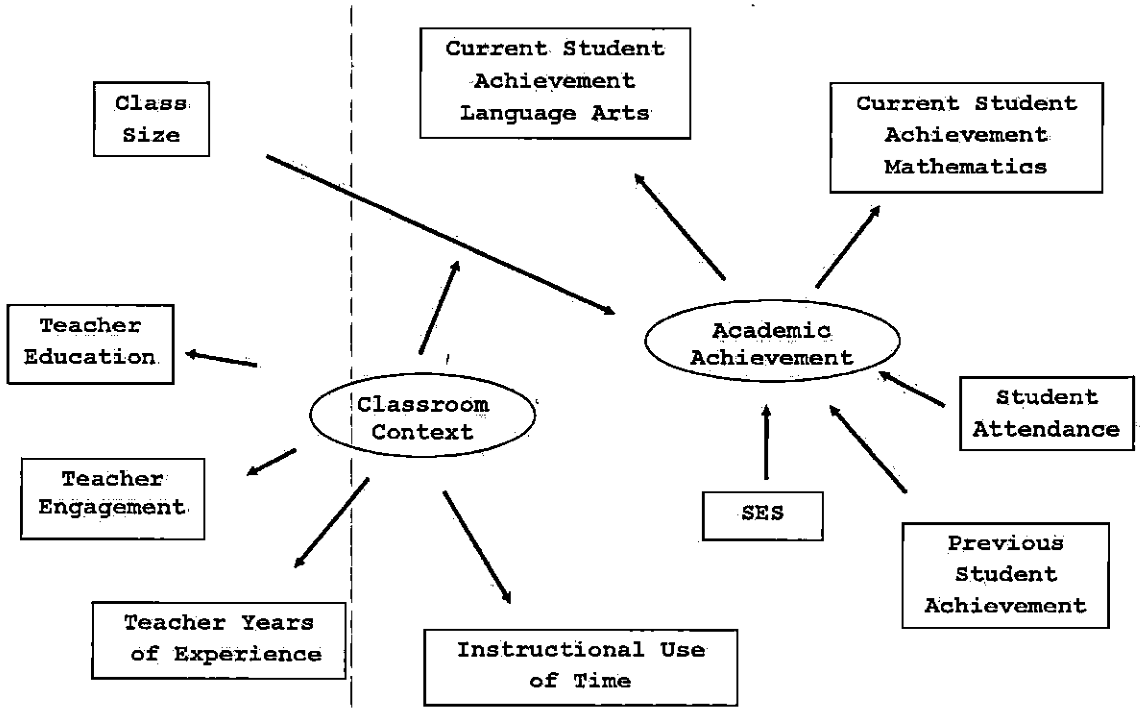

Figure 2.3. Study hypothesized model of the relationship between class size and academic achievement.

It was hypothesized that Instructional Use of Time, Teacher Experience, and Teacher Engagement, moderated the Class size and Academic Achievement relationship. For instance, individual student seatwork assignment denotes a type of instructional activity unlikely to produce greater academic achievement regardless of actual class sizes, be it 25 or 35 students. However, small group instruction or whole group instruction may moderate the impact of smaller class sizes on Academic Achievement.

It was also hypothesized that the smaller class sizes lead to a decrease of the amount of time spent by teachers in administrative and discipline tasks in middle schools. This decrease, in turn, leads to maximizing academic learning time, and thereby potentially increasing academic achievement. Finally, SES and previous level of academic achievement at the student and classroom level must be controlled as confounding constructs. Unlike most of the body of research currently available, the proposed study extended beyond school level analysis to reach both the classroom and student levels of analysis.

Research Hypotheses

Hypothesis 1: Two separate models, one in Language Arts and one in Mathematics, tested whether differences in academic achievement existed between students enrolled in QEIA reduced-size classrooms versus nonreduced classrooms after controlling for specific within-classroom constructs (student level – L1) and between-classroom constructs (class level – L2) after determining the suitability of a multilevel linear model through running an unconditional 2-level analysis without predictors.

Hypothesis 2 tested whether an interaction existed between student socioeconomic status and classrooms fixed effects: Pupil-to-Teacher Ratio, Teacher Engagement, Teacher Experience, Teacher Education, QEIA participation, and Instructional Use of Time. Two full models will be included, one of each subject matter.

Hypothesis 3: Model tested for interaction between previous level of Achievement at student level and the level-2 fixed effects described in question 2: Pupil-to-Teacher Ratio, Teacher Engagement, Teacher Experience, Teacher Education, QEIA participation, and Instructional Use of Time. It was hypothesized that students with lower previous level of Academic Achievement would obtained the greatest gains in Academic Achievement in both English Language Arts and Mathematics.

Hypothesis 4: Similar to questions 2 and 3 but focused on the interaction between student English proficient status and level-2 fixed effects: Pupil-to-Teacher Ratio, Teacher Engagement, Teacher Experience, Teacher Education, QEIA participation, and Instructional Use of Time.

CHAPTER THREE

RESEARCH METHODS

The proposed study targeted five middle schools of a large high-poverty high-minority suburban Southern California K-12 school district. Two of the five schools (schools 1 & 2) were selected for the Quality Education Improvement Act of 2006 (QEIA), a state grant aimed at reducing class size in 488 selected K-12 schools ranking in the lowest two deciles of the 2005 base Academic Performance Index (API) statewide. The remaining three participating schools (schools 3, 4 & 5) did not qualify as their 2005 base API exceeded the second decile criteria set by the QEIA grant requirements. Implementation of class size reduction in both participating middle schools began in school year 2008-2009.

One year later, 2009-2010, both schools 1 & 2 receiving QEIA funding showed academic improvement in Mathematics and Language Art, and moved up to the same deciles as schools 3, 4 & 5. The performance of QEIA school 1 reached the third decile, and matched the performance of schools 3 & 4; the second QEIA school (school 2) ranked in the fourth decile, and matched the performance of school 5. At the onset of the 2009-2010, the academic performance of all five participating schools matched closely. The proposed study used a multilevel model based on quantitative data sources both at the school and student level.

Student Participants

Participants in the study included middle school students continuously enrolled in five Southern California middle schools in 2009-2010 and their assigned Language Arts and Mathematics teachers for the same period as defined by the school master schedule.

Continuous enrolment was defined as participation in the instructional programs of one and only one school, with an enrolment date prior or on October 7th, 2009 and an exit date after April 22, 2010, the beginning date of the California Standards Test (CST) administration time window. Students who entered or exited a school between these two dates were excluded. Other criteria for inclusion and exclusion are described below. Information pertaining to participating students and their respective teachers were de-identified from all records. All data was maintained on a password-protected hard drive.

Language Arts and Mathematics teachers of the seventh graders participating in the study were included. Teachers who taught Language Arts and Mathematics were recruited to take part in a voluntary survey. This information was used both at the student and classroom levels. The relational database primary key linking student and teacher data files was removed prior to analysis, thereby ensuring de-identification of all participants.

Inclusion/Exclusion Criteria

Middle schools are traditionally departmentalized, and school days are divided into five or six periods (also known as sections) according to a master schedule. During each period, departmentalized single subject teachers instruct different groups of students, commonly referred to as sections. Only Mathematics and Language Arts core instruction class offerings were considered in the present study. English Language Development (ELD) and all other supplemental or remedial class offerings were excluded. Classes in QEIA exceeding a PTR of 25:1 were omitted from analysis. Likewise, classes in nonQEIA schools with a PTR equal or lower than 25:1 were not considered in the analysis.

The current study did not include students and teachers in sections designated exclusively for extremely high- and extremely low-achievers in order to preserve the central assumption of uncorrelated error between student variables (such as prior achievement) and class size assignment. For instance, it was likely that lower achievers enrolled in special education program be assigned to smaller class size sections. At the high end of the academic achievement continuum, Gifted and Talented (GATE) students may be organized in sections labeled "Honor," or "GATE." Similarly, sections organized for students with disabilities, may be labeled "Resource," "Resource Specialist Program (RSP)", "Special Day Class (SDC)," "Learning Handicap (LH)," or "Severely Emotionally Disturbed (SED)." Such sections were not considered in the study. Furthermore, students labeled as participating in special education or in the Gifted and Talented (GATE) programs who were instructed in general education classroom were also excluded. Although these students are not the object of the present study, they were still taken into account when reporting class sizes.

The choice of seventh grade as the grade level for participation in the study lies in that seventh grade core curricula are common to all students. Indeed, it is not until eighth grade that students are noticeably segregated according to achievement levels; GATE sections or Honors sections typically cater to high achievers at that grade level. Furthermore, unlike in eighth grade, Mathematics coursework set forth by the California frameworks and standards remains general Mathematics in grade seven as opposed to general Mathematics and algebra in eighth grade.

Recruitment

Procedure for recruitment stressed the voluntary nature of teacher participation. Prior authorization to undertake research was secured with the California State University, San Bernardino Institutional Review Board (IRB). Permission to conduct research in the school district was secured (Appendix B), and volunteer teacher participants gave their informed consent before taking a survey to measure use of instructional time. The California State University, San Bernardino Institutional Review Board previously approved the research and the letter of informed consent (Appendix A).

Measures

Teacher Questionnaire

After informed consent had been granted (Appendix B), volunteer teachers were asked to report instructional time spent on classroom activities by answering multiple-choice questions based on a five-point Likert scale (Appendix C). Betts and Shkolnik (1999) developed the *Instructional Activity Survey* to study the behavioral effects of class size reduction in Mathematics at the high school level. They concluded the potential benefit of smaller class was affected by instructional grouping and differentiation. Similar studies at the middle school level and in Language Arts have not been carried out. The proposed study targeted students and classrooms in these contexts.

The *Instructional Activity Survey* was administered during teacher preparation days. In the middle schools, five or six teachers typically form core subject departments such as for Language Arts and Mathematics. The *Instructional Activity Survey* was printed on optical scan sheets and bar-coded with the section number assigned in the school master schedule for 2009-2010. This allowed for pairing of sections, teachers, and students. Each department holds weekly meetings to discuss curriculum and organization of instruction.

The *Instructional Activity Survey* measured the amount of time devoted to group and individual instruction. Use of time was subdivided between instructional and noninstructional time. Instructional times refer to the academic learning activities carried out in the classroom. Five types of teacher-led activities were considered: lecturing; leading a classroom discussion; working in small groups; doing seatwork; and, providing differentiated instruction. Participants indicated on a 5-point Likert scale the weekly amount of time (0 minutes, thirty minutes, one hour, two hours, and more than two hours) spent on such activities. Noninstructional Use of Time denotes the amount of time spent by teachers on Administrative Tasks or on Discipline.

The *Teacher Engagement Scale,* consisted of ten self-reported items was built on a five-point Likert scale (Appendix D). This questionnaire was developed based on the California Healthy Kids Survey (WestEd, Spring 2011). Every year, California students in grades 5 through 12 are invited to fill in a survey with questions to assess school climate, pro-social and risks behaviors. Key-learning and behaviors such as school connectedness and relations with adults are measured to better understand the impact of these factors on learning. For the purpose of the present study, questions addressed to students regarding teacher engagement were rephrased so that teachers would assess their personal level of Engagement. Question ten ("When I am in class, my mind wanders") was reverse-coded; scale reliability of the Teacher Engagement scale was 0.66 as measured by Cronbach Alpha.

Procedures

In summer 2010, application to conduct research was filed with the district, and permission was granted on August 23, 2010. Copies of the approval (Appendix A) were signed by the district Director of Assessment and Evaluation, and copies were forwarded to middle school site principals. Prior to survey administration to the teachers, site administrators and head of departments of both Language Arts and Mathematics were contacted. The survey was administered at a weekly department meeting in Winter 2011 at each of the five participating schools. Teachers absent or reassigned to sites other than the five schools mentioned in the study were contacted to request participation. Teachers who taught the student participants were identified by matching the 2009-2010 master schedule for 7th grade with the current staff roster in each school. Teachers who left for another school within the district were identified through the searchable district email database. Teachers who were no longer in the district were contacted at their last known address as per the emergency contact files maintained in each school office. As an incentive to participate in the study teacher participants of the ten departments (Language Arts and Mathematics departments at five different schools) were given the chance to win one $25 gift card per department at each site. At the conclusion of the data collection, the gift cards were awarded, using a lottery.

<u>Archived Data</u>

Archived data included student demographic characteristics and achievement in Language Arts and in Mathematics. These data also included faculty years of teaching experience. Permission to conduct research in the district under consideration was granted, and data specifications were submitted to their technology department in order to produce electronic data files. Student, classroom (also known as section), and teacher data were matched prior to being de-identified. The completed data requests with the district technology department and the original student data files released to the district by

Education Testing Services (ETS) after administering the CSTs in the spring of 2010 provided student demographic characteristics and performance achievement. This extracted data file included the following: Academic Achievement (scale score, and performance levels) of the 2010 CST administration; Academic Achievement of the 2009 CST administration; Participation in National School Lunch Program; self-reported Parent Level of Education; and, English proficiency status.

Academic Achievement. Academic Achievement was measured by performance on the California Standards Test (CST). In 1999, the California Board of Education introduced the CST in Language Arts (ELA) and Mathematics (MA) for grade two through 11 as a measure of academic achievement within the more comprehensive educational accountability program coined as the Standardized Testing and Reporting (STAR). The performance levels and scale scores on these high-stake tests were used. Scale scores were aggregated to account for the Academic Achievement for groups of students. Educational Testing Service (ETS) based in Princeton, NJ is the official contractor and publisher of these criterion-reference tests based on a multiple-choice format: 75 questions in ELA, and 65 questions in Mathematics for grade seven. Test questions are aggregated into five or six clusters. No item analysis is made available by the test publisher or California Department of Education (CDE). Finally, scale scores spanning from a low 150 to 600 are divided in five ranges denoting student performances level, from low to high: Far Below Basic; Below Basic; Basic; Proficient; and, Advanced.

Academic Achievement measures for the 2009-2010 7th grade cohort were available for both 2009-2010 (7th grade CST) and 2008-2009 (6th grade CST). 2008-2009 data provided a measure of prior level of academic achievement. Table 3.1 presents the scale score ranges corresponding to the five achievement levels established by CDE for California schools. Only the proficient and advanced levels are considered as at grade level performance. The Ns included in Table 3.1 refer to the number of 7th grade students in participating middle schools only. A total of 1,603 student participants were selected in English/Language Arts sections and 1,591 in Mathematic sections.

Even though test questions are equally weighted, the scale score is more appropriate as a measure as it reflects adjustments to raw scores to account for differences in question difficulties from year to year.

Although standardized testing in ELA and Mathematics has severe limitations as it only refers to one type activity (multiple-choice answers) to measure academic achievement, it remains the one state-wide measure of achievement against which the performance levels of all California schools are being measured.

In a technical report released in 2010, ETS estimated the Cronbach's Reliability Coefficient of the CST to be 0.93 in the 2009 English Language Arts and Mathematics seventh grade test (Educational Testing Service, March 2010).

Socioeconomic Status. Low socioeconomic status students are students who qualified for free or reduced lunch under the National School Lunch Program (NSLP) in 2009-2010 or (nonexclusive) whose parents did not graduate from

Table 3.1

California Standards Test 7th Grade Cohort Performance Level Scale Score Ranges of Proposed Participants

Performance level	English language arts		Mathematics	
	6th Grade	7th Grade	6th Grade	7th Grade
Advanced	394 – 600 (*n* = 108)	401 – 600 (*n* = 123)	394 – 600 (*n* = 127)	414 – 600 (*n* = 158)
Proficient	350 – 393 (*n* = 482)	350 – 400 (*n* = 547)	350 – 393 (*n* = 418)	350 – 413 (*n* = 483)
Basic	300 – 349 (*n* = 564)	300 – 349 (*n* = 613)	300 – 349 (*n* = 507)	300 – 349 (*n* = 592)
Below basic	268 – 299 (*n* = 167)	263 – 299 (*n* = 201)	253 – 299 (*n* = 301)	257 – 299 (*n* = 302)
Far below basic	150 – 267 (*n* = 45)	150 – 262 (*n* = 102)	150 – 252 (*n* = 57)	150 – 256 (*n* = 97)
Missing	*n* = 237	*n* = 16	*n* = 241	*n* = 17
Invalid	-	-	-	*n* = 2
Total *n*	1,603	1,603	1,651	1,651

Note. Educational Testing Service. (2010). 2010 STAR posttest guide. Retrieved November 2, 2010, from http://www.startest.org/reports.html

high school. This information is self-reported by parents upon enrollment of their child in the district. The two indicators of socioeconomic status, NSLP participation and parent level of education, are readily available from the data collected by Education Testing Service (ETS), publisher and administrator of the test. The district data file sent to the test publisher contains these data. Missing data regarding parent level of education is resolved by the district prior to sending preidentification student file to ETS for the purpose of printing individualized answer sheets.

Class Size. The measure of class size was provided by the district technology department for the five participating middle schools. The data allowed to determine the actual number of students enrolled in each section at the end of the second trimester, a time that closely coincide with the administration of the CST's. Therefore, class size in this study was defined as a Pupil-to-Teacher Ratio (PTR) equivalent to the actual number of students who received instruction from one teacher on any given day just prior to the spring administration of the CST's.

