Instructional Assistants, Class Size and Academic Achievement
An Evaluation of Indiana's Prime Time

Daniel K. Lapsley
Ball State University

Katrina M. Daytner
Western Illinois University

With the Assistance of

Ken Kelley and Scott Maxwell

University of Notre Dame

A Report Submitted to the Indiana Department of Education
February, 2002

Project Prime Time
Teachers College 526
Ball State University
Muncie, IN 47306

Please address correspondence to Dr. Lapsley at the above address
Email: dklapsley@bsu.edu
Ph. 765.285.8500
Fax: 765.285.3653

Instructional Assistants, Class Size and Academic Achievement
An Evaluation of Indiana's Prime Time

EXECUTIVE SUMMARY

Overview

In 1999 the Indiana Department of Education funded a three-year program evaluation of the Prime Time funding formula. This effort, called Project Prime Time, was designed to provide critical information regarding program implementation and student achievement outcomes. Three phases were envisioned. Phase 1 was designed to provide descriptive information about Prime Time implementation on the basis of representative samples of teachers, principals and superintendents. The results of these survey studies are summarized in a final report that has been submitted to the Indiana Department of Education (Lapsley & Daytner, 2001a). In addition, the results of the teacher survey were reported at a national conference (Lapsley & Daytner, 2000b).

In this report we summarize the results of our quantitative analyses of third-grade ISTEP+ achievement data, obtained from the September, 2000 administration. We first report an extensive set of descriptive data on third-grade ISTEP+ achievement, broken down by class size, race, the presence or absence of Prime Time instructional assistants, and geographic region, among other variables. We next report the results of multi-level analyses using hierarchical linear modeling. Finally, we discuss the implications of these results for Indiana class size reduction initiatives, and for future program evaluation.

However, in order to put the present work in an appropriate theoretical and empirical context, we prefaced the summary of our quantitative findings with a selective synthesis of the major themes and controversies in the class size reduction literature. We also provide the first critical synthesis of the previous (mostly unpublished) literature on Prime Time.

The Class Size Reduction Literature

1 The Glass and Smith (1978) meta-analysis found that substantial reductions of class size was associated with increases in academic achievement.

2. Subsequent meta-analyses were interpreted as strong support for the benefits of substantial class size reduction on both academic and non-academic outcomes.

3. But a "best-evidence synthesis," and studies reported by the Educational Research Service, suggested that the effects of class size on achievement were weak, non-cumulative, or else moderated by numerous variables.

4. Of particular interest are findings that class size reduction has its strongest effects on

reading and mathematics in the early primary grades, and particularly with ethno-racial minority and lower-income students.

5. Subsequent debate has focused on possible theoretical explanations of class size reduction effects, and on whether there are viable mediating mechanisms. There is some interest in the *fixed instructional resource* and *teacher adjustment to student* explanation of CSR effects.

6. Much of the positive evidence for class size reduction comes from Tennessee's Project Star, which was a state-funded experimental test of the educational benefits of reduced class size. The results of this "longitudinal experiment" provide striking documentation of the beneficial effects of small class size on student achievement. Subsequent studies report "lasting benefits" of class size reduction that endure beyond the early primary grades.

7. Numerous other states have adopted class size reduction initiatives. Although in most cases program evaluation is still incomplete, the available evidence shows that class size reduction has encouraging but mixed effects on student outcomes.

8. Presumably, students do better in smaller classes because teachers alter instructional practices to capitalize on the educational opportunities that smaller classes afford. The results of our teacher and principal surveys do indicate that teachers with Prime Time instructional assistants use smaller grouping structures more often than teachers without assistants, and use more learning centers, and more educational technology. This indicates that the presence of Prime Time instructional assistants appear to have a favorable influence on teacher instructional practices.

9. But not just *how* class size reduction works, but *whether* it works. The "educational production function" literature doubts that class size reduction has any positive influence on student achievement, on the basis of "vote-counting" methods for synthesizing the literature.

10. This is strongly disputed by researchers using meta-analytic techniques to summarize the literature (rather than vote-counting). They conclude that "a broad range of school inputs are positively related to student outcomes, and that the magnitude of the effects are sufficiently large to suggest that moderate increases in spending may be associated with significant increases in achievement" (Greenwald et al., 1996a, p. 362).

11. A recent study (Nye et al., 2000), using hierarchical linear modeling (HLM) similarly concluded that small classes conveyed a statistically significant advantage in reading and mathematics "at every grade level from kindergarten through Grade three...suggesting that small classes benefit students of all types in all kinds of schools" (p. 147).

12. In addition, Nye et al (2000) showed that

- The small class effect is greater for students who have experienced more years in small classrooms.
- The small class effects estimated with HLM are similar to those predicted from meta-analyses of small-scale experiments.
- For example, on the basis of this analysis of STAR data, a reduction of class size from 25 to 15 could be expected to yield an increase in achievement in the range of .13 - .30 standard deviations. The increase reported in the Glass and Smith (1979) meta-analysis was .215 standard deviations.
- Moreover, the small class effect reported in this analysis of STAR is comparable to the effects sizes reported in previous meta-analyses of the production function literature (see Greenwald et al., 1996a).

The Prime Time Literature

1. Although Prime Time began as an exercise in class size reduction, it has gradually evolved into a program that attempts to lower pupil-teacher ratio, largely by employing teachers and instructional assistants. The "Prime Time strategy" then, takes a very different approach than Project STAR, which is a pure class size reduction effort.

2. Prime Time poses special challenges to program evaluation. First, because Prime Time was designed as a demonstration project, insufficient attention was paid to the sort of controls that would permit meaningful and informative evaluation. Second, different achievement tests have been used in different schools, and, over time, different versions of the state mandated test have been used, making it difficult to evaluate the effectiveness of class size reduction using a common metric. Third, Prime Time may be confounded in some school corporations with other educational programs and initiatives (such as Title I).

3. Although the initial evaluation of the pilot implementation was quite favorable, the conclusions of the study should be qualified in light of six important limitations (see page 20).

4. The McDaniel (1985) report was an extensive summary of a large ethnographic study of Prime Time classrooms, which were selected on the basis of administrator nominations. McDaniel (1985, p. 25) draws this portrait of Prime Time classrooms:

"The picture of PRIME TIME classrooms which emerges from the tabulations is of classes characterized by flexible seating patterns and use of space, abundant displays of student work, and one or more learning centers. Within these rooms, the children work in an atmosphere of relaxed control. Paradoxically,...it appears evident that children enjoy the most freedom in rooms where class procedures are well-planned, organized and structured. Almost everywhere, children are on-task and a feeling of industriousness pervades the room."

5. David Gilman and his colleagues at Indiana State University have conducted numerous studies of Prime Time, most of them confined to the North Gibson School Corporation. Although early studies tended to show very striking benefits in Prime Time classrooms, later studies drew quite opposite conclusions, particularly at third-grade. Gilman would later argue that class size reduction conveys little benefits, and that program evaluation of Prime Time, and of Project STAR, has been corrupted by political pressure, bureaucratic resistance, and scientific incompetence.

6. Mueller, Chase and Walden (1988) reported strong support for Prime Time among teachers, principals and parents; positive achievement effects, particularly at first-grade; and observational data that indicated that teachers and aides engaged in instructional practices appropriate for more favorable pupil-teacher ratio.

7. It is clear that Prime Time is still a program that has been insufficiently evaluated over the years. The literature occasionally suggests that lower PTR, or smaller classes, pays off in terms of student academic outcomes, at least in first- and perhaps second-grade, although many of these studies are also hard to evaluate due to methodological concerns. Moreover, the majority of evaluation studies have been conducted in just a single school corporation.

Summary of Empirical Findings

1. An inspection of mean NCE scores indicated that students in classes with Prime Time instructional assistants reported higher achievement scores than students in classrooms without such assistants in reading, language, mathematics, and the total composite score.

2. This Prime Time advantage was evident in every geographic category. That is, achievement scores in reading, language, mathematics and the total composite were higher in Prime Time assisted classrooms than in unassisted classrooms in urban, suburban, rural and township schools, although the advantage was particularly pronounced in suburban and township schools.

3. Indiana school corporations were selected using a stratified (geographic category and educational service region) random cluster sampling procedure. This sampling procedure yielded 61 corporations, a total of 10, 927 students in 163 schools and 573 classrooms, of which 190 had Prime Time instructional assistants (Prime Time assisted), 376 had no Prime Time instructional assistant (unassisted), and 7 were assisted by other means.

4. The sample was nearly equivalent by gender, with 5425 girls, 5457 boys. The sample was predominantly white/Caucasian (84%, N = 9207), followed by black/African-American

(9.1%, N = 995), Hispanic/Latino-American (3.2%, N = 318) and Asian-American (.6%, N = 62).

5. The modal class size in this sample was 21. Approximately 10 % of third-grade pupils were in classes of 17 or fewer; nearly 15% were in classes of 18 or fewer. Moreover, about 37% of Hoosier third-graders were in classrooms of 20 or fewer. This leaves about 63% of pupils in classrooms with enrollment over 20, with approximately 15 % of students in classes of 25 or larger.

6. The ethno-racial background of students was not equally distributed across the range of class size. A much larger percentage of African-American and Hispanic children were in small classes (12-17) than were White children. For example, 16% of black children and almost 24% of Hispanic children were in classes of 12-17, compared to 8.7% white and 6.45% Asian children. This might suggest a greater minority participation in smaller compensatory classes, although subsequent analyses casts some doubt on this interpretation.

7. Although white pupils are approximately equally distributed among urban, suburban, township and rural settings, the non-white population, particularly black and Hispanic pupils, are overwhelmingly located in urban school corporations.

8. Mean achievement scores also tended to increase as class size got larger. This might be expected if at least some small classes in this sample were formed for students in compensatory or remedial programs. Although this question could not be addressed definitively, some aspects of the present data would not support this interpretation.

9. The descriptive analyses appear to demonstrate a clear achievement advantage in Prime Time assisted classrooms. The apparent benefits of Prime Time instructional assistants is not associated with class size. Indeed, the descriptive analyses do not show that achievement is better in classrooms with smaller enrollment.

10. HLM analyses showed that patterns of student achievement were significantly affected by SES and by race. Students in schools in higher SES categories, and white students, tended to report higher ISTEP+ (NCE) scores. For example, there is a 6.9 point differential in the prediction of white (predicted score of 62.12) and black (predicted score of 55.22) students' NCE total composite score.

11. Class size was a significant predictor of mathematics achievement and the NCE total composite score. In both cases, higher achievement was associated with larger class size. For example, adding one student to a class increases NCE total composite achievement .35 points (or, alternatively, adding 10 students to a class raises average achievement 3.5 points).

12. The presence or absence of a Prime Time assistant interacted with SES in the analysis of reading, language and the NCE total composite score. In all cases higher achievement was associated with higher SES and the presence of an assistant. In no analysis did the presence or absence of a Prime Time assistant yield a significant main effect.

13. In addition, better mathematics achievement was associated with higher SES and larger classrooms. Better reading achievement was associated with higher SES and white pupils.

14. The influence of class size on achievement varied by racial category. African-American students, for example, showed better achievement in smaller classes, while white (and Hispanic) students show better achievement in larger classes.

15. The HLM analysis of pupil-teacher ratio is strikingly similar to the results for class enrollment: better achievement was associated with larger pupil-teacher ratio. For example, a unit increase in pupil-teacher ratio was associated with an increase of .38 points in NCE total composite achievement..

16. An important caveat must attend the interpretation of the present data: students sit for the ISTEP+ in the second week of September of the Fall term. This means that a Prime Time instructional assistant would be active in a classroom for just a few weeks prior to the administration of ISTEP+. This would not allow sufficient time for the benefits of an instructional assistant to be evident within a classroom, given the critical importance of duration noted by Finn et al. (2000).

Conclusions and Recommendations

1. The descriptive data show that the presence of a Prime Time instructional assistant is associated with clear advantages with respect to ISTEP achievement at third-grade.

2. The benefits of a Prime Time instructional assistant appear to vary as a function of certain demographic characteristics of school, such as SES. Future research should examine, using ethnographic and survey techniques, how instructional assistants are used across a range of demographic categories.

3. Small class size was not associated with better achievement, except for reading and composite achievement of minority pupils. Future research should compare the instructional practices of teachers and assistants in larger and smaller classes.

4 Future research undertake a longitudinal examination of the influence of Prime Time aides (and pupil-teacher ratio) on student achievement, in order to more fully assess the role of the duration variable. This would require the utilization of standardized assessments at second grade in classrooms with and without Prime Time aides, with assessments administered in the fall and spring terms. These students should then be followed into

third-grade, and their ISTEP+ scores be included in the analysis.

5. Moreover, the benefits of instructional assistants might be particularly evident at preschool and first-grade. Future research should consider using assessments other than the state-mandated test and similar standardized assessments, such as local tests of mastery of state proficiencies, to assess the benefits of class size, pupil-teacher ratio, and the presence or absence of Prime Time instructional assistant.

TABLE OF CONTENTS

LIST OF TABLES

Instructional Assistants, Class Size and Academic Achievement
An Evaluation of Indiana's Prime Time

Chapter 1
Background to the Report

Funding Objectives. Prime Time is a funding formula that is designed to reduce class size (teacher-pupil ratio) in the early primary grades (kindergarten through third grade). The program is predicated on the assumption that reduced class size, or a reduction in the pupil-teacher ratio, will result in better student outcomes. To that end, and since the phase-in implementation in 1984-85, funding has been provided to local school corporations to hire additional teachers or instructional assistants in order to assist school corporations in moving toward the target pupil-teacher ratio. The initial goal of funding was to achieve a *corporation* average of 18 students per teacher (18:1) in kindergarten and first-grade, and an average of 20 students per teacher (20:1) in second- and third-grade. However, during the 1999 Legislative Session, the funding formula was changed. Under the new formula each school corporation has a target pupil-teacher *kindergarten through third-grade ratio* that ranges from 15:1 to 18:1, depending upon the corporation's at-risk index and the amount of tuition support the corporation receives. Moreover, the monies are included in the basic grant that the corporation receives from the state. These monies are still to be used, of course, for hiring teachers or instructional assistants in order to reach these target ratios. Indiana's Prime Time, along with Tennessee's Project STAR program, are often held out as a national model of innovative educational programming (Pate-Bain & Achilles, 1986).

The 2001-2002 school year marks the eighteenth year of implementation of Prime Time. However, during this time, there have been few evaluation studies of Prime Time, and none that have been conducted with representative samples using the current state-mandated standardized achievement test (ISTEP+) as the outcome measure. Moreover, although there is an emerging consensus regarding the educational benefits of reduced class size, particularly in the early primary grades (e.g., Department of Education, May, 1998; Finn, 1998; Robinson, 1990), the evidence is not entirely unequivocal (e.g., Hanushek, 1994, 1996, 1997; Robinson, 1990; Tomlinson, 1990). In addition, the few studies that have examined pupil-teacher ratio and the influence of teacher aides on student achievement outcomes are not encouraging (e.g., Finn, Gerber, Farber & Achilles, 2000). Hence, an evaluation of Prime Time not only has clear policy and funding implications at the state level, but it also promises to make an important contribution to the literature on class size reduction interventions.

Project Prime Time. To this end the Indiana Department of Education funded a three-phase program evaluation of Prime Time, beginning in 1999. This effort, called Project Prime Time, was designed to provide critical information concerning program implementation and student achievement outcomes.

In the first phase we conducted four studies of Indiana educators. The first study was a full population survey of elementary school building principals that was designed to ascertain the extent of off-year testing, the kinds of tests that are used throughout the state, the timing of test administration during the school year, and the availability of test scores for research purposes, among other questions.

We then developed extensive teacher, principal and superintendent surveys for subsequent studies. The teacher survey was designed to better understand the implementation of Prime Time at the classroom level, and included questions regarding the training and educational background of instructional assistants, the extent of their participation in classroom activities, and, importantly, whether the presence of instructional assistants altered teacher instructional practices. The principal and superintendent surveys were designed to obtain information about the implementation of Prime Time at the building and corporation level, respectively. In addition, each survey solicited open-ended responses from teachers and administrators regarding the strengths and weaknesses of Prime Time, and on ways that the funding formula could be better administered. These surveys were distributed to representative samples of teachers, principles and superintendents, using a stratified cluster sampling procedure.

