HAS NO CHILD LEFT BEHIND ENHANCED SCHOOL EFFICIENCY?

By

Meisha Fang

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Leadership and Policy Studies

May, 2010

Nashville, Tennessee

Approved:

Professor R. Dale Ballou

Professor Stephen N. Elliott

Professor Thomas M. Smith

Professor Matthew G. Springer

ACKNOWLEDGEMENTS

I would like to acknowledge many people who have made this dissertation possible. First and foremost, I owe my deepest gratitude to my advisor, Prof. R. Dale Ballou, for his unlimited patience and sage advice throughout my time in graduate school. Thank you for always being available whenever I needed help. Your guidance, encouragement, and support enabled me to make it through this long journey. I consider myself very lucky to have had you as a mentor and role model.

I would also like to thank the other faculty at Vanderbilt who helped me throughout my years in graduate school. I want to thank Prof. Cliff Huang from whom I have learned the essential techniques to conduct the analysis in this dissertation. I am especially grateful to Prof. Steve Elliot, Thomas Smith, and Matthew Springer for their help throughout the dissertation process. Your feedback has been truly invaluable to me.

I must also thank my beloved husband, Wenjun Guo, and my son, Xinche Guo for their persistent love, patience, and support in every step along the way. Last but not least, I thank my parents, 方日品 and 李兰娥, for their unconditional love and support. I never would have made it here without you as my foundation. 爸爸，妈妈，我爱你们。

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

This dissertation investigates whether No Child Left Behind (NCLB) has made public schools operate more efficiently. The interest in measuring technical efficiency in education dates back to the beginning of the twentieth century when educators began to adopt business values and practices into education (Callahan, 1962). In the 1990s an increasing number of researchers began to employ advanced measurement techniques to empirically estimate technical efficiency when a debate began on whether money matters in improving student achievement led by Hanushek, et al. (1996) and Greenwald, et al. (1996). The final conclusion of this debate was that there was no strong evidence to support the hypothesis that educational inputs, such as educational expenditure, have positive impact on educational outcomes, such as students' test scores. Thus, the lack of conclusive evidence on the positive relationship between educational expenditure and student achievement, together with the fact that the substantial rise of educational expenditure over time does not correspond to an equal increase of student test scores (Lips, Watkins, & Fleming, 2008), made some scholars argued that resources in schools are not utilized efficiently (Hanushek, 1986).

During this period, researchers began to adopt measurement methods that are found to adequately estimate inefficiency in private sectors and apply them to public sectors, such as local governments (Geys & Moesen, 2009), hospitals (Ozcan, Luke, & Haksever, 1992), police departments (Parks, 2005), as well as education (Bessent, Bessent, Kennington, & Reagan, 1982; Chakraborty, Biswas, & Lewis, 2001; Grosskopf, Hayes, Taylor, & Weber, 1997; Noulas & Ketkar, 1998). No matter what types of measurement methods are used, parametric or non-parametric, empirical studies reached a similar conclusion that production in educational settings was inefficient. In other words, those studies suggested that schools (or school districts), the unit of producing educational outcomes, did not use resources to their optimal level. For instance, school districts in New Jersey were found to be 81% efficient (Noulas & Ketkar, 1998), implying that school districts could have achieved 19% more outputs if they utilized their resources to the optimal level. Similar results were found for school districts in Utah whose average technical efficiency was 86% (Chakraborty, Biswas, & Lewis, 2001).

Despite the finding that inefficiency in education exists, few studies have offered the determinants of inefficiency, or suggested the remedies for increasing technical efficiency of public education. Some scholars (Grosskopf, Hayes, Taylor, & Weber, 2001; Ruggiero, Duncombe, & Miner, 1995) have attempted to investigate the determinants of inefficiency in education based on several theories of inefficiency in public services, including the X-inefficiency theory of Leibenstein (Leibenstein, 1966), budget-maximising bureaucrats theory of Niskanen (Niskanen, 1968, 1971), and the principal-agent theory. According to these theories, inefficiency in public sectors arise because of the imperfect competition market of public organizations (Leibenstein, 1966), the excessive cost of outputs caused by the budget-maximizing behavior of bureaucrats (Niskanen, 1968, 1971), or the failure to choose best-performing agent due to the asymmetric information between principal and agent (Eisenhardt, 1989).

All in all, the theories of inefficiency in public services imply that lack of incentives and/or monitoring mechanisms for individuals to perform optimally is the primary cause of inefficiency in education. Therefore, it is expected that schools or school districts can improve their technical efficiency if there are accountability incentives that monitor school administrators and teachers' behavior. Currently available empirical studies, however, did not provide strong evidence to support the hypothesis. For instance, studies have considered competition from other schools, including public and private, as one of the monitoring mechanisms for schools to maximize educational outcomes (Grosskopf, et al., 2001; Kang & Greene, 2002; Ruggiero, et al., 1995). The results of the studies were mixed. Ruggiero and colleagues (1995) found that the presence of competition was associated with greater inefficiency in New York State; but other studies (Kang & Greene, 2002; Grosskopf, et al, 2001) indicated that competition had no significant impact on inefficiency.

Currently, with the development of standards-based reform movement in K-12 education in the United States, more and more emphasis is put on developing and implementing school accountability policies. In particular, the NCLB act implemented in 2001 lays out specific goals, requirements, as well as sanctions and rewards for schools to improve student achievement in the whole nation.

The main goal of NCLB is to make every student in public schools proficient in math and reading by the 2013-2014 school year. Under the policy, states are required to set proficiency targets for each year, to test all students annually in grades 3 through 8 in math and reading, and to report to the public student performances on the tests. Test scores are reported in several performance levels, determined by individual state, and disaggregated by race, gender, special education, English-language proficiency, and socio-economic status. To accomplish the goal of having every student proficient by 2014, schools need to make adequate yearly progress (AYP). Schools that do not make AYP for the first year will receive additional assistance, but face sanctions after two or more years of failing to make AYP. Sanctions include providing free tutoring services to students, allowing students to transfer to another school, implementing a corrective action plan, and/or restructuring school.

Thus, it is expected that NCLB, the currently most widely implemented accountability policy in the country, will serve as a monitoring mechanism and provide strong incentives for schools to maximize their productive activities and subsequently to improve their technical efficiency level.

The Problem

Many studies have been conducted to examine the impacts of NCLB on public schools from different angles since its passage. Some have investigated whether student achievement has improved under NCLB (Center on Education Policy, 2005, 2007; Choi, Seltzer, Herman, &

Yamashiro, 2007; Cohen, Murray, & Clewell, 2007; Gribben, Campbell, Mathew, 2008; Miura, 2008; Hoerandner & Lemke, 2005; Mueller & Schmitt, 2006), while others are concerned if NCLB has had impacts on teaching practices (D'Amico, 2008; McMurrer, 2008; Opfer, Henry, & Mashburn, 2008; Russell, 2008). Still others have investigated whether NCLB has propelled schools to put more effort on improving students on the margin of becoming proficient while ignoring those at the highest- and lowest-end (Chakrabarti, 2006; Figlio & Rouse, 2006; Krieg, 2008; Neal & Schanzenbach, 2007; Reback, 2008; Springer, 2008).

Among empirical research on the impacts of NCLB, however, few studies investigate the effect of NCLB on school technical efficiency, defined as the distance between observed outputs schools produced and maximum outputs they are able to produce. Primont and Domazlicky (2006) conducted the only study so far to examine the impact of NCLB on school technical efficiency. The authors used Data Envelopment Analysis (DEA) to estimate technical efficiency of schools districts in Missouri in 2002. They found that the average inefficiency level in Missouri was 22.5%, suggesting that school districts should have produced 22.5% more outcomes if they operated at efficient level. In addition, they found that school districts that failed to make AYP were substantially less efficient than passing ones. However, this study only estimated one year's performance, and this was the first year when NCLB was implemented. It is very likely that the policy has lagged effects which will take time to manifest. Moreover, in the study the effect of NCLB was analyzed by utilizing simulated data at the district level, which may not capture the impact NCLB has actually had on school level technical efficiency.

It is important to examine the influence of NCLB on school technical efficiency for several reasons. First, estimating schools' technical efficiency can be useful to more accurately evaluate schools' performance. Under NCLB and other test-based accountability policies, schools are currently evaluated and ranked according to whether they reach a designated benchmark, which in the case of NCLB is the percentage of students proficient in state assessments, without considering how many inputs are used to achieve the output. Such a method of evaluation is output-based and effectiveness-focused, for it only indicates whether schools are effective in producing the outcome - proficient students. It does not provide information on how well schools utilize their resources during the process of producing the outcomes. It is very likely that effective schools, the ones that made AYP, were inefficient in using the resources while some ineffective ones already used their resources to the maximal level. Therefore, evaluating schools using their technical efficiency level not only shows a better picture of the production process, but also has implications for allocating resources among schools.

Second, it is critical to estimate school technical efficiency because of budget shortfalls in education. The current financial crisis leads to substantial spending cuts in many state services which include K-12 education by state governments (McNichol & Lav, 2008). Therefore, schools have to use the inputs more wisely so that they can achieve more proficient students and make AYP as required by NCLB. The knowledge about the past and current efficiency levels can be valuable for school administrators and policy makers to improve school performance.

Finally, examining the issue of the effect of NCLB on school technical efficiency can contribute to developing a better theory of inefficiency in education. As discussed, few empirical studies have been conducted to investigate the determinants of inefficiency in education, and among existing studies no strong evidence has been found to support the theories of inefficiency. Using NCLB as a monitoring mechanism on the behaviors of school administrators and teachers, we can gain some insight as to whether NCLB provides incentives for them to maximize their productive activities and in turn to produce more proficient students.

The Purpose

The primary purpose of this dissertation is to investigate the effect of NCLB on school efficiency in three states, Indiana, Minnesota, and South Carolina. These three states are chosen as the subjects of investigation because of their availability of test scores in pre-NCLB years, which enables us to compare the pre-NCLB to post-NCLB changes in school efficiency. This dissertation employs two-stage analysis to address the issue. The first-stage analysis is to estimate school efficiency using Data Envelopment Analysis (DEA). The outputs for the first-stage DEA models are percentages of students in three proficiency levels in each tested subject as required by NCLB. The inputs of the first-stage DEA models are per pupil expenditure and three students demographic characteristics, including those who are not eligible for free- and reduced-price lunch program, not minority, and not Limited English proficiency.

At the second stage, a difference-in-differences estimator, which can rule out unobserved influences other than NCLB, is used to examine the effect of NCLB on school efficiency. The difference-in-differences estimator is constructed by classifying schools into two groups: unthreatened and threatened, based on their pre-NCLB test scores. The effect of NCLB is then captured by the difference between two groups of schools in terms of their pre-NCLB to post-NCLB changes in efficiency.

This dissertation hypothesizes that NCLB has helped schools, especially those at risk of not making AYP, to raise their efficiency. Specifically, we ask the following research questions:

1. Has school efficiency changed over time?

Subsidiary questions:

1a. Have public schools in Indiana, Minnesota, and South Carolina improved efficiency over time?

1b. Have different types of schools in three states had different efficiency levels?