Attendance. Attendance was defined as the number of days of student presence divided by the total number of possible days of presence at the same school site. As indicated previously, only the students continuously enrolled from October 2009 to April 2010 were included in the study.

Teacher Length of Service. Teacher length of service was defined as the number of year of teaching service credited by the school district to place the employee on the uniform salary schedule.

Teacher Education. Teacher Education indicated whether a teacher held a Bachelor's or Master's degree.

Rationale for Multilevel Linear Models

Multilevel linear models refer to nested structure analyses also known as hierarchical linear models (HLM) in sociological research, mixed effects models and random-effects models in biometric studies, and random-coefficient regression models in econometric research (Raudenbush & Bryk, 2002). Unlike aggregated regression models, HLM takes into account “within classroom” sources of variance at the student level (Level 1) and “between-classroom” variance at the classroom level (Level 2).

At the lowest level, the model for each classroom is written as:

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{0j}X_{ij} + r_{ij}$$

where *Y* is the dependent variable for i^{th} student in j^{th} classroom. In the present study, the outcome was student Academic Achievement while *X* denoted an independent constructs also acting as covariate at the student level (e.g., SocioEconomic Status, English proficiency, prior level of Achievement, and Attendance); r_{ij} denoted the residual. If all student-level independent constructs are included in this model, Level 1- equation for Mathematics achievement was written as:

$$(CR\_MA)_{ij} = \beta_{0j} + \beta_{1j}(SES) + \beta_{2j}(LEP) + \beta_{3j}(ATT) + \beta_{4j}(PR\_MA) + r_{ij}$$

At the second level, the intercept (β_{0j}) and slope coefficients ($\beta_{1j}, \beta_{2j}, \beta_{3j}, \beta_{4j}$) of independent constructs may become outcomes of a fixed effect (mean) and a random effect (error). Luke (2004) suggested testing the overall need for HLM by testing first the intercept as outcome while assuming fixed slope coefficients. In the current study, this step, taken in hypothesis 1 below determined if mean differences existed between classrooms receiving reducing PTR and classrooms with nonreduced PTR. However, a simpler structure with two independent variables, one at each of the student and classroom levels is written as:

Student Level: $Y_{ij} = \beta_{0j} + \beta_{0j}X_{ij} + r_{ij}$

Classroom Level: $\beta_{0j} = \gamma_{00} + \gamma_{00}w_j + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}w_j + u_{1j}$

In this model, both intercept and slope are allowed to vary. As described above, the slope coefficient β_{1j} can be replaced in the student level equation by its classroom level value, which implies a cross-level interaction between independent variables. $\gamma_{00}, \gamma_{00}w_J, \gamma_{10}$, and $\gamma_{11}w_J$ are fixed effects while, r_{ij}, μ_{0J}, and μ_{1J} are random effects.

For instance, a study aimed at determining if class size influences and possible interactions with only SES and previous level of achievement at the student level could be written as:

Student Level: $(CR\_MA) = \beta_{0j} + \beta_{1j}(SES) + \beta_{2j}(PR\_MA) + r_{ij}$

Classroom Level:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(QEIA) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(QEIA) + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}(QEIA) + u_{2j}$$

If no-cross level interaction is considered in an intercept-as-outcome model, only mean variations would be considered across classroom without interactions with student variables. In this case, the model was written as:

Student Level:

$$(CR\_MA) = \beta_{0j} + \beta_{1j}(SES) + \beta_{2j}(PR\_MA) + r_{ij}$$

Classroom Level:

$$\beta_{0j} = \gamma_{00} + \gamma_{00}(QEIA) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$

Independent constructs at the student level and classroom level are likely to determine both individual and overall test performance on standardized tests, thereby causing a violation of the assumption of uncorrelated error necessary to carry out classical regression model. Multilevel regression models, however, remove the concern of including several students of an identical classroom (Ehrenberg et al., 2001b). Individual students are nested in classrooms.. The interdependent nature of these levels may be exemplified by the influence of the school socioeconomic level on the individual student performance, and led researchers to reexamine the STAR Tennessee large class size reduction experiment with the insight provided by a hierarchical (multilevel) linear model (Nye et al., 2000). The constructs of interest considered in this study (Table 3.2) were divided between the student level (L1) and the classroom Level (L2). The outcome student Academic Achievement level in English Language Arts (CR_LA) and Mathematics (CR_MA) are based on the California Standards Tests (CST) scale scores of the spring 2010 administration. These scores have the same ranges as the previous measures of Academic Achievement from spring 2009.

Previous Academic Achievement level, on the other hand, becomes the mean of all the previous achievement scale scores obtained by student *i* in classroom *j*. Therefore, the two units of analysis will be both present in both levels, representing different measurements.

Hypotheses

Question 1 Hypothesis

Question 1: Two separate models, one in each subject, Language Arts and Mathematics, tested whether differences in academic achievement existed between students enrolled in QEIA reduced-size classrooms versus students instructed in regular-size classrooms after controlling for specific within-classroom constructs (student level – L1) and between-classroom constructs (class level – L2). Building a final multilevel model involves adding student level and classroom level constructs in successive steps in the hope of reducing the error components (unexplained variance).

Table 3.2

Constructs of Interest

Level	Label	Range	Construct acronym
Student-context (L1)	Socioeconomic status	Low, High	SES
	ELA current achievement	150-600	CR_ELA
	Math current achievement	150-600	CR_MA
	ELA prior achievement	150-600	PR_ELA
	Math prior achievement	150-600	PR_MA
	English learner status	Yes, No	LEP
	Attendance (%)	0-100	ATT
Classroom-context (L2)	ELA prior achievement	150-600	PR_LA
	Math prior achievement	150-600	PR_MA
	Use of instructional time	1-5	UIT
	Class size program	Reduced, nonreduced	QEIA
	Pupil-to-teacher ratio	12 - 38	PTR
	Teacher length of service	1-40	TEX
	Teacher engagement scale	1-50	TEG
	Teacher education	BA, MA	TDG

First, the suitability of a multilevel linear model was determined by running an unconditional 2-level analysis without predictors (Kreft & DeLeeuw, 1998; Raudenbush & Bryk, 2002). These unconditional models with no student level and classroom level predictors in Language Arts and Mathematics served as baselines to assess model fit. Multilevel modeling was found appropriate as the intra-class correlation coefficient indicated that both L1 (within class) and L2 (between classes) levels explained variance.

Next, Level-1 equations for Language Arts and Mathematics were generated for each of the 121 classrooms of the study. Student level constructs (SES, Attendance, Previous Academic Achievement, and English Proficiency) acted as covariates to control potentially confounding characteristics measures were group-centered by subtracting the mean of the above measures from student individual scores. Centering helped interpreting constructs such as Attendance, which do not have a true zero. The two models in Language Arts and Mathematics included student level constructs and no classroom level construct; these served as a second baseline upon which improvements by addition of classroom level constructs were considered.

In a third step, classroom level constructs were used, one at the time in separate models with no student level construct) to examine the expected between-classroom variability suggested in the literature. It was hypothesized that individual Academic Achievement scores would vary among classrooms as a function of the following classroom level constructs: Pupil-to-Teacher Ratio (PTR); Teacher Engagement (TEG); Teacher Experience (TEX); Teacher Education (TDG); QEIA class size reduction program (QEIA); and, the seven levels of Instructional Use of Time (Lecturing, Leading Class Discussion, Working in small Groups, Doing Seat Work, Providing Individual Instruction, Student Discipline, and Administrative Tasks). This step examined only the mean student Academic Achievement (intercept-as-outcome) and did not consider interactions between student- and class- level constructs (slopes-as-outcome). Therefore, student level constructs were included, and comparison fits were made with the unconditional models containing no predictors at either level. The Intercept-as-outcome models only focused on the impact of the classroom context on student Academic Achievement regardless of the student characteristics. Only classroom level predictors found significant were retained. Classroom level described the mean level of Academic Achievement β_{0j} of the i^{th} student in j^{th} classroom as an intercept-as-outcome function where β_{0j} (student Academic Achievement intercept) is a function of each classroom level predictor. u_{0j} is a classroom-level error term, labeled as error component in the HML7 statistical software. Beyond significance of the intercept and slopes, the error components were examined closely in an attempt to determine the variance explained.

Student Model

$$CR\_LA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_LA_{ij}) + r_{ij}$$

Classroom Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(PTR_j) + \gamma_{02}*(TEG_j) + \gamma_{03}*(TEX_j) + \gamma_{04}*(TDG_j) + \gamma_{05}*(QEIA_j) + \gamma_{06}*(IUT) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$
$$\beta_{4j} = \gamma_{40} + u_{4j}$$

Gamma intercept γ_{00} is the adjusted grand mean of the average level of classroom achievement in Mathematics, or the mean of the averaged scores of students each in class. Although all γ parameter estimates were tested for statistical significance, coefficients γ_{01} and γ_{05}, associated with the Pupil-to-Teacher Ratio and QEIA participation, are the two parameters central to the study. The full model for Mathematics is:

Student Model

$$CR\_MA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_MA_{ij}) + r_{ij}$$

Classroom Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(PTR_j) + \gamma_{02}*(TEG_j) + \gamma_{03}*(TEX_j) + \gamma_{04}*(TDG_j) + \gamma_{05}*(QEIA_j) + \gamma_{06}*(IUT) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$
$$\beta_{4j} = \gamma_{40} + u_{4j}$$

<u>Question 2 Hypothesis</u>

Question 2 tested the relationship between student Socioeconomic Status and the classrooms fixed effects. It was hypothesized that students with low Socioeconomic Status would demonstrate a greater level of Academic Achievement in smaller classrooms as defined by the QEIA program participation and the PTR. Two full models were included, one of each subject matter. In Language Arts, the model design was

Student Model

$$CR\_LA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_LA_{ij}) + r_{ij}$$

Classroom Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(PTR_j) + \gamma_{02}*(TEG_j) + \gamma_{03}*(TEX_j) + \gamma_{04}*(TDG_j) + \gamma_{05}*(QEIA_j) + \gamma_{06}*(IUT) + u_{0j}$$
$$\beta_{1j} = = \gamma_{10} + \gamma_{11}*(PTR_j) + \gamma_{12}*(TEG_j) + \gamma_{13}*(TEX_j) + \gamma_{14}*(TDG_j) + \gamma_{15}*(QEIA_j) + \gamma_{16}*(IUT) + u_{0j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$
$$\beta_{4j} = \gamma_{40} + u_{4j}$$

In order to compare difference in variance components,, the intercept outcome as function of the eight classroom effects (Pupil-to-Teacher Ratio [PTR], Teacher Engagement [TEG], Teacher Experience [TEX], Teacher Education [TDG], QEIA class size reduction program [QEIA], and Instructional Use of Time [IUT]) was maintained with the general model of Question 1 (Raudenbush & Bryk, 2002). In Mathematics, the model design was:

Student Level Model

$$CR\_LA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_MA_{ij}) + r_{ij}$$

Classroom Level Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(PTR_j) + \gamma_{02}*(TEG_j) + \gamma_{03}*(TEX_j) + \gamma_{04}*(TDG_j) + \gamma_{05}*(QEIA_j) + \gamma_{06}*(IUT) + u_{0j}$$
$$\beta_{1j} = = \gamma_{10} + \gamma_{11}*(PTR_j) + \gamma_{12}*(TEG_j) + \gamma_{13}*(TEX_j) + \gamma_{14}*(TDG_j) + \gamma_{15}*(QEIA_j) + \gamma_{16}*(IUT) + u_{0j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$
$$\beta_{4j} = \gamma_{40} + u_{4j}$$

β_{1j} is the coefficient of a slope-as-outcome function modeling the interaction of SES and eight effects of the classroom contexts.

Questions 3 Hypothesis

Question 3 tested the interaction effect between student previous level of achievement and the eight classrooms fixed effects: Pupil-to-Teacher Ratio (PTR); Teacher Engagement (TEG); Teacher Experience (TEX); Teacher Education (TDG); QEIA class size reduction program (QEIA); and, Instructional Use of Time (IUT). Special attention was given to random components as it was believed that this model would improve on the general model by a decrease the error term at classroom level. It was hypothesized that students with lower previous level of achievement status would demonstrate a greater level of Academic Achievement in smaller classrooms as defined by the QEIA program participation and the Pupil-to-Teacher Ratio. Two full models were included, one of each subject matter. In Language Arts the model design was:

Student Level Model

$$CR\_LA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_LA_{ij}) + r_{ij}$$

Classroom Level Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(PTR_j) + \gamma_{02}*(TEG_j) + \gamma_{03}*(TEX_j) + \gamma_{04}*(TDG_j) + \gamma_{05}*(QEIA_j) + \gamma_{06}*(IUT) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}*(PTR_j) + \gamma_{42}*(TEG_j) + \gamma_{43}*(TEX_j) + \gamma_{44}*(TDG_j) + \gamma_{45}*(QEIA_j) + \gamma_{46}*(IUT) + u_{4j}$$

In Mathematics, the model was:

Student Level Model

$$CR\_MA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_MA_{ij}) + r_{ij}$$

Classroom Level Model

$$\begin{aligned}
&\beta_{0j} = \gamma_{00} + \gamma_{01}*(PTR_j) + \gamma_{02}*(TEG_j) + \gamma_{03}*(TEX_j) + \\
&\gamma_{04}*(TDG_j) + \gamma_{05}*(QEIA_j) + \gamma_{06}*(IUT) + u_{0j} \\
&\beta_{1j} = \gamma_{10} + u_{1j} \\
&\beta_{2j} = \gamma_{20} + u_{2j} \\
&\beta_{3j} = \gamma_{30} + u_{3j} \\
&\beta_{4j} = \gamma_{40} + \gamma_{41}*(PTR_j) + \gamma_{42}*(TEG_j) + \gamma_{43}*(TEX_j) + \\
&\gamma_{44}*(TDG_j) + \gamma_{45}*(QEIA_j) + \gamma_{46}*(IUT) + u_{4j}
\end{aligned}$$

Question 4 Hypothesis

It was hypothesized that the Academic Achievement of students identified as English learners would be greater for those enrolled in smaller classes as defined per PTR and QEIA program participation as compared to those enrolled in nonreduced classes. It was also inferred that differences in achievement for both groups of students would be moderated by the Instructional Use of Time. The model design in Language Arts was:

Student Level Model

$$CR\_LA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_LA_{ij}) + r_{ij}$$

Classroom Level Model

$$\begin{aligned}
&B_{0j} = \gamma_{00} + \gamma_{01}*(PTR_j) + \gamma_{02}*(TEG_j) + \gamma_{03}*(TEX_j) + \gamma_{04}*(TDG_j) + \gamma_{05}*(QEIA_j) + \\
&\gamma_{06}*(IUT) + u_{0j} \\
&\beta_{1j} = \gamma_{10} + u_{1j} \\
&\beta_{2j} = \gamma_{20} + \gamma_{21}*(PTR_j) + \gamma_{22}*(TEG_j) + \gamma_{23}*(TEX_j) + \gamma_{24}*(TDG_j) + \gamma_{25}*(QEIA_j) + \\
&\gamma_{26}*(IUT) + u_{2j}
\end{aligned}$$

$\beta_{3j} = \gamma_{30} + u_{3j}$
$\beta_{4j} = \gamma_{40} + u_{4j}$

In Mathematics, the model design was:

Student Level Model

$$CR\_MA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_MA_{ij}) + r_{ij}$$

Classroom Level Model

$B_{0j} = \gamma_{00} + \gamma_{01}*(PTR_j) + \gamma_{02}*(TEG_j) + \gamma_{03}*(TEX_j) + \gamma_{04}*(TDG_j) + \gamma_{05}*(QEIA_j) + \gamma_{06}*(IUT) + u_{0j}$
$\beta_{1j} = \gamma_{10} + u_j$
$\beta_{2j} = \gamma_{20} + \gamma_{21}*(PTR_j) + \gamma_{22}*(TEG_j) + \gamma_{23}*(TEX_j) + \gamma_{24}*(TDG_j) + \gamma_{25}*(QEIA_j) + \gamma_{26}*(IUT) + u_{2j}$
$\beta_{3j} = \gamma_{30} + u_{3j}$
$\beta_{4j} = \gamma_{40} + u_{4j}$

CHAPTER FOUR

RESULTS AND ANALYSES

The data collection proceeded as presented in the preceding chapter. However, the population retained for final analysis varied from the proposal for two reasons: (a) not all teachers participated, causing missing cases in classes (level 2) and students (level 1); (b) some classes in the schools not benefitting from the class size reduction QEIA grant (group labeled herein “nonreduced”) had PTR ratios equal or lower to those found in classes of QEIA schools, and were omitted from the analysis.