The results of these survey studies are summarized in a final report that has been submitted to Indiana Department of Education (Lapsley & Daytner, 2001a). In addition, the results of the teacher survey was reported at a national conference (Lapsley & Daytner, 2000b).

In the second phase we attempted to assess the relationship between pupil-teacher ratio, class size and third-grade performance on the state-mandated standardized achievement test (ISTEP+). A stratified cluster sampling procedure identified a sample of nearly 11,000 third-grade students. This document is a report of the findings of that study. We note in passing that a third-phase of Project Prime Time was anticipated. The third-phase of the evaluation would have attempted to provide rich observational data from purposively selected Prime Time and non-Prime Time assisted classrooms. This study was not attempted.

The Present Report. In this report we summarize the results of our quantitative analyses of third-grade ISTEP+ achievement data, obtained from the September, 2000 administration. We first report an extensive set of descriptive data on third-grade ISTEP+ achievement, broken down by class size, race, the presence or absence of Prime Time instructional assistants, and geographic region, among other variables. We next report the results of multi-level analyses using hierarchical linear modeling. Finally, we discuss the implications of these results for Indiana class size reduction initiatives, and for future program evaluation.

However, in order to put the present work in an appropriate theoretical and empirical context, we preface the summary of our quantitative findings with a selective synthesis of the major themes and controversies in the class size reduction literature.

outcomes in the early primary grades. Second, these effects are probably moderated by pupil, teacher, and contextual variables. As one Project STAR research team put it:

> There is a sense is which class size is a psychologically empty concept. We are confident that reducing the number of students in a class does not impact performance directly but instead affects the *processes* that mediate achievement (Finn, Fulton, Zaharias & Nye, 1989, p. 83)

Mediating Mechanisms. This suggests that reduced class size may be necessary, but not sufficient, for improving academic achievement (Varble, 1990). One set of mediating processes that link class size to student outcomes is presumably teacher instructional practices. Indeed, it is generally agreed that class size reduction "is unlikely to have a direct and beneficial effect on student achievement if teaching practices do not change" (Evertson & Randolph, 1989, p. 103).

Mitchell and Beach (1990) reviewed four possible mechanisms that link smaller classes to student achievement. The *classroom overhead* theory suggests that large classes demand more "overhead", that is, require an increase in non-instructional activities, such as passing out and collecting papers, taking attendance, and monitoring compliance and misbehavior.

The *student interaction time* model assumes that smaller classrooms multiply opportunities for student interactions with the teacher, and with each other. When more students are added to a class, student interaction time diminishes accordingly, which detracts from overall learning.

The *teacher adjustment to student ability* model argues that large classes diminishes achievement because teachers "pitch it low" in order to reach the least able students. In large classes, even in classes that are randomly assembled, there are more likely to be less able students than in more homogenous smaller classes. Hence, with more low ability students, teachers of large classes will tend to slow the pace and limit the depth of instruction.

Finally, the *fixed instructional resource* model assumes that a teacher's instructional attention is a fixed resource that is spread thinner with each new pupil added to a class. Consequently, as class size increases, a pupil receives a smaller portion of the teacher's instructional attention. In a class of 10 pupils, for example, each student can expect 1/10th of the teacher's fixed instructional resources. In a class of 30, each student can expect 1/30th of the teacher's instructional attention. Hence, the loss of instructional attention is more keenly felt in a smaller class than in a larger class.

Which mechanism is better supported by research? The Glass and Smith (1978) meta-analysis suggests that the *fixed instructional resource* theory best fits the data, followed closely by the *teacher adjustment to student* model. Hence, according to Mitchell and Beach (1990, p. 3):

report, for example, the Nevada State Department of Education concluded that although principals, teachers and parents were positive about the class size reduction effort, the program had only a marginal (statistically non-significant) influence on student math and reading achievement. Similar mixed results were reported in a 1995 evaluation. For example, at second-grade, smaller classrooms (15 students) were an advantage for mathematics scores, but not for reading scores. What's more, a longitudinal analysis of gain scores showed that mathematics and reading scores were higher for third-graders who attended second-grade in *larger* classrooms (> 15:1) than for those who attended second-grade in small classrooms. Moreover, there were a number of factors that were "overwhelmingly more important" (Sturm, 1997, p. 9) than class size in predicting pupil performance, including special education status, ESL status, free lunch eligibility and class configuration. Even with these significant predictors 90% of the variance in achievement scores was still left unexplained.

California. In 1996-97 California initiated its Class Size Reduction Program. The aim of this program is to reduce student-teacher ratio to 20:1 in kindergarten through grade 3. Nearly $1 billion was allocated in 96-97 to support the reduction of class size, fund facilities, and train teachers. This was increased to $1.5 billion in 1997-98. In the first year of implementation 18,400 new classes were added. By 1997-98, over 97% of the eligible school districts were participating in the program, with 99% of all first-graders and 95% of all second-graders enrolled in smaller size classrooms. Approximately 18,400 new teachers were hired, half of whom were inexperienced; 30% had no credentials, and 21% were hired on emergency permits (WestEd Policy Brief, August, 1998; *DOE*, State Initiatives, 1998). State-wide evaluation of program effectiveness is not yet available, although smaller scale and local studies have been attempted.

One study (PACE-WestEd, 1998) selected a stratified random sample of 12 school districts, and then interviewed various district officials and teachers *concerning their perceptions* of the effect of the Class Size Reduction Program (CSR) on special populations; on staff development; on classroom practices; on parental involvement; and on facilities and space.

With respect to special populations, most districts and schools reported a number of advantages for Limited-English-Proficient (LEP) students, although nearly 40% of newly-hired teachers have no state teaching credential, and more than 50% are not certified to teach LEP students. Special education specialists also reported that smaller class sizes were more conducive for inclusion programs and that 'mainstreaming' was easier to accomplish. But some (mostly urban) districts have reported a movement of special education teachers to general education, and a loss of space, as special education classrooms are pressed into general education service.

With respect to staff development, 70% of teachers noted that in-service training addressed smaller class size, yet district documents related to professional development rarely made reference to class size. Moreover, depleted substitute pools made it difficult for teachers to avail themselves of training. Most teachers reported that the CSR program allowed them to pick up the instructional pace, and to cover topics in greater depth. A vast majority (84%)

report fewer student disruptions, and greater student motivation. But nearly 60% still report using whole-classroom instruction, and one-third report no change at all in their instructional techniques as a result of smaller class size. Although most parents are "enthusiastic" and "supportive" of the CSR program, it has not increased parental involvement. Finally, most districts reported shortages of portable classroom and other facilities constraints.

The San Diego County Office of Education has developed a useful set of class size reduction surveys for parents, teachers and principals, for purposes of program evaluation. These "perception" surveys were recently distributed to parents, teachers (grades K-2) and school site administrators in the Coronado Unified School District (San Diego County). Nearly 90% of parents (with a 55% return rate) thought that CSR improved their child's learning of reading, writing and mathematics. Teachers were unanimous (with a 100% return rate!) that CSR improved student learning (California School News, 1997) Although these data are encouraging, direct assessment of student outcomes, in addition to "perception" data, will also be necessary in the evaluation of the CSR program.

Finally, a recent study was reported comparing teacher practices in class size reduced and non-reduced classrooms (Stasz and Stecher, 2000). In general, the authors found that instruction of mathematics and language arts did not look much different in smaller and larger classes, indicating that teachers tend not to alter their instructional practices to capitalize on smaller class size. We discuss this study in more detail below.

<u>Wisconsin,</u>. The Wisconsin class size reduction program is called Student Achievement Guarantee in Education (or, SAGE). It was initiated in 1996-97 to improve academic outcomes in schools that had a high proportion of students living in poverty. The program requires participating schools to implement four interventions: 1) reduce pupil-teacher within a classroom to 15 students per teacher; 2) provide before- and after-school programs; 3) develop rigorous curriculum; and 4) create a system of staff development and professional accountability.

The reduction of pupil-teacher ratio was the first intervention to be uniformly and immediately implemented. However, this was accomplished in a number of ways. For the most part, the target pupil-teacher ratio was accomplished by creating "small class size" of 15 students to one teacher within a classroom. (Note that Project STAR would refer to this as a "small class"). But other configurations were possible, too. For example, some classrooms of 30 pupils would be assigned two teachers (what SAGE called a "two teacher team" and what STAR would regard as a "large class" with a pupil-teacher ratio of 15:1). A "floating teacher" arrangement would involve a classroom of 30 students, except during critical instruction in mathematics, reading, and language arts, when the class would be joined by a second teacher (thereby effecting the target pupil-teacher ratio for only part of the instructional day). A "shared space" arrangement was also possible, whereby a single classroom included a temporary partition that created two teaching spaces, each with one teacher and 15 pupils. Other arrangements were also possible.

In the first annual evaluation, first-grade students in SAGE classrooms outperformed a matched comparison sample of first-graders on various areas of the Comprehensive Test of Basic Skills (*DOE*, State Initiatives, 1998).

The results of the first two years of SAGE for first grade were recently reported by Molnar, Smith, Zahorik, Palmer, Halbach and Ehrle (1999), and we report on this effort in some detail insofar as we adopt a similar analytic strategy in our evaluation of Prime Time.

This study utilized a quasi-experimental comparative change design. The Comprehensive Test of Basic Skills (CTBS) complete battery (Terra Nova edition, Level 10) was administered to first-grade students in SAGE schools and comparison schools in October 1996 and May 1997 (i.e., near the beginning and end of the school year). In 1997-98, first-grade students were administered Level 10 in October and Level 11 in May. Second-graders were administered Level 12. Student demographic profiles, the organization of the classroom and teacher questionnaire and interview data, were also acquired. Individual achievement data were analyzed using ordinary least squares linear regression (controlling for student individual differences). Classroom effects, including pupil-teacher ratio, were analyzed using hierarchical linear models (HLM). Only first-grade results were reported.

The linear regression results suggested that the number of days absent, and a poverty index (school lunch eligibility), were significant predictors of achievement on the CTBS (the pre-test score was also a significant predictor). However, attending a SAGE school was also a significant predictor of posttest scores on the CTBS. The SAGE influence on student achievement ranged from 3 to 7 points, depending on the CTBS subtest. Moreover, African-American pupils in SAGE schools "scored significantly higher than African-American comparison school students on every subtest and had significantly higher total scale scores" (p. 170). Although the gap between White and African-American students remained statistically significant (favoring White pupils) in both SAGE and comparison schools, African-American students in SAGE schools showed more gains than did White SAGE students, "closing the achievement gap" (p. 170).

Hierarchical linear modeling allows one to analyze nested, multi-level data structures (e.g, individual level data nested within classroom; Bryk, & Raudenbush, 1992). At issue here is whether results would differ when Level 1 (individual student) and Level 2 (classroom) variables are included. Level 1 variables were pretest achievement scores and the SES index (school lunch eligibility). Three different Level 2 models were tested: one with actual pupil-teacher ratio; the second with pupil-teacher ratio and a SAGE participation variable (dummy coded), and the third includes actual pupil teacher ratio and class SES.

The Level 1 results showed that test scores were significantly predicted by SES (lower SES was associated with lower test scores) and by the pretest scores. Classroom pupil-teacher ratio was a significant predictor when added at Level 2, to the extent that an increase of one in pupils per teacher produced a .29 to 1.17 loss in average posttest performance. Participation in

SAGE, when added along with pupil-teacher ratio at Level 2, did not produce a significant effect over and above the contribution of pupil-teacher ratio alone, indicating, according to the authors, "that the other SAGE interventions (e.g., rigorous curriculum, lighted schoolhouse, and staff development) do not have a significant achievement in SAGE classrooms" (p. 172).

The third model showed that class SES had a large effect: a 1-point class average gain in SES (measured on a 3-point scale) predicted a 10-point and 21-point gain on average posttest scores (depending on the subtest). The interview data showed that teachers attributed more individualized instruction and class participation to smaller class size, and that smaller classes did *not* necessarily result in more student-centered classroom instruction.

These results document, then, a significant achievement benefit to lower pupil-teacher ratio. It also demonstrates the pervasive and powerful influence of socio-economic factors on measures of student academic achievement.

Two further points should be made. First, it should be noted that the "SAGE strategy" attempts to bring about a lower pupil-teacher ratio by class size reduction (among other arrangements), which is considered an unfortunate confound by some researchers (e.g., Achilles, 1997; Achilles, n.d.). Hence, it is not clear whether the salutary benefits noted in this study are the result of smaller classes achieved by one teacher and 15 pupils (a "small class effect") or the result of aides or other teachers assisting in the classroom in order to reach the target pupil-teacher ratio of 15:1 (a "pupil-teacher ratio effect"). Second, the authors suggest that SAGE interventions other than reducing pupil-teacher ratio did not have a significant effect in this study, since SAGE participation did not significantly predict achievement over and above the contribution of pupil-teacher ratio. This conclusion, however, must be considered tentative until the impact of the other interventions can be more adequately assessed.

Texas, Utah Indeed, it has been suggested that class size reduction strategies are most effective when they are part of a comprehensive framework of program initiatives (WestEd Policy Brief, 1998). One study of 15 schools in Austin, Texas showed, for example, that schools that demonstrated the most impressive gains in student achievement were ones that combined class size reduction with other initiatives, such as implementing new curricula and teaching methods, and increasing individualized attention, parental involvement and health services (Murnane & Levy. 1996; WestEd Policy Brief, 1998).

Similarly, a study of five school districts in Utah showed that the most successful schools combined class size reduction with professional development and attempts to improve instruction and instructional resources (Evans-Stout, Fleming, Johnson, Jr., Erlinger, Gray, Merrill, Parket, Roberts, & Stewart, 1997). The theme here is that reform initiatives are most successful when the effort is to improve achievement through broad based interventions that have multiple components, "rather than just getting the numbers down" (WestEd Policy Brief, 1998, p. 6).

Florida. Although Florida does not have a class size reduction program, the Office of

Policy Research of the Florida Department of Education has conducted a statewide analysis of the relationship between class size(and school size) and student achievement. The average class data from the 1993-94 Florida School Report was used. Class size data was divided into quartiles. Grades were grouped into three levels: K-5, grades 6-8, and grades 9-12. The results were not encouraging. According to the report, "The analysis did not support the popular belief that there is a relationship between class size and student achievement" (p. 12). The report does note, however, that school-level data may obscure important within-school variations that might favorably influence achievement. It should also be noted that grouping the early primary grades into a K-5 block may also have obscured a small class effect on achievement.

Mediating Mechanisms Revisited

There is clear evidence in favor of the educational benefits of class size reduction, although there are also equivocal findings in the literature as well. As noted previously, one controversy that surrounds the evaluation of Tennessee's STAR program concerns whether the benefits of class size reduction accumulate or endure over time from first-grade to third-grade, or whether they "wash out"in subsequent grades. In a Project STAR's Final Executive Summary, Word and her colleagues (Word et al., 1990) suggested that STAR's kindergarten results "showed a definite advantage for students in small classes in achievement and no significant advantage for the use of a teacher aide" (p. 13). Moreover, it was reported that at the end of first grade, Project STAR children were outperforming children in regular classes and in regular classes with an aide. According to the authors, "This pattern continued in grades two and three" (p. 13).

Other commentators have suggested, however, that class size effects were only evident at first grade, and did *not* accumulate over the remaining primary grades (Tomlinson, 1990). Indeed, Tomlinson (1990) interprets this finding as an indication that the positive first grade results were more likely the result of *Hawthorne effects*, given the fact that teachers were recruited as volunteers and were likely motivated, therefore, to "prove one of their cherished beliefs: Small classes are better than larger ones" (p. 19). Although Tennessee's Lasting Benefits study does document a carry-over effect up to the middle-school years, the effect sizes are much attenuated over time. Indeed, Hanushek (1997, p. 153) has argued that although small class size may be decisive in kindergarten, "class size reductions after kindergarten have little potential impact on achievement" and that adding resources to effect class size reduction "appear to add nothing to student performance."

Instructional Practices. It should also be noted, finally, that the positive achievement results noted in various Project STAR reports are typically attributed to class size reduction, *simpliciter*, rather than to altered instructional practices of teachers or other mediating mechanisms noted, for example, by Mitchell and Beach (1990)..