2. Has NCLB influenced the change of school efficiency over time?

Outline

This dissertation is constructed as follows. Chapter two reviews the empirical studies on the effects of NCLB on public schools and those on technical efficiency in K-12 education in the United States. The first part of chapter two analyzes the effects of NCLB on (1) student test scores, and (2) teachers' strategic instruction. The second part of the chapter examines (1) the methods used to measure technical efficiency in K-12 education, (2) the measures of technical efficiency in education, and (3) the determinants of inefficiency in education. The third part of the chapter reviews studies that investigated the effect of NCLB on technical efficiency, followed by the final part that summarizes the findings.

Chapter three discusses the methodology of conducting analysis. It begins by presenting the concept of economic efficiency and the method of measuring efficiency in this dissertation. The second part of this chapter defines the outputs used at the first stage analysis, followed by the discussion of model specification in the third part. Strategies of handling the threats to validity are presented in the fourth part. The final part of chapter three discusses the data used to conduct the analysis.

Chapter four presents the results. This dissertation is ended with chapter five which discusses the findings, policy implications, the limitations of the study, as well as the recommendations for future research.

CHAPTER II

LITERATURE REVIEW

The Effects of NCLB on Public Schools

Many empirical studies have been conducted to examine the influence of NCLB on public schools since its passage into law. Generally, the majority of studies focus on two broad areas: student achievement and teachers and principals' behavior. Overall the average student test scores are consistently found to increase after the introduction of NCLB. However, different groups of student have experienced distinct rates of improvement. Furthermore, some scholars have argued that the increased student test scores are not the result of the improvement of real learning; rather they are caused by teachers' strategic instruction. In the following sections, I review the studies on these areas separately.

NCLB and Student Achievement

The center of the debate with respect to the effects of NCLB is whether the policy has positive impacts on student achievement. Studies have consistently shown that average student test scores have increased since NCLB was implemented (Center on Education Policy (CEP), 2007; Choi, Seltzer, Herman, & Yamashiro, 2007; Cronin, et al., 2005; Gribben, Campbell, Mathew, 2008; Mueller & Schmitt, 2006).

Using national data, the Center on Education Policy (CEP) examined the trends in state test scores since NCLB was enacted (CEP, 2007). Their study showed that math and reading scores have increased since the policy was implemented. In addition, among 13 states which have sufficient data to make the comparison before and after NCLB, nine states showed that average annual gains in two subjects' test scores have been greater in post-NCLB years than in pre-NCLB years. Another study that examined national trends in achievement (Cronin, et al., 2005) reached a similar conclusion that student test scores in reading and math have improved in the post-NCLB era.

Furthermore, the study by Mueller and Schmitt (2006) also found evidence that student achievement has increased under NCLB. Using Kansas data from 2000 - 2005, the authors examined effects of AYP requirement on student achievement by calculating effect sizes of the change in proportion of students in different achievement levels before and after NCLB. They found that effect sizes were large when comparing results two years before NCLB with three years after NCLB, suggesting that student achievement gains were higher in post-NCLB than in pre-NCLB years.

Although there is evidence of improvement in average student achievement since NCLB was implemented, the effects of the policy are not identical for different student groups. Generally, there are two types of achievement gaps that researchers are interested in. The first is the achievement gap among different racial groups, usually between black/Hispanic and white/Asian students. Studies indicated that different racial groups experienced different rates of improvement in achievement, resulting in gaps as great as or greater than in the past. Lee (2006) examined national reading and math achievement trends using NAEP data. The author found that although the achievement gap between black and white students in math narrowed to some extent immediately after NCLB was enacted, the narrowing did not persist. Furthermore, the gap in reading did not decline under NCLB. Another study even found that the gap in the percentage of students in the top performance level between black/Hispanic and white students widened over time (Gibben, et al., 2008), using states' assessment results.

The second is the achievement gap among students with different levels of ability. The minimum proficiency requirement of NCLB gives schools strong incentives to target resources to students at the margin of becoming proficient and ignore those at the high- and low-end of the achievement distribution. There is mixed evidence on this issue.

Several studies indicated that the improvement of low-achieving students was realized at the expense of high-achieving peers (Figlio & Rouse, 2006; Reback, 2008; Neal & Schanzenbach, 2007; Krieg, 2008). Using Texas data from 1992-93 to 1997-98 school year, Reback (2008) examined the impacts of the incentives from the school accountability system on the distribution of student achievement among different ability groups of students. The author found that low-achieving students experienced larger than expected gains in achievement, but high-achieving students had lower gains if their test scores were irrelevant to school's rating.

Moreover, Krieg (2008) found that in Washington high-performing students had less than expected gains in achievement in schools that did not make AYP, suggesting that schools facing the threats of sanctions under NCLB ignored the high-performing students and that more resources were devoted toward those with a high probability of being proficient.

However, other studies reached a different conclusion: that the improvement of low-achieving students was accomplished without hurting that of high-achieving counterparts (Chakrabarti, 2006; Mullier, & Schmitt, 2006; Springer, 2008). Chakrabarti (2006) examined the impacts of Florida school accountability on public schools. She found that schools that faced threats of vouchers did put more emphasis on improving the achievement for students below the minimum proficiency cutoffs. However, the improvement was not to the detriment of high-performing students. Similarly, Springer (2008) found that both low- and high-performing students experienced larger gains in test scores in non-AYP schools than their corresponding peers in AYP schools, suggesting that the improvement of low-achieving students did not hurt high-performing students.

NCLB and Teachers' Strategic Instructions

Although the overall test scores have increased since the accountability policy was introduced, some scholars argued that the rise of test scores might not be the result of the improvement of real learning (Jacob, 2005; Klein, Hamilton, McCaffrey, & Stecher, 2000; Koretz, 2008). Rather, it is claimed that strategic behavior exhibited by teachers and school leaders was the reason for the inflation of test scores. Specifically, two behaviors are considered in the literature. First, the increase of test scores is believed to be the result of teachers spending more time on high-stakes subjects. Studies that examined pre-NCLB school accountability found evidence that teachers indeed spent more time on high-stakes subjects, such as math and reading (Deere & Strayer 2001; Koretz & Barron, 1998; Stecher, 2002). Moreover, among high-stakes subjects, teachers were more likely to focus on the one that was relatively easier to improve, such as writing in the case of Florida, in order to improve their school's rating and avoid sanctions (Chakrabarti, 2006; Goldhaber & Hannaway, 2004). Additionally, researchers in CEP also showed that teachers spent more time on high-stakes subjects, i.e. reading/English and math, while reducing instructional time on other low-stakes subjects, such as science, social studies, and the arts, since NCLB was launched (CEP, 2007).

Second, it is believed that the inflation of test scores is due to the exclusion of low-achieving students from testing pools. There is mixed evidence on this argument. Some studies indicated that more students were classified into special education after school accountability systems were introduced (Cullen & Reback, 2002; Figlio & Getzler, 2002; Jacob, 2005; Mintrop, 2003). However, other studies showed that there was no evidence to suggest schools displayed gaming behavior by classifying more students into special education category (Chakrabarti, 2006). Since NCLB requires at least 95% of students should participate in state assessments, otherwise schools will be considered as not making AYP, it is expected that schools have little opportunity to exclude low-performing students from taking the tests.

Technical Efficiency in Education

A majority of studies on technical efficiency in education have emphasized how to measure technical efficiency per se, without considering the sources of inefficiency. Thus, much attention has been paid to the methods of measuring technical efficiency in education. In the following sections, I review the methods of measuring, the results of measured technical efficiency, and the determinants of inefficiency, respectively.

Methods of Measurements

Generally, two approaches are adopted to estimate technical efficiency in different settings: one is parametric and the other non-parametric. Both approaches have their disadvantages and advantages. Thus, which approach to use is a source of debate in the literature. However, there is mixed evidence on this question. Some researchers stated that the choice of approaches affected the empirical estimation (Cummins & Zi, 1998; Ferrier & Lovell, 1990), while other showed that both approaches yielded similar ranking of efficiency scores (Chakraborty, et al., 2001; Forsund, 1992; Murillo-Zamorano & Vega-Cervera, 2001).

In educational settings, several studies have compared two approaches (Bifulco & Bretschneider, 2001; Chakraborty, et al., 2001; Sengupta & Sfeir, 1986). Bifulco and Bretschneider (2001) simulated 12 datasets, with different assumptions on measurement errors, sample sizes, and correlations between error terms and inputs, to compare corrected ordinary least square (COLS) with Data Envelopment Analysis (DEA). They found that if there are no measurement errors or no correlations between error terms and inputs in the data, both methods perform well, in particular COLS, as the estimated efficiency is close to the true efficiency. However, when it is assumed there are measurement errors that are correlated with inputs, as is common in educational data, efficiency measures from both approaches deviate substantially from true efficiency. Under this circumstance, COLS slightly outperformed DEA when sample sizes were bigger.

Another study (Chakraborty, et al., 2001) empirically measured the efficiency level of school districts in Utah using both stochastic frontier analysis and DEA. The authors found two approaches yield similar rankings of school districts in terms of efficiency measures, suggesting that choosing either method will not substantially influence the empirical results.

Measures of Technical Efficiency in Education

Despite the fact that the estimation methods vary from studies to studies, researchers have generally reached the same conclusion, that public schools (or school districts) are fairly inefficient in producing desirable educational outcomes (Bessent, et al., 1982; Chakraborty, et al., 2001; Grosskopf, et al., 1997; Noulas & Ketkar, 1998). For instance, Grosskopf and colleagues (1997) found that the average efficiency level, estimated by using indirect output distance function, is .71 for school districts in Texas, implying that school districts should be able to produce almost 30% more output if they operated efficiently. Another study (Chakraborty, et al.,

2001) also showed that the average technical efficiency for school districts in Utah is .86, estimated by using stochastic frontier analysis.

However, some studies concluded that at the elementary level, schools or school districts operated at efficient level (Deller & Rudnicki, 1993; Anderson, Walberg, & Weinstein, 1998; Chalos, 1997). For instance, Anderson and colleagues (1998) concluded that elementary schools in Chicago were relatively efficient in producing student achievement. Similarly, Deller and Rudnicki (1993) and Chalos (1997) indicated that the average technical efficiency was .91 for elementary schools in Maine and Illinois, considered to be fairly efficient in educational settings,

Determinants of Technical Inefficiency in Education

Although inefficient production in education is acknowledged, no conclusive evidence is available to explain the sources of inefficiency. Generally, two arguments are offered in the literature to identify the determinants of inefficiency. The first argument follows the education production function literature in noting that family backgrounds and community characteristics play a more important role in student achievement than school inputs. Inefficiencies in schools are therefore due to the severity of the environment they face, such as having more at-risk students from less favorable socio-economic environments. However, there is disagreement among researchers with respect to the ways of handling non-discretionary environmental factors. One group of researchers favors a single-stage approach in which environmental variables are included in the production frontier models along with discretionary inputs (Bessent, et al, 1982;

Chalos, 1997; Cooper & Cohen, 1997; Deller & Rudnicki, 1993). However, the single-stage approach assumes environmental factors are endogenous and discretionary by incorporating them in the frontier models, which is not the case in reality.