A total of 51 teachers of Languages Arts and Mathematics teaching 1,685 students organized in 121 classes formed the initial participant population. 44 teachers (86.3%) took part in the study and completed the *Teacher Survey* (see Appendix C). Out of those the teachers who did not participate, three (5.9%) could not be contacted, two (3.9%) refused to participate, one (2.0%) retired from the district, and one (2.0%) resigned his position prior to the data collection. Consequently, only cases with information at both the student and classroom levels were retained for final analysis. Thirteen classes and their students instructed by the missing teachers were deleted listwise prior to analysis. In addition, six classes were omitted from analysis for not meeting inclusion criteria: three “reduced” classes for exceeding the maximum allowable Pupil-to-Teacher Ratio (PTR ≤ 25:1); and, three “nonreduced” classes for falling below the 26:1 PTR.

Prior to data screening for parametric assumptions, the participant population had decreased from the initial figures to: 44 teachers (86.3%), 102 classes (84.3%), and 1,645 students (97.6%) (see Table 4.1). Of these 1,645 students, 1,481 (90.0%) attended classes in Language Arts and 1,298 (78.9%) in Mathematics (Tables 4.1, 4.2 and 4.3).

Table 4.1

Data Collection Results: Number of Participants

	Language arts		Mathematics	
	Initial	Retained	Initial	Retained
Teacher	26	24	26	21
Class	59	54	62	48
Student	1,603	1,481	1,651	1,298

Note. One teacher surveyed taught both subject matters

Achievement means between students of participating and nonparticipating teachers were evaluated to determine if values were missing at random. In Language Arts, differences in achievement means for students in participating teachers' classrooms (N = 1481, M = 338.00, SD = 46.50) and nonparticipating teachers' classrooms (N = 95, M = 351.46, SD = 42.16) were found ($t\ (1584)$ = -2.75, p = .01). However, the two groups were considerably different in size. No difference between students in participating teachers' classrooms (N = 1401, M = 339.82, SD = 59.06) and nonparticipating teachers'

Table 4.2

Language Arts: Participants by Class Size Type

	Reduced		Nonreduced	
	Initial	Retained	Original	Retained
Teacher	12	11	14	13
Class	31	26	28	26
Student	683	617	920	864

Table 4.3

Mathematics: Participants by Class Size Type

	Reduced		Nonreduced	
	Initial	Retained	Initial	Retained
Teacher	11	10	15	11
Class	30	26	32	22
Student	687	595	964	703

classrooms (N = 230, M = 332.38, SD = 55.76) in Mathematics were found (t (1629) = 1.78, p = .08). No meaningful differences were found and it was concluded data were Missing at Random (MAR).

Data Screening

Missing Values

Missing values in the previous and current academic measures of achievement in Language Arts accounted for 14.6% (*n* = 216) and 1.1% (*n* = 16), respectively (see Table 4.4). Missing values in the previous and current academic measures of achievement in Mathematics accounted for 15.1% (*n* = 196) and 1.3% (*n* = 17), respectively. Mertler and Vannatta (2005) suggest replacing missing values for no more than 15% of total number of cases within a dataset.

Table 4.4

Missing Values by Student Level (Level 1)

Student level indicators	Missing values	%
Current language arts achievement	16	1.1
Previous language arts achievement	216	14.6
Current mathematics achievement	17	1.3
Previous mathematics achievement	196	15.1
Attendance (%)	1	0.0

A regression analysis used Current Academic Achievement in Language Arts to predict Previous Academic Achievement scores in the same subject matter. The missing values replacement method was chosen as it preserves the variance that would otherwise be lost with mean replacement while remaining objective. Likewise, the same procedure was applied to missing values in for Previous Academic Achievement in Mathematics. Another reason for selecting this method of replacing missing values was that current and previous measures of achievement were highly correlated in Language Arts (r = .78, p <.001) and Mathematics (r = .71, p <.001). 16 cases (1.1%) containing missing data in Current Academic Achievement in Language Arts (CR_LA), 17 (1.3%)cases with missing data in Current Academic Achievement in Mathematics (CR_MA), and one case missing value in the Attendance were omitted from all further analysis. No other missing data was noted.

Outliers

For previous and current Language Arts Academic Achievement, seven univariate outliers (z-scores ≥ 3.0) were identified and omitted from further analysis. Likewise, 16 univariate outliers (z-scores ≥ 3.0) were found in previous and current Mathematics Achievement; associated cases were also omitted from further analysis. Student Attendance was negatively skewed (see Tables 4.5 and 4.6). Nine cases were omitted listwise from further analysis due to skewed attendance.

The criteria set for multivariate outlier was a χ^2 value set at 5.99 with two degrees of freedom (p = .05). 12 and 18 student cases exceeded this critical value and were omitted from further analysis for Language Arts and Mathematics Achievement measures, respectively.

Teacher Experience (TEX), used at classroom level (L2) had one outlier at 37 years of experience. This outlier was assigned the nearest continuous value of 31 to reduce a positive skew.

Parametric Assumptions

Parametric assumptions and linearity and homoscedasticity were examined using QQ plots and scatter plots. After removing outliers, replacing missing values, and omitting cases, all parametric assumptions were met.

Descriptive Statistics: Level-1 (Students)

Tables 4.5 and 4.6 present the descriptives before and after data screening. The final participant population after data screening was: 54 classes and 1,441 students in Language Arts; and, 48 classes and 1,242 students in Mathematics.

In Language Arts, 396 students (27.5%) were identified as English Learners (EL). The 1,045 (72.5.0%) remaining students were identified as bi-literate, exited from second language program, or native English speakers. SES status was derived from two sources: participation in the National School Lunch Program (NSLP); and, self-reported parent level of education. 959 students (66.6%) were identified as low socioeconomic students.

In Mathematics, 327 students (26.3%) were identified as English Learners (EL). The 915 (73.7%) remaining students were identified as bi-literate, exited

Table 4.5

Student Level Constructs Before Data Screening

Construct	*N*	Min	Max	Mean	*SD*	Skewness
Current language arts achievement	1465	197	541	338.40	46.40	-.03
Previous language arts achievement	1269	234	485	341.10	40.50	.13
Attendance(%) in language arts classes	1464	0	1.00	.94	.07	-7.73
Current mathematics achievement	1280	8	600	340.01	59.50	.54
Previous mathematics achievement	1102	207	561	337.58	54.95	.59
Attendance(%) in mathematics classes	1279	0	1.00	.95	.06	-7.69

from second language program, or native English speakers. 730 students (58.8%) were identified as low-socio economic students.

Correlation of Level 1 Student Variables

Multilevel modeling is sensitive to multicollinearity. The presence of linearly dependent predictors may produce unstable models. Table 4.7 and 4.8 indicate moderate linear dependency between current and previous levels of Academic Achievement in Language Arts and Mathematics.

Table 4.6

Student Level Constructs After Data Screening

Construct	*N*	Min	Max	Mean	*SD*	Skewness
Current language arts achievement	1441	214	464	338.16	45.10	-.08
Previous language arts achievement	1441	240	436	338.95	38.87	-.02
Attendance(%)in language arts classes	1441	.71	1.00	.95	.05	-2.33
Current mathematics achievement	1242	196	516	337.11	54.68	.46
Previous mathematics achievement	1242	206	468	333.96	49.62	.21
Attendance(%)in mathematics classes	1242	.71	1.00	.96	.04	-2.35

Table 4.7

Language Arts: Correlations Between Student Constructs

	Attendance (%)	Current level of achievement
Current level of achievement	0.13**	
Previous level of achievement	0.09**	0.78**

**p* ≤ 0.05; ***p* ≤ 0.01

Table 4.8

Mathematics: Correlations Between Student Level Predictors

	Attendance (%)	Current level of achievement
Current level of achievement	0.12**	
Previous level of achievement	0.06*	0.71**

*$p \leq 0.05$; **$p \leq 0.01$

Descriptive Statistics: Level-2 (Classrooms)

Level-2 unit of analysis was the classroom contexts retained for final analysis in which student cases were nested at Level 1: 54 classes in Language Arts; and, 48 classes in Mathematics (see Table 4.9).

Table 4.9

Language Arts and Mathematics Classes by Student Enrollment

	N (classes)	Min	Max	Mean	*SD*	Skewness
Language arts nonreduced class	26	26	37	33.23	3.17	-.95
Language arts reduced class	28	17	25	22.43	2.59	-1.01
Mathematics nonreduced class	22	26	37	31.95	3.00	-.38
Mathematics reduced class	26	18	25	23.00	1.96	-.86

Figures 4.1 and 4.2 indicates the descriptive statistics for these classes according to QEIA participation.

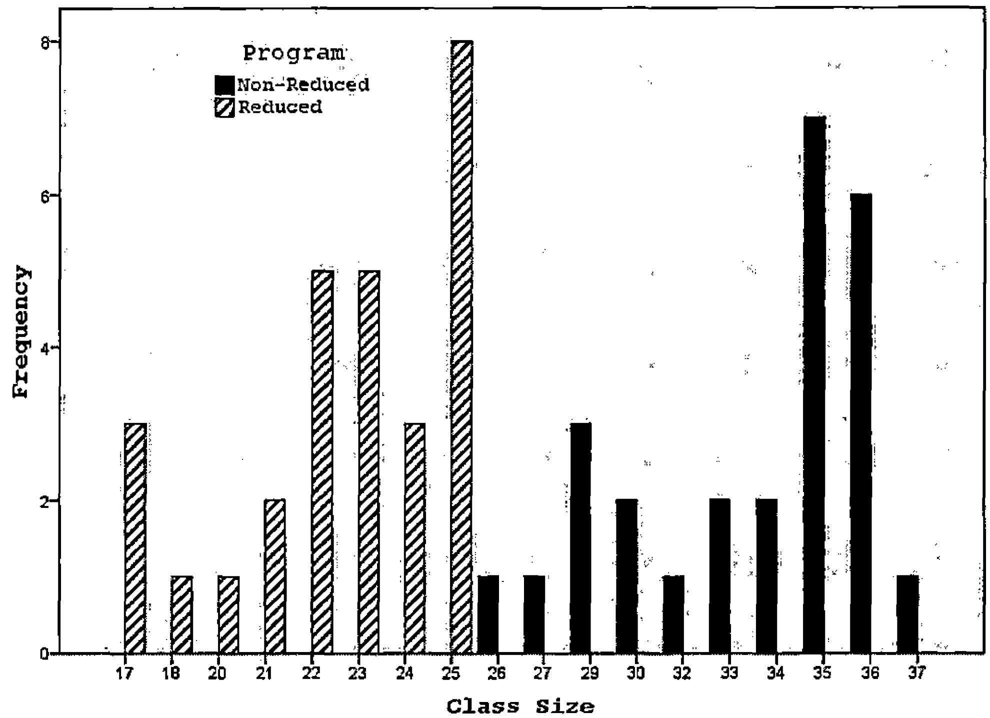

Figure 4.1. Language arts classroom pupil-teacher ratios by Quality Education Improvement Act Class Size Reduction Program.

Teacher Experience expressed in years was self-reported and denoted the number of years teachers had been practicing. The professional experience of participant teachers ranged from 3 to 31 years (M =13.00, SD = 8.17) in 54 Language Arts classes, and from 3 to 21 years (M =8.67, SD = 4.73) in 48 Mathematics classes.

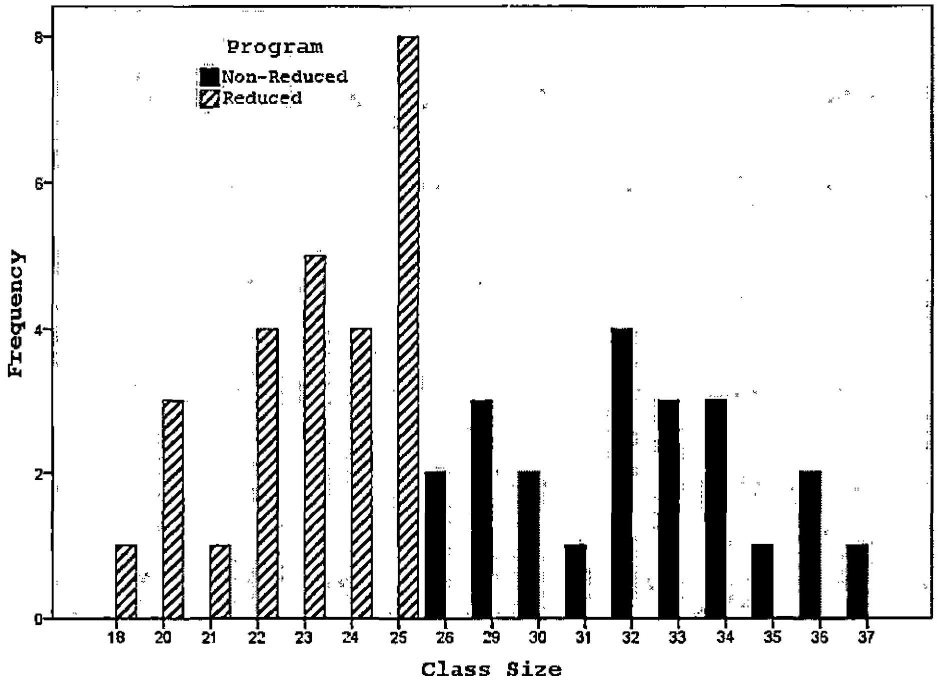

Figure 4.2. Mathematics classroom pupil-teacher ratios by Quality Education Improvement Act Class Size Reduction Program.

Teacher Engagement (see Appendix C) described the degree of involvement of teachers in the instructional process, with higher values indicating greater levels of involvement. Results are reported in Table 4.10. Descriptives for Instructional Use of Time are displayed in Table 4.11 (Language Arts) and 4.12 (Mathematics).

Fourteen teachers (31.8%) reported holding a Bachelor's degree, while the remaining 30 teachers (68.2%) reported they held a Master's degree. No teachers indicated they had a doctorate. Language Arts classes were instructed by 17 (31.5%) teachers holding a Bachelor's degree and 37 (68.5%) teachers

Table 4.10

Teacher Engagement in Classroom

	N	Min	Max	Mean	*SD*	Skewness
Teacher engagement in language arts classes	54	37	48	44.41	2.85	-.96
Teacher engagement in mathematics classes	48	34	49	41.77	4.18	-.20

Table 4.11

Instructional Use of Time in Language Arts

Instructional Activity	*N*	Min	Max	Mean	*SD*	Skewness
Lecturing	54	2	5	3.50	1.02	.06
Leading class discussion	54	1	5	3.11	.84	.58
Working in small groups	54	1	5	2.83	1.06	.74
Doing seat work)	54	1	5	2.94	1.07	.50
Providing individual instruction	54	1	5	2.67	.91	.73
Student discipline	54	1	5	2.07	1.10	.92
Administrative tasks	54	1	4	2.11	.82	.22

Note. Likert scale: amount of time spent weekly on the activity. 1: 0 minutes; 2: 30 minutes; 3: one hour; 4: two hours; 5: more than two hours.

Table 4.12

Instructional Use of Time in Mathematics

Instructional Activity	*N*	Min	Max	Mean	*SD*	Skewness
Lecturing	48	2	5	3.65	.96	-.44
Leading class discussion	48	1	5	2.96	.99	.09
Working in small groups	48	1	5	3.10	1.29	-.20
Doing seat work)	48	1	5	3.06	1.12	-.60
Providing individual instruction	48	1	5	2.77	1.02	-.02
Student discipline	48	1	5	1.90	1.04	1.17
Administrative tasks	48	1	5	2.48	.95	.30

Note. Likert scale: amount of time spent weekly on the activity 1: 0 minutes; 2: 30 minutes; 3: one hour; 4: two hours; 5: more than two hours

holding a Master's degree. In Mathematics classes, 19 (39.6%) teachers held a Bachelor's degree and 29 (60.4%) a Master's degree.

Analysis

Hierarchical Linear Modeling (HLM) was used as students (level-1) were nested within classrooms (level-2). The use of hierarchical modeling helped prevent both the ecological fallacy where inferences on individual Academic Achievement are based on aggregated data assuming homogeneous groupings, and the atomistic fallacy that suggests inferences on group characteristics based on individual student achievement results.

To confirm this choice, two unconditional models, one in each subject matter, were used to determine the amount of variance in the student Achievement (outcome) between classroom and within classroom. For each of the four hypotheses, model testing proceeded in four steps: intercept-only model; means-as-outcome model; random-regression coefficients model; and, intercepts-and slopes-as-outcomes (Luke, 2004). Predictors at student and classroom levels were added or subtracted to improve model fit.

Accounting for Variance and Model Fit

In a typical regression analysis, the amount of variance explained (R^2) is used to estimate how well the model fit the data. In multilevel analysis, assessment of model fit is not directly observable. Instead, unconditional or unconstrained models were used as baseline against which all suggested improved models were compared. The same way traditional regression models are based on R^2 (variance explained). Model fit in HLM is assessed by the proportional reduction of prediction error from a comparison model over the unconditional null model. Kreft and DeLeeuw (1998) suggested the following formula to calculate R^2 in HLM: (unconditional error – restricted error)/unconditional error), where unconditional error is the variance component of the One-way null ANOVA (i.e., the model without predictors) and restricted error the variance component of the suggested final model.