Presumably, when class size reduction is associated with improvements in achievement and non-achievement outcomes, it is because the smaller class provides an opportunity to alter classroom practices. Indeed, as Mitchell and Beach (1990) point out, policy makers should be as concerned with how classrooms are organized as they are with how they are staffed. But the extent to which teachers adjust their instruction in smaller classes is not entirely clear. In the PACE-WestEd study of twelve California districts, for example, 65% of the teachers reported still using whole-classroom instruction in their smaller classes, and one-third of the teachers reported making no change whatsoever in their instructional practices (but rather doing *more* of what they have always done, and doing it, somehow, *better*).

RAND Study. This was demonstrated in a recent RAND study of California's Class Size Reduction CSR) initiative. In this study Stasz and Stecher (2000) compared instructional practices in reduced size ($M_N = 20$) and non-reduced size ($M_N = 29$) third-grade classrooms. They found that teaching practices evident in both reduced and non-reduced classrooms to be "quite similar" (p. 313). For example, topic coverage in mathematics and language arts was similar in reduced and non-reduced classrooms, as was the total amount of time devoted to instruction on these topics. The amount of homework assigned to students was similar, as was grouping practices and time spent giving students individual attention. Moreover, Stasz and Stecher (2000, p. 318) reported that teachers in the reduced and non-reduced classes "did not differ significantly in terms of teacher background, student demographic factors, professional development or instructional support." Differences were evident, however, in three areas. Teachers in reduced size classes reported fewer discipline problems; they spent more time discussing a student's personal problems, and more time providing individual attention to the poorest readers.

Other Research. Similarly, Robinson (1990; also, Kickbusch, 1996) concluded, on the basis of his "clustered" synthesis of the class size reduction literature, that teachers do not routinely exploit the opportunities provided by smaller classes. In addition, he reports evidence that the self-perception of what teachers claim they do in the classroom (e.g., "provide more individualized attention") does not always square with what they are observed to do. In a two-year Toronto study, for example, a vast majority of teachers reported that they changed their instructional practices as a result of class size reduction, a claim that was nonetheless difficult to document upon observation of their classrooms (Shapson, Wright, Eason & Fitzgerald, 1980). Even in Project STAR classrooms, it was observed that teachers taught much the same way, whether they had larger or smaller classrooms (Odden, 1990). As Shapson et al. (1980, p. 151) concluded, class size "makes a large difference to teachers but little difference to the students or to the instructional methods used."

Education Production Functions

Hence when educational benefits are noted for class size reduction, it is often difficult to attribute the positive results to anything teachers do (which is why it is class size, *simpliciter*, that is touted, for example, by Project STAR). Moreover, as noted earlier, the putative benefits of class size reduction are disputed by economic models of education production functions.

The production function approach is a research strategy primarily associated with economists. It utilizes multiple regression procedures to estimate the functional relationship between certain educational "outputs," such as student achievement, and various "input" variables, while controlling for student characteristics (such as social class or prior achievement). This approach adopts an economic or factory model in the sense that schools are viewed as producing achievement much the way a factory produces a commodity. A successful factory operation is able to specify the functional relationship between the production of a commodity (the "output") and variables such as raw materials and labor costs (the "input").

Similarly, a successful educational operation is one where a precise functional relationship can be established between student achievement (treated as the output commodity produced the by school-factory) and input variables. The input variables typically include student variables, such as race, and school variables, such as PTR, class size, and expenditures-per pupil. The input-output relationship (the regression equation) is called a "production function."

Hanushek's Synthesis. A synthesis of the education production function literature typically shows that SES has a strong influence on academic achievement, but not school inputs, such as class size, PTR, and expenditures-per-pupil (Hanushek, 1989; 1996, 1997). Consequently, Hanushek and his colleagues (Hanushek and others, 1994) conclude that increasing expenditures to lower the size of classes, for example, or hiring teachers and paraprofessionals to bring about lower PTR, or raising the rate of expenditures-per-pupil more generally, is to "throw money at the problem." In their view, the amount of monies required for class size reduction (by any means) is not only exorbitant and unrealistic, but also empirically unjustified. Instead school financing is better directed towards increasing educational "efficiency" (also a matter of economics and factories) by establishing a system of incentives and sanctions for teachers and administrators, along with other mechanisms of accountability (Hanushek & Jorgenson, 1996).

We note in passing that student attendance is an "input" variable that is positively related to achievement outcomes. In an examination of standard production functions in Baltimore elementary schools, Lamdin (1996) found little support for pupil-teacher ratio and expenditures-per-pupil, findings that are common to this literature, but significant support for school attendance (and for SES). He concluded that "the average level of attendance at a school does have a positive influence on student performance" (p. 160), and that a prima facie case could be made for targeting resources to increase attendance rates provided that attendance is amenable to cost effective intervention.

Hedges and Meta-Analysis. In a series of articles Hedges and his colleagues have strongly challenged the received view among production function researchers that school resources are unrelated to student achievement (e.g., Greenwald, Hedges & Laine, 1996a; Greenwald, Hedges & Laine, 1996b; Hedges, Laine & Greenwald, 1994). In particular, this research team dismisses Hanushek's method of synthesizing the production function evidence. Hanushek's "vote-counting" method categorizes findings as to direction and statistical significance. That is, after a catalogue of findings, one simply counts up the number of statistically significant positive and negative findings for a particular production function. And, since econometric modeling of production functions so often finds statistically non-significant effects for class size, Hanushek (1989) concludes that investing resources in class size reduction initiatives is ill-advised. According to Hedges and Olkin, 1980), however, vote counting is an insensitive method for summarizing a research literature, and is not considered "best practice" in research synthesis.

In contrast to vote counting Hedges and his colleagues synthesize the literature using meta-analysis (Hedges & Olkin, 1985), which focuses on the magnitude of the regression coefficients regardless of their statistical significance. Indeed, after re-analyzing Hanushek's evidence, and after a more thorough examination of this literature, this research team concluded that "a broad range of school inputs are positively related to student outcomes, and that the magnitude of the effects is sufficiently large to suggest that moderate increases in spending may be associated with significant increases in achievement"(Greenwald et al., 1996a, p. 362).

Validity Issues. Recently Nye, Hedges and Konstantopoulos (2000) noted that extant approaches to class size reduction research have important strengths and limitations. For example, the production function approach probably has more external validity than experimental evaluations of class-size reduction, insofar as econometric analyses model naturally-occurring variation in class sizes in operating schools. But such analyses, they argue, are weak with respect to internal validity. The lack of internal validity can be traced to the fact that large-scale econometric modeling typically has limited information about student background characteristics and hence is unable to account for individual differences among students. This might lead to a mis-specification of the direction of effect in the class size-achievement relationship.

For example, the lack of statistically significant findings in production function studies is typically interpreted as indicating that small classes do not result in better achievement. But this conclusion is premature in the absence of more detailed information about student individual differences. For example, it may be the case that students are assigned to small classes *just because* of their poor achievement, as would be the case in remedial and compensatory programs.

In contrast, Nye et al. (2000) argued that experimental studies of the effect of class size reduction on achievement are probably lacking with respect to external validity largely because most studies are small-scale, local and short-term. Consequently, it would not be clear just how safely one could generalize findings from these interventions to natural school settings.

Project STAR and Validity. Nye et al. (2000) point out that data from Project STAR could be exploited to address the class size-achievement relationship in a way that maximizes internal and external validity. For example, internal validity is addressed by the randomization procedures adopted by STAR. External validity is maximized by the fact that STAR includes a range of schools, large and small, across all geographic categories. As a result "Project STAR includes essentially the entire range of educational conditions that occur in American education and it is more likely to be generalizable than smaller, more circumscribed studies conducted in only one location" (Nye et al., 2000, p. 126)

Problems with STAR. But a number of problems plague previous analysis of STAR data. One is that previous studies, using conventional statistical procedures (ANOVA, MANOVA, ordinary least squares regression), make a number of assumptions that are implausible. These statistical procedures assume, for example, that individual scores (i.e., achievement scores of individual students) are independent of each other. Yet there are clear dependencies evident within classrooms. Students within classrooms are more similar than students in different classrooms. In addition, there are clear dependencies among students in the same school. Achievement scores of students within a school are more similar than the scores of students in different schools. Conventional statistical procedures do not, however, account for the hierarchical nature of the data (viz., that students are nested within classrooms and schools). However, HLM procedures (Bryk & Raudenbush, 1992) address these concerns.

A second problem concerns possible implementation errors in the experimental design of STAR. The basic design called for the creation of small classes, regular classes, and regular classes with a full-time instructional assistant. But were the intended class sizes realized in the actual implementation of STAR? In addition to imperfect implementation of treatment, there was perhaps also imperfect randomization of students to condition (classroom), given the fact that students invariably drop out of classrooms and schools, or switch classrooms or schools. These problems pose possible threats to the validity of STAR

HLM Analysis of STAR. To address these issues Nye et al. (2000) examined fidelity of implementation, attrition and switching among class types in Project STAR. They did find, for example, that larger classes were sometimes smaller than intended, but that this overlap would tend to underestimate the size of the "small class effect." There was also substantial attrition, although the treatment effect for those who dropped out and those who remained was the same, "suggesting that the observed differences in achievement between small and large classes were not due to differential attrition" (p. 147). Similarly, estimates of the class size effect in "initial assignment" vs "actual" class type did not vary, suggesting that differences between small and large classes could not be attributed to class switching.

Nye et al (2000) also concluded, on the basis of HLM analyses, that small classes conveyed a statistically significant advantage in reading and mathematics "at every grade level from kindergarten through Grade three...suggesting that small classes benefit students of all types in all kinds of schools" (p. 147).

In addition, Nye et al (2000) report these findings:

- The small class effect is greater for students who have experienced more years in small classrooms.

- The small class effects estimated with HLM are similar to those predicted from meta-analyses of small-scale experiments. For example, on the basis of this analysis of STAR data, a reduction of class size from 25 to 15 could be expected to yield an increase in achievement in the range of .13 - .30 standard deviations. The increase reported in the Glass and Smith (1979) meta-analysis was .215 standard deviations.
- Moreover, the small class effect reported in this analysis of STAR is comparable to the effects sizes reported in previous meta-analyses of the production function literature (see Greenwald et al., 1996a).

Hence the authors conclude that

> "the magnitude of effects in this study is quite consistent with that obtained in small-scale randomized experiments (whose generalizability might be questioned) and with the results of econometric studies (whose internal validity might be questioned). Together, all of this evidence points to positive effects of small classes on achievement that are large enough to be educationally significant"

It is important to note, though, that the authors are not able to say *how* small classes lead to achievement.

Chapter 3

The Prime Time Literature

Overview

History of Prime Time. In 1981, at the urging of Governor Robert D. Orr, the Indiana General Assembly appropriated $300,000 for the '81-82 and '82-83 school years to implement Prime Time as a pilot project in nine schools across the state. The goal of Prime Time was to improve basic skills in reading, writing and arithmetic among Kindergarten, first- and second-graders by reducing class size. Local schools were given the option of hiring teachers to reduce class size, or else hiring instructional assistants to lower pupil/teacher ratio (PTR). (Note that this is to confound class size reduction with PTR). In 1983, the General Assembly, at the request of Gov. Orr, increased Prime Time funding to $2 million for the '83-84 and '84-85 academic years, expanding the funding to 24 school corporations and 28 school sites. In 1986, third-grade was included in Prime Time funding targets. Today virtually every school corporation in the state participates in Prime Time funding.

PTR and Instructional Assistants. Although Prime Time began as an exercise in class size reduction, it has gradually evolved into an intervention to lower pupil-teacher ratio, largely by employing teachers and instructional assistants. By statute instructional assistants funded by Prime Time may be assigned instructional, clerical and auxiliary duties in their assistance of teachers in delivery of the core curricular program. They are not to be used exclusively for non-instructional activities, nor are they to be used as substitute teachers. A full-time Prime Time assistant must be present for at least 5 hours in a classroom. Moreover, aides cannot be shared with more than two teachers. Prime Time instructional assistants must have at least a high school diploma (or equivalent), and one of the following:

- post high school work in education or related field;
- previous work experience in an early childhood program;
- previous experience as an elementary classroom assistant;
- completion of a 15 contact hour training program conducted by the local school corporation.

As noted earlier, each corporation has a target pupil-teacher kindergarten through third grade ratio that ranges from 15:1 to 18:1, depending upon the corporation's at-risk index and the amount of tuition support it receives. For purposes of calculating PTR, a full time teacher is counted as 1 FTE; a full time instructional assistant is counted as .33 FTE, and a part-time instructional assistant is counted as .167 FTE.

Challenge for Program Evaluation. Finn (1998) noted a number of reasons why Prime Time poses special challenges to program evaluation. First, because Prime Time was designed as a demonstration project, insufficient attention was paid to the sort of controls that would permit meaningful and informative evaluation. For example, no attempt was made to equalize or match smaller and larger classes on relevant variables. Second, different achievement tests have been used in different schools, and, over time, different versions of the state mandated test have been used, making it difficult to evaluate the effectiveness of class size reduction using a common metric. Third, Prime Time interventions may be confounded in some school corporations with other educational programs and initiatives (such as Title I). Moreover, as noted earlier, Prime Time is implemented in varying ways across school corporations, some emphasizing the reduction of actual class size, others emphasizing reduced teacher-pupil ratio. Consequently, *actual* class sizes in some Prime Time classrooms varies from 12 to 31 (Finn, 1998). Because of variations in implementation, then, the extant evaluations of Prime Time "cannot be interpreted as confirming or refuting a class-size effect" (Finn, 1998, p. 6).

The early expansion of Prime Time was influenced by "early studies confirming the basic theory that reduced class size does lead to increased learning" (Indiana State Department of Public Instruction, 1983, p. 5). In the following section we critically review the extant literature on Prime Time.

Previous Evaluation Studies

The 1982-83 Report. The initial evaluation of Prime Time focused on the achievement of students in the 9 schools that were part of the pilot project, for the school year 1982-83. Prime Time classes of 14 or fewer students were compared with regular classes of an average of 23 students. There were 3 Kindergarten sites, 3 sites for first-grade and 3 sites for second-grade. Students were given math and reading tests (otherwise unspecified) after the first and second semester of the school year.

The evaluation of the results of this pilot intervention was said to be favorable. Three major results were claimed. First, Prime Time students showed a greater incidence of improvement on test results than did students in larger classes. Second, teachers reported that discipline and behavior problems were reduced in classes with lower pupil-teacher ratio. Third, teachers reported greater productivity and effectiveness in classes with fewer pupils. The study concluded that

> ...project PRIMETIME has provided Indiana educators, parents and children with real evidence in favor of the positive effects of smaller class size. Small class size during the crucial early years of a child's academic development strongly enhances the quality of that child's initial educational experiences (p. 10).

What was the 'real evidence" in favor of small class size? According to the report after two semesters, 61% of Prime Time students improved ("exceed normal growth") in reading, compared to 47% of "other" students. Similarly, 53% of Prime Time students improved in math after two semesters (1982-83), compared to 49% of other students.

When the data was examined at each grade level, the percentage of students who improved in reading after two semesters was higher in Prime Time classrooms in both the 1981-82 and 1982-83 academic years. A somewhat different pattern was evident with respect to math scores. The percentage of students who improved after two semesters favored Prime Time classrooms at each grade for the academic year 1981-82. But in the academic year 1982-83, the percentage of students who improved in math after two semesters favored Prime Time classrooms *only in Grade 1* (stated claims notwithstanding). "Other" students showed more improvement in math than Prime Time students in Kindergarten, and equivalent improvement in second-grade.

Limitations. Hence the data is not as consistent as is claimed. Five other limitations qualify the conclusions of this report. First, using "percent who improved" is an insensitive indicator of program effectiveness. One wonders, for example, if significant mean differences were evident in Prime Time and comparison classrooms. Second, little is known about the nature of the tests that were used, and of their general comparability. Third, although the study is discussed in terms of pretest and post-test, there is little discussion of issues that speak to the internal validity of the evaluation. Fourth, insofar as the variable pattern of results is derived from a cross-sectional comparison of grades at each academic year, one cannot rule out cohort effects as a rival explanation of the observed improvement in Prime Time classrooms. Fifth, the evidential basis for the second and third conclusions of the report appears to be anecdotal.

On the basis of this initial evaluation project Prime Time was expanded in 1984, involving 286 of 303 school corporations. Subsequent evaluations of *Prime Time* have yielded mixed results (Chase, Mueller & Walden, 1986; Gilman, Swan & Stone, 1988; Malloy & Gilman, 1989; McGiverin, Gilman & Tillitski, 1989; Mueller, Chase & Walden, 1988; Varble, 1990).