Therefore, researchers also separate exogenous environmental factors from discretionary inputs in a widely used two-stage approach in which estimated efficiency scores obtained from first stage are regressed on non-discretionary factors in a second stage (Chakraborty, et al., 2001; Kang & Greene, 2002; Noulas & Ketkar, 1998; Ray, 1991). For instance, Noulas and Ketkar (1998) adopted a two-stage analysis to investigate the impact of environmental factors on school districts' efficiency level in New Jersey. They used DEA to estimate district level efficiency. The predicted efficiency measures were then regressed on several environmental factors, including median value of homes, crime rate, proportions of population below poverty, and proportions of minority students in the districts. They found that all environmental factors have significant influence on school districts' efficiency level, with median home value having a positive impact on efficiency and the other three factors a negative impact. Nevertheless, single-stage and two-stage approaches do not give substantially different ranking of efficiency; the efficiency level estimated from the single-stage method is highly correlated to that adjusted for the influence of environmental factors from the two-stage approach (McCarty & Yaisawarng, 1993).

The second argument explains the source of inefficiency in education from the viewpoint of the principal-agent theory, which postulates that individuals do not always behave in a maximizing manner in their productive activities; instead, they tend to make compromises between their personal motives and the goals of the firm, which results in the deviation from the frontier (Leibenstein, 1978; Niskanen, 1968, 1971). In other words, inefficiency in public education is hypothesized to be caused by lack of incentives and/or monitoring mechanisms in the system (Arrow, 1985; Downes, 1996). Empirically researchers tend to operationalize the incentives and monitoring mechanisms into following categories: (1) competition from other public schools or private schools, which is expected to motivate schools to maximize their productivity and subsequently improve technical efficiency (Grosskopf, et al., 2001; Kang & Greene, 2002; Ruggiero, et al., 1995), (2) the socio-economic status of the community, including the shares of homeowners and households with school age children, as such households tend to improve efficiency by monitoring school performance (Grosskopf, et al., 2001; Kang & Greene, 2002; Ruggiero, et al., 1995), (3) size of school districts, which is expected to be positively correlated with inefficiency as bigger bureaucracies introduce more inefficiency (Ruggiero, et al., 1995), and (4) internal characteristics of schools or school districts such as the proportion of tenured teachers, given that tenured teachers are expected not to perform to their maximal productivity level (Ruggiero, et al., 1995).

Nonetheless, empirical studies do not find strong evidence to support these hypotheses. Take competition for example. Ruggiero and colleagues (1995) found that the presence of competition was associated with greater inefficiency in New York State; but other studies (Kang & Greene, 2002; Grosskopf, et al, 2001) indicated that competition had no significant impact on inefficiency.

Additionally, higher socio-economic status of the community was found to be associated with greater inefficiency, while school district size was negatively correlated with inefficiency (Ruggiero, et al., 1995), which contradicts with the predictions of the inefficiency theory. Only teacher tenure had the expected positive relationship with inefficiency (Ruggiero, et al., 1995).

Impacts of NCLB on Technical Efficiency in Education

As discussed, few studies have investigated the impacts of NCLB or other test-based accountability policies on how efficiently schools produce educational outcomes. The study by Primont and Domazlicky (2006) is thus far the only one that examined the question. The authors used DEA to estimate the efficiency level for school districts in Missouri in 2002. They found that the average inefficiency level for school districts was 22.5%, suggesting that school districts would have produced 22.5% more output if they operated at the efficient level. Moreover, the authors indicated that school districts that failed to make AYP were substantially less efficient than the ones that made AYP. Additionally, the authors simulated the effects of two sanctions of NCLB, providing tutoring services and transferring students from failing schools to passing ones, on the performance of failing school districts. They found that the diversion of resources to these functions made failing schools more inefficient.

However, this study has several limitations. First, it is a cross-sectional analysis; thus it is unclear about the change of efficiency over time. Second, the findings were based on simulated data, which did not capture the real impacts of sanctions from NCLB. Since the data used in this study were from the first year of NCLB (i.e. 2002), in which the effects of the policy were not yet manifested, actual changes in student achievement were not taken into account. Third, the analyses were at the district level; therefore, the findings may not fully reflect the impacts of NCLB on school level efficiency.

Why Further Research is Needed

As discussed in the previous sections, most empirical studies that examined the influence of NCLB on public schools have focused on student achievement. Those studies have two limitations. First, they have only emphasized outputs without taking into consideration inputs. This type of output-only research has resulted to some extent in bias in assessing schools' performance as the same output can be produced using less input in some schools than in others. Second, although many studies have indicated that average student test scores were higher in post-NCLB years compared to pre-NCLB years, it is unclear whether the improvement of test scores can be attributed to the policy. The reason is that the current studies used simple pre- to post-NCLB comparison (CEP, 2007; Mueller & Schmitt, 2006), which did not control for other possibilities, such as changes in test difficulty or in student and teacher familiarity with tests, or in other accountability policies implemented over the same period.

Furthermore, the literature on technical efficiency in education does not provide adequate evidence on the determinants of inefficiency. Although it is proposed in the literature that lack of incentives and monitoring mechanisms is the source of inefficiency in education, empirical studies have not find strong support for this hypothesis. Nowadays, NCLB is an accountability policy that influences almost all public schools in the nation. However, few studies have examined whether NCLB and its associated incentives have influenced technical efficiency in schools.

Therefore, this dissertation fills the gap in the literature by investigating the impact of NCLB on school efficiency. NCLB requires that schools reach Annual Measurable Objectives (AMO), as set by individual states, to make AYP and thus avoid sanctions. With state budgets shrinking (McNichol & Lav, 2008), schools have to use their resources more wisely so that more output (in this case, the percentage of proficient students) can be produced. Thus, this dissertation hypothesizes that the threats of facing sanctions under NCLB will work as an incentive for schools to improve their efficiency.

CHAPTER III

METHODOLOGY

The Production Frontier and Data Envelopment Analysis (DEA)

The idea of economic efficiency was introduced as early as the 1950s by Koopmans (1951), Debreu (1951), and Farrell (1957). Generally, technical efficiency is defined as the situation in which it is impossible to increase one output of a decision-making unit (DMU) without increasing its inputs or reducing one or more of its other outputs (Cooper, Seiford, & Zhu, 2004). Schools that produce less than they could, given their production frontier, are deemed inefficient.

There are two broad approaches to the estimation of the production frontier and technical efficiency, parametric and non-parametric. The parametric approach, of which stochastic frontier analysis (SFA) is the most widely used, assumes specific functional forms for technology and for inefficiency, which causes problems related to model specification and estimation. The non-parametric approach is represented by Data Envelopment Analysis (DEA). DEA does not impose any functional form, which makes it possible to disclose relationships that are unseen by other approaches. Furthermore, DEA can easily accommodate multiple outputs and inputs.

Which approach to use is a source of debate in the literature. However, no decisive evidence is offered with respect to the question which method more accurately measures technical efficiency. Stochastic frontier analysis has been conducted in the literature, but our preliminary analysis was unsuccessful because of the orientation of the data distribution. Minnesota, one of the states that will be analyzed in this dissertation, can be taken as an example. There are two outputs: the percentage of students at the proficient level or higher (Y1), and the percentage of advanced students (Y2)[1]. Figure 1 shows the scatter plot of these two outputs, holding constant the level of inputs, using Minnesota data. Y1 and Y2 are positively associated. However, stochastic frontier analysis assumes a negative relationship between two outputs. Consequently, it is inappropriate to use this method to conduct the analysis.

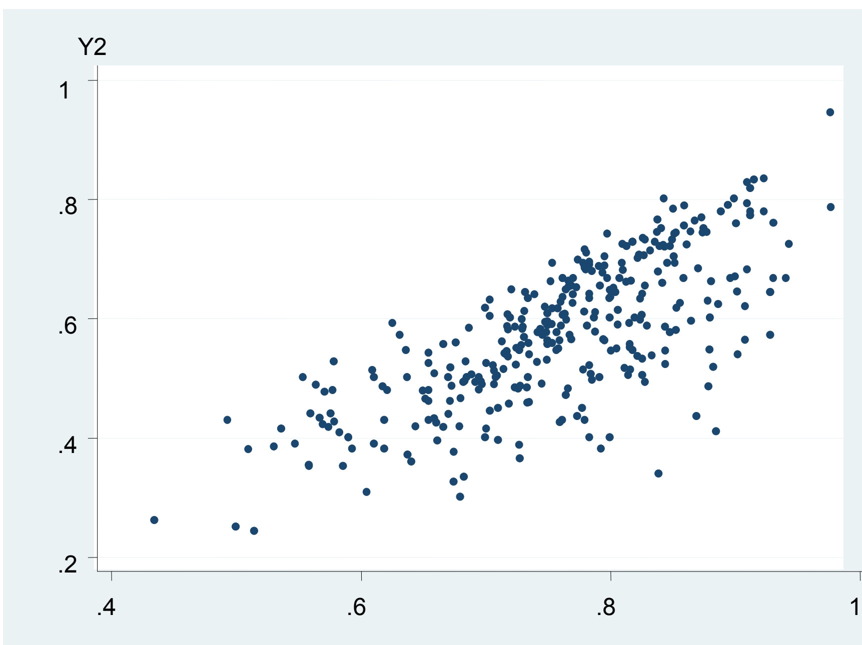


Figure 1, Scatter Plot of Y1 and Y2, Minnesota

1 The reason of constructing these two measures of outputs will be explained in more details in later sections. They are mentioned here only to illustrate the distribution of the data.

Unlike stochastic frontier analysis, DEA assumes no functional form for production and for inefficiency, which makes it possible to unveil the unseen relationship, such as the one in this case. Additionally, DEA's advantages in easily handling multiple outputs and inputs make it a better choice for conducting efficiency analysis in the education sector. Thus, this study uses DEA to estimate efficiency in schools.

DEA was introduced by Charnes, Cooper, and Rhodes (1978, 1981). The basic output-oriented CCR model, assuming constant returns to scale (CRS), determines technical efficiency for individual DMU *i* that uses a set of inputs $X_i = (x_{1i}, x_{2i}, ..., x_{ki})$ to produce a vector of outputs $Y_i = (y_{1i}, y_{2i}, ..., y_{mi})$ by using the following linear programming problem:

$$\underset{\phi,\delta}{Max} \ \phi$$

$$\text{Subject to} \quad \delta_1 y_{t1} + \delta_2 y_{t2} + ... + \delta_n y_{tn} \ \geq \phi \ y_{ti}, \text{ t = 1, 2, …, m}$$

$$\delta_1 x_{s1} + \delta_2 x_{s2} + ... + \delta_n x_{sn} \ \leq \ x_{si}, \ \text{ s = 1, 2, …, k}$$

$$\delta_1, \ \delta_2, …, \ \delta_n \ \geq \ 0 \qquad (1)$$

in which *n* is the numbers of DUMs, *m* the numbers of outputs, *k* the numbers of inputs, and δ_1, δ_2, …, δ_n are non-negative weights assigned to inputs and outputs. Problem (1) asks, given the same or lower input level, how much more output a DMU can produce and still remain within the production possibility set, as determined by the outputs and inputs of other DMUs as well as itself. Technical efficiency is the reciprocal of ϕ, whose value ranges from 0 to 1. A value of 1 means the DMU operates efficiently, less than 1 that it is inefficient. For instance, if ϕ equals 1.5, the DMU can produce 1.5 times as much of each output without using more inputs. Thus, the technical efficiency of the DMU is 1/1.5 = .67. This linear programming problem is then solved *n* times (because there are *n* DMUs), yielding efficiency estimates for all DMUs.