Alternatively, Snijders and Bosker (1999) suggested an alternative method to compute R^2 at both levels:

Student Level

$$R^2 = 1 - [(\sigma^2_r + \tau^2)_{\text{comparison model}} / (\sigma^2_r + \tau^2)_{\text{baseline model}}]$$

Classroom Level

$$R^2 = 1 - [(\sigma^2_r / n + \tau^2)_{\text{comparison model}} / (\sigma^2_r / n + \tau^2)_{\text{baseline model}}]$$

where n is the number of student level units in any classroom level. τ^2 is the variance of classroom level error, also noted as σ^2_u in the literature (Luke, 2004; O'Connell & McCoach, 2008; Raudenbush & Bryk, 2002). The predictive ability of the model will be expressed as a range between student and classroom R^2 estimates.

The Kreft and DeLeeuw (1998) R^2 formula may experience difficulties in the event of a residual being large in the restricted model versus the unconditional model without predictors (R^2 values may become negative). This may occur especially with random coefficients of models specifying cross-level interactions between student level and classroom level predictors.

The Kreft and DeLeeuw (1998) L1 and L2 R^2 calculation was chosen for use in the present study as it is most commonly reported in multilevel analysis

Question 1 (Model1)

To determine the source of variability, two general unconditional null models (one-way random-effect ANOVA), one in Language Arts and another in

Mathematics, evaluated between-group effects with the Intra-class Correlation Coefficient (ICC).

In Language Arts (LA), the intercept-only model written: Current LA Academic Achievement$_{ij}$ = $\beta_{0j} + r_{ij}$ (student level) and $\beta_{0j} = \gamma_{00} + u_{0j}$ (classroom level). The mixed equation is Current Academic Achievement$_{ij}$ = $\gamma_{00} + u_{0j} + r_{ij}$, where $r_{ij} \sim (0, \sigma^2)$ is the level-1 residual and $u_{0j} \sim (0, \tau)$ is the deviated mean achievement of a particular classroom from the grand mean of all classrooms. Similarly, the intercept-only model for Mathematics is: Current Mathematics Academic Achievement$_{ij}$ = $\beta_{0j} + r_{ij}$ (student level) and L2: $\beta_{0j} = \gamma_{00} + u_{0j}$ (classroom level). The mixed equation is: Current Mathematics Academic Achievement$_{ij}$ = $\gamma_{00} + u_{0j} + r_{ij}$, where classroom level equation for β_{0j} is placed into student level equation.

The null hypothesis (H_0) is $\gamma_{00} = \beta_{0j}$. Thus, this suggested no variance existed at the classroom level, with no classroom context effects on individual student (L1) Academic Achievement.

When running the unconditional model or one-way random-effects ANOVA model, all additional predictors were removed in order to reveal how level-2 Classroom-context factors impacted level-1 Academic Achievement scores of individual students.

Intra-class Correlation coefficients (ICC) were interpreted as the variance explained by classroom level 2 between-class components in the model. ICC = $\sigma^2_{u0} / (\sigma^2_r + \sigma^2_{u0})$ or the level-2 variance component divided by the sum of the student level-1 and classroom level-2 variance components.

In Language Arts, variance components (see Table 4.13) were used to calculate R^2 both within and between classrooms. For Language Arts, ICC = 312.15/ (1752.66 + 312.15) = .144. Thus, in Language Arts 14.4 percent of the variance in Academic Achievement was between-classes and 85.6 percent of the variance in Academic Achievement was found at the student-level. The weighted least square estimate centered on the grand mean was 337.63 (*SD* = 2.63). The 95% confidence interval for the Academic Achievement estimate in Language Arts was 337.63 ± 1.96(2.63) = (332.48, 342.78).

Table 4.13

Language Arts: One-Way Random Effects ANOVA

Fixed effects		Coefficient	*SE*	
Average class mean, γ_0		337.63	2.63	
Random effects	Variance components	*df*	χ^2	*p*-value
Class mean, u_{00}	312.15	53	286.28	<.001
Level-1 effects, *r*	1752.66			

In Mathematics, the ICC = 773.76/ (2291.20 + 773.76) = .252, using the variance components reported in Table 4.14. For Mathematics, class level explained 25.2 percent of the variance in Academic Achievement while 74.8 percent of the variance was explained at the student level. The weighted least square estimate centered on the grand mean was 338.45 (*SD* = 4.20). The 95%

confidence interval for the Academic Achievement estimate in Mathematics was $338.45 \pm 1.96(4.20) = (334.25, 342.65)$.

Table 4.14

Mathematics: One-Way Random Effects ANOVA

Fixed effects		Coefficient	*SE*	
Average class mean, γ_0		338.45	4.20	
Random effects	Variance components	*df*	χ^2	*p*-value
Class mean, u_{00}	773.76	47	426.66	<.001
Level-1 effects, r	2291.20			

These results supported the use of multilevel models such hierarchical level modeling as appropriate. The relatively strong Intra-class Correlation Coefficients (ICC) revealed the nested nature of the observations. Individual student Academic Achievement results were not independent observations but interdependent within each classroom.

<u>Means as Outcomes Models</u>

The second step in building HLM models involved creating means-as-outcomes models, where classroom level-2 predictors were added one at a time and analyzed. In Language Arts, class size (nonreduced/reduced) was added (see Table 4.15) to examine the possible impact of the class size reduction on Academic Achievement at the student level. The means-as-outcome student level model: Current LA Academic Achievement$_{ij}$ = $\beta_{0j} + r_{ij}$ and L2:$\beta_{0j} = \gamma_{00}$ +

γ_{01}*(QEIA) + u_{0j}. The regression coefficient related to participation in class reduction program was not significant (γ_{01} = 4.60, *t* (52) =.88, *p* = .38).

Table 4.15

Classroom Level Means-as-Outcome: Quality Education Improvement Act Class Size Program for Language Arts

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	337.72	2.64	127.79	52	<.001
QEIA slope, γ_{01}	4.60	5.22	.88	52	.38
Random effect	***SD***	**Variance component**	***df***	**χ^2**	***p*-value**
Class mean, u_0	17.69	313.06	52	280.71	<.001
Student level effect, r	41.87	1752.82			

Note. QEIA: QEIA class size program.

In Mathematics, a similar model was tested at student level: Current Mathematics Academic Achievement$_{ij}$ = β_{0j} + r_{ij} and at classroom level:β_{0j} = γ_{00} + γ_{01}*(QEIA) + u_{0j} (see Table 4.16). Model statistics for β_{0j} grand mean were γ_{01} = 14.91, *t* (46) = 1.89 (*p* = .07).

No difference in Academic Achievement was found for class size in Language Arts (reduced class [*M* = 22.4, *SD* = 2.6] or nonreduced class [*M* = 33.2, *SD* = 3.2]). Similarly, no differences were found in Mathematics between QEIA reduced classes (*M* = 23.0, *SD* = 2.0) and nonreduced classes (*M* = 32.0, *SD* = 3.0).

Table 4.16

Classroom Level Means-as-Outcome: Quality Education Improvement Act Class Size Program for Mathematics

Fixed effects	Coefficient	*SE*	t-ratio	*df*	*p*-value
Class mean, γ_{00}	338.59	4.10	82.57	46	<.001
QEIA slope, γ_{01}	14.91	7.88	1.89	46	.07
Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	27.09	733.81	46	395.68	<.001
Student level effect, r	47.87	2291.31			

Note. QEIA: QEIA class size program.

Table 4.17

Classroom Level Means-as-Outcome: Pupil-to-Teacher Ratio for Language Arts

Fixed effects	Coefficient	*SE*	t-ratio	*df*	*p*-value
Class mean, γ_{00}	337.59	2.65	127.29	52	<.001
PTR slope, γ_{01}	.14	.36	.38	52	.71
Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	17.87	319.46	52	286.84	<.001
Student level effect, r	41.86	1752.54			

Note. PTR: Pupil-to-teacher ratio.

Table 4.18

Classroom Level Means-as-Outcome: Pupil-to-Teacher Ratio for Mathematics

Fixed effects	Coefficient	*SE*	t-ratio	*df*	*p*-value
Class mean, γ_{00}	338.58	4.15	81.63	46	<.001
PTR slope, γ_{01}	-1.20	.76	-1.57	46	.12
Random effect	***SD***	**Variance component**	***df***	**χ^2**	***p*-value**
Class mean, u_0	27.44	752.91	46	404.05	<.001
Student level effect, r	47.87	2291.32			

Note. PTR: Pupil-to-teacher Ratio.

To confirm these findings, Pupil-to-Teacher Ratio (PTR) was substituted for QEIA class size in both models (see Tables 4.17 and 4.18). Results of the model analyses revealed classroom size as measured by PTR was not a level-2 predictor in Language Arts (γ_{01} = .14, $t(52)$=.38, p = .71) nor Mathematics (γ_{01} = -1.20, $t(46)$= -1.57, p = .12).

Neither participation in QEIA class size program nor PTR reduced the variance at classroom context level. Thus class size reduction in either method failed to explained student Academic Achievement. Other classroom context constructs were considered in later analyses when accounting for full intercepts-and-slopes-as-outcomes models.

Question 2 (Model 2)

The second hypothesis examined the impact of SES as a covariate at the student level, and the possible cross-level interaction with school context predictors. Its purpose was to test the hypothesis that smaller class size helps reduce the achievement gap between students identified as socioeconomically disadvantaged and not socioeconomically disadvantaged.

Prior to investigating this hypothesis, a full student level random-coefficient model was built to determine the statistical significance and magnitude of all student level constructs. This constitutes the third step of building multilevel models (Raudenbush & Bryk, 2002).

Student level constructs were group-centered in order to ease interpretation. These were socioeconomic Status (SES), Limited English Proficiency (LEP), Attendance (ATT), and previous level of achievement in Language Arts (PR_LA) or Mathematics (CR_MA). Two separate equations were generated, one each in subject matter. The residual r_{ij} is interpreted as the variance remaining unexplained after accounting for the predictors in the models:

$$CR\_LA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_LA_{ij}) + r_{ij}$$

$$CR\_MA_{ij} = \beta_{0j} + \beta_{1j}*(SES_{ij}) + \beta_{2j}*(LEP_{ij}) + \beta_{3j}*(ATT_{ij}) + \beta_{4j}*(PR\_MA_{ij}) + r_{ij}$$

Group-centered variables coefficients denote:

β_{0j}: mean achievement in class j

β_{1j}: mean difference in achievement between students classified as high- and low socioeconomic status in class j

β_{2j}: mean difference in achievement between students classified as limited English proficient and students not identified as limited English proficient in class j

β_{3j}: degree to which attendance contributes to differences in achievement between students in class j

β_{4j}: degree to which previous level of achievement contributes to differences in achievement in the same subject for students enrolled in class j

Each coefficient β is composed of a fixed effect γ and a random error, where γ represent the mean value for each class predictor.

The parameter estimates of the random-coefficient model were tested (see Table 4.19). The classroom intercept for Language Arts was 337.49 (SE = 2.68). Students of different socioeconomic levels (SES) did not differ on Language Arts scores (t = .64, p = .52). The average difference between proficient and nonproficient students in English was significant (t = -5.26, p <.001). Attendance (t = 3.85, p <.001), and previous Academic Achievement (t = 33.54, p <.001) were related to current Achievement (see Table 4.19).

As compared to the unconditional null model (Table 4.12), the revised student level random-coefficient regression model (Table 4.19) found a proportion of variance explained of (1752.66 – 743.74)/ 1752.66) ≈ 0.576 or about 57.6%. The reliability of the estimate of classroom scores increased from 0.81 to 0.93 as compared to the unconditional model.

In Mathematics, the parameter estimates of the model random coefficients were also tested (see Table 4.20). The average classroom mean in Mathematics was 338.52 (SE = 4.27). Students of different socioeconomic levels (SES) did

Table 4.19

Student Level Random-Coefficient Regression Model for Language Arts

Fixed effects	Coef.	*SE*	*t*-ratio	Approx. df	*p*-value
Class mean intercept, γ_{00}	337.49	2.68	125.70	53	<.001
Socioeconomic status slope, γ_{01}	.96	1.50	.64	1383	.52
English proficiency slope, γ_{02}	-10.42	1.98	-5.26	1383	<.001
Attendance slope, γ_{03}	65.75	17.09	3.85	1383	<.001
Previous achievement, γ_{04}	.80	0.02	33.54	1383	<.001
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	19.16	367.05	53	674.92	<.001
Student level effect, *r*	27.27	743.74			

not differ on Math scores (t = -.23, p =.82). Proficient and nonproficient students in English did not differ on Mathematics scores (t = -1.40, p =.16). Attendance (t = 4.37, p <.001) and previous level of Academic Achievement (t = 30.96, p <.001) were related to current Achievement.

A revised student level random-coefficient model was calculated with Attendance and previous level of Academic Achievement in Mathematics. The proportion of variance explained by the revised random-coefficient regression model as compared to the null unconditional model (Table 4.14) was (2291.20-1154.88)/2291.20 ≈ 0.496 or about 49.6%. The reliability of the estimate of

Table 4.20

Student Level Random-Coefficient Regression Model for Mathematics

Fixed effects	Coef.	*SE*	*t*-ratio	Approx. *df*	*p*-value
Class mean intercept, γ_{00}	338.52	4.22	80.15	47	<.001
Socioeconomic status, γ_{01}	-.42	1.78	-.23	1190	.82
English proficiency, γ_{02}	-4.71	3,37	-1.40	1190	.16
Attendance, γ_{03}	120.78	27.64	4.37	1190	<.001
Previous achievement, γ_{04}	.69	.02	30.96	1190	<.001
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	28.77	827.86	47	846.67	<.001
Student level effect, *r*	33.98	1154.88			

classroom scores in Mathematics increased from 0.89 to 0.95 as compared to the unconditional model.

In summary, English proficiency, Attendance and previous level of Academic Achievement explained 57.6 percent of the variance of student Academic Achievement in Language Arts at the student level. In Mathematics, Attendance and previous level of Academic Achievement explained 49.6 percent of the variance at the student level.

SES was not found significant and was removed from the two models. Thus, hypothesis 2 that student Academic Achievement for students of different socioeconomic levels varies with class size, a classroom context, was rejected.

Hypotheses 3 and 4

It was hypothesized that students with lower Academic Achievement would benefit most from reduced class sized. Similarly, Hypothesis 4 examined the effect of class size on the Academic Achievement of English proficient and nonproficient students.

Hypothesis 1 had found neither Pupil-to-Teacher Ratio (PTR) nor QEIA Class Size Program (QEIA) to be significant. Hypotheses three and four were tested as stated and a main effect was not found for either PTR or QEIA (see results for hypothesis 1). Although no main effects were found for hypothesis 3 or 4, Baron and Kenny (1986) recommend testing for moderating effects (see Ancillary Analysis).

Ancillary Analysis

In hypothesis 1, the unconditional one-way ANOVA with no student or classroom constructs revealed 14.4 percent of the variance in student Academic Achievement was found between-classrooms, and 85.6 percent among students within classrooms. The second step of model building was limited to examining the QEIA and PTR predictors only. As no variance explained at the classroom-level was removed by PTR or QEIA, the author then examined the interactions of PTR and QEIA with the student level constructs Socioeconomic Status (SES), Language Proficiency Status (LEP), Attendance (ATT), and previous level of Achievement in both Language Arts and Mathematics. All potential moderator terms were small in magnitude and statistically insignificant (see Appendix F).

The author also considered the remaining constructs for main effects and interactions: Teacher Engagement (TEG); Teacher Experience (TEX); Teacher Education (TDG); Instructional Use of Time (Lecturing, Leading Class Discussion, Working in Small Groups, Doing Seat Work, Providing individual Instruction, Student Discipline, and Administrative Tasks).

Since variances existed at both the context- and student-levels both in Language Arts (LA) and Mathematics (MA), classroom level constructs were added to the model one at the time at level 2, with the aim of reducing (thereby explaining) the variance through the building of improved models.

After evaluating all level-2 predictors (see Tables 4.21 and 4.22), Teacher Engagement (TEG) and Teacher Experience (TEX) were found significant. However, none of the variables related to classroom Instructional Use of Time were found significant (see summary results in Appendix E).

Table 4.21

Classroom Level Means-as-Outcome: Teacher Engagement for Language Arts

Fixed Effects	Coefficient	*SE*	*t*-ratio	df	*p*-value
Class mean, γ_{00}	337.63	2.57	131.47	52	<.001
Teacher engagement slope, γ_{01}	-2.03	0.91	-2.24	52	.03
Random effect	***SD***	**Variance Component**	***df***	χ^2	***p*-value**
Class mean, u_0	16.94	287.12	52.00	268.75	<.001
Student level effect, *r*	41.86	1752.27			

Table 4.22

Classroom Level Means-as-Outcome: Teacher Experience for Language Arts

Fixed Effects	Coefficient	*SE*	*t*-ratio	df	*p*-value
Class Mean, γ_{00}	337.64	2.46	137.39	52	<.001
Teacher experience slope, γ_{01}	-.89	.29	-3.06	52	<.001
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	16.43	269.78	52	258.34	<.001
Student level effect, *r*	41.86	1751.99			

For Language Arts, the regression coefficients were Teacher Engagement (TEG) (β = -2.14, p = .01) and Teacher Experience (TEX) (β = -.87, p <.001). These were included in the classroom constructs in the final model. As compared to the unconditional one-way random effect ANOVA (Table 4.12) set as a reference base, the inclusion of these estimated parameters in the means-as-outcome model (Table 4.23)reduced unexplained variance by 22.5 percent, (312.15 – 242.06)/312.15.