McDaniel Report (1985). An extensive ethnographic evaluation of Prime Time was reported by Ernest McDaniel of Purdue University (1985). An evaluation team of six researchers, drawn from six different Indiana universities, conducted site visits and interviews at schools that were "representative of the geographic areas of the State" (p. 3). School corporations were selected by the state department of education, superintendents selected schools within the corporation, and building principals selected classrooms within a school. A total of 59 classrooms were observed (totaling 97 classroom visits and nearly 1000 students). Observers made detailed notes of classroom activities and interviewed teachers. Questionnaires (which solicited teachers' retrospection over their experience with Prime Time) and surveys of pupil attitudes were also employed. The general analytic strategy, common to qualitative, observational research, was to identify convergence and agreement among many observers (including parents). The written observer narratives, their "thick" descriptions of classrooms, were then organized around some common themes.

For example, 55% of Prime Time teachers eschewed seating students in rows (although 20% of classrooms had no recorded information about the physical setting). Nearly 75% of teachers exhibited student work (although this information was not recorded for 17% of the classrooms). Most Prime Time classrooms were characterized by "relaxed control" (74%) rather than "tight control" (26%). Most teachers were "flexible" (69%) rather than inflexible (19%). Most classrooms were "on-task" (92%) rather than "off-task" (8%); most emphasized process (64%) over "product (29%) work orientations; most emphasized intrinsic (42%) or mixed (20%) motivational techniques, as opposed to extrinsic motivational techniques (32%). A vast majority of classrooms (92%) opted for small class grouping for reading. Over half of the teachers (51%) used predominantly divergent questions, while 39% used predominantly convergent questioning techniques. Student questioning and inquiry behavior was observed in 42% of the classes, not observed in 46% of classes (and not recorded for 11% of classes). McDaniel (1985) also reported that Prime Time had "a significant energizing effect on teachers" (p. 49). Moreover, parents were reportedly "sold" on the benefits of Prime Time, "and they are strong supporters of the program" (p. 49).

In summary, McDaniel (1985, p. 25) draws this portrait of Prime Time classrooms:

> "The picture of PRIME TIME classrooms which emerges from the tabulations is of classes characterized by flexible seating patterns and use of space, abundant displays of student work, and one or more learning centers. Within these rooms, the children work in an atmosphere of relaxed control. Paradoxically,...it appears evident that children enjoy the most freedom in rooms where class procedures are well-planned, organized and structured. Almost everywhere, children are on-task and a feeling of industriousness pervades the room."

Moreover, Prime Time was said to afford at least four opportunities:

- to explore ways of reducing extensive use of worksheets (still a common practice when teachers are engaged in small group instruction);
- to accelerate the trend to broaden activities at learning centers (from paper-and-pencil drill to divergent, creative and imaginative activities);
- to explore ways to encourage student questioning and inquiry (through projects);
- to offer opportunities for a balanced, integrated learning environment.

McDaniel (1985) concluded with several recommendations, including the need for continuous staff development on how to adapt instruction to capitalize on small class size. Clearly this ambitious project provided a rich portrait of what goes on in Prime Time classrooms. What it does not do is show that what goes on in Prime Time classrooms is the result of class size reduction or lower PTR. Perhaps what observers recorded was general best practice, and its variations, rather than practices attributable to Prime Time. Morever, the report has the drawbacks of all ethnographic research, particularly with respect to its representativeness and generalizability. Schools were not selected randomly. There was no comparison group (non-Prime Time classes were not observed). Observers did not appear to be trained in the use of a common observation protocol.

The Gilman Studies. David Gilman and his colleagues at Indiana State University and North Gibson School Corporation (Princeton, Indiana) reported a series of studies that admirably capitalized on the three year phase-in of Prime Time. These studies are the most sustained attempt at evaluating a "state supported class size reduction program," which is how Prime Time is described by this research team. The first report (Swan, Stone & Gilman, 1987) compared the achievement scores of first-graders in three successive academic years:1983-1984, 1984-1985, and 1985-1986 (see also Gilman, Swan & Stone, 1988, who report similar data over the first two successive years).

During this time Prime Time class size reduction was being implemented. Consequently, the mean class size of first-grades dropped accordingly, from 23.7 in 1983-1984, to 19.9 in 1984-1985, to 16.1 in 1985-1986. Swann et al. (1987) also tested two successive samples of second-graders, once in 1984-1985, when the mean class size was 20.5, and once more the following year when the mean class size was 17.4. Students submitted to locally constructed tests of basic math and reading skills (administered in May), to the Iowa Test of Basic Skills (administered in March), and to affective measures of "attitudes toward school" and "self-concept." The general strategy was to show that mean scores on the various achievement and affective measures were higher in classrooms with lower enrollment, or, alternatively, that first- and second-grade scores were higher after the introduction of Prime Time than before.

The results at first-grade were clear and dramatic, and, according to the authors, "spectacular" (p. 23). The mean scores on all of the achievement and affective measures (with the single exception of "school attitude") were significantly lower in the first-grade class of 1984 (prior to Prime Time) than in the first-grade classes of 1985 and 1986. The results at second-grade were attenuated but "favorable." There was no difference in second-grade ITBS reading scores before and after class size reduction. Hence, reducing class size from 20 (1985) to 17 (1986) did not appear to pay off with respect to reading. However, statistically significant results were evident in ITBS math scores and in the ITBS composite, favoring Prime Time classrooms.

This general analytic strategy was again followed when Prime Time was extended into third-grade (Gilman, Tillitski, Swan, & Stone, 1987). In addition to the comparisons made in the prior study, two samples of third-grade students were added. One sample sat for testing in 1985-86, prior to the implementation of Prime Time at third-grade. The second sample of third-grade students were tested in 1986-87, after the introduction of Prime Time. Accordingly, the mean third-grade class size dropped from 24.0 to 18.0 after Prime Time was introduced in 1986-87.

Table 1 reports our tabulation of mean achievement scores in the time series for first-graders.

Table 1
Mean Achievement Scores for the First-Grade Time Series (Gilman et al., 1987)

	1983-1984	1984-1985	1985-1986	1986-1987
Mean Class Size	**23.4**	**19.9**	**16.1**	**16.6**
Local Test: Math (# correct)	49.11^{12}	56.01^{134}	54.3^{2}	52.4^{4}
Local Test: Reading (# correct)	71.9^{1}	79.9^{1}	78.4	77.1^{1}
Local Test: Math Skills	8.5^{123}	11.8^{1}	11.3^{2}	10.9^{3}
Local Test: Reading Skills	10.6^{123}	13.6^{1}	13.3^{2}	13.1^{3}
ITBS Math Subtest	59.3^{123}	70.0^{1}	79.1^{34}	77.6^{34}
ITBS Reading Subtest	71.4^{12}	80.7^{13}	79.9^{4}	76.5^{234}
ITBS Composite	52.4^{124}	76.3^{13}	80.6^{4}	76.5^{4}

Note. Means that share a common superscript are said to be statistically different from each other.

In general, first-grade mean achievement scores are significantly higher in the first year of Prime Time (1984-1985) than they are in 1983-1984. But there is some attenuation in the years that follow. Indeed, in the case of the three ITBS scores (Math, Reading, Composite), the decline in scores from 1986 to 1987, when mean class size was about 16, is statistically significant (the authors' claims notwithstanding). Local math and reading scores (# correct) were also lower in 1987 than they were in the first year of Prime Time (1985). Still, the overall pattern of the data was not discouraging to the authors.

Table 2 reports the mean achievement scores for the second-grade time series. One should note first that the actual mean class size reduction over the course of this time series was negligible. This makes the statistically significant improvements in ITBS achievement rather inexplicable, and certainly difficult to attribute confidently to the Prime Time intervention. Nonetheless, the authors concluded that "the significant achievement gains in mathematics, reading and composite scores after the first year of Prime Time have been maintained in 1986-87" (Gilman et al, 1987, p. 14).

Table 2
Mean Achievement Scores for the Second-Grade Time Series (Gilman et al., 1987)

	1984-1985	1985-1986	1986-1987
Mean Class Size	**20.5**	**20.0**	**19.1**
ITBS Math	63.1[1]	66.4	67.8[1]
ITBS Reading	54.5[12]	68.9[1]	68.5[2]
ITBS Composite	63.2[12]	71.4[1]	70.8[2]

Note. Means that share a common superscript are said to be statistically different from each other.

Table 3 reports achievement data for Grade 3. After Prime Time was implemented class size dropped, on average, from 24 students in a classroom in 1985-86 to18 students in a classroom in 1986-87. As Table 3 illustrates, however, a class size reduction of this magnitude had virtually no consequences for academic achievement, with the possible exception of Iowa Test math scores.

Table 3
Mean Achievement Scores for the Third-Grade Time Series (Gilman et al., 1987)

	1985-1986	1986-1987
Mean Class Size	**24**	**18**
Iowa Test: Math	67.3[1]	71.9[1]
Iowa Test: Reading	61.1	60.5
Iowa Test: Composition	67.5	69.8
Indiana Competency Test: Reading	58.5	59.4
Indiana Competency Test: Writing	29.8	29.3
Indiana Competency Test: Composite	122.2	121.2

Note. Means that share a common superscript are significantly different from each other.

Although the authors noted that "the third-grade performance is particularly disappointing," and that "many of the early gains experienced as a result of Prime Time being introduced in Grades 1 and 2 have seemed to disappear in Grade 3" (p. 16), the authors nonetheless concluded that "the continuation of Prime Time continues to produce its desired results on all measures utilized in this study" (p. 17).

Yet the evidence would also support a more discouraging conclusion. It is quite clear, certainly, that class size reduction conveyed no benefit at third-grade. Moreover, the positive results reported for second-grade are also suspect. Indeed, the data for the second-grade time series cannot be properly interpreted, given the rather large class size evident in these classes. This leaves first-grade data as supporting the putative benefits of class size reduction. But even here the positive benefits of Prime Time implementation are not pervasive over the course of the time series (and can be attributed perhaps to other factors, see below). The safest conclusion, on the basis of these data, is that the benefits of class size reduction are most evident in first-grade, but significantly attenuate thereafter

Several methodological issues complicate a clean interpretation of these data. One difficulty with time series designs is that one is tempted to draw inappropriate conclusions about intra-individual developmental change from the observation of inter-individual differences. For example, the authors noted that "The third grade performance is particularly disappointing *since these same students* scored much higher as Prime Time students when they were in Grades 1 and 2" (p. 16, our emphasis). But this "longitudinal" conclusion does not follow from the time series analyses presented here. The same individuals were not tested over time, so one cannot infer how the "same students" would have done over time. Indeed, it is quite possible that students would indeed show developmental (intra-individual) gains in achievement as a result of the Prime Time class size intervention, but that cannot be discerned by the time series analysis presented here. Clearly there is a need for longitudinal evaluation of Prime Time (e.g., McGivern et al., 1989; Weiss, 1990).

A second difficulty is that the time series analysis examined mean performance of first-graders (for example) in successive years. But these students, although of the same age when tested, are drawn from different cohorts. Consequently, age and cohort are necessarily confounded. Third, there is the possibility that the present data are also confounded by time-of-testing effects.

In subsequent reports Gilman and his colleagues reached progressively more critical conclusions about the efficacy of Prime Time. In 1988 Malloy and Gilman compared results of the Indiana Basic Competency Skills Test (IBCST) of two groups of third-graders. One group consisted of 67,987 third-grade students who received no "Prime Time instruction" in 1986. Another group consisted of 65,911 third-grade students who, in 1987, had three years experience in Prime Time classrooms (or so they assumed).

Recall that the first year of implementation of Prime Time at third-grade was in the 1986-1987 school year. Hence, students who were in third-grade in the 1985-1986 school year would not have been exposed to Prime Time class size reduction. Of course, it is possible for these students to have been in Prime Time classrooms at second- and first-grade. Hence it is not at all clear that the intended comparison (between third-grade students who had no experience with Prime Time vs third-grader students with three years experience) was achieved. Nor can it be certain that all of the students in the "experimental" group were in fact in Prime Time assisted classrooms, given variation in implementation even within corporations and schools. Note that these data are also confounded by cohort and time-of-testing effects.

That said, Malloy and Gilman (1988) reported only a .7 point difference in the mean IBCST composite score between the 1986 cohort and the 1987 cohort of third-graders. They concluded that the "results of this large scale comparison of PRIME TIME and regular classes suggest that the long-term effects of a state sponsored reduced class size program are negligible" (Malloy & Gilman, 1988, p. 9). This argument was also advanced in a paper written in the same year (Gilman, Harder & Tillitski, 1988). In this sometimes tendentious paper, Gilman, Harder and Tillitski (1988) argued that:

- class size is a smokescreen that draws attention from factors that really do influence achievement, such as teacher competence;
- class size is an issue embraced by teacher unions as a strategy to improve working conditions, and that unions take up the cause of smaller class size "as a tool for union victories" (p. 4);
- class size reduction, and the hiring of additional teachers and aides, forces administrators to engage in more "paper shuffling activities" (p. 5);
- millions of dollars have been poured into class size reduction programs without any concern for cost effectiveness;
- that research on the alleged benefits of class size reduction is inconclusive
- and may be harmful to students in the long run to the extent that reduced PTR makes students more teacher dependent, making them ill-prepared to learn in larger classes in later grades;

They concluded with the complaint that "Politicians, teachers, and parents are not asking questions and nobody else seems to notice...or care...whether reducing class size is a cost effective and beneficial advantage or a meaningless administrative exercise" (Gilman, Harder & Tillitski, 1988, p. 15).

But in the same year Gilman (1988) reported findings from a "longitudinal" analysis of first-graders in the North Gibson School Corporation. Four cohorts of first-graders were examined: the class of 1984-85 (mean class size = 19.9), 1985–86 (mean class size = 16.1), 1986-87 (mean class size = 16.6), and 1987-88 (mean class size = 17.5). Their performance on locally constructed "basic skills tests" and "affective measures" (attitude toward school and self-concept) were compared with the cohort of first-graders in 1983-84. This cohort was in first-grade prior to Prime Time (with an average class size of 23.7), and thus served as the comparison group by which to compare the subsequent Prime Time cohorts.

Hence this report is of a "four year" evaluation of first-grade performance. The three-year evaluation was reported in Gilman, Tillitski, Swan and Stone (1987), and discussed earlier (see e.g., Table 1). This is not, of course, a longitudinal evaluation, as was claimed, but rather a time series study of successive cohorts. Consequently, variations in mean scores across time could also be attributed reasonably to cohort or to time-of-testing effects. Nonetheless, the results seemed to indicate that the performance of first-graders from 1984 to1988 on the various measures was significantly better than the performance of first-graders in 1983. Gilman (1988) noted that while the effect of Prime Time apparently loses its impact by Grade 3, it "can be concluded that the gains experienced by PRIME TIME students during the early years of the project have not only been maintained by they have also been strengthened" (p. 18) and that "the continuation of PRIME TIME in Grade 1 continues to produce its desired results on all measures" (p. 15).

The only meta-analysis of Prime Time data was reported by McGiverin, Gilman and Tillitski (1989). The data for this meta-analysis were ten studies for six randomly selected schools for which second-grade data was available. These students were presumably in a Prime Time classroom for two years. These results were compared to four studies from three school corporations where second-graders had not experienced Prime Time class size reduction. The results indicated a significant advantage for students in Prime Time classrooms. The average effect size for academic improvement in Prime Time classrooms was .34 standard deviations (approximately 6 points) which is considered educationally significant. Such gains were not evident in the comparison groups. Hence, the authors concluded "that second-grade students in smaller classes (19.1 students, as defined by project PRIME TIME) had significantly higher achievement test scores than did students in larger, pre-PRIME TIME classrooms" (McGiverin et al., 1989, p. 54).

Clearly the Gilman studies provide a very mixed picture of Prime Time effectiveness. Significant methodological issues notwithstanding, these studies tended to conclude that reduced class size does convey benefits in first- and second-grade, although not for third-grade. Of course, most of these studies were conducted within a single school corporation, and hence any conclusion must be considered tentative.

Finally, we will note in passing a report by Sanogo and Gilman (1994) that attempted to account for the different conclusions about class size to be derived from Project STAR and from Prime Time. The "contradiction" was that studies of Prime Time do not support the formation of smaller classes, but that studies of Project STAR do. Moreover, "the cause of the contradiction must have been the methodologies and design" (Sanogo & Gilman, 1994, p. 9). And the authors found the design of STAR to be wanting.