Let us consider a simple example where there are 5 DMUs (A, B, C, D, and E) and they use one input (x) to produce two outputs (y_1, y_2). The example can be illustrated in Figure 2. As indicated, the frontier is constructed as the piecewise linear combination of efficient DMUs, A, B, and E. Relative to A, B, and E, C and D are inefficient and lie within the frontier line.

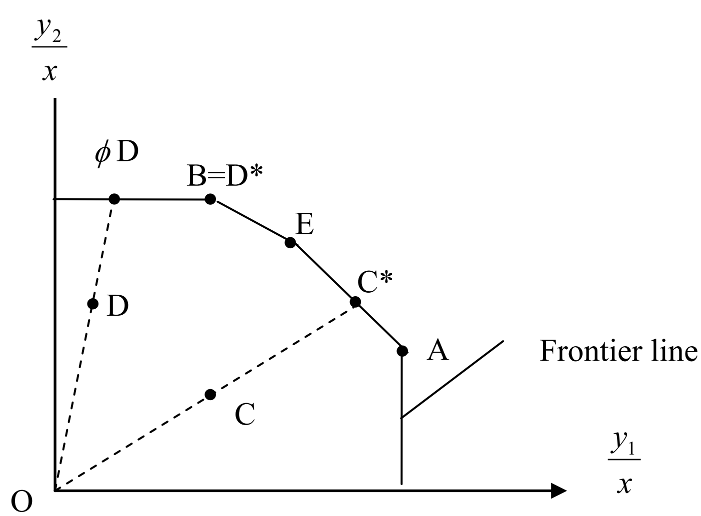


Figure 2, Output-Oriented Efficiency and Slack

If we radially extend C until it reaches the frontier line, we can obtain point C*. C* is considered to be the efficient point for C, meaning that the DMU can expand its y_2 and y_1 to C* without using more x. Therefore, ϕ equals the ratio OC*/OC. The efficiency for C is the reciprocal of ϕ, that is, OC/OC*. Additionally, C* lies between A and E, indicating that DMU's A and E are the reference group for C.

Similarly, D is inefficient and its y_2 and y_1 can be radially expanded to point ϕD. Thus, the efficiency of DMU D is the ratio OD/OϕD. However, ϕD is not a truly efficient point for D. Compared to ϕD, given the same x B has the same amount of y_2 but more y_1, suggesting that D can further expand its output of y_1 without increasing x. Consequently, B is the truly efficient point for D. The distance between ϕD (weak efficiency) and B (true efficiency) is called *output slack* (slack = B - ϕD).

Many DEA models assume constant returns to scale (CRS). However, this assumption can produce misleading measures of efficiency if returns to scale are variable and some schools are operating at suboptimal scales, for the effect of scale can then be confounded with inefficiency[2] (Coelli, Rao, O'Donnell, & Battese, 2005). To take this possibility into account, this study employs a DEA model assuming variable returns to scale (VRS). As a result, problem (1) is modified slightly into the following linear programming problem:

$$\max_{\phi,\delta} \ \phi$$

$$\text{Subject to} \quad \delta_1 y_{t1} + \delta_2 y_{t2} + ... + \delta_n y_{tn} \ \geq \phi \ y_{ti}, \ t = 1, 2, \ldots, m$$

$$\delta_1 x_{s1} + \delta_2 x_{s2} + ... + \delta_n x_{sn} \ \leq \ x_{si}, \ \ s = 1, 2, \ldots, k$$

$$\sum_{1}^{n} \delta_i \ = 1 \qquad (2)$$

[2] If not all DMUs operate at the optimal scale, technical efficiency estimated from DEA models that assume CRS will be confounded with scale efficiencies. Models that assume VRS can solve this problem. The difference between CRS and VRS models is explained in more detail in the appendix.

Note that the constraint of $\sum_{1}^{n} \delta_i = 1$ in problem (2) ensures that DMU_i is compared with other DMUs of a similar size.

Defining Outputs given Characteristics of NCLB Data

It is a common practice in education research to treat student achievement in standardized tests as the measure of educational outputs. We follow that tradition. However, unlike other research that uses student-level achievement or school/district average scores as outputs, our study defines outputs based on the percentage of students at various performance levels. Under NCLB, schools generally report student test scores in three broad levels: below proficient, proficient, and advanced[3]. Given that these three outcomes add up to one hundred percent, one is redundant and can be dropped. We drop the middle group, the percent proficient.

The reason for dropping the percent proficient as an output is that the other two outputs, percent advanced and percent below proficient, will show the tradeoffs among three proficiency levels. As required by NCLB, the percentage of proficient students is one of the most important determinants of schools' AYP status. The percentage of proficient students can be increased by two ways: moving students either from below proficient up to proficient, or from advanced down to proficient. The movements of students from the top and the bottom level toward the middle level imply two different responses of schools toward the provisions of NCLB. First, by

[3] Some states report test scores in four (below basic, basic, proficient, and advanced) or five performance levels (below basic, basic, proficient, advanced, and advanced plus). Still, the performance levels can be classified into three groups: not proficient, proficient, and above proficient. Therefore, we consider there are three broad performance levels.

providing high quality instruction and other effective strategies schools will try to improve the achievement for low performing students and help them reach higher levels, which is consistent with the ultimate goals of the policy. On the other hand, it is possible that schools, in particular those with a large proportion of low performing students, will be so eager to increase the percentage of proficient students that they will sacrifice the number of students in the advanced group (Figlio & Rouse, 2006; Reback, 2008).

Unfortunately, empirical studies have not found suggestive evidence on whether schools trade off advanced students for those at lower proficiency levels (Chakrabarti, 2006; Krieg, 2008; Neal & Schanzenbach, 2007; Reback, 2008; Springer, 2008). Using the top and the bottom proficiency levels as two outputs will shed some light on the possible tradeoffs among three proficiency levels. However, the percent below proficient should not be regarded as an output in the conventional sense, as no school wants to produce more of these students. Thus it is necessary to make some adjustment to the percent below proficient so that it becomes a positive, desirable output. The adjustment is made by subtracting the percent below proficient from 100. The result equals the total of percent proficient and percent advanced. As a result, the two outputs used at the first stage models are the percent advanced (ADV) and the total of the percent proficient and the percent advanced (PRF+ADV).

As shown in Figure 3, these measures of output define a meaningful efficiency frontier. If a school moves from point E to point A, it reduces its share of advanced students but increases the total of proficient and advanced students—thus the reduction in advanced students occurs at the same time as an increase in the overall percentage that are proficient or above – a plausible trade-off. For such a school, the increase of (PRF+ADV) implies that students have moved out of the bottom category. In a school that is operating efficiently, this can only occur if something else is given up, in this case, the percentage of students in the top category.

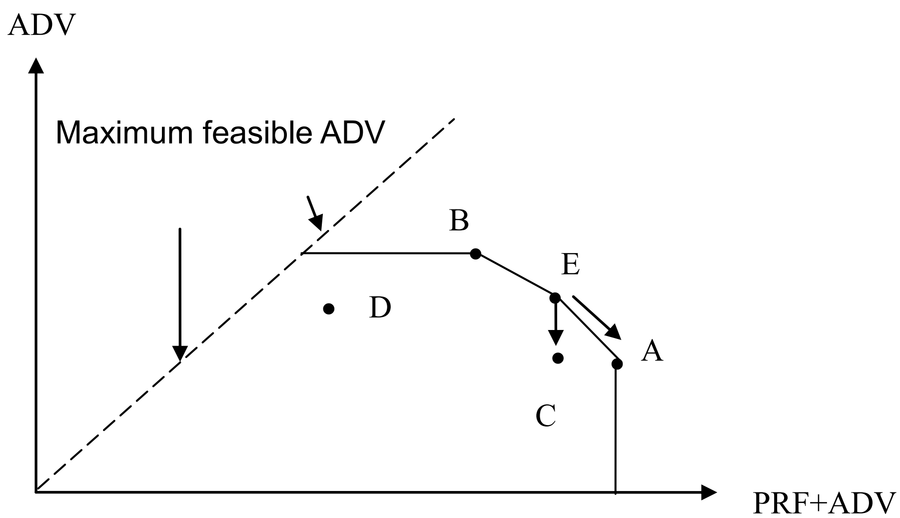


Figure 3, The Production Frontier using ADV and (PRF+ADV) as Outputs

If a school simply produces fewer advanced students (they drop down to the proficient level) without also reducing the proportion of students below proficient, ADV declines while (PRF+ADV) is unchanged. Such a school becomes less efficient, moving from a point such as E on the frontier to C within the frontier.

As far as inefficient schools are concerned, there are three possible movements toward the frontier line, which is illustrated in Figure 4. The first possible movement is to increase ADV without changing (PRF+ADV), implying students are moved from proficient to advanced level. The second possible movement is to increase (PRF+ADV) without changing ADV, suggesting students are moved from below proficient to proficient. The third possible movement is to increase ADV and (PRF+ADV) simultaneously, meaning students are moved from the bottom category to proficient or from proficient to advanced.

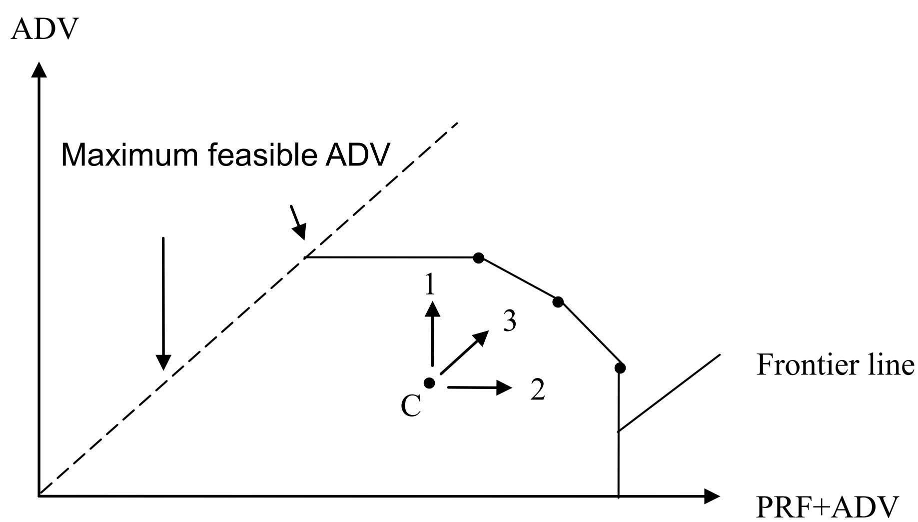


Figure 4, Possible Movements for Inefficient Schools

Consequently, it is possible to tell the movement of students among three performance levels by looking at the path taken by an individual school across time. In addition, this path reveals whether there are tradeoffs among low-, average- and high-achieving students.

Model Specifications

First-Stage Models

Technical efficiency is measured at the first-stage analysis using DEA. When it comes to specifying DEA models, three issues need to be considered: first, to what extent the production of math and reading outcomes should be treated as independent or as substitutes for one another; second, the possibility that schools in different locations have substantially different production processes; and third, the assumption that technology in education does not change greatly over a short period of time.

It is uncertain to what extent the production of math and reading outcomes should be treated as independent or as substitutes for one another. In the elementary level they are more apt to be substitutes, as both subjects tend to be taught in the same classroom by the same teacher, who controls the allocation of time across subjects. At the middle and high school levels they are more likely to be independent, as instruction is departmentalized.