The same analysis was repeated for Mathematics (see Table 4.24 and 4.25). Teacher Engagement (TEG) (β = 1.77, p = .01) and Instructional Use of Time: Administrative Tasks (β = -8.64, p <.001) were included in the class context constructs in the final model in Mathematics. None of the other classroom context predictors associated to the other Instructional Use of Time, as well as Teacher Experience and Teacher Education were found significant (see Appendix F).

Table 4.23

Classroom Level Means-as-Outcome: Final Model for Language Arts

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	337.64	2.33	144.69	51	<.001
Teacher engagement slope, γ_{01}	-2.14	.77	-2.77	51	.01
Teacher experience slope, γ_{02}	-.87	.29	-3.02	51	<.001
Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	15.56	242.06	51	236.09	<.001
Student level effect, *r*	41.85	1751.66			

Table 4.24

Classroom Level Means-as-Outcome: Teacher Engagement – Teacher Engagement for Mathematics

Fixed effects	Coefficient	*SE*	t-ratio	*df*	*p*-value
Class mean, γ_{00}	338.55	3.95	85.74	46	<.001
Teacher engagement slope, γ_{01}	2.46	.68	3.61	46	<.001
Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	26.14	683.26	46	376.20	<.001
Student level effect, *r*	47.87	2291.29			

Table 4.25

Classroom Level Means-as-Outcome: Instructional Use of Time: Administrative Tasks for Mathematics

Fixed Effects	Coefficient	*SE*	t-ratio	*df*	*p*-value
Class mean, γ_{00}	338.43	3.92	86.36	46	<.001
Administrative tasks slope, γ_{01}	-11.30	2.97	-3.80	46	<.001
Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	26.03	677.39	46	382.20	<.001
Student level effect, *r*	47.86	2290.93			

Teacher Engagement and Administrative Tasks were both used as classroom constructs in the Means-as-Outcome model described in Table 4.26. As compared to the unconditional one-way random effect ANOVA (Table 4.13) set as a reference base, the inclusion of these estimated parameters in the means-as-outcome model (Table 4.26) reduced unexplained variance by 10.2 percent, or (773.76 – 694.86)/773.76.

Language Art: Intercepts-and-Slopes-as-Outcomes Model

This model included the student level constructs identified in the random-coefficient regression model (see Table 4.19), and the classroom level constructs from the means-as-outcome model (see Table 4.23). This intercepts-and-slopes-as-outcomes model then was revised to reduce variance components and improve model fit.

Table 4.26

Classroom Level Means-as-Outcome: Final Model for Mathematics

Fixed Effects	Coefficient	se	*t*-ratio	df	*p*-value
Class mean, γ_{00}	338.55	3.81	88.88	45	<.001
Teacher engagement slope, γ_{01}	1.77	0.64	2.76	45	.01
Administrative tasks slope, γ_{01}	-8.64	2.89	-2.99	45	<.001
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	26.36	694.86	45	709.02	<.001
Student level effect, *r*	34.01	1156.46			

The means-as-outcomes model retained only Teacher Engagement and Teacher Experience, while class size related predictors QEIA and PTR, Teacher Education, and Instructional Use of Time did not enter the model. The regression coefficient relating Teacher Engagement to student Academic Achievement in Language Arts was negative which suggested that student LA Academic Achievement is lower in classrooms where teachers demonstrated a higher level of engagement. The coefficient relating Teacher Experience to student Academic Achievement in Language Arts was also negative. Teacher Engagement and Teacher Experience together (see Table 4.23) resulted in producing a decrease of classroom level variance from 312.15 to 242.06 between unconditional and restricted models, therefore explaining (312.15-

242.06)/312.15 or 22.5 percent of the variance between classrooms per Kreft and DeLeeuw (1998) R^2 formula.

Using the Kreft and DeLeeuw (1998) formula, the random-coefficient regression model explained (1752.66 - 745.85)/1752.66, or 57.5 percent of the variance in student Academic Achievement in Language Arts. Using an alternative R^2 formula (Snijders & Bosker, 1999), the random-coefficient regression model decreased in unexplained variance was estimated at 1 - [(L1 restricted error + L2 restricted error)/(L1 unrestricted error + L2 unrestricted error)], or 1 – [(366.93 + 745.85)/(312.15 + 1752.66)]. The random-coefficient regression model explained 46.1 percent of the variance in student Academic Achievement in Language Arts.

The choice between random and fixed effects in the model was made with the aim of maximizing R^2 multilevel equivalent measures. However, defining variance explained in multilevel modeling is difficult, and stems from the cross-level interaction: changes at the student level impact the meaning of the intercept at the classroom context level. The principle of parsimony has been applied to defining a final model.

Student Level Model

$$CR\_LA_{ij} = \beta_{0j} + \beta_{1j}*(LEP_{ij}) + \beta_{2j}*(ATT_{ij}) + \beta_{3j}*(PR\_LA_{ij}) + r_{ij}$$

Classroom Level Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(TEG_j) + \gamma_{02}*(TEX_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(TEG_j) + \gamma_{12}*(TEX_j)$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}*(TEG_j) + \gamma_{22}*(TEX_j)$$
$$\beta_{3j} = \gamma_{30} + \gamma_{31}*(TEG_j) + \gamma_{32}*(TEX_j)$$

Although some cross-level moderating effects were significant between student and classroom constructs (see Table 4.27), these were not retained in the final model as they did not further explain the relationship between student LA Academic Achievement, Teacher Engagement, and Teacher Experience.

The full model with all cross-level moderating terms (Table 4.27) was revised in a final model (Table 4.28) that only retained constructs improving model fit.

Student Level Model

$$CR\_LA_{ij} = \beta_{0j} + \beta_{1j}*(LEP_{ij}) + \beta_{2j}*(ATT_{ij}) + \beta_{3j}*(PR\_LA_{ij}) + r_{ij}$$

Classroom Level Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(TEG_j) + \gamma_{02}*(TEX_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{01} + \gamma_{11}*(TEG_j)$$
$$\beta_{2j} = \gamma_{02}$$
$$\beta_{3j} = \gamma_{03}$$

The revised model included the regression coefficients related to the effect of Teacher Engagement and Teacher Experience and one coefficient related to the cross-level moderation between Teacher Engagement and student English Proficiency improved model fit. The cross-level coefficient suggested Teacher

Table 4.27

Language Arts Full Model

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Intercept β_{0j}					
Intercept, γ_{00}	337.49	2.36	142.98	51	<.001
Teacher engagement, γ_{01}	-2.20	.78	-2.81	51	<.001
Teacher experience, γ_{02}	-.90	.30	-2.99	51	<.001
English proficiency β_{1j}					
Intercept, γ_{01}	-9.61	1.78	-5.41	1378	<.001
Teacher engagement, γ_{11}	2.27	0.51	4.44	1378	<.001
Teacher experience, γ_{12}	-0.03	0.16	-0.19	1378	.85
Attendance β_{2j}					
Intercept, γ_{02}	48.19	11.02	4.37	1378	<.001
Teacher engagement, γ_{21}	-8.01	4.37	-1.84	1378	.07
Teacher experience, γ_{22}	-1.07	0.76	-1.40	1378	.16
Previous level of achievement β_{3j}					
Intercept, γ_{03}	0.81	0.02	36.09	1378	<.001
Teacher engagement, γ_{21}	0.02	0.01	2.23	1378	.03
Teacher experience, γ_{32}	-0.01	0.00	-2.26	1378	.04

Random effect	*SD*	Variance component	*df*	χ^2	p-value
Class mean, u_0	17.00	289.02	51	561.73	<.001
Student Level Effect, *r*	27.16	737.79			

Table 4.28

Language Arts Full Model (Revised)

Fixed Effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Intercept β_{0j}					
Intercept, γ_{00}	337.50	2.34	143.97	51	<.001
Teacher engagement, γ_{01}	-2.23	0.77	-2.88	51	.01
Teacher experience, γ_{02}	-0.96	0.29	-3.26	51	<.001
English proficiency β_{1j}					
Intercept, γ_{01}	-10.29	1.77	-5.82	1383	<.001
Teacher engagement, γ_{11}	1.68	.62	2.70	1383	.01
Attendance β_{2j}					
Intercept, γ_{02}	67.72	17.03	3.98	1383	<.001
Previous level of achievement β_{3j}					
Intercept, γ_{03}	.81	0.024	33.87	1383	<.001
Random effect	***SD***	**Variance Component**	**df**	**χ^2**	**p-value**
Class mean, u_0	16.87	284.60	51	555.21	<.001
Student level effect, *r*	27.21	740.23			

Engagement positively impacts to a greater extent non-English proficient students as compared to their English proficient peers. Cross-level moderating effects between Teacher Engagement and Attendance as well as between Teacher Experience and Attendance in the initial full model were not included in the revised model. Finally, it was also decided not to include in the final model the moderating terms between classroom level constructs and previous level of

LA Academic Achievement as their inclusion failed to reduce the variance components and improve model fit.

R^2 was recalculated at the classroom-context level in order to account for a reduction in the variance component. The random One-way ANOVA used as baseline could not be used since the inclusion of constructs at L1 changed the nature of the intercept β_{0j}. Instead, the random-coefficient regression model was used as reference. The proportion reduction in variance based on classroom level constructs Teacher Engagement and Teacher Experience for the class mean for LA Academic Achievement intercept β_{0j} and the student English proficiency intercept β_{1j} after controlling for English proficiency, Attendance, and previous level of LA Academic Achievement is [Var L2 (random regression) – Var L2 (Intercepts- and slopes-)]/ Var L2 (random regression) or (366.93 - 284.60)/ 366.93. For β_{0j}, Teacher Engagement and Teacher Experience explained 21.4 percent of the variance in student LA Academic Achievement when controlling for Teacher Education, Class size indicators, and Instructional Use of Time.

Mathematics: Intercepts-and-Slopes-as-Outcomes Model

This model included the student level constructs identified in the random-coefficient regression model (see Table 4.20), and the classroom level constructs from the means-as-outcome model (see Table 4.24). The same procedure was followed for Mathematics as described above for Language Arts. Therefore, this intercepts-and-slopes-as-outcomes model was subsequently revised to reduce variance components and improve model fit.

The means-as-outcomes model retained only Teacher Engagement and Instructional Use of Time: Administrative Tasks, while class size related constructs QEIA and TPR and other classroom level constructs did not enter the initial full model. The regression coefficient relating Teacher Engagement to student Mathematics Academic Achievement was $\beta = 1.77$ ($p = .01$). A one-unit increase in Teacher Engagement was predicted to improve student Mathematics Academic Achievement in Mathematics by 1.77 scaled score points on the California Standards Test. The coefficient relating Instructional Use of Time: Administrative Tasks to student Academic Achievement was negative ($\beta = -8.64$, $p < .001$), indicating that a one-unit increase in Administrative Tasks negatively impacted student Mathematics Academic Achievement in Mathematics by lowering CST Mathematics scores by 8.64 scaled score points. The two constructs combined resulted in producing a decrease of classroom level variance between unconditional and restricted models, therefore explaining (773.76 – 694.86)/ 773.76 or 10.2 percent of the variance between classrooms per Kreft and DeLeeuw (1998) R^2 formula.

In the random-coefficient regression model created in the third step (Table 4.19), constructs were added to the student level with no constructs at the classroom level. Two constructs were retained in the final model: Attendance ($\beta = 120.78$, $p \leq .00$), and Previous Mathematics Academic Achievement ($\beta = .69$, $p \leq .00$). Socioeconomic level, SES, ($\beta = -.42$, $p = .82$) and English Proficiency ($\beta = -4.71$, $p = .16$) did not enter the final model.

Using Kreft and DeLeeuw (1998) formula, the random-coefficient regression model explained (2291.20 - 1154.88)/ 2291.20, or 49.6 percent of the variance in student Mathematics Academic Achievement. Using an alternative R^2 formula (Snijders & Bosker, 1999), the random-coefficient regression model decreased in unexplained variance in Mathematics was estimated at 1 - [(L1 restricted error + L2 restricted error)/(L1 unrestricted error + L2 unrestricted error)], or 1 – [(1154.88+ 827.86)/(2291.20 + 773.76)]. The random-coefficient regression model explained 35.3 percent of the variance in student Mathematics Academic Achievement.

Finally, the intercepts-and-slopes-as-outcomes model was created where coefficients at the student levels were allowed to behave with fixed effects or become random with the addition of interaction with classroom level constructs (see Table 4.29).

Student Level Model

$$CR\_MA_{ij} = \beta_{0j} + \beta_{1j}*(ATT_{ij}) + \beta_{2j}*(PR\_LA_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(TEG_j) + \gamma_{02}*(IUT_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(TEG_j) + \gamma_{12}*(IUT_j)$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}*(TEG_j) + \gamma_{22}*(IUT_j)$$

The choice between random and fixed effects in the model was made with the aim of maximizing R^2 multilevel equivalent measures and the principle of parsimony was applied to defining a final model (see Table 4.30).

Table 4.29

Mathematics Full Model

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Intercept β_{0j}					
Intercept, γ_{00}	338.55	3.81	88.88	45	<.001
Teacher engagement, γ_{01}	1.77	.64	2.76	45	.01
Administrative tasks, γ_{02}	-8.64	2.89	-2.99	45	<.001
Attendance β_{1j}					
Intercept, γ_{01}	115.76	29.38	3.94	1188	<.001
Teacher engagement, γ_{11}	-2.14	8.91	-.24	1188	.81
Administrative tasks, γ_{12}	16.38	27.45	.60	1188	.55
Previous level of achievement β_{2j}					
Intercept, γ_{02}	.70	.02	29.76	1188	<.001
Teacher engagement, γ_{21}	.00	.01	-.78	1188	.44
Administrative tasks, γ_{22}	.01	.03	-.49	1188	.62

Random effect	*SD*	Variance component	*df*	χ^2	p-value
Class mean, u_0	26.35	694.75	45	707.60	<.001
Student level effect, *r*	34.04	1158.77			

Student Level Model

$$CR\_MA_{ij} = \beta_{0j} + \beta_{1j}*(ATT_{ij}) + \beta_{2j}*(PR\_MA_{ij}) + r_{ij}$$

Classroom Level Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(TEG_j) + \gamma_{02}*(IUT7) + u_{0j}$$

Table 4.30

Mathematics Full Model (Revised)

Fixed Effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Intercept β_{0j}					
Intercept, γ_{00}	338.58	3.79	89.24	45	<.001
Teacher engagement, γ_{01}	1.77	0.64	2.76	45	.01
Administrative tasks, γ_{02}	-8.64	2.89	-2.99	45	<.001
Attendance β_{1j}					
Intercept, γ_{01}	70.67	21.84	3.24	1192	<.001
Previous level of achievement β_{2j}					
Intercept, γ_{02}	.71	.02	30.30	1192	<.001
Random effect	***SD***	**Variance component**	***df***	χ^2	**p-value**
Class mean, u_0	26.24	688.28	45	693.15	<.001
Student level effect, *r*	34.14	1165.36			

The regression coefficients related to the effect of Teacher Engagement (β = 1.77, p = .01) and Instructional Use of Time Administrative Tasks (β = -8.64, p <.001) on student Academic Achievement in Mathematics were significant. While an increase in Teacher Engagement was predicting a minimal increase in student Mathematics Academic Achievement, an increase in Administrative Tasks was associated with a decrease in student Mathematics Academic Achievement.

R^2 was recalculated at the classroom-context level in order to account for a reduction in the variance component. As it was the case when examining

Language Arts, the random One-way ANOVA used as baseline could not be used since the inclusion of constructs at the student level changed the nature of the intercept β_{0j}. Instead, the random-coefficient regression model was used as reference. The proportion reduction in variance based on classroom level constructs Teacher Engagement and Teacher Experience for the class mean Mathematics Academic Achievement intercept β_{0j} and the student English proficiency intercept β_{1j} after controlling for Attendance, and previous level of Mathematics Academic Achievement was [Var L2 (random regression) – Var L2 (Intercepts- and slopes-)]/Var L2(random regression) or (827.41 – 688.28)/ 827.41. For β_{0j}, Teacher Engagement and Administrative Tasks explained 16.8 percent of the variance in student Academic Achievement in Mathematics when controlling for Attendance and Previous level of Mathematics Academic Achievement included in the student level of the model.