This paper insinuated that the "amazing" results reported in Project STAR were corrupted by politics. The study results could have been influenced by the Tennessee teachers union, who were keen to demonstrate a small class effect "in order to convince the Governor, and the Tennessee Legislature to drop the Better Schools Program" (p. 19). The Better Schools Program, opposed by the teacher union, was an initiative to establish a "master teacher"program, which included a system of teacher evaluation and merit pay. As a consequence, teachers were motivated (according to the authors) to demonstrate the effectiveness of small classes (hence vitiating the need for teacher evaluation, merit pay, and graduated ranks). This resulted in a set of conclusions about class size that are little more than a Type 1 error!

The erroneous basis of STAR conclusions was also linked to the "strong probability" (p. 19) that the STAR evaluation was corrupted by Hawthorne effects. Project STAR teachers, knowing that they were in the "small class size condition," worked harder to produce the outcomes desired by their professional union. The authors even insinuated that perhaps teachers in control group classrooms, knowing that they were in control groups, did not work their students hard so as to not outperform students in the experimental groups (knowing how dear the notion of reduced class size is to the teaching profession). No evidence was in sight for this peculiar "John Henry" effect, however.

Finally, Sanogo and Gilman (1994) argued that Project STAR and Prime Time were (in some sense) equivalent, but only because "both were not implemented with any scientific attitude" (p. 21). The methodological details of STAR and Prime Time are very different, of course. But, with respect to STAR, "the whole effort became worthless because the sample of the study was biased" (p. 22). With respect to Prime Time, "the Indiana Board of Education was biased since the beginning of Prime Time and failed to evaluate the project seriously" (p. 22). Hence, both projects, although methodologically very different, are equivalent in their biases and special pleadings. "Neither were true experimental research"(p. 22).

This commentary is, however, strikingly idiosyncratic and far from the consensus opinion about Project STAR or Prime Time.

Mueller, Chase and Walden (1988). This paper reports on a full-scale evaluation of first- and second-grade classrooms soon after implementation of Prime Time in 1986. Survey data from two schools in each of 29 school districts (out of 304) was collected. In addition, structured observations were obtained in 300 classrooms. Finally, achievement data was obtained that permitted a comparison of scores before and after Prime Time implementation in all first- and second-grade classrooms in ten school districts.

The survey data indicated strong support for Prime Time among teachers, parents and principals. For example, about 90% of teachers were convinced that students in Prime Time classrooms received individual attention and received immediate feedback. About 80% asserted that below- and above-average students were achieving more. A vast majority of teachers also believed that pupils were on-task (77%); that a variety of instructional methods were being used (76%); and that pupils could progress at their own rate (71%). Teachers who actually had smaller, Prime Time classrooms were particularly convinced of the positive educational benefits of Prime Time. Moreover, parents of children in small classes (vs parents of children in larger classes) reported that their child's general school progress and reading progress was above their expectations; that their child's teacher was available for consultation; that their child received adequate or more than adequate individual attention; that their child had frequent homework; and that Prime Time was an important factor in their child's learning.

What about achievement differences between small and large classes? Mean classroom (reading and mathematics) achievement data for first- and second-grade was available in 10 school districts for the first year of Prime Time and for the year preceding implementation. The average decrease in class size was slight: from 22 to 19 students in first grade; from 21 to 20 students in second-grade. Nonetheless, half of the districts reported higher reading scores in first-grade after implementation of Prime Time, and 30% reported higher scores in mathematics. At second-grade, only 20% of the districts reported more gains in reading after implementation, and 10% for mathematics. As the authors put it, "Clearly, the Prime Time program affected 1st grade more than 2nd grade, and reading more than mathematics" (p. 50).

The observational data from 300 classrooms allowed the authors to reach some conclusions about teacher-student interactions and how aides are used. They concluded that a greater proportion of teacher and instructional assistant interactions in Prime Time classrooms were in small groups of two to five students, compared to larger classrooms. Moreover, instructional assistants were typically used for small groups, or for working with individual children, or tutoring, findings also reported in the recent study by Lapsley and Daytner (2001a, b).

General Conclusion

It is clear that Prime Time is still a program that has been insufficiently evaluated over the years. The literature occasionally suggests that lower PTR, or smaller classes, pays off in terms of student academic outcomes, at least in first- and perhaps second-grade, although many of these studies are also hard to evaluate due to methodological concerns. Moreover, the majority of evaluation studies have been conducted in just a single school corporation.

Project Prime Time: Third-Grade Achievement

In subsequent chapters we report on our quantitative analysis of how Prime Time assisted classrooms differ from classrooms that are not assisted by Prime Time aides, using a statewide representative sample of third-graders. Our criterion measures were derived from the state-mandated achievement test (ISTEP+) administered in September of the 2000-2001 academic year.

Our general analytic strategy was to provide information at two levels. First, we wanted to provide a rich descriptive summary of how Hoosier third-graders performed on the ISTEP+ test, as a function of geographic category, class size, race and the presence or absence of instructional assistants. Second, we wanted to examine the quantitative relationship between PTR, class size and achievement using "best practice" statistical procedures (HLM) appropriate for analyzing nested data structures. Although we were interested in providing helpful information to state educators and program planners, we were also mindful of the various controversies in the class size reduction literature. Consequently, our analyses were also directed towards pressing empirical issues in this literature.

Chapter 4

Summary of the Empirical Findings

Sampling

Stratified Cluster Sampling Rules. School corporations were sampled using a stratified cluster sampling procedure. Our original goal was to include 25% of corporations within each of 9 educational service regions. For example, Region 1 includes a total of 32 school corporations, hence 8 corporations (or 25%) were to be randomly selected to participate in this project. Our second sampling rule was that at least one urban corporation must be included from each region, with the remaining corporations per region determined by proportional representation within geographic category (urban, suburban, town, rural). Once again using Region 1 as an example: 8 corporations were to be randomly selected (the 25% rule); 1 corporation was urban, 4 of the 8 were rural (because 50% of corporations within Region 1 are rural), 3 of the 8 were town (because 38% of corporations within Region 1 were township schools). There are no suburban school corporations within Region 1. This sampling strategy would have yielded 78 school corporations, distributed across region and geographic category, as illustrated in Table 4.

Table 4
Target (and Actual) Sampling Distribution of School Corporations, by Educational Service Region and Geographic Category

Educational Service Region	Geographic Category				
	Rural	Suburban	Town	Urban	Total
1	4 (3)	0 (0)	3 (2)	1 (1)	8 (6)
2	3 (3)	1(1)	2 (2)	1 (1)	7(7)
3	3 (2)	1 (1)	4 (4)	1 (2)	9 (9)
4	5 (4)	2 (2)	4 (3)	1 (1)	12 (10)
5	6 (3)	1 (1)	3 (1)	1 (1)	12 (6)
6	3 (2)	1 (1)	2 (1)	2 (1)	9 (5)
7	4 (4)	1 (0	2 ((1)	1 (1)	8 (6)
8	3 (4)	1 (0)	3 (3)	1 (0)	8 (7)
9	0 (0)	4 (4)	0 (0)	1 (1)	5 (5)
Total	31 (25)	14 (10)	23 (17)	10 (9)	78 (61)

For the most part we were reasonably successful in our planned target sample. The number in parentheses in Table 4 indicates the actual number of school corporations who agreed to participate in this project, sampled within educational service region and geographic category. This stratified sampling procedure yielded 61 corporations (or 78% of the target sample), totaling 10, 927 students in 163 schools and 573 classrooms. Of these classrooms, 190 were Prime Time assisted (or about 1/3 of the total sampled), 376 were unassisted, and 7 were assisted by other means.

We asked school corporations to provide us with the ISTEP+ results of individual children in all third-grade classrooms within the corporation, with the exception of large urban school corporations, where we randomly selected a certain number of schools (typically 7 or 8 schools) to represent the corporation.

The Sample. A total of 10, 927 third-graders (5,425 females, 5,457 males; gender was not indicated for 45 students) were included in this sample. Table 5 shows the breakdown by region and geographic category.

Table 5
Number of Third-Grade Students, by Educational Service Region and Geographic Category

Educational Service Region	N of Third-Graders	Percent
1	1028	9.4
2	1438	13.2
3	1861	17.0
4	1244	11.4
5	1013	9.3
6	1095	10.0
7	1184	10.8
8	881	8.1
9	1183	10.8
Total	10927	100.0
Geographic Category		
Urban	3936	36.0
Suburban	2019	18.5
Town	25253	23.1
Rural	2449	22.4
Total	10927	100.0

The ethno-racial background of the third-grade students is summarized in Table 6. As can be seen, approximately 84% of the sample was white; 9% was African-American; 3% was Hispanic, with the remaining students representing Asian, multi-racial and native American Indian status. Ethno-racial information was not available for 111 (or 1%) of the sample.

Table 6
Ethno-Racial Background of Third-Grade Students

Category	Frequency	Percent
Caucasian American/White	9207	84.3
African American/Black	995	9.1
Hispanic-American	348	3.2
Asian American	62	.6
Multi-Racial	188	1.7
Native American Indian	16	.1
Total	10816	99.0
Missing	111	1.0
Total	10927	100.0

Test Scores

Class Record Sheet. The "Class Record Sheet" was provided by school corporations. The class record sheet reports National Percentile, Grade Equivalent, Normal Curve Equivalent and Anticipated Normal Curve Equivalent scores, for the following ISTEP+ tests: Reading, Vocabulary, Reading Composite, Language, Language Composite, Math, Math Composite and the Total Composite Score. These scores were indicated for each child (we instructed school personnel to blacken the name of each child in order to protect anonymity). The class record sheet also indicated each child's birthday. School personnel indicated race and gender for each child, and also whether the class had a Prime Time assistant (full time or part-time).

Normal Curve Equivalent Scores. The normal curve equivalent (NCE) is a normalized standard score developed primarily for use with federal program evaluation efforts. The normal curve equivalent is found by the following formula:

$$NCE = 21.06z_n + 50$$

Hence an NCE score is a linear transformation of the z-score (which itself is a linear transformation of the raw score). NCE scores have a mean of 50 and a standard deviation of 21.06. The choice of standard deviation was to allow the range of scores to span the range of 1 to

99. The NCE was used in all analyses. We restricted our analyses to the NCE composite scores for reading, language and math, and the total NCE composite score.

School Variables. We culled information about each school, and each corporation, from the state Department of Education website. This information was gathered to facilitate our test of hierarchical linear models. This information will be described in more detail below.

Descriptive Analyses: Demographic Characteristics

We first wanted to provide a comprehensive descriptive account of how our representative sample of third-graders performed on the ISTEP+ exam, broken down by stratification (educational service region, demographic category), classroom variables (class size, presence of PT instructional assistant) and by student variables (race). In a subsequent section we report the results of HLM analyses.

Enrollment and Class Size. Table 7 presents the frequency of enrollment by class size.

Table 7: Enrollment by Class Size

Class Size	Frequency	Percent
12-13 -14	24 + 37 + 26 = 87	.8
15 -16	224 + 331 = 555	5.0
17	443	4.1
18	540	4.9
19	1167	10.7
20	1292	11.8
21	1572	14.4
22	1377	12.6
23	1325	12.1
24	966	8.8
25	871	8.0
26	315	2.9
27	264	2.4
28-29	40 + 28 = 68	.7
30 -31	63 + 22 = 85	.8
Total	10927	100

As Table 7 indicates, the modal class size in this sample is 21. Approximately 10 % of third-grade pupils are in classes of 17 or fewer (what Project STAR would denote as a "small class") and nearly 15% are in classes of 18 or fewer. Moreover, about 37% of Hoosier third-graders are in classrooms of 20 or fewer. This leaves about 63% of pupils in classrooms with enrollment over 20. Of course, the presence of Prime Time assistants would reduce PTR to the intended target. At the other end, approximately 15 % of students are in classes of 25 or larger. To simplify subsequent analyses we grouped class enrollment into four categories: classes of 12-17, 18-21, 22-26, and 27 or more pupils. These categories were chosen to accord partly with Project STAR classifications (which considered 22-26 a "regular" class enrollment).

Table 8 reports the frequency and mean enrollment of third-grade pupils by class size.

Table 8
Frequency and Mean Enrollment, by Class Size

Class Size	Frequency	Percent	Mean Enrollment	Std. Dev.
12-17	1085	9.9	15.96	1.17
18-21	4571	41.8	19.85	1.02
22-26	4854	44.4	23.47	1.25
> 26	417	3.8	27.89	1.33

Although Prime Time is designed primarily to reduce pupil-teacher ratio rather than class size, it is of interest to note that nearly half of third-grade pupils are in classrooms of 22 or greater (almost 4% in classrooms of 27 or greater).

Class Size and Race. Table 9 reports the distribution of students by racial background and class size. The most striking pattern evident in Table 9 is the fact that a much larger percentage of African-American and Hispanic children are in small classes (12-17) than are White children. For example, 16% of black children and almost 24% of Hispanic children are in classes of 12-17, compared to 8.7% white and 6.45% Asian children. This perhaps reflects a greater minority participation in smaller compensatory classes, an issue we take up in more detail below.

Table 9
Students by Racial Background and Class Size

Race	Class Size								
	12-17		18-21		22-26		> 26		Total
	N	%	N	%	N	%	N	%	
White	804	8.70	3833	41.63	4204	45.66	366	3.97	9207
Black	160	16.08	437	43.92	383	38.49	15	1.51	995
Hispanic	82	23.56	135	38.79	109	31.03	22	6.32	348
Total Black-Hispanic	242	18.02	572	42.59	492	36.63	37	2.75	1343
Asian	4	6.45	31	50.00	25	40.32	2	3.22	62
Other	17	8.33	107	52.45	71	34.80	9	4.41	204
Total Non-White	263	16.34	710	44.12	588	36.54	48	2.98	1609

Class Size and Geographic Category Table 10 reports the frequency of third-grade enrollment by class size for each of the four geographic categories.

Table 10
Frequency of Enrollment by Class Size and Geographic Category

Class Size	Geographic Category				
	Urban	Suburban	Town	Rural	Total
12-17	429	182	299	175	1085
18-21	1598	680	1092	1201	4571
22-26	1622	1133	1107	992	4854
> 26	287	24	25	81	417
Total	3936	2019	2523	2449	10927

This table shows that nearly 69% of pupils in very large classrooms (> 26) are in urban school corporations. On the other hand, nearly 40% of pupils in small classrooms (12-17) are also from urban districts.

Geographic Category and Race. Finally, in Table 11 we chart the distribution of race among the geographic categories. Although white pupils are approximately equally distributed among urban, suburban, township and rural settings, the non-white population, particularly black and Hispanic pupils, are overwhelmingly located in urban school corporations. For example, 71%

of black students in our sample, and 78% of Hispanic youngsters, were in urban schools, compared to about 30% of white students. Most of the remaining black and Hispanic youngsters were in suburban schools. Indeed, the incidence of non-white school membership in township and rural schools is relatively slight.

Table 11
Students by Racial Background and Geographic Category

	Geographic Category								
Race	Urban		Suburban		Township		Rural		Total
	N	%	N	%	N	%	N	%	
White	2743	29.79	1620	17.59	2451	26.62	2393	25.99	9207
Black	709	71.26	276	27.74	5	.50	5	.50	995
Hispanic	272	78.16	33	9.48	26	7.47	17	4.88	348
Total Black-Hispanic	981	73.04	309	23.01	31	2.31	22	1.64	1343
Asian	23	37.09	29	46.77	7	11.29	3	4.84	62
Other	122	59.80	37	13.23	22	10.78	23	11.27	204
Total Non-White	1126	69.98	375	23.31	60	3.72	48	2.98	1609

Summary. The preceding tables provide a descriptive summary of our sample, including student demographic characteristics (race), classroom variables (class size, enrollment) and geographic categories. The key points can be summarized in the following way:

- About 10% of third-graders in our sample were in "small" classes of 17 or fewer.
- About 37% of students were in classrooms of 20 or fewer, although 15% were in classrooms of 25 or greater.
- Nearly half of the students in our sample was in a classroom of 22 or greater students.
- The average class size was 21.
- A larger percentage of black and Hispanic (vs white) youngsters was in a small class.
- Black and Hispanic students were overwhelmingly concentrated in urban school corporations.