To explore the issue of substitution/independence, this study begins with a restrictive model before relaxing the assumptions. The most restrictive model assumes there is no substitution at all across subjects. Each subject is produced independently in each grade. As a result, technical efficiency is estimated separately for each subject in each grade. We then relax this assumption to allow for substitution across subjects, estimating models in which reading and math achievement appear as joint outputs. In addition, it is assumed that there is no substitution across grades. Therefore, the DEA models are run separately by grades.

The second issue to be considered is the possibility that schools in different locations have substantially different characteristics which result in different production processes. Schools in rural areas tend to be smaller and contain more grades than those in urban area. In addition, urban and suburban schools are more likely than rural schools to have highly-qualified teachers (Ballou & Podgursky, 1998). Thus, it is expected that schools in urban and rural area face distinct process of producing educational outputs. To take into account this possibility the DEA models are run separately by location.

Finally, technology in education seems not to change substantially over a short period of time. In sectors like agriculture, technology changes often and quickly, pushing outward the production frontier. However, in educational settings, usually technology does not change dramatically over a short period of time, which results in a relatively constant production frontier. Consequently, the sample is pooled over all years.

To summarize, four DEA models are constructed based on different pairs of outputs. One uses ADV and (PRF+ADV) in math as outputs, one ADV and (PRF+ADV) in reading, one (PRF+ADV) in both math and reading, and one ADV and (PRF+ADV) in both math and reading[4]. These four models are estimated separately by grade and by location (urban or rural).

Our analysis does not investigate the allocation of resources within schools directly, e.g., how many computers and books schools possess, what degrees teachers hold, and how big classes are. All such decisions remain inside a black-box model of production that treats

[4] In states that tested more than math and reading, such as South Carolina, the numbers of DEA models constructed will equal numbers of tested subjects+1.

outcomes as a function of the level of schools' non-discretionary inputs. The nondiscretionary inputs in the first-stage analysis are per-pupil expenditure (PPE), the percentage of students who are not eligible for free- and reduced-priced lunch program (non-FRL), the percentage that are not limited English proficient (non-LEP), and the percentage that are white (non-MINORITY). Note that an increase in any input tends to push the frontier out from the origin.

Because of the assumption of black-box production, estimates of technical efficiency in this study in fact encompass both technical and allocative efficiency. Allocative efficiency refers to the situation in which inputs are utilized in optimal proportions, given input prices. Figure 5 illustrates how allocative efficiency is estimated by considering a simple case which two inputs X1 and X2 are used, where w is the input prices. As indicated in the graph, point A* is a technically efficient but not allocatively efficient, because at A* the ratio of marginal productivity of X1 to that of X2 does not equal the ratio of price of X1 to that of X2. Given the prices of two inputs, a DMU (a school in this study) can gain allocative efficiency by using less X1 and more X2 (i.e. at point A') during the production process, which subsequently improves the total efficiency for the school.

It is very likely that NCLB may influence schools' decisions on reallocating resources within schools, such as reassigning more highly qualified teachers to classrooms or grades which have more low-achieving students, which in turn will boost the overall efficiency level. Therefore, the measures of efficiency in this dissertation in fact include both technical and allocative efficiency.

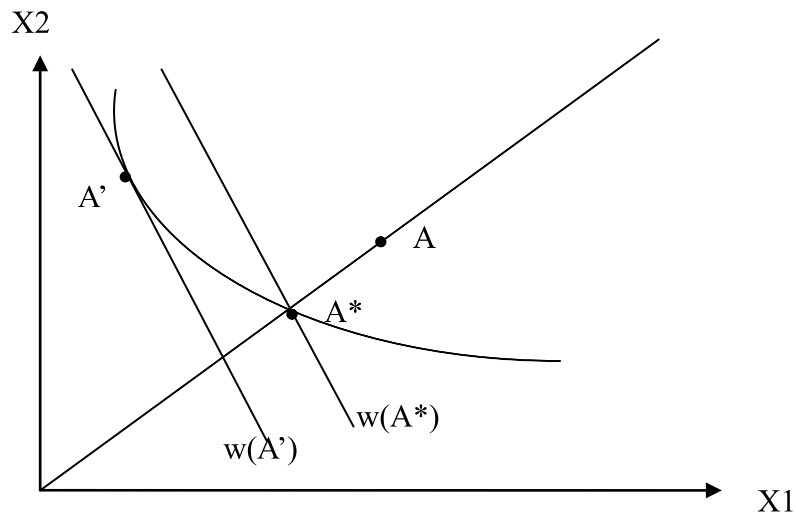


Figure 5, Allocative Efficiency

Second-Stage Models

In the second stage, estimated efficiency and slack from the first-stage analysis are used as dependent variables. A difference-in-differences approach is utilized to identify NCLB effects on school efficiency and on slack. Difference-in-differences is often used to detect the effect of policy change. Suppose there are two groups, A and B. Group A is affected by the new policy in the second period of time but not in the first period and group B is not affected during either period. The effect of the new policy can be estimated three possible ways. The first method is to compare group A in the second time period to itself in the first time period. However, this simple before-and-after comparison does not remove the bias due to the trends; it is possible that the change that happens in the second time period would be the natural development of group A in the absence of the new policy. The second method of estimating the effect of the new policy is to compare group A to group B after the new policy is introduced. Nonetheless, the effect of the policy estimated by this method is biased in that it may be caused by the permanent difference between group A and B rather than the new policy itself.

The third method is difference-in-differences estimation. This method estimates the effect of the new policy by subtracting group B's change from the first to the second period from group A's change. Mathematically this can be expressed as $(\bar{y}_{A,2} - \bar{y}_{A,1}) - (\bar{y}_{B,2} - \bar{y}_{B,1})$, in which y is the output of interest and 1 and 2 refer to the first and second time period, respectively. Since group B is not affected by the policy, its difference before and after the introduction of the policy (i.e., $\bar{y}_{B,2} - \bar{y}_{B,1}$) will illustrate what would have happened in the absence of the policy. Then by subtracting the change between the first and second time period for group B from that for group A, we can obtain the effect of the new policy. Consequently, the difference-in-differences estimator can control for confounding factors that are due to the trends and the permanent differences between two groups. Thus, it is more appropriate to be used to identify the effect of NCLB.

However, NCLB is a policy that applies to all public schools in the whole nation, making it impossible to distinguish a real control group of schools. Nonetheless, not every school responds to NCLB's sanctions in the same way. Schools with high student test scores may not have the same reactions to NCLB as the ones with low test scores. Therefore, it is possible to construct a pseudo-treatment and pseudo-control group based on schools' pre-NCLB test performances. Schools with high pre-NCLB test scores do not have incentives to improve their efficiency since they already make AYP and face little or no threat of sanctions. They are termed *unthreatened schools* and treated as a control group. By contrast, schools with a non-negligible probability of failing to make AYP have stronger incentives to improve their efficiency in order to avoid sanctions. These are the *threatened schools* and considered as the treatment group.

Since *unthreatened schools* are assumed not to change their behavior, our estimate of NCLB effects is a difference-in-differences estimator: pre-NCLB to post-NCLB changes in the performance of threatened schools, less the same change in the performance of unthreatened schools. A school fixed effect is included so that these are within-school changes. To the extent that even unthreatened schools have responded to NCLB, our estimates will fail to capture the full effects of the policy. However, given the sharp difference in incentives faced by the two sets of schools, we believe that our measures will capture an important part of those effects.

Test scores from a pre-NCLB year, when schools were not influenced by NCLB yet, are used to define unthreatened and threatened schools. I calculate the percentage of students who would have been proficient had NCLB been in effect and construct a 90% confidence interval for that percentage for each subgroup in each tested grade in each school.[5] Next, the minimum of the lower bounds of these intervals is compared to state proficiency target as established in the first year of NCLB. If this minimum exceeds the target (known as the Annual Measurable

[5] We also calculate the confidence interval at 95%. The preliminary results suggest that two measures of confidence interval do not make final results significantly different. Thus, the final analysis uses the 90% confidence interval.

Objective), the school faces little threat of being sanctioned under NCLB. These schools are classified as *unthreatened*. The rest are deemed to be *threatened*.[6]

In addition to the impact of NCLB on school efficiency, the second-stage analysis also investigates whether NCLB reduces output slack. As noted, schools may improve student achievement in the middle group (proficient) at the expense of high-achieving students (advanced). In terms of school efficiency, schools may improve their efficiency, but the output slack related to ADV[7] also increases at the same time, as illustrated in Figure 6. The school moving from C to C' has increased efficiency, but has more slack. Thus, investigating output slack can shed light on the trade-off between low- and high-performing students. As a result, output slack in ADV is also used as a dependent variable at the second-stage analysis to test the hypothesis that NCLB has influenced the tradeoff between proficient and advanced students.

[6] To justify our classification of *unthreatened* and *threatened* schools, a cross-tabulation is reported in Table 1 in the appendix which presents the AYP status for two types of schools in Minnesota, Indiana, and South Carolina. It is shown that all unthreatened schools in Minnesota and Indiana made AYP, and only a very small fraction of unthreatened schools did not make AYP across years. Thus, our classification of *unthreatened* and *threatened* schools is reasonable.

[7] We only examine the slack in ADV. The slack in (PRF+ADV) is not practically meaningful as there are hardly any schools would have large percentage of advanced students but small percentage of proficient ones, except for the first stage models that use (PRF+ADV) in tested subjects as outputs.

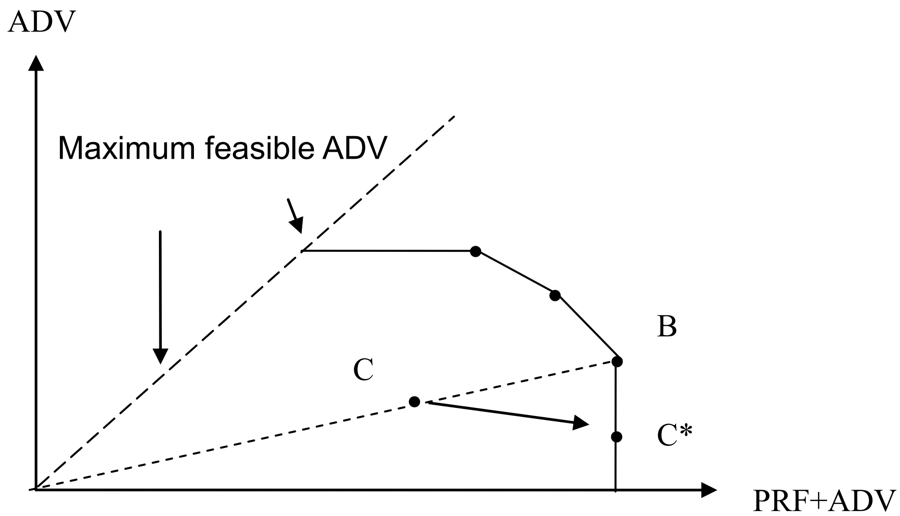


Figure 6, Slack in ADV as the Second Measure of inefficiency

Overall, a total of ten models[8] are analyzed at the second stage analysis. As mentioned earlier, four DEA models are run at the first stage analysis using different outputs. At the second stage, two types of school-fixed effect models are estimated using results from each DEA model as dependent variables: one in which the dependent variable is *efficiency*, the other in which it is *slack*. The ten baseline school-fixed effect models are run separately by school level (elementary vs. middle[9]) using same set of right-hand side regressors: a vector of year dummies, a vector of grade dummies, and interactions between the threatened school indicator and the year dummies. These models can be expressed in a more general form:

$$y_i = \alpha + \beta YEAR + \gamma GRADE + \lambda YAER * THREATENED + \eta SCHOOL + \varepsilon \qquad i = TE, Slack \quad (3)$$

[8] Again the actual numbers of second stage models for each state are determined by the numbers of DEA models that are estimated at the first stage (=2 times numbers of DEA models).