Summary

<u>Language Arts</u>

14.4 percent of the variance in LA Academic Achievement was found between classrooms (classroom level) and 85.6 percent of the variance in LA Academic Achievement was attributed to students within classrooms (student level). The inclusion of Teacher Experience and Teacher Engagement at the classroom level indicated that students taught by more experienced and engaged teachers had slightly lower LA Academic Achievement. Teacher Experience and Engagement accounted for 22.5 percent of the variance found between schools.

English Proficiency, Attendance and Previous LA Academic Achievement accounted for 57.5 percent of the variance explained within classes, among students. The full model including cross-level moderator effects was subsequently revised to include only significant main effects and moderator effects. Teacher Engagement and Experience at the classroom level continued to impact LA Academic Achievement. The cross-level moderation between Teacher Engagement and English Proficiency status suggested that students with a lower level of English Mastery experience a greater level of LA Academic Achievement with more engaged teacher. After controlling for English Proficiency, Attendance, and previous level of achievement, Teacher Experience and Engagement accounted for 21.4 percent of the variance. Class size (as measured by participation in the QEIA or by PTR) was not found to impact LA Academic Achievement.

Mathematics

In Mathematics, the unconditional model suggested 25.2 percent of the variance in Mathematics Academic Achievement lay between classroom (classroom level) and 74.8 percent of the variance in Mathematics Academic Achievement was found at the student level within classrooms (student level). Teacher Engagement and Instructional use of Time: Administrative Task predicted Mathematics Academic Achievement between classrooms. Teacher Engagement predicted higher Mathematics Academic Achievement at the student level. On the contrary, an increase in the amount of time spent on Administrative Tasks during class would be associated with lower Mathematics

Academic Achievement. Combined Teacher Engagement and Instructional Use of Time: Administrative tasks accounted for 10.2 percent of the variance between classrooms.

Within classroom, students with a higher socioeconomic status and higher level of Previous Mathematics Academic Achievement were found to reach a higher level of academic achievement. These student level constructs accounted for 49.6 percent of the variance within classrooms. Following the same procedure as described for Language Arts, a revised full model suggested only main effects without cross-level interactions. When controlling for Attendance and Previous Mathematics Academic Achievement at the student level, Teacher Engagement and Instructional Use of Time: Administrative Tasks accounted for 16.8 percent of the variance in student achievement. In Mathematics, as in Language Arts, class size was not found to predict Mathematics Academic Achievement at the student level.

CHAPTER FIVE

DISCUSSION

The aim of the study was to examine the relationship between smaller classes and student Academic Achievement in middle school grades. The award of the California Quality Education Improvement Act (QEIA) in 2006 to two of five middle schools within a large school district provided an opportunity to set up a quasi-experiment within the context of a homogeneous suburban school district. The importance of the present study is critical in that, at the time of this writing, K-12 and postsecondary education funding is dramatically decreasing in California. Approximately $2.7 billion were slated to be dispersed by the QEIA grant over a seven year period.

The literature review revealed no direct relationship which satisfactorily linked class size reduction to student Academic Achievement. Meta-analyses (Glass, 1976; Hedges & Stock, 1983; Robinson & Wittebols, 1986; Shapson et al., 1980; Slavin, 1984) indicated marginal effect size gains not exceeding .10 to .20 standard deviations, and most studies only examined the impact of class size reduction in the first three years of elementary education. Studies which investigated class size reduction in later elementary grades or in secondary school were often inconclusive as the classroom student average level of ability was uncontrolled (Shkolnik, 1997).

Starting in the 1990s, partly due to the availability of increasingly sophisticated tools, researchers no longer viewed the relationship between class size and Academic Achievement as a direct pathway. Instead, moderating factors were increasingly considered in an attempt to explain the mechanisms linking class size and Academic Achievement. Classroom-context factors such as teacher characteristics, teacher practices, and organizational setups were considered while controlling within class variations due to student characteristics. This evolution in class size research along with the progress in statistical tools designed to provide multilevel analyses provided justification for this study.

The results of multilevel analyses marginally confirmed the fit of the suggested model for both Language Arts and Mathematics. Most notably, class size did not impact Academic Achievement for either Language Arts or Mathematics. On the other hand, ancillary analysis revealed Teacher Engagement and Teacher Experience explained approximately 21 percent of the explained variance in Language Arts Academic Achievement. Whereas in Mathematics, Teacher Engagement and Instructional Use of Time: Administrative Tasks explained 16.8 percent of the variance after controlling for student characteristics. Ancillary analysis also revealed no moderating effects of QEIA or PTR on student-level or classroom-level constructs.

Discussion

The present study departed from most prior research studies on class size effects in at least two meaningful ways: (a) middle school was chosen as the school grade level of the participants; and, (b) multilevel analysis was used to analyze the differences in Academic Achievement between reduced and nonreduced class sizes.

The first hypothesis which stated students enrolled in smaller classes in QEIA-recipient schools would have higher levels of Academic Achievement in both Language Arts and Mathematics was not supported. A posthoc analysis, using the PTR failed to support the contribution of smaller class size to Academic Achievement.

From the literature review, it became evident that class size reduction impacts student Academic Achievement differentially. Much has been written on the successes of class size reduction in the early elementary grades (K-3). Yet, the literature review revealed that similar studies were scarce in either the middle school or secondary school grade levels. In the present study, the gains found for class size reduction on academic achievement in early elementary grades (K-3) were not also found in middle school grades (grade 7). Based on the researcher's experiences during 20 years in education as a teacher and administrator in grades K-8, class size reduction success in early elementary grades (K-3) may be due to the nature of learning. In the formative years of early elementary grades, students learn to read. At this stage of literacy development and skill acquisition, students rely heavily on teacher explicit instruction and feedback. This may no longer be the case the middle grades, fourth grade and fifth grade, when students now read to learn the content. In the middle grades students are more independent and draw learning not only from direct contact with their teachers but also with their peers, be it during small group activities or pair work. This observation may shed some light as to why class size reduction was not shown to increase academic achievement in the present study.

Another possible reason for the lack of impact of class size on student Academic Achievement may be that the mean class size of the reduced classes was 23 students, an average class size that exceeds the 17 threshold suggested by the large meta-analysis by Glass and Smith (1977). Although the idea of decreasing class size had scientific bases as discussed earlier in the literature review, the participating schools the QEIA initiative failed to lower class size to the recommended levels. This initiative was conceptually flawed from its inception since research clearly suggested class size needed to be reduced to levels at least below the 20:1 Pupil-to-Teacher Ratio (Achilles et al., 2002; Finn, 1998; Glass & Smith, 1979; Robinson, 1990; Shapson et al., 1980; Slavin, 1989a).

The second hypothesis stated differences in Academic achievement associated with smaller class size for students of low- and high- socioeconomic status would be found. This hypothesis was also not supported. SES level was a composite determined by either of two proxy variables: participation in the National School Lunch Program (NSLP), also called free/reduced lunch program; or, self-reported parental education level (not being a high school graduate was an indicator of low SES). The use of a proxy is only an approximation, and may not be as accurate as if the data the proxy is representing were able to be collected. Unlike prior research findings (Achilles et al., 1997; Caldas & Bankston, 1997; Robinson, 1990; White, 1982), the present study did not reveal differences in Academic Achievement gains for low socioeconomic students enrolled in reduced sized classes. One possible reason for this was students at both ends of the academic ability continuum (e.g., special education and gifted students) were not considered in the study. It is possible this exclusion criterion may have reduced the variability of SES in the participants.

The third hypothesis stated that previous level of Academic Achievement and class sizes would be linked to Current Academic Achievement; however, this hypothesis was not supported despite what was suggested in prior research (Robinson & Wittebols, 1986; Slavin, Karweit, & Madden, 1989). It is plausible the high correlation between Previous and Current Academic Achievement did not allow any of the other constructs of interest to meaningfully enter the models after removing the variance accounted for by Previous Academic Achievement. Again, as in the case of socioeconomic status, it would have been beneficial to include students with a broader range of Academic Achievement. The classrooms under consideration in the present study were mainstreamed were also less likely to benefit from reduced class size. Nonmainstreamed students may indeed be less likely to learn independently.

The fourth hypothesis stated there would be a cross-level moderation effect between student English proficiency status and classroom level fixed effects (PTR; Teacher Engagement; Teacher Experience; Teacher Education; QEIA participation; and, Instructional Use of Time). Only in Language Arts was this hypothesis supported, confirming a greater level of Teacher Engagement positively impacted the Academic Achievement of non-English proficient students as compared to their English fluent classmates. Judging by the low beta weights, the moderation of English proficiency status and Teacher Engagement on Academic Achievement in Language Arts has far reaching implication for educators. Moving beyond class size, administrators and counselors must assign to classes with predominantly English learners those teachers who can motivate and make personal connections with the students. Class size in this context may favor pro-social behavior (Finn et al., 2003).

In a posthoc analysis, a revised Language Arts multilevel model retained English Proficiency status, Attendance, and previous level of Academic Achievement at student level while the intercept (or class achievement) of the random-coefficient regression was a function of two fixed effects: Teacher Engagement; and, Teacher Experience. However, the resulting beta weights were contrary to intuitive beliefs that greater Teacher Engagement or Teacher Experience would be associated with in greater Academic Achievement. Also, the classroom level coefficients were small to the extent that these constructs had little impact on current Academic Achievement.

For Mathematics, previous Achievement and Attendance were retained as predictors at the student level. Students with different levels of English proficiency and SES were not found to achieve differently, and this student level construct was dropped from the model. Interestingly, Teacher Engagement positively moderated the relationship between English Proficiency and student Mathematics Academic Achievement. Teachers who demonstrated more engagement contributed to increasing the Mathematics Academic Achievement of English learners. Also of interest was the impact of the instructional time spent by teachers on Administrative Tasks. Increased time on these noninstructional activities was associated with slightly lower Academic Achievement in Mathematics at the student level.

Although it is possible that Teacher Experience reaches a point of diminishing academic return, the practical implication of this finding is to focus on increasing Teacher Engagement. Pro-social behaviors among staff and students are associated to school climate. Savvy administrators will take every opportunity to genuinely demonstrate to teachers they are valued by offering training, recognition, and praise.

Limitations

The limitations of this study include: temporal relevance; external validity; assessment of the Teacher Engagement construct; the use of multilevel modeling; the lack of a qualitative component; and, characteristics of the student population observed. Each of these will be discussed along with recommendations to address these limitations in future research.

Despite the *Teacher Engagement Scale* being reliable, there was really little variation among responses. The operationalization of the construct Teacher Engagement may not have completely or accurately measured the facets which make up this construct. The addition of more items, rewording detailed questions, and reverse-coding similar items may have further improved the assessment of Teacher Engagement. Additionally, the broader construct of Engagement may need to be expanded to the school community at large, which would provide a different lens on Engagement in the classroom. It is therefore recommended that specific questionnaires be developed to measures the Engagement construct at the student and administrator levels.

Although the revised multilevel models in Language Arts and Mathematics were reliable, it may be argued that a one-year interval time-series data analysis of class size as a mode of academic intervention was too short to reveal differences. Time series over a two-year to three-year may considerably increase the possibility of detecting differences in Academic Achievement. Furthermore, measures other than standardized testing (e.g., teacher-designed tests, projects, or local district benchmarks) would greatly enhance capturing the construct of Academic Achievement.

The limited amount of variance explained by the two revised models impedes external validity and generalization beyond the participating school district. The present study took into account this possibility by eliminating classes strictly designated for students identified as gifted or in need of special education services. It is recommended future studies include more classrooms, and include students with a wider range of ability, from special education to gifted and talented. Furthermore, the current study was not able to address difference in previous and current Academic Achievement longitudinally. It is recommended that future studies include Academic Achievement data spanning over several academic years.

Multilevel modeling analysis inherently has limitations which impacted the study. Results in multilevel analysis are seldom interpreted in the literature in terms of prediction. The concept of variance explained itself is elusive since a comparison to the unconditional model holds as valid when student- and classroom-level equations are considered separately. Once constructs are entered at the student level, the nature of the classroom level equation has changed as the intercept of student level equation has been modified.

The model would also be improved by adding a qualitative component in the analysis. For example, student classroom activities (cooperative learning, project-based instruction, student-to-student interactions) and student/teacher Engagement activities could be observed to allow for a more richness of the true classroom context. Interviews with students and teachers would provide teachers and administrators with greater insight on student and teacher Engagement. It is critical to refrain from implementing "magic bullet" policies such as class size reduction, and recognize efforts must be tailored to fit the circumstances of each school and each classroom. For instance, higher achieving students could be scheduled in larger classes, allowing for lower achievers to receiving more individual attention, at least for a portion of the day.

Lastly, the limited variability in SES and the overall low Academic Achievement level of the population under consideration may have prevented detecting moderating effects. It is recommended future studies considered a large sample, including more than one school district.

Research Contribution and Implications

The results suggested in this study contribute to the body of knowledge on class size and the educational community in several ways.

First, the findings reframe class size reduction as only one of many instructional interventions available. Popular though class size reduction may be among parents and teachers, other forms of instructional intervention have shown better returns on the instructional dollar.

Does spending $2.7 billion dollars on class size reduction make educational sense? This question can only be answered when economic resources are matched to educational outcomes for all forms of intervention, class size reduction included. Hanushek (1997), Jepsen & Rivkin (2002) and Shkolnik (1997) have long questioned the use of school resources to close the achievement gaps (Hanushek, 1997; Jepsen & Rivkin, 2002; Shkolnik, 1997).

Secondly, the present study, like prior research, questioned the added value of class size reduction intervention as compared to interventions aimed at increasing academic achievement. Hattie (2005) calculated effect sizes of 46 types of interventions in 500 meta-analyses summarizing some 300,000 studies of factors associated with academic achievement. Class size as a form of intervention fell well below the average mean effect size for all 46 factors. While class size reduction seems a logical path towards closing the achievement gap and improving academic achievement, the educational community, parents and educators, must not consider it in isolation as an easy remedy, but rather as a choice of intervention options which must be considered in concert. School board members and policy makers must first consider promoting effective classroom contexts, and refrain from the temptation of "fixing" education with one blanket decision. Too often, class size reduction has been implemented because policy makers could edict it with the stroke of pen.

The current study highlighted that educational decisions are not made in a political vacuum, but in the light of personal and group interests (Graue & Rauscher, 2009). For instance, QEIA was ordered by the Sacramento Superior Court (*California Teachers Association et al. v. Schwarzenegger et al., 2006*) to remedy the budget shortfall California schools experienced after Governor Schwarzenegger suspended Proposition 98, which requires a fixed portion of state budget be spent in K-14 education. It is possible that the state preferred settling for a dubious class-size reduction program for fear of seeing this windfall turn into teacher salary increase.

Results of this study suggested Teacher Engagement may raise student Academic Achievement by several points on the California Standards Tests. District and school administrators should promote self-efficacy and empowerment among teachers. Professional recognition, praise, but also opportunities for professional development are powerful incentives which promote and foster engaged teachers. Teacher Experience was found to impact Academic Achievement in Language Arts, but did not explain most of the variability in student Academic Achievement. Beyond teacher self-efficacy, Teacher Engagement and Teacher Experience are areas administrators and policy makers have the power to make a difference. Programs that would address these constructs can and should be developed for the benefit of both new and veteran teachers.

The present research also pointed at the importance of decreasing tasks which are not directly associated with student learning. Administrative Tasks were associated with lower Academic Achievement in Mathematics. Therefore, teachers and administrators must ensure that every minute spent in class productively contribute to learning. At the time of this writing, teachers often complain, and rightly so, that much of instructional time is devoted to testing. All testing that does not produce feedback formative data should be reduced to a minimum.

Students are not passive in the educational process. If previous Academic Achievement is a strong predictor of future performance, so is Attendance. Students who exceeded the average student attendance also demonstrated higher Academic Achievement. It is recommended that administrators and teachers keep students accountable for attending school regularly through engaged activities and self-actualization.

Finally, the current research highlighted the importance of accounting for nested phenomena with a multilevel analytical tool since student characteristics within the same unit of analysis, in this instance the classroom, are no longer independent observations (Raudenbush & Bryk, 2002).

Directions for Future Research

Class size alone does not change outcomes academic achievement. Nor are students affected in the same fashion. Concurring with Ehrenberg et al. (2001a), this study points out that not all students are affected in the same way. English learners and possibly lower achievers may benefit the most from reduced class sizes. Teacher Engagement and effective Use of Instructional Time should be given more attention for their potential to improve achievement and close the gap between English learners and lower achievers, on one hand, and mainstream students, on the other. Perhaps, one of the most compelling reasons explaining the low added value of class size reduction is that teachers may not actually change and maximize their teaching strategies when class sizes are reduced. Additional research in the area of instructional practices linked to class size should further clarify the potential of this strategy

Conclusions

Academic Achievement varies widely among classrooms, a fact which cannot be explained by class size alone, but rather by a wider range of classroom-context and student-context factors. Smaller class size strategy alone as a tool for school reform is not directly associated with an increase academic achievement. Although student achievement is overwhelmingly dependent upon past student achievement, the present study highlighted some important components that impact learning: Teacher Engagement; Teacher Experience; and, student Attendance.