Descriptive Analyses: Prime Time Assistance and Achievement

The previous descriptive analyses focused on background characteristics of students and classrooms. In the analyses that follow we introduce the Prime Time assistance variable to the descriptive account of our sample. Hence, we revisit the classifications of race, class size, and geographic location, but in the context of Prime Time assistance and academic achievement.

Class Size and Prime Time Assistance. We were first interested in charting the use of instructional assistants in our sample schools. Table 12 reports the number of students who are exposed to Prime Time assistants, by class size.

Table 12
Number of Third-Grade Students With and Without Prime Time Instructional Assistants, by Class Size

Class Size	Prime Time Assistant	No Instructional Assistant	Other Assistant	Total
12-17	116	954	15	1085
18-21	639	3900	32	4571
22-26	2987	1797	70	4854
> 26	279	138	0	417
Total	4021	6789	117	10927

We noted earlier that one-third of the third-grade classrooms in this sample were Prime Time assisted. Similarly, Table 12 shows that 36% of third-grade students were exposed to a Prime Time assistant. Moreover, it is apparent that a relatively large number of students (1,935, or 18% of the sample) are in large class rooms ($\geq$ 22) without the benefit of Prime Time instructional assistants.

Class Size and ISTEP+ Achievement. Table 13 reports mean NCE achievement scores for composite reading, language and math and the total composite score, for classrooms with and without Prime Time assistants, and also by class size.

Table 13
Means and Standard Deviations of NCE Achievement Scores (ISTEP+) by Prime Time and Class Size

	Reading		Language		Mathematics		Total Composite	
PT Status	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Assistant	59.26	16.82	62.46	19.47	66.14	21.99	62.64	16.75
Regular	56.34	17.18	60.66	19.36	63.13	21.89	60.10	16.78
Class Size								
12-17	53.25	16.51	58.55	18.78	61.71	21.68	57.92	16.10
18-21	56.85	16.92	60.96	19.33	63.13	21.56	60.39	16.52
22-26	58.66	17.21	62.11	19.53	65.65	22.29	62.16	17.07
> 26	60.21	17.38	63.49	19.93	66.87	22.44	63.50	17.22

As Table 13 indicates achievement scores, in every case, were higher for students who are exposed to Prime Time assistants than for students in unassisted classrooms, particularly in reading. Note, too, however, that achievement scores tended to increase as class size got larger. This dual pattern with respect to Prime Time assistants and class size is further examined in Table 14 and Table 15. Table 14 reports mean NCE achievement scores in Prime Time assisted and unassisted regular classrooms by geographic category.

Table 14
Mean NCE Achievement Scores in Prime Time Assistant (PT) and Unassisted Regular (R) Classrooms, by Geographic Category

	Urban		Suburban		Township		Rural	
NCE	PT	Regular	PT	Regular	PT	Regular	PT	Regular
Reading	56.91	54.52	63.78	58.06	58.30	57.86	58.58	56.73
Language	60.83	59.33	68.57	63.01	60.99	62.60	59.93	59.03
Math	66.2	62.39	73.02	67.01	64.05	64.54	61.76	59.76
Total	61.29	58.82	68.51	62.75	66.11	61.75	60.09	58.52

In nearly every comparison (Table 14), in every geographic category, achievement scores were higher in Prime Time classrooms than in unassisted classrooms, with the Prime Time advantage particularly evident among children in suburban and township schools. However, when class size is examined (Table 15), it is clear that better student achievement is associated with large class sizes in all four geographic categories.

Table 15
Means and Standard Deviations of NCE Achievement Scores by Class Size and Geographic Category

	Urban		Suburban		Township		Rural	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Reading								
12-17	50.26	16.19	53.03	16.84	55.81	15.55	56.31	17.35
18-21	54.97	17.85	58.88	16.25	58.11	16.55	57.03	16.13
22-26	56.18	17.65	63.19	17.27	58.24	16.89	58.01	15.74
> 26	59.65	17.32	57.17	16.55	61.95	20.02	62.59	17.02
Language								
12-17	56.22	19.43	59.57	18.21	61.26	18.03	58.52	18.42
18-21	59.94	20.16	63.47	19.47	62.28	18.86	59.69	18.31
22-26	60.07	20.02	67.95	18.92	61.99	19.13	58.89	18.38
> 26	63.50	20.69	63.33	19.86	67.68	18.41	62.21	17.74
Math								
12- 17	60.14	21.63	63.84	21.34	64.48	22.64	58.61	19.32
18-21	63.39	22.13	66.08	21.83	64.47	21.22	59.92	20.54
22-26	64.13	22.68	73.39	21.25	63.93	21.29	61.19	21.79
> 26	66.43	23.16	53.71	19.67	82.80	18.85	67.40	18.74
Total								
12-17	55.71	16.23	58.86	15.50	60.54	16.17	57.80	15.60
18-21	59.53	17.36	62.93	16.46	61.75	16.03	58.87	16.63
22-26	60.16	17.53	68.18	16.53	61.39	16.36	59.38	16.05
> 26	63.13	18.11	58.58	16.10	70.72	15.53	64.01	14.13

The fact that achievement scores tended to be lowest in classrooms with the smallest enrollments struck us an oddity that required further analysis. As noted by Nye et al (2000), large scale studies such as this one typically have satisfactory external validity, but oftentimes lack sufficient internal validity if one is unable to understand how particular classrooms are formed and with which kinds of students. For example, in the present analyses, we grouped classroom data into broad enrollment categories (12-17, 18-21, 22-26, > 26). These categories were chosen to allow certain comparisons with Project STAR classrooms (particularly with respect to small and regular classrooms). However, it is possible that at least some small classrooms in our analyses were formed for remedial or compensatory purposes just because students demonstrated low achievement. Given achievement differences commonly noted between white and non-white minority children, for example, and the relatively larger incidence of minority children in the small class category in this sample (see e.g., Table 9), it is indeed possible that many low-achieving students were disproportionately represented in small classes. In this case it would not be surprising that our smallest enrolled classrooms would show the weakest profile of ISTEP+ achievement.

Although we are unable to explore this question with precision, we did analyze the data with a more differentiated grouping of small class enrollment under the assumption that if classroom grouping was guided by children's individual differences (e.g, remedial and gifted programs), then differences in achievement would be most pronounced among children in classrooms with the smallest enrollment. Hence we further divided children in our 12-17 classroom grouping into classrooms of 12-14, and 15-17, and compared their performance against the remaining classroom groupings (18-21, 22-26, and > 26). Our results are reported in Table 16.

Table 16
Means and Standard Deviations of Achievement Scores, by Differentiated Small Class Size

	Reading		**Language**		**Math**		**Total**	
Class Size	Mean	SD	Mean	SD	Mean	SD	Mean	SD
12-14 N = 87	51.29	17.16	59.93	18.78	65.11	20.37	58.82	16.14
15-17 N = 991	53.42	16.44	58.43	18.78	61.42	21.70	57.84	16.10
18-21 N = 4557	56.85	16.92	60.96	19.33	63.13	21.56	60.39	16.53
22-26 N = 4848	58.66	17.21	62.11	19.53	65.65	22.29	62.16	17.07
> 26 N = 415	60.21	17.38	63.49	19.93	66.87	22.44	63.50	17.22

It is apparent that the smallest class grouping in this analysis (12-14) is not associated with a uniformly poor profile of achievement. Although it reports the lowest mean score in reading, it shows broadly comparable achievement with other categories of class size in language and in the total composite score, and strong achievement in math. Hence this data would not support a claim that the smallest classes in our sample were primarily remedial or compensatory (and hence the association of better achievement with higher class enrollment). We explored this further in Table 17, which included Prime Time assistance as a classification variable

Table 17
Mean ISTEP+ Achievement by Class Size and Prime Time Status

	Composite ISTEP+ Achievement (NCE)							
Class Size/ PT Status	Reading		Language		Math		Total	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
12-14								
PT Assistant	–	–	–	–	–	–	-	-
Regular	51.29	17.16	59.93	18.78	65.11	20.37	58.82	16.14
15-17								
PT Assistant	57.16	16.80	60.98	20.02	60.73	21.98	59.59	17.43
Regular	52.80	16.28	58.06	18.59	61.67	21.70	57.61	15.92
18-21								
PT Assistant	58.71	16.52	61.48	19.30	64.81	21.25	61.72	16.22
Regular	56.54	16.94	60.93	19.32	62.88	21.58	60.20	16.55
22-26								
PT Assistant	59.42	16.92	62.58	19.42	66.47	22.07	62.84	16.82
Regular	57.49	17.67	61.22	19.72	64.16	22.62	60.97	17.49
≥ 26								
PT Assistant	59.71	16.52	63.95	20.03	68.00	22.35	63.91	17.07
Regular	61.24	19.07	62.54	19.76	64.58	22.54	62.65	16.83

There is, of course, no Prime Time assistant associated with classrooms as small as 12-14 pupils. But for the remaining groups, it is clear that the pattern noted in Table 13 and Table 14 is present here as well. In virtually every comparison, students with a Prime Time instructional assistant report higher mean achievement scores than students without an assistant. The advantage is particularly striking for reading in small classes of 15-17.

Quantitative Analyses: Hierarchical Linear Modeling

The descriptive analyses reported above appear to demonstrate a clear achievement advantage in classrooms that are assisted by Prime Time instructional assistants, in contrast to classrooms that are not assisted by Prime Time assistants. Moreover, the apparent benefit of Prime Time assistants is not associated with class size. Indeed, the descriptive analyses do not show that achievement is better in classrooms with smaller enrollment, although this latter finding will require further qualification.

These conclusions are only tentative, of course, insofar as the pattern of mean differences do not adequately reflect the influence of qualifying variables, such as race and socioeconomic status. Until recently it was common to use ordinary least squares multiple regression or analysis of variance procedures to analyze questions of this sort. But these analyses require a number of assumptions that are implausible given the multi-level nature of the school achievement data. Empirical studies of the influence of classroom or school variables on student achievement are inherently hierarchical in nature just because individual students are nested within classrooms which are nested within schools which are nested within districts and regions, and so on (Lee, 2000). This nested, multi-level nature of the data cannot, however, be exploited by ordinary least squares analysis which focuses on just a single level (e.g, at the level of the individual, or the level of the school).

For example, to analyze data at the individual level requires the assumption that observations are independent, but there are clear dependencies among students in the same class (leading to a mis-estimation of standard errors and a mis-estimation of the statistical significance of the group level effect). To analyze data at the level of the classroom or school requires an aggregation that ignores substantial individual variability. Moreover, a variable can take on different meaning at different levels of aggregation (*aggregation bias*). What's more, the relationship between student characteristics and achievement may vary across schools as a function of group level variables (*heterogeneity of regression slopes*) that may be of substantial theoretical interest (although it is typically treated as a nuisance in ordinary least squares analyses).

Hierarchical linear modeling (HLM) is a set of procedures that allows one to consider multiple units of analysis in multi-level, hierarchical data structures. "With hierarchical linear models, each of the levels in this structure is formally represented by its own submodel. These submodels express relationships among variables within a given level, and specify how variables at one level influence relations occurring at another (Bryk & Raudenbush, 1992, p. 4).

The Level 1 variables considered in this project were variables associated with <u>students</u>: gender and race. Preliminary analyses did not indicate any effects attributable to gender. Hence only race was included as a Level 1 variable. Our analysis of race was restricted to those categories with sufficiently large N (white/Caucasian, African-American, Hispanic), excluding Asian (N = 62), American Indian (N = 16) and Multi-Racial (N = 188) students. Level 2 variables included variables associated with the <u>classroom</u>: class size (and pupil-teacher ratio) and presence or absence of a Prime Time instructional assistant. Socio-economic status (SES) was modeled as a Level 3 <u>school</u> variable. The class size variable in these analyses is enrollment, or the number of students in a classroom, rather than the class size groupings noted in Table 8. The criterion variables in all analyses were the NCE composite scores for reading, language, mathematics, and the total NCE composite score.

The first set of analyses below includes class enrollment (class size) and the presence or absence of a Prime Time instructional assistant as the classroom variables of interest. In subsequent analyses we replaced class enrollment with pupil-teacher ratio.

Test of the General Models: Reading. A summary of the test of fixed effects for the composite NCE achievement scores for Reading is reported in Table 18. As can be seen, significant effects were evident for SES (higher reading achievement was associated with higher socio-economic status) and race (higher reading achievement was evident in white students). A Race by Class Enrollment interaction was also significant, indicating that white pupils did better in larger classrooms while minority (black and Hispanic) pupils did better in smaller classrooms. Note that the Prime Time Assistant x SES interaction was also marginally significant. This suggests that the influence of a Prime Time instructional assistant on reading achievement was contingent on the SES status of the school. Children in higher SES schools profited more by the presence of an instructional assistant than did children in lower SES schools.

Table 18
Summary of Effects: HLM Tests of the Model for Composite NCE Reading Scores, by Race, Class Enrollment, Presence or Absence of a Prime Time Assistant, and SES

Fixed Effects	Coefficient	Standard Error	t (159)	p-value
Intercept	56.7476	.4185	135.60	.000**
SES	.1466	0.020	7.218	.000**
Class Enrollment	.4496	0.134	3.359	.001**
Class Enrollment x SES	-.0019	0.006	-.320	.749
PT Assistant	-1.349	0.878	-1.535	.124
PT Assistant x SES	.0858	0.047	1.797	.072*
Race	6.727	0.091	7.35	.000**
Race x SES	.0194	0.034	.575	.565
Race x Class Enrollment	.7020	0.368	1.907	.056**
Race x Class Enrollment x SES	.0036	0.010	.351	.725
Race x PT Assistant	-.8413	2.017	-0.417	.676
Race x PT Assistant x SES	-.059	.0835	-.706	.40

Table 19
Unadjusted Means and Standard Deviations for NCE Composite Reading Scores, by Race and Class Enrollment

	White		Total Non-White Minority		Black		Hispanic	
Class Size	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
12-17	55.05	16.45	47.54	14.86	47.97	15.44	46.72	13.73
18-21	58.19	16.66	47.25	15.36	46.39	15.75	49.99	13.73
22-26	59.80	16.96	48.58	16.13	47.88	16.43	51.29	14.91
> 26	61.46	17.33	48.51	11.47	49.13	10.88	48.09	12.09

Table 19 reports mean NCE composite reading scores, by race and class enrollment. As can be seen in Table 19, there is a monotonic increase in reading achievement scores with increasing class size for white pupils. In contrast, reading scores for non-white minority students are more stable (or show slight fluctuation) across different class sizes. See Table 24 for additional evidence regarding the benefits of small classes for African-American pupils, relative to white (and Hispanic) students.

Test of the General Models: Language. A summary of the test of fixed effects for the composite NCE achievement scores for Language is reported in Table 20.

As can be seen, significant effects were evident for SES (higher language achievement was associated with higher socio-economic status) and race (higher reading achievement was evident in white students). A significant interaction for the presence or absence of Prime Time assistants and SES was also observed. Hence the influence of a Prime Time assistant on achievement was contingent on the SES status of the school. Children in higher SES schools profited more by the presence of an instructional assistant than did children in lower SES schools.

Table 20
Summary of Effects: HLM Tests of the Model for Composite NCE Language Scores, by Race, Class Enrollment, Presence or Absence of a Prime Time Assistant, and SES

Fixed Effects	Coefficient	Standard Error	t (159)	p-value
Intercept	60.7946	.4810	126.379	.000*
SES	.1298	0.023	5.55	.000*
Class Enrollment	.1839	0.1582	1.162	.246
Class Enrollment x SES	.0039	0.007	.551	.581
PT Assistant	-1.3885	1.0049	-1.382	.167
PT Assistant x SES	.1108	0.0547	2.023	.043*
Race	6.4219	1.0698	6.003	.000*
Race x SES	.0060	0.0399	.151	.881
Race x Class Enrollment	.4422	0.4391	1.007	.314
Race x Class Enrollment x SES	.0115	0.0122	.946	.345
Race x PT Assistant	-1.608	2.474	-0.650	.515
Race x PT Assistant x SES	-.1189	.1034	-1.150	. 250

Test of the General Models: Mathematics. A summary of the test of fixed effects for the composite NCE achievement scores for Reading is reported in Table 21. As can be seen, significant effects were evident for SES (higher reading achievement was associated with higher socio-economic status), race (higher reading achievement was evident in white students) and class enrollment (higher mathematics achievement was associated with classrooms of increasing enrollment). Note that the Class Enrollment x SES interaction was marginally significant, indicating that the achievement benefits of higher class enrollment was more likely to be evident in higher SES schools.