[9] High schools are excluded from our analysis in that high schools are tested in different tests (such as high school exit exam) and are hold accountable using different measurements.

In the models, the grade dummies (γ) capture differences that are constant across grades, while a school fixed effect (η) is included so that the changes examined are within school. Given the assumption that unthreatened schools do not change their behavior under NCLB, the coefficients on the year dummies (β) will capture change of efficiency (or slack in ADV) for unthreatened schools or effects due to other confounding factors. The effect of NCLB can be estimated by subtracting the change of efficiency (or slack in ADV) for unthreatened schools from that for threatened schools. Thus, the coefficients on the interaction of the threatened school dummy with the year dummies (λ) will indicate the real effect of NCLB on efficiency (or slack in ADV). Positive coefficients during the post-2002 period from models using *efficiency* as the dependent variable will provide evidence that NCLB has led schools to improve technical efficiency. Positive coefficients from the models using *slack* as the dependent variable will suggest NCLB has *decreased* school efficiency, as improved achievement for low- and middle-group students has come at the expense of high-achieving students.

A detailed summary of the models for each state that are analyzed in this dissertation is presented in Table 1.

Table 1, Summary of Inputs and Outputs at the First- and Second-Stage Analysis

First -Stage		
	Outputs	Inputs
Model 1	ADV and (PRF+ADV) in math	Per-pupil expenditure (PPE), % of Non-poor students (NON-FRL), % of Non-minority students (NON-MIN), % of Non-limited English proficient students (NON-LEP)
Model 2	ADV and (PRF+ADV) in reading	
Model 3	(PRF+ADV) in math & in reading	
Model 4	ADV and (PRF+ADV) in math, and ADV and (PRF+ADV) in reading	
Second-Stage		
	Outputs	Inputs
Model 1a	Efficiency in math,	Vector of year dummies, Vector of grade dummies, interaction of the year dummies with threatened dummy, school fixed effect
Model 1b	Slack related to ADV in math	
Model 2a	Efficiency in reading	
Model 2b	Slack related to ADV in reading	
Model 3a	Efficiency in math & reading	
Model 3b	Slack related to (PRF+ADV) in math	
Model 3c	Slack related to (PRF+ADV) in reading	
Model 4a	Efficiency in math & reading	
Model 4b	Slack related to ADV in math	
Model 4c	Slack related to ADV in reading	

Strategies of Handling Threats to Validity

There are several possibilities that may introduce threats to validity of the analysis in this dissertation, including: (a) student test scores are inflated, (b) resources in schools are diverted from low-stakes subjects/grades to high-stakes subjects/grades, and (c) outputs other than math and reading are omitted from the analysis. These issues are discussed below.

Inflated Test Scores

As discussed previously, the outputs in this dissertation are measured by student test scores. However, test scores do not necessarily reflect students' real improvement of learning (Koretz, 1988). The rise of test scores could be the result of easier assessments or of students getting more familiar with the assessments. Under this circumstance, using inflated test scores as outputs will contribute to an overestimated efficiency.

The solution to this threat to validity is difference-in-differences estimation. As described above, a difference-in-differences estimator is constructed by classifying schools into two groups: unthreatened and threatened. These two types of schools are expected to have distinct responses toward the threat of sanctions under NCLB. However, if the state assessments get easier or students become more familiar with the assessments, students in both unthreatened and threatened schools are expected to be affected alike, with similar if not identical changes in test scores for both groups of schools. Thus, a change of test scores due to easier assessments or students' growing familiarity with them can be controlled for by subtracting the trends of test scores for unthreatened schools (i.e., control group) from those for threatened schools (i.e., treatment group), that is, difference-in-differences estimation.

Changes in Resources

It is argued that one of the reasons for increased math and reading achievement is that schools divert resources from untested subjects like science and social studies or from untested grades. When facing the threats of being sanctioned, schools have strong incentives to focus more effort and resources on subjects and grades that are tested and held accountable under the requirements of NCLB. For instance, teachers may allocate more instructional time on math and reading while ignoring untested subjects, especially in lower grades. Or schools may assign better teachers to tested grades.

To control for the possibility that resources are extensively devoted to tested grades, this dissertation creates and includes in the second-stage models a new variable that measures the ratio of students in tested grades, or high-stakes grades, to school's total enrollment. Note that Indiana and South Carolina have tested all students through third to eighth grade from the beginning of NCLB. Therefore, this new variable is constant and so not included in the second-stage models for these two states.

Next, the possibility that resources are drawn from other subjects can be examined by using a state that tests low-stakes subjects as well as high-stakes subjects. South Carolina is a good choice as it has administered tests in science and social studies to all students in third through eighth grade from the beginning of NCLB. If schools improve math and reading achievement at the expense of science and social studies, estimates of Model 1 and Model 2 may show a positive association between NCLB and efficiency, but estimates of model in which all four subjects are outputs will not.

Omitted Outputs

This dissertation focuses on two main outputs: math and reading achievement[10]. However, schools produce many outputs, of which math and reading achievement are only a small fraction. Although the South Carolina analysis will show whether estimates are sensitive to the exclusion of achievement in science and social studies, there are may other schooling outputs, including attitudes and behaviors, that are omitted from these models. As a result, it could be argued that this analysis does not include all relevant outputs required to determine whether schools are operating efficiently.

Nevertheless, math and reading are the most fundamental subjects, especially in elementary grades. Students need the basic knowledge learned in math and reading in order to understand the materials in science and social studies. Furthermore, math and reading are especially important for low-achieving students, on whom NCLB is focus. In their case, it is reasonable to argue that math and reading are the critical school outputs and that valid measures of school efficiency can be obtained by focusing on those subjects. Additionally, discipline outcomes and

[10] In the case of South Carolina, four main outputs are considered in this dissertation: math, reading, science, and social studies.

academic outcomes are complementary; students who perform well academically are usually also well-behaved in classrooms. As a result, it is more important to examine schools' performance on math and reading relative to other subjects and exclusions of other outcomes from the analysis will not greatly affect the evaluations of schools' performance.

Data

This study uses data from three states: Indiana, Minnesota, and South Carolina. All data are collected from school report cards available on state Department of Education websites. These three states represent different types of accountability systems under NCLB. As far as Indiana is concerned, students in third, sixth, eighth, and tenth grade were tested since 1998. However, in 2002 the state introduced new content standards and modified the assessments to meet the NCLB accountability requirements, which makes it inappropriate to compare test scores before and after 2002. Therefore, test scores after 2002 are used in this study. In the 2003-2004 school year[11] Indiana began to administer the assessments in reading and math to all students in third through tenth grade. Overall, the Indiana data set contains information related to test scores, finance, and demographics from 2002 to 2006.

In the case of Minnesota, at the beginning of NCLB (i.e., 2001), students in third and fifth grade were tested in math and reading, with seventh grade added in 2004. Minnesota has tested

[11] Indiana administered its assessments in Septembers before 2008, meaning that the tests administered in the 2003-2004 school year actually assessed how students learned in the 2002-2003 school year. Thus, in the remaining of this dissertation, whenever a year is mentioned, it refers to the spring term of the school year, i.e., 2002 in the text refers the 2001-2002 school year.

all students in grades three through eight in math and reading, plus grade 10 reading and grade 11 math since 2006. In the same year the state changed its assessments to align with new academic standards. The Minnesota data set contains information related to test scores, finance, and demographics from 2001 to 2007.

South Carolina has its own school accountability system other than NCLB. The state has administered assessments in math and reading to all students in grades three through eight since 2001 and in science and social studies since 2003, making it a good choice to analyze the tradeoffs among different subjects.

When it comes to specifying models using each state's data, it is essential to make some adjustments based on each state's data characteristics. As mentioned earlier, only grade 3 and 5 were tested prior to 2004 in Minnesota. Therefore, using test scores prior to NCLB (i.e. 2001 and 2002) to classify unthreatened and threatened schools makes most middle schools fall into the threatened group since they do not have third and fifth grade. Thus, it is necessary to use different criteria to classify unthreatened and threatened groups for elementary and middle schools. For the elementary level, schools are classified as unthreatened schools (threatened dummy = 0) if their third and fifth test scores[12] in 2001 and in 2002 are higher than 2003 state proficiency targets, the first year of state proficiency targets; otherwise they are threatened schools. Note that schools that do not have third and fifth grades have missing values of *threatened*. At the secondary level, schools are deemed to be unthreatened if their seventh grade

[12] Test scores here refer to the minimum of the lower bounds of confidence intervals of the observed test scores.

test scores[13] in 2004 are higher than state proficiency targets in that year; otherwise they are threatened schools.

Additionally, Minnesota changed its assessments in 2006, making it inappropriate to compare test scores of 2006 and 2007 to those of previous years. The solution to this problem is to analyze 2006 and 2007 data separately from 2001-2005 data (including the fitting of the DEA models) as if schools in these two periods faced different production frontiers.

South Carolina models also need to be adjusted based on the availability of the data. The South Carolina Department of Education does not provide enrollment information on Limited English proficiency (LEP) students. Thus, it is not feasible to run the original DEA models, one of whose inputs is the percentage of LEP students, as discussed in the previous section. I have replaced the percentage of LEP students with the percentage of Hispanic students in the South Carolina DEA models. It is found that a large proportion of LEP students are Hispanic (Development Associates, Inc., 2003). Thus, the percentage of Hispanic students can be a good substitute of LEP students in DEA models. Because of the inclusion of Hispanic students in the models, the percentage of minority students was replaced by the percentage of African American students accordingly. Consequently, the inputs used in the first stage DEA models for South Carolina include per-pupil expenditure (PPE), the percentage of students eligible for free and reduced-priced lunch (FRL), the percentage of African American students (Black), and the percentage of Hispanic students (Hispanic).

[13] Similarly, test scores here refer to the minimum of the lower bounds of confidence intervals of the observed test scores.

CHAPTER IV

RESULTS

Have Schools Improved Efficiency over Time?

As discussed, at the first stage four DEA models for Indiana and Minnesota and five for South Carolina are estimated using different pairs of outputs, based on different assumptions. Generally, the DEA models within each state yield similar results in terms of the magnitudes of efficiency and slack. Thus, this dissertation only reports results for Indiana and Minnesota of Models 1 and 2, which give estimates for math and reading, respectively, and results for South Carolina of Models 1 to 4, which provide estimates for math, reading, science, and social studies, respectively, as well as a fifth model containing all four subjects. The full results from other models are presented in the appendix.