If the decision is made to implement class size reduction, specific groups should be targeted such as low socioeconomic students and English learners, and Pupil-to-Teacher ratios (PTR) should be below 20:1. The question is not whether schools should implement class size reduction or not. Rather, the wise administrator and policy maker will consider class size reduction as only one among other interventions to shape educational policies with the intent of maximizing available resources. Teachers do not fundamentally change the way they teach when they are given smaller classes (Cahen, Filby, McCutcheon, & Kyle, 1983; Slavin, 1989b). Therefore, staff development stressing effective teaching strategies within small classrooms should become part of any proposed class size reduction program.

APPENDIX A

PERMISSION TO CONDUCT RESEARCH INSTITUTIONAL

REVIEW BOARD AUTHORIZATION

Fontana Unified School District

9680 Citrus Avenue · P.O. Box 5090 · Fontana · CA 92334-5090 (909) 357-5000

February 10, 2011

To: J. J. Francoisse

From: Kimberly MacKinney, Director Secondary/Assessment & Evaluation

Subject: Research Request-The effects of smaller class sizes on academic achievement in middle school

---

Thank you for completing an Application to Conduct Research in the Fontana Unified School District. Your application has been accepted and you may conduct the research at the middle schools. Please note that the Technology Department may be calling you for further information with assistance on the data format that you are requesting. School and student records must be accessed thru the Fontana Unified School District's Technology Department.

If you have any other questions, please do not hesitate to call my office at 909-357-5000 ext. 7103.

Sincerely,

Kimberly MacKinney
Director Secondary Education/Assessment & Evaluation

/ib Research Request-Francoisse

Cc: Middle School Principals
Technology

CALIFORNIA STATE UNIVERSITY
SAN BERNARDINO

Academic Affairs
Office of Academic Research • ~~Institutional Review Board~~

March 21, 2011

CSUSB INSTITUTIONAL REVIEW BOARD
Administrative Review
IRB# 10058
Status
APPROVED

Mr. Jean-Jacques Francoisse
c/o: Prof. Marita Mahoney
Department of Education Administration
c/o: Prof. Matt Riggs
Department of Psychology
California State University
5500 University Parkway
San Bernardino, California 92407

Dear Mr. Francoisse:

Your application to use human subjects, titled, "A Multilevel Model of the Relationship between Smaller Class Sizes and Academic Achievement in Middle Schools" has been reviewed and approved by the Chair of the Institutional Review Board (IRB) of California State University, San Bernardino and concurs that your application meets the requirements for exemption from IRB review Federal requirements under 45 CFR 46. As the researcher under the exempt category you do not have to follow the requirements under 45 CFR 46, which requires annual renewal and documentation of written informed consent which are not required for the exempt review category. However, exempt status still requires you to attain consent from participants before conducting your research.

The CSUSB IRB has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval notice does not replace any departmental or additional approvals which may be required.

Although exempt from federal regulatory requirements under 45 CFR 46, the CSUSB Federal Wide Assurance does commit all research conducted by members of CSUSB to adhere to the Belmont Commission's ethical principles of respect, beneficence and justice. You must, therefore, still assure that a process of informed consent takes place, that the benefits of doing the research outweigh the risks, that risks are minimized, and that the burden, risks, and benefits of your research have been justly distributed.

You are required to do the following:

1) Protocol changes must be submitted to the IRB for approval (no matter how minor) before implementing in your prospectus/protocol. Protocol Change Form is on the IRB website.
2) If any adverse events/serious adverse/unanticipated events are experienced by subjects during your research. Form is on the IRB website.
3) And, when your project has ended.

Failure to notify the IRB of the above, emphasizing items 1 and 2, may result in administrative disciplinary action.

If you have any questions regarding the IRB decision, please contact Michael Gillespie, IRB Compliance Coordinator. Mr. Michael Gillespie can be reached by phone at (909) 537-7588, by fax at (909) 537-7028, or by email at [illegible]. Please include your application identification number (above) in all correspondence.

Sincerely, *Sharon L Ward, Ph.D*

Sharon Ward, Ph.D, Chair
Institutional Review Board

SW/mg

cc: Prof. Marita Mahoney, Department of Education Administration and Prof. Matt Riggs, Department of Psychology

909.537.7588 • fax: 909.537.7028 • http://irb.csusb.edu/
5500 UNIVERSITY PARKWAY, SAN BERNARDINO, CA 92407-2393

APPENDIX B

LETTER OF INFORMED CONSENT

CALIFORNIA STATE UNIVERSITY
SAN BERNARDINO

College of Education
Office of the Dean

INFORMED CONSENT

The study in which you are being asked to participate is designed to investigate instructional practices in middle schools. Your thoughtful participation is greatly appreciated. Your participation in this survey will make you eligible to win a $25 Target certificate.

This study is being conducted by J.J. Francoisse under supervision of Dr. Matt Riggs in the Department of Psychology, and Dr. Marita Mahoney, Director, Office of Assessment & Research, at California State University, San Bernardino (CSUSB). This study has been approved by the Institutional Review Board, California State University, San Bernardino.

PURPOSE: The purpose of the study is to investigate instructional practices in middle schools.

DESCRIPTION: Data collection consists of distribution of surveys to approximately 50 Language Arts and Mathematics middle schools teachers, who taught seventh-grade in a large suburban southern California school district during the academic year 2009-10.

PARTICIPATION: Participation in the survey is voluntary. Refusal to participate will involve no penalty or loss of benefits to which the participant is otherwise entitled. Also, the participants may discontinue participation at any time without penalty of loss of benefits, to which the participant is otherwise entitled.

CONFIDENTIALITY: Participants have a right to privacy and all information identifying participants will remain confidential. Information will be recoded and confidentiality of the participants will be maintained by storing data on a password protected computer.

DURATION: The time to complete the survey will be approximately 10 minutes.

RISKS: There are no foreseeable risks or discomforts to the participants who consent to participate in the study.

BENEFITS: There is no particular benefit to the participants other than an opportunity to win a $25 Target certificate: one name per school site will be drawn for completing the survey. The aggregated results of this research may be published in a professional journal after it has been completed thereby contributing to the body of empirically-based educational research. The study will potentially help improve student academic outcomes in middle schools. Participant confidentiality will continue to be maintained.

CONTACT: If you have any question about the research and research participants' rights, you may contact Dr. Matt Riggs, Professor Department of Psychology at (909 537-5574, mriggs@csusb.edu or Dr. Marita Mahoney, Director, Office of Assessment & Research, at California State University, San Bernardino (CSUSB), at (909) 537-3621, mmahoney@csusb.edu. This study has been approved by the Institutional Review Board, CSUSB (909) 537-5315.

RESULTS: Results will be available by July 2011. The results will be presented during a public defense and a bound copy of the dissertation will available in the CSUSB library.

CONSENT: I understand that I am participating in research and the research has been explained so that I understand my involvement. I understand that I may stop participating at any time without any consequences or penalty for so doing. I understand that I must be 18 years of age or older to participate in the study.

Signature: ______________________ Date: ______________

909.537.5600 • fax: 909.537.7011
5500 UNIVERSITY PARKWAY, SAN BERNARDINO, CA 92407-2393

APPENDIX C

TEACHER SURVEY

Instructional Survey

Dear teachers,

This study is being conducted by J.J. Francoisse, doctoral student at California Sate University, San Bernardino, under supervision of Dr. Matt Riggs in the Department of Psychology, and Dr. Marita Mahoney, Director, Office of Assessment & Research, at California State University, San Bernardino (CSUSB). This study has been approved by the Institutional Review Board, California State University, San Bernardino.

The study in which you are being asked to participate is designed to investigate instructional practices in middle schools. Your thoughtful participation is greatly appreciated. This study has been authorized by [redacted]. All information will remain confidential.

Your participation in this survey will make you eligible to win a $25 Target certificate. Upon completion of the survey, your name will be entered in a drawing under the supervision of the your principal. The winner of the certificate will be notified by email, and the gift card will be available at the principal's office.

Thank you for your participation in this survey.

J.J. Francoisse
Ed.D. Doctoral Candidate
California State University, San Bernardino

School [redacted]

Teacher [redacted]

The winner of the $25 Target gift certificate at Wayne Ruble Middle will be notified via email. Please write your email address below i you wish to participate in the raffle.

__________________@____________

School [redacted]

Teacher [redacted]

Period 6 Lang Arts 7

For each item, please indicate how strongly you agree with the statement by circling your response.

Please think of your period 6 Lang Arts 7 class of last year 2009-10 at [redacted] about how much class time did you allocate weekly to the following activities:

0 Minutes	30 Minutes	One Hour	Two Hours	More than Two Hours
1	2	3	4	5

1.	LECTURING	1	2	3	4	5
2.	LEADING DISCUSSION	1	2	3	4	5
3.	WORKING IN SMALL GROUPS	1	2	3	4	5
4.	DOING SEAT WORK	1	2	3	4	5
5.	PROVIDING INDIVIDUAL INSTRUCTION	1	2	3	4	5
6.	STUDENT DISCIPLINE	1	2	3	4	5
7.	ADMINISTRATIVE TASKS	1	2	3	4	5

7415

For each item, please indicate how strongly you agree with the statement by circling your response.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

1.	I think what I teach is interesting	1	2	3	4	5
2.	I actively encourage participation and discussion	1	2	3	4	5
3.	I listen to my students when they have something to say	1	2	3	4	5
4.	Most mornings, I look forward to teaching	1	2	3	4	5
5.	I enjoy teaching new topics to my students	1	2	3	4	5
6.	I keep teaching in different ways until my students understand	1	2	3	4	5
7.	I encourage my students to raise questions about what they learn	1	2	3	4	5
8.	I volunteer to help with school activities	1	2	3	4	5
9.	My students will understand the concepts if I spend energy explaining..	1	2	3	4	5
10.	When I am in class, my mind wanders	1	2	3	4	5

How long have you been a teacher? ___________ Years

Highest degree earned: B.A./B.S. MA/MS Ed.D./Ph.D.
(circle one)

7415

Developed by Jean-Jacques Francoisse.

APPENDIX D

TEACHER ENGAGEMENT SCALE: MATHEMATICS AND LANGUAGE ARTS

	N	Mean	*SD*	Cronbach Alpha if item deleted
1. I think what I teach is interesting.	102	4.47	.54	.66
2. I actively encourage participation/discussions.	102	4.66	.52	.63
3. I listen to my students when they have something to say.	102	4.60	.69	.58
4. Most mornings, I look forward to teaching.	102	4.13	.98	.53
5. I enjoy teaching new topics to my students.	102	4.65	.48	.60
6. I keep teaching in different ways until my students understand.	102	4.24	.79	.61
7. I encourage my students to raise questions about what they learn	102	4.50	.63	.61
8. I volunteer to help with school activities.	102	3.80	.97	.71
9. My students will understand the concepts if I spend energy explaining.	102	3.95	.93	.62
10. When I am in class, my mind wanders. (reverse-coded)	102	4.18	.94	.72

Developed by Jean-Jacques Francoisse.

APPENDIX E

NONSIGNIFICANT MEANS AS OUTCOMES FOR

LANGUAGE ARTS AND MATHEMATICS

Table E1

Means-as-Outcome: Teacher Education - TDG (Language Arts)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	337.66	2.61	129.61	52	<.001
Teacher education slope, γ_{01}	-7.09	5.91	-1.20	52	.24
Random effect	***SD***	**Variance component**	***df***	**χ^2**	***p*-value**
Class mean, u_0	17.62	310.56	52	285.31	<.001
Student level effect, *r*	41.86	1752.28			

Table E2

Means-as-Outcome: Lecturing - IUT1 (Language Arts)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	337.65	2.63	128.56	52	<.001
Lecturing slope, γ_{01}	1.96	2.12	.93	52	0.36
Random effect	***SD***	**Variance component**	***df***	**χ^2**	***p*-value**
Class mean, u_0	17.77	315.64	52	282.86	<.001
Student level effect, *r*	41.86	1752.62			

Table E3

Means-as-Outcome: Leading Class Discussion – IUT2 (Language Arts)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	337.61	2.59	130.20	52	<.001
Leading class discussion slope, γ_{01}	-4.48	3.51	-1.28	52	.21
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	17.54	307.48	52	282.11	<.001
Student level effect, *r*	41.86	1752.267			

Table E4

Means-as-Outcome: Working in small Groups – IUT3 (Language Arts)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class Mean, γ_{00}	337.60	2.57	131.11	52	<.001
Working in small groups slope, γ_{01}	-4.20	2.62	-1.60	52	.12
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	17.35	301.15	52	272.90	<.001
Student level effect, *r*	41.86	1752.40			

Table E5

Means-as-Outcome: Doing Seat Work – IUT4 (Language Arts)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	337.63	2.63	128.22	52	<.001
Doing seat work slope, γ_{01}	-.49	2.43	-.20	52	0.84
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	17.88	319.75	52	286.64	<.001
Student level effect, *r*	41.86	1752.57			

Table E6

Means-as-Outcome: Providing individual Instruction – IUT5 (Language Arts)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	337.62	2.63	128.15	52	<.001
Providing indiv. instruction slope, γ_{01}	-.99	2.58	-.38	52	.70
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	17.88	319.65	52	419.12	<.001
Student level effect, *r*	41.86	1752.50			

Table E7

Means-as-Outcome: Student Discipline – IUT6 (Language Arts)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	337.62	2.63	128.16	52	<.001
Student discipline slope, γ_{01}	.58	2.54	.23	52	.09
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	17.88	319.54	52	285.33	<.001
Student level effect, *r*	41.86	1752.58			

Table E8

Means-as-Outcome: Administrative Tasks – IUT7 (Language Arts)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	337.69	2.58	131.13	52	<.001
Administrative tasks slope, γ_{01}	-5.16	2.97	-1.74	52	.09
Random effect	***SD***	**Variance component**	***df***	χ^2	***p*-value**
Class mean, u_0	17.36	301.24	52	274.59	<.001
Student level effect, *r*	41.86	1752.68			

Table E9

Means-as-Outcome: Teacher Experience – TEX (Mathematics)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	338.45	4.22	80.26	46	<.001
Teacher experience slope, γ_{01}	-.06	.69	-.08	46	.94
Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	28.16	793.21	46	426.80	<.001
Student level effect, *r*	47.87	2291.13			

Table E10

Means-as-Outcome: Teacher Education - TDG (Mathematics)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	338.51	4.05	83.52	46	<.001
Teacher education slope, γ_{01}	-15.61	8.48	-1.84	46	.07
Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	27.00	729.05	46	387.75	<.001
Student level effect, *r*	47.87	2291.51			

Table E11

Means-as-Outcome: Lecturing - IUT1 (Mathematics)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	338.45	4.20	80.64	46	<.001
Lecturing slope, γ_{01}	1.63	5.22	0.31	46	.76

Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	28.11	790.25	46	423.77	<.001
Student level effect, *r*	47.87	2291.18			

Table E12

Means-as-Outcome: Leading Class Discussion – IUT2 (Mathematics)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	338.44	3.86	87.62	46	<.001
Leading class discussion slope, γ_{01}	11.67	5.82	2.01	46	.06

Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	25.60	655.29	46	360.32	<.001
Student level effect, *r*	47.87	2291.36			

Table E13

Means-as-Outcome: Working in small Groups – IUT3 (Mathematics)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	338.46	4.19	80.83	46	<.001
Working in small groups slope, γ_{01}	1.99	2.90	0.69	46	.50

Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	28.03	785.80	46	421.85	<.001
Student level effect, *r*	47.87	2291.20			

Table E14

Means-as-Outcome: Doing Seat Work – IUT4 (Mathematics)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	329.58	8.58	38.43	46	<.001
Doing seat work slope, γ_{01}	2.90	2.96	0.98	46	.33

Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	27.96	781.52	46	418.98	<.001
Student level effect, *r*	47.87	2291.22			

Table E15

Means-as-Outcome: Providing individual Instruction – IUT5 (Mathematics)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	338.47	4.17	81.27	46	<.001
Providing individual instruction slope, γ_{01}	4.09	3.75	1.090	46	.28

Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	27.85	775.57	46	419.12	<.001
Student level effect, *r*	47.87	2291.14			

Table E16

Means-as-Outcome: Student Discipline – IUT6 (Mathematics)

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
Class mean, γ_{00}	338.47	4.13	81.99	46	<.001
Student discipline slope, γ_{01}	-5.59	3.22	-1.74	46	.09

Random effect	*SD*	Variance component	*df*	χ^2	*p*-value
Class mean, u_0	27.56	759.42	46	413.65	<.001
Student Level effect, *r*	47.86	2291.04			

APPENDIX F

NONSIGNIFICANT INTERACTIONS BETWEEN CLASS SIZE AND STUDENT-LEVEL CONSTRUCTS

Table F1

QEIA Participation and Student Level Interactions for Language Arts

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
	For Socio economic Status Slope β_1				
Intercept, γ_{10}	300.53	8.64	34.80	1433	<.001
QEIA, γ_{11}	6.60	11.39	.58	1433	.56
	For English proficiency status slope β_2				
Intercept, γ_{20}	107.93	15.02	7.19	1433	<.001
QEIA, γ_{21}	7.92	22.75	.35	1433	.73
	For attendance slope β_3				
Intercept, γ_{30}	-90.06	120.59	-.75	1433	.46
QEIA, γ_{31}	-122.28	199.16	-.61	1433	.54
	For previous level of achievement slope β_3				
Intercept, γ_{30}	1.53	.23	6.57	1433	<.001
QEIA, γ_{31}	-.03	.31	-.08	1433	.94