Table 21
Summary of Effects: HLM Tests of the Model for Composite NCE Mathematics Scores, by Race, Class Enrollment, Presence or Absence of a Prime Time Assistant, and SES

Fixed Effects	Coefficient	Standard Error	t (159)	p-value
Intercept	63.6708	.6584	96.707	.000*
SES	.1245	0.0319	3.907	.000*
Class Enrollment	.3738	0.1861	2.009	.044*
Class Enrollment x SES	.0147	0.0083	1.770	.076*
PT Assistant	-1.3993	1.3694	-1.022	.307
PT Assistant x SES	.0894	0.0760	1.175	.240
Race	7.4409	1.1954	6.225	.000*
Race x SES	.0222	0.0454	.490	.624
Race x Class Enrollment	.4119	0.4979	.827	.408
Race x Class Enrollment x SES	.0105	0.0141	.744	.457
Race x PT Assistant	-.5488	2.8112	-0.195	.846
Race x PT Assistant x SES	-.0811	.1188	-.683	. 494

Test of the General Models: Total NCE Composite. A summary of the test of fixed effects for the total NCE composite score is reported in Table 22. As with previous analyses, significant effects were evident for SES (higher total achievement was associated with higher socio-economic status), race (higher achievement was evident in white students) and Class Enrollment (higher total achievement was associated with classrooms of increasing enrollment). The Prime Time Assistant x SES interaction was also marginally significant, indicating that the influence of a Prime Time assistant on achievement was contingent on the SES status of the school. Children in higher SES schools profited more by the presence of an instructional assistant than did children in lower SES schools.

Table 22
Summary of Effects: HLM Tests of the Model for Total Composite Scores, by Race, Class Enrollment, Presence or Absence of a Prime Time Assistant, and SES

Fixed Effects	Coefficient	Standard Error	t (159)	p-value
Intercept	60.4642	.4882	123.85	.000*
SES	.1316	0.0234	5.619	.000*
Class Enrollment	.3509	0.1468	2.391	.017*
Class Enrollment x SES	.0049	0.0065	.767	.443
PT Assistant	-1.3825	.9742	-1.419	.156
PT Assistant x SES	.0959	0.0535	1.192	.073*
Race	6.9749	.9333	7.473	.000*
Race x SES	.0062	0.0358	.172	.864
Race x Class Enrollment	.4337	0.3850	1.127	.260
Race x Class Enrollment x SES	.0064	0.0108	.591	.554
Race x PT Assistant	-.4426	2.202	-0.201	.841
Race x PT Assistant x SES	-.0978	.0931	-1.051	. 294

Summary. Table 23 reports a summary of the significant effects that emerged in the previous analyses. (See Table 32 for a summary of effect sizes across all main effects). It is clear that the pattern of student performance on the third-grade ISTEP+ test is uniformly affected by SES and by Race. Schools in higher SES categories, and white students tended to report higher

Table 23
Summary of Significant Effects Across HLM Analyses of NCE Reading, Language, Mathematics and Total NCE Composite Achievement Scores, Race, Class Enrollment, Presence or Absence of a Prime Time Assistant, and SES

	Significant Effects					
NCE Scores	SES	Race	Class Enrollment (CE)	PT Assistant x SES	CE x SES	Race x CE
Reading	X	X		X		X
Language	X	X		X		
Mathematics	X	X	X		X	
Composite	X	X	X	X		

ISTEP+ achievement scores. Class size was a significant predictor of mathematics achievement, and of the total NCE composite score. In both cases, higher mathematics and composite achievement was associated with larger classes. Several significant interactions emerged as well. In three analyses (for Reading, Language, and the total NCE composite score) the presence or absence of a Prime Time instructional assistant interacted with SES, although the effects were often of marginal significance. In all cases higher achievement was associated with higher SES and the presence of a Prime Time instructional assistant. Two additional interactions were observed: class enrollment and SES (better mathematics achievement associated with higher SES and larger classrooms) and race x SES (better reading achievement was associated with higher SES and white pupils).

Supplementary Analyses Within Race. As Table 23 illustrates, there were pervasive racial differences in the pattern of achievement. Moreover, race interacted with class size, at least in the case of reading achievement. Similar effects have been previously reported in the class size reduction literature. Robinson and Wittebols (1986) have shown, for example, that minority youngsters might differentially profit from class size reduction in the early primary grades, as opposed to white pupils. In light of these considerations we elected to analyze achievement data within racial categories.

Table 24 reports means and standard deviations for the NCE total composite by race and class size. (These means are not adjusted for SES). We note, first of all, that this pattern of achievement differences by racial category is typically reported in the achievement literature. Generally, on average, Asian children outperform white/Caucasian children, who outperform Hispanic and African-American children. Second of all, it would appear that the influence of class size on achievement varies by racial category. African-American students, for example, show the best achievement in the small classes (12-17), while Caucasian (and Hispanic) students show better achievement in larger classes. Consequently, in subsequent analyses, we explored the relationship between classroom (class size, presence of a Prime Time instructional assistant) and school (SES) variables on ISTEP+ achievement) within White and Black students. For reasons of economy we restrict our analysis to the total composite NCE score.

Table 24
Means and Standard Deviations of NCE Total Composite Scores, By Racial Category and Class Size

	African-American		**Asian**		**Caucasian**		**Hispanic**	
Class Size	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std. Dev.
12 - 17	53.59	1.31	80.25	8.21	59.70	.580	50.61	1.81
18 - 21	51.16	.79	72.90	2.95	61.51	.266	54.53	1.42
22 - 26	52.67	.84	81.48	3.28	63.17	.250	54.40	1.57
> 26	50.73	4.24	74.50	11.61	64.80	.862	53.22	3.50

Test of Achievement Effects: White and Black Students. We conducted subsequent HLM analyses within the samples of white and black students. Table 25 reports the final estimate of fixed effects (with robust standard errors) for white students. Table 26 reports the same analysis for black students. In both analyses SES and Class Enrollment are significant predictors of student achievement. But the effect of class size cuts in different directions for white and black pupils. White students tend to do better in classrooms with larger enrollment, whereas the opposite effect is found for black students. That is, black students report better achievement in classrooms with smaller enrollment. White students report better achievement in classrooms with larger enrollment. A Prime Time Assistant by SES interaction was also evident in the black student sample..

Table 25
Estimate of Fixed Effects (with robust standard errors): White Students

Fixed Effect	Coefficient	Standard Error	t-ratio	d.f.	p-value
Intercept (grand mean)	61.5265	.4947	124.37	156	.000*
SES	.1245	.0292	4.257	156	.000*
Class Enrollment	.3869	.1528	2.53	156	.012*
Class Enrollment x SES	.0102	.0093	1.086	156	.278
PT Assistant Status	-1.27	.9697	-1.31	156	.190
PT Assistant Status x SES	.0754	.0639	1.178	156	.239

Table 26
Estimate of Fixed Effects (with robust standard errors): Black Students

Fixed Effect	Coefficient	Standard Error	t-ratio	d.f.	p-value
Intercept (grand mean)	53.600	.9148	58.590	960	.000*
SES	.1204	.0265	4.55	74	.000*
Class Enrollment	-.6728	.3211	-2.09	74	.036*
Class Enrollment x SES	-.0089	..0073	-1.226	74	.221
PT Assistant Status	-.3372	2.28	-.148	74	.813
PT Assistant Status x SES	.1985	.0784	2.53	74	.012*

Analysis of Pupil-Teacher Ratio. In the previous analyses we treated class enrollment as the Level 2 classroom variable of interest, largely because of its prominence in the class size reduction literature. However, we also wanted to examine the influence of pupil-teacher ratio (PTR) on student achievement, insofar as the Prime Time strategy is largely one of reducing PTR in the interest of increasing student achievement. We adopted the HLM strategy in these analyses as well, which are summarized in Table 27(NCE Reading), Table 28 (NCE Language), Table 29 (NCE Mathematics) and Table 30 (NCE Total Composite Score). Table 31 reports a summary of these analyses.

Table 27
Summary of Effects: HLM Tests of the Model for Reading, by Race, Pupil-Teacher Ratio (PTR), Presence or Absence of a Prime Time Assistant, and SES

Fixed Effects	Coefficient	Standard Error	t (159)	p-value
Intercept	56.7354	.4104	138.251	.000*
SES	.1412	.0204	6.923	.000*
Pupil-Teacher Ratio (PTR)	.5160	.1312	3.933 (573)	.000*
PTR x SES	.0016	.0061	.270 (573)	.787
PT Assistant	.6054	.8466	.715 (573)	.474
PT Assistant x SES	.0594	.0472	1.259 (573)	.208
Race	6.969	.8936	7.799	.000*
Race x SES	.0196	.0314	.624	.532
Race x PTR	.5836	.3687	1.583 (583)	.113
Race x PTR x SES	.5837	.3688	.635 (573)	.525
Race x PT Assistant	2.6763	1.8566	1.441 (573)	.149
Race x PT Assistant x SES	-.0551	.0749	-.736 (573)	.462

Table 28
Summary of Effects: HLM Tests of the Model for NCE Language, by Race, Pupil-Teacher Ratio (PTR), Presence or Absence of a Prime Time Assistant, and SES

Fixed Effects	Coefficient	Standard Error	_t_ (159)	_p_-value
Intercept	60.814	.4886	124.462	.000*
SES	.1229	.0235	5.226	.000*
Pupil-Teacher Ratio (PTR)	.1970	.1677	1.175	.240
PTR x SES	.0051	.0075	.682	.495
PT Assistant	-.5685	.9742	-.584	.559
PT Assistant x SES	.1168	.0528	2.211	.027
Race	6.452	1.077	5.990	.000*
Race x SES	.0006	.0401	.014	.989
Race x PTR	.4030	.4596	.877	.381
Race x PTR x SES	.0166	.0126	1.312	.190
Race x PT Assistant	.0253	2.362	.011	.992
Race x PT Assistant x SES	-.0769	.1017	-.756	.450

Table 29
Summary of Effects: HLM Tests of the Model for NCE Mathematics, by Race, Pupil-Teacher Ratio (PTR), Presence or Absence of a Prime Time Assistant, and SES

Fixed Effects	Coefficient	Standard Error	t (159)	p-value
Intercept	63.7160	.6567	97.012	.000*
SES	.1187	.0318	3.726	.000
Pupil-Teacher Ratio (PTR)	.4049	.2029	1.995	.046*
PTR x SES	.0159	.0089	1.730	.083
PT Assistant	.3469	1.309	.265	.791
PT Assistant x SES	.1551	.0712	2.144	.032*
Race	7.5866	1.231	6.160	.000*
Race x SES	.0211	.0464	.454	.649
Race x PTR	.0512	.5252	.098	.923
Race x PTR x SES	.0097	.0148	.654	.513
Race x PT Assistant	.1815	2.703	.067	.947
Race x PT Assistant x SES	-.0419	.1164	-.360	.719

Table 30
Summary of Effects: HLM Tests of the Model for Total Composite Scores, by Race, Pupil-Teacher Ratio (PTR) Presence or Absence of a Prime Time Assistant, and SES

Fixed Effects	Coefficient	Standard Error	t (159)	p-value
Intercept	60.5108	.4863	124.420	.000*
SES	.1252	.2339	5.35	.000*
Pupil-Teacher Ratio (PTR)	.3763	.1553	2.423	.016*
PTR x SES	.0052	.0068	.763	.446
PT Assistant	.0703	.9390	.075	.941
PT Assistant x SES	.1131	.0514	2.199	.028*
Race	6.9923	.9474	7.381	.000*
Race x SES	.0016	.0360	.045	.964
Race x PTR	.3295	.3988	.826	.409
Race x PTR x SES	.0093	.0112	.832	.406
Race x PT Assistant	1.002	2.071	.483	.628
Race x PT Assistant x SES	-.0674	.0906	-.744	.457

Table 31
Summary of Significant Effects Across HLM Analyses of NCE Reading, Language, Mathematics and Total NCE Composite Achievement Scores, by Race, Pupil-Teacher Ratio (PTR) Presence or Absence of a Prime Time Assistant, and SES

	Significant Effects			
NCE Scores	SES	Race	Pupil-Teacher Ratio (PTR)	PT Assistant x SES
Reading	X	X	X	
Language	X	X		X
Mathematics	X	X	X	X
Composite	X	X	X	X

As these tables make clear, the general profile of achievement effects for PTR are strikingly similar to the results for classroom enrollment. As expected, achievement scores for reading, language, mathematics and the total composite score are strongly influenced by SES and by race.

Similarly, the PT Assistant x SES interaction was significant for 3 of the 4 comparisons (language, mathematics and the composite score), indicating that better achievement was associated with classrooms with Prime Time instructional assistants in schools of higher socio-economic status. Finally, PTR is a significant predictor of reading, mathematics and the composite score. As with class enrollment, this effect suggests that better achievement is associated with larger pupil-teacher ratio.

<u>General Summary of Coefficients for Main Effects</u>. Table 32 presents a summary of coefficients across all analyses of the main effects SES, race, class enrollment, the presence or absence of a Prime Time instructional assistant, and pupil-teacher ratio. These coefficients can be meaningfully interpreted as effect sizes. For example, a percentage increase in school-wide SES results in a .1316 increase in NCE total composite achievement. A white student's predicted NCE total composite achievement score is 1.66 points above the mean intercept score of 60.46 (or 62.12); a black student's predicted score is 5.24 points below the intercept (or 55.22). Adding one student to a class increases average NCE total composite achievement .35 points (or, alternatively, adding 10 students to a class raises average achievement 3.5 points). The presence of a Prime Time instructional assistant lowers average NCE total composite achievement 0.92 points; the absence of a Prime Time instructional assistant increases average NCE total composite achievement 0.45 points (although this main effect was not statistically significant for any comparison). Finally, a unit increase in pupil-teacher ratio was associated with higher average NCE total composite achievement of .3763 points.

<u>General Summary of Coefficients for Interactive Effects</u>. Table 33 presents a summary of coefficients across all analyses for interactive effects with SES. As this table illustrates, there is a general absence of significant interactions with SES, particularly with respect to class enrollment (with the possible exception of mathematics, $p < .076$) and with pupil-teacher ratio (with the possible exception of language, $p < .08$). Hence, an increase of one student is associated with an increase of .0147 points in mathematics in higher SES schools; a unit increase in pupil-teacher ratio is associated with an increase of .0159 points in mathematics in higher SES schools. These effects are quite trivial, however. In contrast, the Prime Time Assistant x SES interaction was statistically significant for language, and marginally significant for reading and for the total composite. In all cases the benefits of a Prime Time instructional assistant on achievement was more pronounced in schools with higher SES.

Table 34 presents a summary of coefficients across all analyses for interactive effects with race. Most effects are not statistically significant, with the exception of the Class Enrollment x Race interaction for reading.