The results of mean efficiency from 2001 to 2007 for each state are presented in Table 2 – Table 4. Overall, schools in three states are fairly inefficient in producing math and reading achievement, which is consistent with other studies' conclusions. The average efficiencies in both math and reading for South Carolina are substantially less than those for other two states. Since each state has its own assessments and production frontier, however, it is meaningless to compare the results across three states. Thus, in the following discussion, each state's results will be reported separately, in the order of Minnesota, Indiana, and South Carolina. However, it should be cautious when interpreting results of mean efficiency because of the threats to validity discussed in the early sections. Because of the possibility that student test scores do not reflect the real learning, the estimated mean efficiency scores reported here may not necessarily illustrate the actual efficiency. As noted above, the difference-in-differences estimator presented later deals with many of these threats.

Minnesota

Measures of mean efficiency for Minnesota schools are presented in Table 2, with 2a reporting results of math and 2b results of reading. In the case of Minnesota, the average efficiency in math ranges from .636 in eighth grade to .737 in third grade, and that in reading ranges from .71 in eighth grade to .778 in fifth grade. Thus, on average schools achieve 73.7% of the feasible level of math achievement (i.e., percentage of ADV and percentage of (PRF+ADV) students) and 75.6% of the feasible level of reading achievement in third grade. Schools would have had 26.3% more ADV and (PRF+ADV) students in math and 24.4% more in reading in third grade had they operated efficiently. In addition, Table 2 shows that efficiency is higher for reading than for math, especially in grades six and eight, implying that resources are relatively better utilized in producing reading achievement.

Generally, schools in Minnesota have raised their efficiency in producing math and reading achievement over time in the grades that tested in several years (i.e., third, fifth, and seventh grades). However, these grades show different patterns of improvements. In third grade, efficiency in math and reading has increased steadily from pre-NCLB years to post-NCLB years, increasing from .68 in 2001 to .79 in 2007. Fifth and seventh grade increased efficiency from 2001 to 2005, but dropped substantially in 2006 and 2007.

Table 2, Mean Efficiency by Grade and Year, Model 1 and Model 2, Minnesota

2a, Efficiency in Math (Model 1)								
	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	0.677	0.665	0.733	0.719	0.790	0.784	0.789	0.737
Grade 4						0.709	0.717	0.713
Grade 5	0.691	0.720	0.766	0.753	0.826	0.642	0.686	0.726
Grade 6						0.637	0.678	0.658
Grade 7				0.714	0.825	0.646	0.686	0.718
Grade 8						0.641	0.631	0.636

2b, Efficiency in Reading (Model 2)								
	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	0.694	0.683	0.739	0.749	0.798	0.825	0.806	0.756
Grade 4						0.787	0.739	0.763
Grade 5	0.761	0.768	0.785	0.768	0.831	0.784	0.753	0.778
Grade 6						0.759	0.724	0.742
Grade 7				0.761	0.819	0.744	0.697	0.755
Grade 8						0.723	0.697	0.710

One possible explanation of decreasing efficiency in fifth and seventh grade was the increase in the number of grades tested in 2006 when fourth, sixth, and eighth grade were added to the testing program. When more grades are tested, schools need to redistribute their resources among tested grades to make every tested student proficient and make AYP. Given the fixed amount of resources, each tested grade will get fewer resources when more grades are tested. This possibility will be tested in the second stage analysis.

Indiana

Mean efficiencies for Indiana schools are reported in Table 3, with 3a showing results of math and 3b results of reading. Schools in Indiana have an average efficiency in math ranging from .685 in fourth grade to .785 in sixth and eighth grade, and in reading from .742 in seventh grade to .782 in third and fifth grade. The average efficiencies across grades do not differ significantly, except for fourth grade math. On average, fourth grade in Indiana only achieves 68.5% of the feasible level of math outputs under current operations, while in other grades efficiency exceeds 74%.

When we compare the average efficiency in math with that in reading, it is found that elementary grades, third, fourth, and fifth grade, do a better job in reading, but perform less satisfactorily in math than the other three grades. The elementary grades' average efficiency in reading was around .78, while the numbers for the middle school grades (six to eighth) were .74 or .76. The pattern in math is switched: as the middle school grades' efficiency in math was approximately .78, the elementary grades' was .69 to .76. However, the difference across the six grades was not substantial.

The estimates of efficiency in math and in reading increased from 2002 to 2006, except for third grade math. The biggest increase happened in sixth grade math. In 2002 the estimate of efficiency in math of sixth grade was .741, suggesting that sixth grade achieved 74.1% of the feasible output level in 2002. It increased to .82 in 2006, which is transformed into an 8%

increase in ADV and (PRF+ADV) students in math. As for third grade, it did not improve its efficiency in math over time. However, the decrease in efficiency in math for third grade was very small, only 1%.

Table 3, Mean Efficiency by Grade and Year, Model 1 and Model 2, Indiana

3a, Efficiency in Math (Model 1)						
	2002	2003	2004	2005	2006	Mean
Grade 3	0.741	0.754	0.752	0.744	0.728	0.744
Grade 4		0.678	0.684	0.685	0.694	0.685
Grade 5		0.733	0.757	0.761	0.771	0.755
Grade 6	0.741	0.760	0.791	0.815	0.820	0.785
Grade 7		0.761	0.775	0.789	0.801	0.781
Grade 8	0.777	0.780	0.781	0.776	0.810	0.785
3b, Efficiency in Reading (Model 2)						
	2002	2003	2004	2005	2006	Mean
Grade 3	0.779	0.789	0.778	0.773	0.794	0.782
Grade 4		0.772	0.774	0.789	0.783	0.780
Grade 5		0.766	0.774	0.793	0.797	0.782
Grade 6	0.725	0.733	0.737	0.741	0.752	0.737
Grade 7		0.748	0.735	0.730	0.758	0.742
Grade 8	0.739	0.762	0.755	0.756	0.769	0.756

South Carolina

Results of mean efficiency for schools in South Carolina are reported in Table 4, which consists of four panels that present results of math, reading, science, and social studies, respectively. Schools in South Carolina were very inefficient in using their resources to produce educational outcomes. On average, they only achieved less than 50% of the feasible level of outputs, for all four subjects tested, with the exception of third grade reading. Take third grade for example, the average efficiency in math was .37 and that in reading was .528 across seven years, meaning that without using more inputs third grade in South Carolina would have achieved 63% more ADV and (PRF+ADV) students in math and 47.2% more ADV and (PRF+ADV) students in reading if they had operated at the efficient level.

The estimates of efficiency in math for the higher grades (sixth through eighth) increased steadily from 2001 to 2005 but decreased afterward. The lower grades did not have clear patterns in the changes of estimates of efficiency; they went up and down differently across seven years. Similarly, there are no obvious patterns in the estimates of efficiency in reading, as the estimates for each grade fluctuate greatly over time. Nonetheless, the estimates of efficiency for all grades, except for third grade, reached the lowest point in 2003, the year when NCLB took effect as well as when science and social studies were tested.

As noted, South Carolina tested not only math and reading, the two mandatory tested subjects under NCLB, but also science and social studies, which makes it possible to look at the issue of tradeoffs between high-stakes subjects, i.e. math and reading, and low-stakes subjects, i.e. science and social studies. One possible way to examine the tradeoffs among subjects was to estimate schools' efficiency in producing low-stakes test scores. Panel 4c and 4d in Table 4 report the results of mean efficiency in these two low-stakes subjects.

Overall, the average efficiency in science and social studies was lower than that in math and reading, ranging from .292 in third grade to .459 in seventh grade for science and .309 in fifth grade to .441 in eighth grade for social studies. In addition, the higher grades, in particular seventh and eighth grade, have higher efficiency in science and social studies than the elementary grades by approximately 10%.

Each grade has different patterns in terms of the change of efficiency in science and social studies. Take third grade for example, the estimates of efficiency in science stayed relatively stable from 2003 to 2006 and increased in 2007, while those in social studies increased steadily over time. The estimates of efficiency in both subjects for fourth grade also continuously increased over time. Fifth and sixth grade raised their estimates of efficiency in both subjects from 2003 to 2005, dropped in 2006, and then moved up again in 2007. The estimates of efficiency in science for seventh grade went up after 2003, stayed relatively stable from 2004 and 2006 and then increased again in 2007. Additionally, seventh grade social studies and eighth grade science have the same patterns as fifth and sixth grade. However, the pattern of efficiency in social studies for eighth grade was different from others; the estimate was increased substantially from 2003 to 2004 and 2005, but dropped greatly in 2006 and 2007.

Despite the different patterns of changes in efficiency in two low-stakes subjects, overall the estimates of efficiency across grades, except for eighth grade social studies, are increased when we compare the first and the last year's results. For instance, the middle school grades had greater efficiency in science by around 22%. In all grades except eighth, the estimate of efficiency in social studies increased by more than 10%. The increase in both low-stakes subjects, together with the relatively flat trends for high-stakes subjects, implies that low-stakes subjects in South Carolina were not ignored.

Table 4, Mean Efficiency by Grade and Year, Model 1 to Model 4, South Carolina

4a, Efficiency in Math (Model 1)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	0.420	0.367	0.385	0.348	0.346	0.385	0.336	0.370
Grade 4	0.322	0.408	0.385	0.405	0.449	0.442	0.447	0.408
Grade 5	0.365	0.352	0.331	0.391	0.383	0.384	0.371	0.368
Grade 6	0.307	0.322	0.413	0.424	0.431	0.403	0.409	0.387
Grade 7	0.404	0.402	0.446	0.482	0.507	0.476	0.483	0.457
Grade 8	0.333	0.347	0.350	0.400	0.424	0.416	0.353	0.375

4b, Efficiency in Reading (Model 2)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	0.474	0.442	0.480	0.599	0.589	0.557	0.552	0.528
Grade 4	0.464	0.392	0.373	0.447	0.414	0.463	0.472	0.432
Grade 5	0.394	0.329	0.268	0.361	0.384	0.439	0.390	0.366
Grade 6	0.410	0.432	0.352	0.366	0.350	0.393	0.397	0.386
Grade 7	0.509	0.443	0.386	0.423	0.398	0.419	0.465	0.435
Grade 8	0.419	0.470	0.346	0.462	0.530	0.436	0.424	0.441

4c, Efficiency in Science (Model 3 for South Carolina)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3			0.282	0.267	0.296	0.278	0.336	0.292
Grade 4			0.295	0.354	0.374	0.395	0.439	0.371
Grade 5			0.295	0.332	0.362	0.344	0.404	0.347
Grade 6			0.281	0.388	0.416	0.339	0.457	0.376
Grade 7			0.343	0.466	0.464	0.458	0.565	0.459
Grade 8			0.329	0.375	0.467	0.428	0.541	0.428

4d, Efficiency in Social Studies (Model 4 for South Carolina)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3			0.242	0.289	0.373	0.421	0.455	0.356
Grade 4			0.239	0.337	0.362	0.371	0.398	0.342
Grade 5			0.261	0.303	0.327	0.298	0.359	0.309
Grade 6			0.205	0.327	0.368	0.358	0.437	0.339
Grade 7			0.374	0.440	0.461	0.403	0.495	0.434
Grade 8			0.381	0.508	0.512	0.481	0.325	0.441

It is worth examining why schools in South Carolina were substantially less efficient than counterparts in Indiana and Minnesota. One way to examine the issue is to look at the scatter plot of two outputs, ADV and (PRF+ADV), for each state. Since DEA measures individual school's efficiency based on the comparison of the individual school to other schools with similar input level, I selected a subset of observations that have the same reference group as estimated by DEA, ensuring that those observations have similar input levels[14]. Figure 7 contains scatter plots of ADV and (PRF+ADV) in reading in third grade for three states.