Table F2

Pupil-to-Teacher Ratio (PTR) and Student Level Interactions for Language Arts

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
	For socioeconomic status slope β_1				
Intercept, γ_{10}	302.39	5.76	52.47	1433	<.001
PTR, γ_{11}	.71	.95	.75	1433	.46
	For English proficiency status slope β_2				
Intercept, γ_{20}	111.68	10.85	10.29	1433	<.001
PTR, γ_{21}	-.51	1.94	-.26	1433	.79
	For attendance slope β_3				
Intercept, γ_{30}	-145	99.54	-1.46	1433	.14
PTR, γ_{31}	1.81	14.67	.12	1433	.90
	For previous level of achievement slope β_3				
Intercept, γ_{30}	1.52	.15	9.91	1433	<.001
PTR, γ_{31}	.01	.02	.04	1433	.97

Table F3

QEIA Participation and Student Level Interactions for Mathematics

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
		For socioeconomic status slope β_1			
Intercept, γ_{10}	302.57	9.25	32.72	1234	<.001
QEIA, γ_{11}	10.73	13.75	.78	1234	.44
		For English proficiency status slope β_2			
Intercept, γ_{20}	91.90	12.68	7.25	1234	<.001
QEIA, γ_{21}	10.87	22.58	.48	1234	.63
		For attendance slope β_3			
Intercept, γ_{30}	-134.16	126.62	-1.06	1234	.29
QEIA, γ_{31}	-18.49	211.40	-.09	1234	.93
		For previous level of achievement slope β_3			
Intercept, γ_{30}	1.05	.13	7.87	1234	<.001
QEIA, γ_{31}	-.23	.19	-1.21	1234	.22

Table F4

Pupil-to-Teacher Ratio (PTR) and Student Level Interactions for Mathematics

Fixed effects	Coefficient	*SE*	*t*-ratio	*df*	*p*-value
	For socioeconomic status slope β_1				
Intercept, γ_{10}	308.31	6.96	44.28	1234	<.001
PTR, γ_{11}	-.71	1.29	-.55	1234	.58
	For English proficiency status slope β_2				
Intercept, γ_{20}	96.00	11.46	8.38	1234	<.001
PTR, γ_{21}	.31	2.05	.15	1234	.88
	For attendance slope β_3				
Intercept, γ_{30}	-137.98	103.61	-1.33	1234	.18
PTR, γ_{31}	-18.36	21.94	-.84	1234	.40
	For previous level of achievement slope β_3				
Intercept, γ_{30}	.92	.10	9.47	1234	<.001
PTR, γ_{31}	.02	.02	1.24	1234	.22

REFERENCES

Achilles, C. M. (n.d.). The difference between class size and pupil/teacher ratio Retrieved from http://www.heros-inc.org/pupil-teacher%20ratio.pdf

Achilles, C. M., Finn, J. D., Bain, H. P., Boyd-Zaharias, J., Johnston, J., Folger, J. . . . Nye, B. A. (1997). Using class size to reduce the equity gap. *Educational Leadership, 55*(4), 40–43.

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

Adcock, E. P., & Winkler, L. (1999, April). *Jump higher or else! Measuring school reconstitution.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Addonizio, M. F., & Phelps, J. L. (2000). Class size and student performance: A framework for policy analysis. *Journal of Education Finance, 26*(2), 135–156.

Allington, R. L., Stuetzel, H., & Shake, M. (1986). What is remedial reading? A descriptive study. *Reading Research and Instruction, 26*, 15–30.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Betts, J. R. & Shkolnik, J. L. (1999). The behavioral effects of variations in class size: The case of math teachers. *Educational Evaluation and Policy Analysis, 21*(2), 193–213.

Biddle, B. J., & Berliner, D. C. (2002). Small class size and its effects. *Educational Leadership, 59*(5), 12.

Biddle, B. J., & Berliner, D. C. (2003). *What research says about unequal funding for schools in america.* San Francisco, CA: West Ed.

Blatchford, P. (2005). A multi-method approach to the study of school class size differences. *International Journal of Social Research Methodology, 8*(3), 195–205.

Blatchford, P., Baines, E., Kutnick, P., & Martin, C. (2001). Classroom contexts: Connections between class size and within class grouping. *British Journal of Educational Psychology, 71*(2), 283.

Blatchford, P., Edmonds, S., & Martin, C. (2003). Class size, pupil attentiveness and peer relations. *British Journal of Educational Psychology, 73*(1), 15.

Blatchford, P., Russell, A., Bassett, P., Brown, P., & Martin, C. (2007). The effect of class size on the teaching of pupils aged 7–11 years. *School Effectiveness and School Improvement, 18*(2), 147–172.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Bohrnstedt, G. W., & Stecher, B. M. (Eds.). (1999). *Class size reduction in California : Early evaluation findings, 1996–1998.* Sacramento, CA: California Department of Education.

Bolander, S. F. (1973). Class size and levels of student motivation. *The Journal of Experimental Education, 42*(2), 12–17.

Bourke, S. (1986). How smaller is better: Some relationships between class size, teaching practices, and student achievement. *American Educational Research Journal, 23*(4), 558–571.

Burch, P. (2007). Educational policy and practice form the perspective of institutional theory: Crafting a wider lens. *Educational Researcher, 36*(2), 84–95.

Cahen, L., Filby, N., McCutcheon, G., & Kyle, D. (1983). *Class size and instruction*. New York, NY: Longman.

Caldas, S. J., & Bankston, C. L. (1997). Effect of school population socioeconomic status on individual academic achievement. *The Journal of Educational Research, 90*, 269–277.

California Teachers Association et al. v. Schwarzenegger et al., No. 05CS01165 (Sacramento County Superior Court 2006).

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *The Journal of Human Resources, 41*(4), 778–820.

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*(2), 237–248.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). Equality of educational opportunity. Washington, D.C.: U.S. Govt. Printing Office: U.S. Dept. of Health, Education, and Welfare, Office of Education.

Deutsch, F. M. (2003). How small classes benefit high school students. *NASSP Bulletin, 87*(635), 35–44.

Downer, J. T., Rimm-Kaufman, S. E., & Pianta, R. C. (2007). How do classroom conditions and children's risk for school problems contribute to children's behavioral engagement in learning? *School Psychology Review, 36*(3), 413–432.

Education Research Services. (1980). Class size research: A critique of recent meta-analyses. *The Phi Delta Kappan, 62*(4), 239–241.

Educational Testing Service. (March 2010). California standards tests technical report spring 2009 administration. Retrieved December from http://www.cde.ca.gov/ta/tg/sr/documents/csttechrpt09.pdf

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001a). Class size and student achievement. *Psychological Science*, *2*, 1–30.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001b). The class size controversy (cheri working paper #14). Retrieved October 5, 2009, from Cornell University, ILR School site http://digitalcommons.ilr.cornell .edu/workingpapers/25/

Evertson, C. M., & Burry, J. A. (1989). Capturing classroom context: The observation system as lens for assessment. *Journal of Personnel Evaluation in Education, 2*(4), 297–320. doi: 10.1007/BF00139647

Finn, J. D. (1998). Class size reduction and student at risk: What is know? What is next? (U.S. Dept. of Education, Trans.). Washington, DC: U.S. Department of Education.

Finn, J. D. (2002). Small classes in American schools: Research, practice, and politics. *Phi Delta Kappan, 83*(7), 551–560.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*(2), 97–109.

Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology, 97*(2), 214–223.

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "why's" of class size: Student behavior in small classes. *Review of Educational Research, 73*(3), 321–368.

Gardner, D., Larsen, Y., Baker, W., & Campbell, A. (1983). Nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the secretary of education. Washington, DC: Government Printing Office.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.

Glass, G. V., Cahen, L., Smith, M. L., & Filby, N. (1982). *School class size: Research and policy*. Beverly Hills: Sage.

Glass, G. V., & Smith, M. L. (1978). Meta-analysis of research on the relationship of class-size and achievement. The class size and instruction project. San Francisco, CA: Far West Laboratory.

Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2–16.

Graue, E., Hatch, K., Rao, K., & Oen, D. (2007). The wisdom of class-size reduction. *American Educational Research Journal, 44*(3), 670–700.

Graue, E., & Rauscher, E. (2009). Researcher perspectives on class size reduction. *Education Policy Analysis Archives, 17*(9), 1–26.

Graue, E., Rauscher, E., & Sherfinski, M. (2009). The synergy of class size reduction and classroom quality. *Elementary School Journal, 110*(2), 178–201.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). Interpreting research on school resources and student achievement: A rejoinder to hanushek. *Review of Educational Research, 66*(3), 411–416.

Grissmer, D. (1999). Conclusion: Class size effects: Assessing the evidence, its policy implications, and future research agenda. *Educational Evaluation and Policy Analysis, 21*(2), 231–248.

Grissmer, D. W., Kirby, S. N., Berends, M., & Williamson, S. (1994). *Student achievement and the changing American family.* Santa Monica, CA: Rand Corporation.

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data *American Economic Review, 61*(2), 280–288.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141–1177.

Hanushek, E. A. (1996). A more complete picture of school resource policies. *Review of Educational Research, 66*(3), 397–409.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2), 141–164.

Hanushek, E. A. (1998). The evidence on class size (Occasional Paper no. 98–1; Vol. 98). Rochester, NY: University of Rochester, W. Allen Wallis Institute of Political Economy.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*(2), 143–163.

Hanushek, E. A., & Rivkin, S. G. (1997). Understanding the twentieth-century growth in U.S. School spending. *The Journal of Human Resources, 32*, 35–68.

Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *The Review of Economics and Statistics, 78*(4), 611–627.

Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research, 43*(6), 387–425.

Hedges, L. V., & Greenwald, R. (1996). Haves times changed? The relationship between school resources and student performance. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 74–92). Washington, DC: The Brookings Institution.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). An exchange: Part I: Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5–14.

Hedges, L. V., & Stock, W. (1983). The effects of class size: An examination of rival hypotheses. *American Educational Research Journal, 20*(1), 63–85.

Hussar, W. J., & Bailey, T. M. (2008). Projections of education statistics to 2017. (36th ed.). Washington, DC: National Center for Education Statistics.

Illig, D. C. (1997). *Early implementation of the class size reduction initiative.* Sacramento, CA: California Research Bureau, California State Library.

Indiana Department of Education. (2010). Prime time: An overview. Retrieved Feb 28, 2010, from http://www.doe.in.gov/primetime/overview.html

Jepsen, C., & Rivkin, S. G. (2002). *Class size reduction, teacher quality, and academic achievement in California public elementary schools.* San Francisco, CA: Public Policy Institute of California.

Johnson, K. A. (2002). The downside to small class policies. *Educational Leadership, 59*(5), 27.

Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from project STAR. *The Elementary School Journal, 108*(4), 275–291.

Kreft, I., & DeLeeuw, J. (1998). *Introducing multilevel modeling.* London, England: Sage.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics, 114*(2), 497–532.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal, 113*(485), F34–F63.

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *The Economic Journal, 111*(468), 1–28.

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.

McGiverin, J., Gilman, D., & Tillitski, C. (1989). A meta-analysis of the relation between class size and achievement. *The Elementary School Journal, 90*(1), 47–56.

Mertler, C., & Vannatta, R. (2005). *Advanced and multivariate statistical methods* (3rd ed.). Glendale, CA: Pyrczak.

Mitchell, D. E., & Mitchell, R. E. (1999). The impact of california's class size reduction initiative on student achievement: Detailed findings from eight school districts. Riverside: University of California, California Educational Research Cooperative.

Mitchell, D. E., & Mitchell, R. E. (2003). The political economy of education policy: The case of class size reduction. *Peabody Journal of Education, 78*(4), 120–152.

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis, 21*(2), 165–177.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*(2), 113–127.

Nye, B. A., Hedges, L. V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*(1), 123–151.

Nye, B. A., Hedges, L. V., & Konstantopoulos, S. (2001). The long-term effects of small classes in early grades: Lasting benefits in mathematics achievement at grade 9. *The Journal of Experimental Education, 69*(3), 245–257.

Nye, B. A., Hedges, L. V., & Konstantopoulos, S. (2004). Do minorities experience larger lasting benefits from small classes? *The Journal of Educational Research, 98*(2), 94–100.

O'Connell, A. A., & McCoach, D. B. (2008). *Multilevel modeling of educational data*. Charlotte, NC: Information Age Publishing.

O'Connell, J. (2007). State schools chief jack o'connell announces list of schools selected to receive nearly $3 billion investment to accelerate student achievement. [press release] . Retrieved from http://www.cde.ca.gov/nr/ne /yr07/yr07rel62.asp

Odden, A., Picus, L., Archibald, S., Goetz, M., Mangan, M., & Aportela, A. (2007). Moving from good to great in Wisconsin:Funding schools adequately and doubling student performance. Madison: Consortium for Policy Research in Education. University of Wisconsin-Madison.

Pate-Bain, H., & Achilles, C. M. (1986). Interesting developments on class size. *The Phi Delta Kappan, 67*(9), 662–665.

Pate-Bain, H., Achilles, C. M., Boyd-Zaharias, J., & McKenna, B. (1992). Class size does make a difference. *The Phi Delta Kappan, 74*(3), 253–256.

Porwoll, P. J. (1978). Class size: A summary of research. Arlington, VA: ERS.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Rice, J. M. (1902). A test in arithmetic. *The Forum, 34*, 281–297.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Robinson, G. E. (1990). Synthesis of research on effect of class size. *Educational Leadership, 47*(7).

Robinson, G. E., & Wittebols, J. H. (1986). Class size research: A related cluster analysis for decision making. ERS research brief.

Rockoff, J. (2009). Field experiments in class size from the early twentieth century. *The Journal of Economic Perspectives, 23*(4), 211–230.

Schussler, D. L. (2009). Beyond content: How teachers manage classrooms to facilitate intellectual engagement for disengaged students. *Theory Into Practice, 48*(2), 114–121.

Schwartz, J., & Warren, P. (1997). Class size reduction *Policy brief.* Sacramento: California Legislature. Joint Legislative Budget Committee. Legislative Analyst's Office.

Shapson, S. M., Wright, E. N., Eason, G., & Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal, 17*(2), 141–152.

Shkolnik, J. L. (1997). *School resource allocation and the production of education.* Ph.D. 9804533, University of California, San Diego. Retrieved from http://libproxy.lib.csusb.edu/login?url=http://proquest.umi.com /pqdweb?did=736591401&Fmt=7&clientId=17861&RQT=309&VName=P QD

Slavin, R. E. (1984). Meta-analysis in education: How has it been used? *Educational Researcher, 13*(8), 6–15.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5–11.

Slavin, R. E. (1989a). Achievement effects of substantial reductions in class size. In R. Slavin (Ed.), *School and classroom organization* (pp. 247–258). Hillsdale, NJ: Lawrence Erlbaum Associates.

Slavin, R. E. (1989b). Class size and student achievement: Small effects of small classes. *Educational Psychologist, 24*(1), 99–110.

Slavin, R. E., Karweit, N. L., & Madden, N. A. (1989). *Effective programs for students at risk.* Needham Heights, MA: Allyn and Bacon.

Smith, P., Molnar, A., & Zahorik, J. (2003). Class-size reduction a fresh look at the data. *Educational Leadership, 61*(1), 72–74.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Stecher, B., Bohrnstedt, G., Kirst, M., McRobbie, J., & Williams, T. (2001). Class-size reduction in California. *Phi Delta Kappan, 82*(9), 670–674.

United States. Congress Senate. Committee on Health Education Labor and Pensions. (1999). *Esea, title vi and class size reduction : Hearing before the committee on health, education, labor, and pensions*. (0160593395). Washington, DC: U.S. Printing Office.

Wenglinsky, H. (1977). When money matters: How educational expenditures improve student performance and how they don't. A policy information perspective *Policy Issue Perspectives*. Princeton, NJ: Policy Information Center, Educational Testing Service.

Wenglinsky, H. (2000). How teaching matters: Bringing the classroom back into discussions of teacher quality *ETS Policy and Research Reports* Policy Information Center, Educational Testing Service.

WestEd. (Spring 2011). California health kids survey. Retrieved from http://chks.wested.org/resources/chks-hs-core-1011.pdf

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461–481.

Wossmann, L., & West, M. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review, 50*(3), 695–736.

Zahorik, J., Halbach, A., Ehrle, K., & Molnar, A. (2003). Teaching practices for smaller classes. [Article]. *Educational Leadership, 61*(1), 75–77.

Zahorik, J. A. (1999). Reducing class size leads to individualized instruction. *Educational Leadership, 57*(1), 50–53.