Table 32
General Summary of Coefficients for Main Effects Across All Analyses

Main Effects	Coefficient	p-value	Table Number
Socio-Economic Status			
Reading	.1466	.000	18
Language	.1298	.000	20
Mathematics	.1245	.000	21
Composite	.1316	.000	22
Race			
Reading	6.727	.000	18
Language	6.422	.000	20
Mathematics	7.440	.000	21
Composite	6.975	.000	22
Class Enrollment			
Reading	.4496	.001	18
Language	.1839	.246	20
Mathematics	.3738	.044	21
Composite	.3509	.017	22
Prime Time Assistant			
Reading	-1.349	.124	18
Language	-1.388	.167	20
Mathematics	-1.399	.307	21
Composite	-1.382	.156	22
Pupil-Teacher Ratio			
Reading	.5160	.000	27
Language	.1970	.240	28
Mathematics	.4049	.046	29
Composite	.3763	.016	30

Table 33
General Summary of Coefficients for Interactive Effects with SES, Across all Analyses

Interactive Effects	Coefficient	p-value	Table Number
Class Enrollment x SES			
Reading	-.019	.749	18
Language	.0039	.551	20
Mathematics	.0147	.076	21
Composite	.0049	.443	22
Prime Time Asistant x SES		.	
Reading	.0858	.072	18
Language	.1108	.043	20
Mathematics	.0894	.240	21
Composite	.0959	.073	22
Pupil-Teacher Ratio x SES			
Reading	.0016	.787	27
Language	.0051	.495	28
Mathematics	.0159	.083	29
Composite	.0052	.446	30

Table 34
General Summary of Coefficients for Interactive Effects with Race, Across all Analyses

Interactive Effects	Coefficient	p-value	Table Number
Class Enrollment x Race			
Reading	.7020	.056	18
Language	.4422	.314	20
Mathematics	.4419	.408	21
Composite	.4337	.260	22
Prime Time Assistant x Race			
Reading	-.8413	.676	18
Language	-1.608	.515	20
Mathematics	-.5488	.846	21
Composite	-.4426	.841	22
Pupil-Teacher Ratio x Race			
Reading	.5836	.113	27
Language	.4030	.381	28
Mathematics	.0512	.923	29
Composite	.3295	.409	30

Chapter 5

Conclusion

The purpose of this project was to examine patterns of achievement in third-grade ISTEP+ scores as a function of class size, the presence or absence of Prime Time instructional assistants, and pupil-teacher ratio. To this end we analyzed the Normal Curve Equivalent scores (for language, mathematics, reading, and the NCE composite) of a stratified random cluster sample of nearly 11,000 third-graders. Two kinds of analyses were conducted. We first described the pattern of achievement, broken down by demographic and classroom variables. We then attempted to predict student achievement by testing a series of hierarchical linear models.

The descriptive results (which do not control for covariates) show clearly that achievement scores are higher in Prime Time assisted classrooms than in classrooms without Prime Time instructional assistants, particularly in reading. A similar pattern is evident when the presence or absence of a Prime Time instructional assistant is examined across urban, suburban, township and rural school corporations. In nearly every comparison, in every geographic category, achievement scores were higher in classrooms with a Prime Time assistant than in unassisted classrooms, with the Prime Time advantage particularly evident among children in suburban and township schools. Hence it would appear that student achievement is consistently and pervasively associated with the presence of Prime Time instructional assistants.

Although this pattern is encouraging, we hasten to add that it invariably masks important sources of variability, such as socioeconomic status and student race. Indeed, the aim of the HLM analyses was to disentangle nested effects in order derive a more precise estimate of the predictive relationship between classroom variables (class size, presence or absence of a Prime Time assistant, PTR) and student achievement.

As expected, student achievement in reading, language and mathematics (and the composite score) was strongly influenced by socioeconomic status and by race. Schools in higher SES categories, and white students, reported higher ISTEP+ achievement scores. Socioeconomic status was not only a uniformly significant main effect, it also interacted with the presence or absence of a Prime Time assistant (better achievement was associated with higher socioeconomic status and the presence of an instructional assistant), with class size (mathematics achievement was associated with higher SES and larger class size), and with race (reading achievement was associated with higher SES and white students). To put it differently, the presence of a Prime Time instructional assistant was a stronger predictor of

Class size was not a significant predictor of reading or language achievement, although it did predict mathematics achievement, and the composite score, but in the opposite direction. That is, higher mathematics and composite achievement was associated with larger, not smaller, class size. But this result must be interpreted with caution. It does not support any conclusion to the effect that class enrollments should be allowed to swell in order to improve ISTEP achievement scores, for two reasons. First, the apparent benefits of larger class size effects were not evident for reading and language. Second, the class size effect interacted with other variables, and hence any general conclusion about class size must be qualified by reference to these interactions. For example, class size interacted with SES, indicating that the apparent benefits of larger classes for mathematics achievement was evident only in higher SES schools (and not lower SES schools). Moreover, class size also interacted with race, which introduces a significant qualification that should be emphasized.

For example, the class size reduction literature has often shown that the benefits of small class size may be greater for minority children in urban school settings. In light of this literature, and in light of the race effect noted earlier, we conducted a set of HLM analyses within samples of white and black students. We found that SES and class enrollment were significant predictors of student achievement, although the influence of class enrollment cuts in different directions for black and white pupils. For example, white students tended to do better in classrooms with larger enrollments, whereas black students tended to do better in classrooms with smaller enrollments. This pattern broadly supports the previous literature on the differential benefits of class size reduction for white and black students.

Hence, although we do find some evidence that higher class enrollment is associated with better achievement in mathematics (and on the composite score), there are two important qualifications. First, this effect may be mostly restricted to higher SES schools. Second, this effect does not hold for minority pupils, who report better achievement in smaller classes, a finding that has been reported elsewhere in the literature (e.g., Robinson & Wittebols, 1986).

Although the Prime Time funding formula also underwrites the hiring of teachers, the "Prime Time strategy" is often considered an intervention that attempts to reduce pupil-teacher ratio rather than class size, and to do so by hiring instructional assistants (in addition to hiring teachers). In no analysis did the presence or absence of a Prime Time assistant predict achievement as a main effect, although it did emerge in a number of significant interactions, as we have noted. In contrast, pupil-teacher ratio was a significant predictor of reading, mathematics and the composite achievement score, but in the opposite direction. That is, better achievement in these domains was evident in classrooms that reported larger pupil-teacher ratio.

Pupil-teacher ratio (PTR) has a controversial status in the class size reduction literature. For example, although the addition of paraprofessionals to the classroom is often touted as a low-cost alternative to class size reduction (because it obviates the need to hire more teachers), evidence for the effectiveness of teacher aides for raising student achievement is "bleak" (Finn, Gerber, Farber & Achilles, 2000). Indeed, Finn et al. (2000, p. 152) concluded, on the basis of an HLM analysis, that the "addition of a teaching assistant to a primary-grade classroom does not affect students' achievement any differently from classes of similar size without an aide." Moreover, much like the present study, Finn et al. (2000) also reported that, in some instances, students in aide classes performed more poorly than students without aides. We should note that pupil-teacher ratio was inferred, in the Finn et al. (2000) study, from a dummy contrast (aide, no aide), rather than the FTE (full-time equivalent) approach adopted here.

But Finn et al. (2000) also reported that the presence of an aide (note: rather than pupil-teacher ratio) might have a positive influence on student achievement as a function of *duration*. For example, in their data, first-grade students who had an aide for one year score nearly 6 points higher than first-grade students without an aide on the SAT reading test. Students who had an aide for two years scored 11 points higher than students without aides. As the authors put it, "The differences suggest that duration may be a factor in the effects of teaching assistants on student achievement" (Finn et al., 2000, p. 152).

This illustrates an important caveat that must attend the interpretation of the present data. Students sit for the ISTEP+ in the second week of September of the Fall term. This means that a Prime Time assistant would be active in a classroom for just a few weeks prior to the administration of ISTEP+. This may not allow sufficient time for the benefits of an instructional assistant to be evident within a classroom, given the critical importance of duration noted by Finn et al. (2000).

We conclude then, with the recommendation that future research undertake a longitudinal examination of the influence of Prime Time instructional assistants (and pupil-teacher ratio) on student achievement, in order to more fully assess the role of the duration variable. This would require the utilization of standardized assessments at second grade in classrooms with and without Prime Time assistants, with assessments administered in the fall and spring terms. These students should then be followed into third-grade, and their ISTEP+ scores be included in the analysis.

We also recommend a research strategy that examines the influence of instructional assistants, and other classroom variables, on the achievement of first-grade pupils, given the fact that the benefits of instructional assistants (and class size reduction) might be particularly evident in the early primary grades. Future research should also consider using assessments other than the state-mandated test and similar standardized assessments. This would include, for example, locally-constructed tests of proficiencies. Such an approach is necessary if the Prime Time evaluation is to be extended to first-grade, but these assessments would also be an informative adjunct to standardized achievement data of second- and third-grade pupils.

Finally, the present data would support additional research on the intriguing possibility that a differentiated Prime Time strategy might be appropriate, one that seeks to reduce class size in lower SES schools, but utilizes instructional assistants in larger classrooms in higher SES schools.

References

Achilles, C.M. (1997). Exploring class-size research issues. *The School Administrator, 54*, 1-3. http://www.aasa.org/SchoolAdmin/oct9702.htm

Achilles, C.M. (n.d.) Issues in the class size and pupil-teacher ratio (PTR) confusion: A classic example of "apples and oranges." Unpublished research synopsis.

Bryk, A. & Raudenbush, S. (1992). *Hierarchical linear models*. Newbury Park: CA: Sage.

Chase, C.I., Mueller, D.J. & Walden, J.D. (1986). Prime Time: Its impact on instruction and achievement. Final Report. Indianapolis: Indiana Department of Education.

Department of Education (May, 1998). *Reducing class size: What do we know? State initiatives.* http://www.ed.gov/pubs/ReducingClass/state.html

Department of Education (May, 1998). *Reducing class size: What do we know? Research on class size.* http://www.ed.gov/pubs/ReducingClass/state.html

Educational Testing Service (1980a). Class size research: A critique of recent meta-analyses. *Phi Delta Kappan, 62*, 239-241.

Educational Testing Service (1980b). *Class size research: A critique of recent meta-analyses.* Arlington, VA.

Evans-Stout, K., Fleming, N., Johnson, Jr. B., Ehleringer, E., Gray, J.M., Merrill, R., Parket, M., Roberts, R., & Stewart, T. (1997). *The use of class size reduction funds in five Wasatch Front Districts.* A study for the Utah Education Consortium. Salt Lake City.

Evertson, C. M. & Randolph, C.H. (1989). Teaching practices and class size: A new look at an old issues. *Peabody Journal of Education, 67*, 85-105

Finn, J.D., Gerber, S.B., Farber, S.L. & Achilles, C.M. (2000). Teacher aides: An alternative to small classes? In M.C. Wang & J.D. Finn (Eds.), *How small classes help teachers do their best* (pp. 131-174). Philadelphia, PA: Temple University Center for Research in Human Development and Education.

Finn, J.D. (1998). *Class size and students at risk: What is known? What is next?* Washington, D.C. : U.S. Department of Education.

Finn, J.D. & Achilles, C.M. (1998). *Tennessee's class-size study: Questions answered, questions posed.* Presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Finn, J.D, Fulton, D., Zaharias, J. & Nye, B.A. (1989). Carry-over effects of small classes. *Peabody Journal of Education, 67*, 75-84.

Finn, J.D. & Achilles, C.M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*, 557-577.

Florida Department of Education. *The relationship of school and class size with student achievement in Florida.* Office of Policy Research. http:/www.firn.edu/doe/bin00048/schosize.htm

Folger, J. & Breda, C. (1989). Evidence from Project STAR about class size and student achievement. *Peabody Journal of Education, 67*, 17-33.

Gilman, D.A. (1988). *PRIME TIME in the first grade at the North Gibson School Corporation: The first four years. A longitudinal evaluation of Indiana's state-supported reduced class size program.* ERIC Document Reproduction Service No. 310 886.

Gilman, D., Tillitski, C., Swan, E. & Stone, W. (1987). *Prime Time at North Gibson School Corporation: A three-year study.* ERIC Document Reproduction Service No. 299 056

Gilman, D., Harder, H. & Tillitski, C. (1988) *Why state sponsored reduced class size programs aren't working: A qualitative research study.* ERIC Document Reproduction Service No. 313 115

Gilman, D.A., Swan, E. & Stone, W. (1988). The educational effects of a state supported reduced class size program. *Contemporary Education, 59,* 112-116.

Glass, G.V. (1979). Does class size make a difference? *Instructor, 89,* 22.

Glass, G.V. (1980). On criticism of our class size/student achievement research: No points conceded. *Phi Delta Kappan, 62,* 242-244.

Glass, G.V. & Smith, M.L. (1978). *Meta-analysis of research on the relationship of class size and achievement.* San Francisco: Far West Laboratory for Educational Research and Development.

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

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

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

Hanushek, E.A. (1989) The impact of differential school expenditures on school performance. *Educational Researcher, 18,* 45-65.

Hanushek, E.A. (with others). (1994). *Making schools work: Improving performance and controlling costs.* Washington, D.C.: Brookings Institution.

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

Hanushek, E.A. & Jorgenson, D.W. (Eds.,1996). *Improving America's schools: The role of incentives.* Washington, D.C., National Academy Press.

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

Hedges, L.V. & Olkin, I. (1980). Vote counting methods in research synthesis. *Psychological Bulletin, 88, 359-369.*

Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

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

Indiana State Department of Public Instruction (1983). *Project Primetime: 1982-83 Report.* ERIC Document 239 765.

Kickbusch, K. (1990). Class sizes. *Educational Issues Series.* Wisconsin Educational Association Council. http//www.weac.org/resource/may96/classize.htm

Kreuger, A.B. (1999) Experimental estimates of education production functions. *Quarterly Journal of Economics, 114,* 497-532.

Lamdin, D.J. (1996). Evidence of student attendance as an independent variable in education production functions. *Journal of Educational Research, 89,* 155-162.

Lapsley, D.K. & Daytner, K.M. (2001a). *An evaluation of Indiana's Prime Time: Final report of the survey results.* Muncie: IN: Ball State University Teachers College and Indiana Department of Education.

Lapsley, D.K. & Daynter, K.M. (2001b). *Indiana's "class size reduction" initiative: Teacher perspectives on training, implementation and pedagogy.* Paper presented at the annual meeting of the American Educational Research Association, April, Seattle.

Lee, V.E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist, 35*, 125-141.

Malloy, L. & Gilman, D. (1989). The cumulative effects of basic skills achievement of Indiana's *Prime Time*: A state sponsored program of reduced class size. *Contemporary Education, 60*, 169-172.; also ERIC Document Reproduction Service No. 300 139.

McDaniel, E. (1985). *Prime Time 1984-1985.* Report submitted to the Indiana Department of Public Instruction.

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

Mitchell, D.E. & Beach, S.A. (1990). How changing class size affects classrooms and students. *Policy Briefs* (No. 12). Far West Laboratory. Office of Educational Research and Development, U.S. Department of Education.

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*, 165-177.

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

Mueller, D.J., Chase, C.I. & Walden, J.D. (1988). Effects of reduced class size in primary grades. *Educational Leadership, 45,* 48-53.

Murnane, R.J. & Levy, F. (1996). Evidence from fifteen schools in Austin, Texas. In *Does money matter*? Washington, D.C.: Brookings Institution Press.

Nye, B. (1998). *The lasting benefits of small classes: A five-year follow-up of students in the Tennessee class size experiment.* Paper presented at the annual meeting of the American Educational Research Association, April 19, Montreal, Quebec, Canada.

Nye, B., 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*, 123-151.

Odden, A. (1990). Class size and student achievement: Research-based policy alternatives. *Educational evaluation and Policy Analysis, 12*, 213-227.

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

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

Robinson, G.E. (1990). Synthesis of research on the effects of class size. *Educational Leadership, 47*, 80-90.

Robinson, G.E. & Wittebols, J.H. (1986). *Class size research: A related cluster analysis for decision-making.* Arlington, VA: Educational Research Service.

Sanogo, Y. & Gilman, D.A. (1994). *Class size and student achievement: Tennessee's*

STAR and Indiana's Prime Time projects. ERIC Document Reproduction No. 370 680

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*, 141-152.

Smith, M.L. & Glass, G.V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal, 17*, 419-433.

Stasz, C. & Stecher, B.M. (2000). Teaching mathematics and language arts in reduced size and non-reduced size classrooms. *Educational Evaluation and Policy Analysis, 22*, 313-329.

Sturm, P. (1997). *Nevada's class-size reduction program: Background paper 97-7.* Senate Committee on Human Resources. Available from: http://www.leg.state.nv.us/lcb/research/Bkground/97-07.HTM

Swann. E., Stone, W. & Gilman, D.A. (1987). The educational effects of a state supported reduced class size program. *ERS Spectrum, 5(4)*, 20-23.

Tomlinson, T.M. (1990). Class size and public policy: The plot thickens. *Contemporary Education, 62, 17-23.*

Varble, M.E. (1990). Smaller class size = higher achievement scores? *Contemporary Education, 62*, 38-45.

Weiss, T. (1990). Indiana's *Prime Time*. *Contemporary Education, 62*, 31-32.

WestEd (1998). Class size reduction: Lessons learned from experience. *Policy Brief, 23*, 1-12.

Word, E., Achilles, C.M., Bain, H., Folger, J., Johnston, J. & Lintz, N. (1990). Project STAR final executive summary: Kindergarten through third grade results (1985-1989). *Contemporary Education, 62,* 13-16.