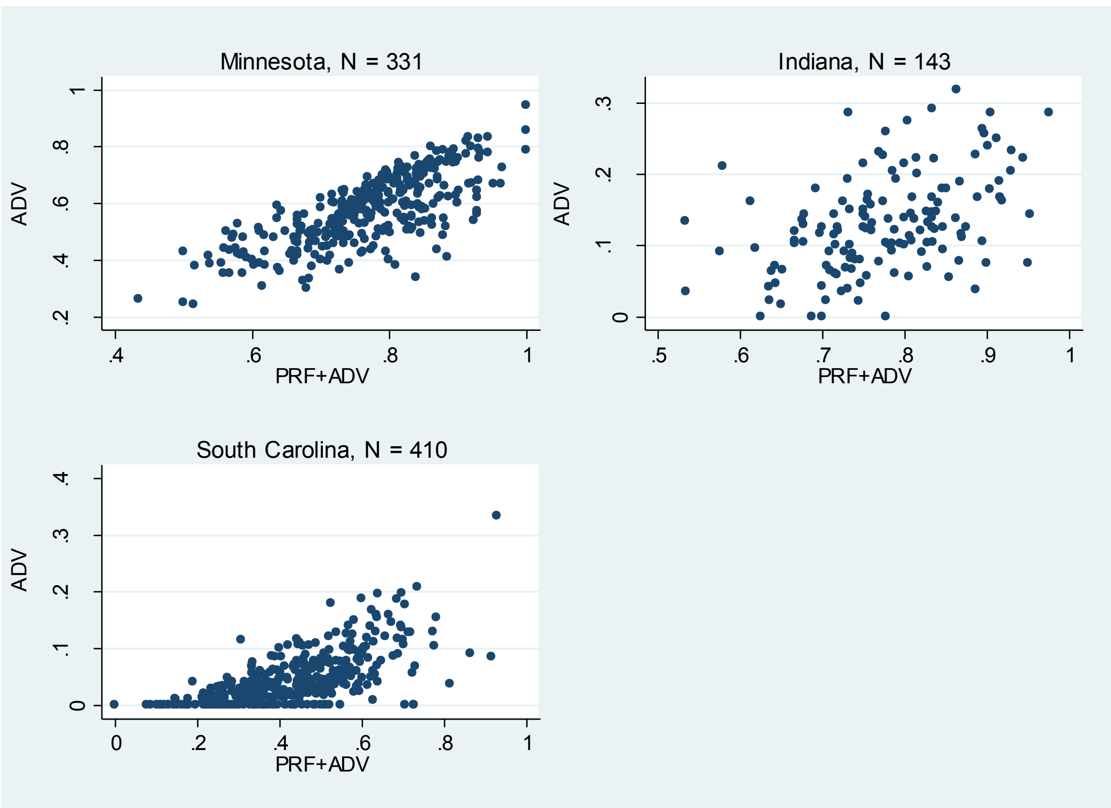


Figure 7, Scatter Plot of Two Outputs (ADV and PRF+ADV), Reading for Grade 3, by States

[14] I have tried to examine observations with different sets of reference group for three states and the similar patterns are found. Therefore, the Figures reported here can be considered as the representative of the data distribution across states.

Every state has a distinctive scatter plot, which in part explains the differences in mean efficiency among them. Minnesota has a positive relationship between ADV and (PRF+ADV), indicating that schools that have fewer below proficiency students tend to have more advanced students. As far as Indiana concerned, the scatter plot of two outputs is more spread out relative to that of Minnesota and does not show a clear pattern of the relationship, implying that schools that have fewer below proficiency students do not necessarily have more advanced students.

Unlike the other two states, a majority of schools in South Carolina are concentrated on the lower left corner of the graph, suggesting that many schools have very low percentage of advanced students. Virtually no schools have more than 20% of advanced students in reading in third grade. Thus, a few high-achieving schools determined the production frontier, with most schools lying far below it, accounting for to their low efficiency scores.

One reason for the overall low percentage of advanced students in South Carolina is that South Carolina is one of the states with the most difficult proficiency standards in the nation (Cronin, Dahlin, Adkins, & Kingsbury, 2007). Because of the high proficiency cut scores in South Carolina, it is natural that fewer students can reach the advanced level. Minnesota and Indiana's proficiency cut scores are around the national median, which partially explains the difference in average efficiency between them and South Carolina.

Has NCLB Influenced the Change of School Efficiency over Time?

Although average estimates of efficiency have increased since NCLB was introduced, it is unclear whether the increase reflects real improvement in efficiency because of the threats to validity discussed in the methodology section. It is possible that easier tests or students' familiarity with tests will result in higher test scores without real improvement in students' real learning. If so, estimates of efficiency are inflated. However, valid inferences about the effect of NCL B on efficiency can still be made using the difference-in-differences estimator that compares trends in efficiency between unthreatened and threatened schools, as discussed in this section.

Overall, as expected, unthreatened schools on average have higher efficiency than threatened schools across three states. However, differences between two groups of schools were not the same in Minnesota, Indiana, and South Carolina. Thus, similar to the previous section, this section reports three states' results separately.

Minnesota

Mean efficiency for unthreatened and threatened schools for each tested grade from 2001 to 2007 are depicted in Figures 8 – 10, with Figure 8 illustrating results for third grade, Figure 9 fifth grade, and Figure 10 seventh grade, respectively[15]. The gaps in efficiency between unthreatened and threatened schools were fairly large in all three grades when the assessments

[15] Since 4th, 6th, and 8th grade were only tested for two years and did not have apparent trends, they are not included in the Figures. Rather, their results, together with other three grades', are reported in Table B.1 in the Appendix B.

were first administered. For instance, both third and fifth grade had a gap of 20% in 2001 and seventh grade had a gap of 16% in 2004 between two groups of schools with regard to efficiency in math. The corresponding gaps in reading were 18%, 15%, and 13%.

Generally, the gap in efficiency is bigger in math than that in reading. It is possible that math teachers in different schools have substantially different quality. Given the fact that public schools are short of highly-qualified math teachers (Fetler, 1999; Ingersoll, 1999; National Commission on Teaching and America's Future, 1996), low-performing schools may have more difficulties in recruiting good math teachers, which results in even lower math achievement.

The gaps in efficiency for both math and reading narrowed over time, though the patterns varied across the three grades. In third grade, unthreatened schools are more efficient in producing math achievement than threatened counterparts by 20.7% in 2001. But the gap between them is reduced to 8.7% in 2007. Similarly, the discrepancy in efficiency with regard to producing reading achievement for third grade drops to 9.2% in 2007 from 18% in 2001. In the case of fifth grade, the gap in efficiency does not close as much as that for third grade, declining substantially from 19.6% in 2001 to 8.6% in 2005 but increasing again to 13.7% in 2006 and 11.1% in 2007 with respect to math and from 15.3% in 2001 to 9% in 2005 through 2007 with respect to reading.

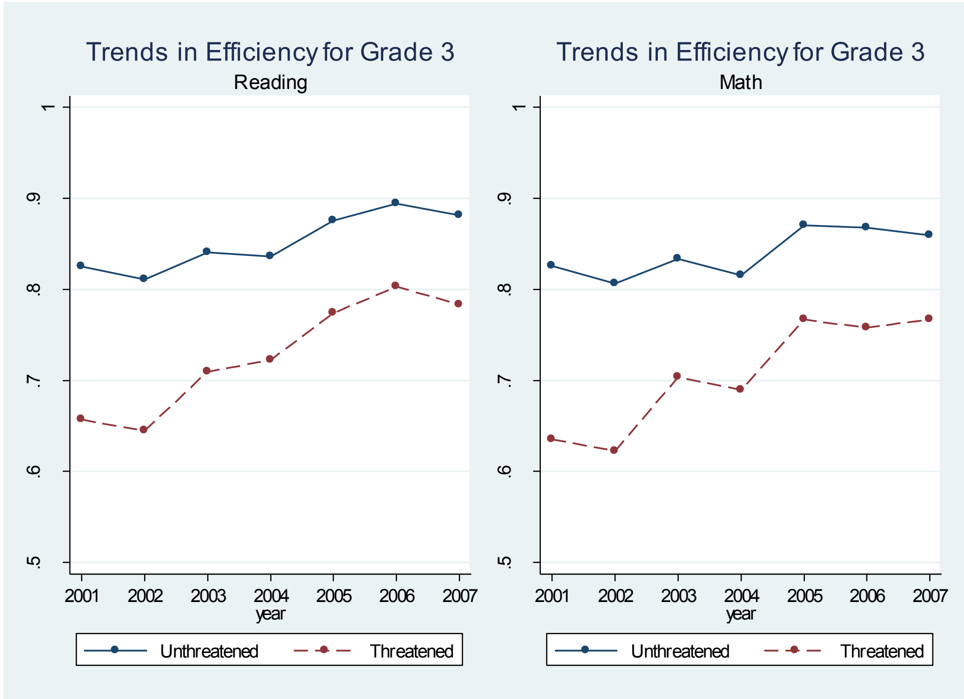


Figure 8, Trends in Efficiency, Grade 3, Minnesota

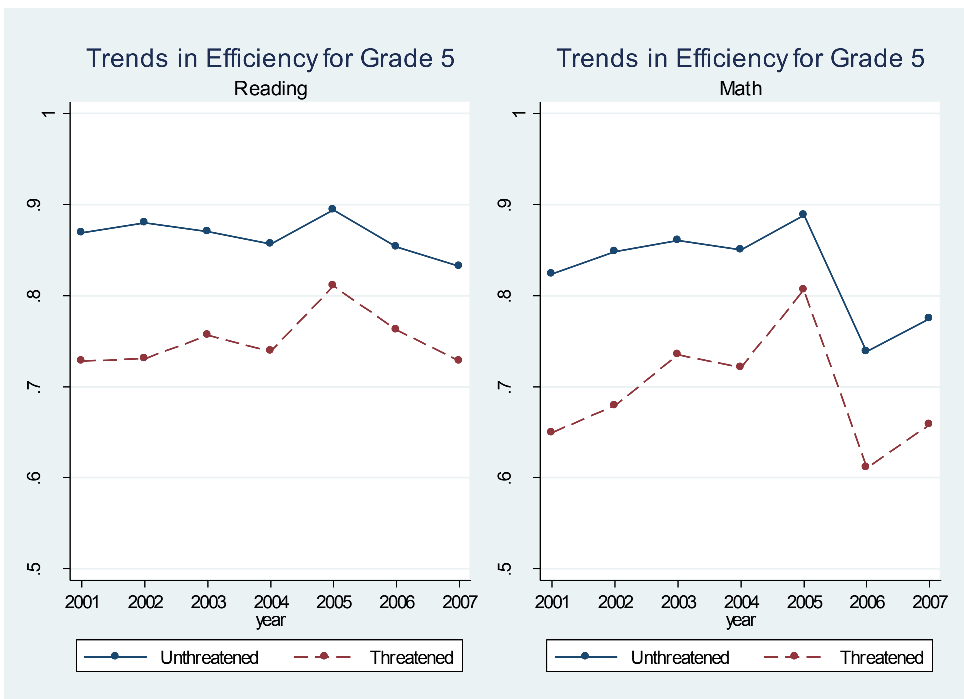


Figure 9, Trends in Efficiency, Grade 5, Minnesota

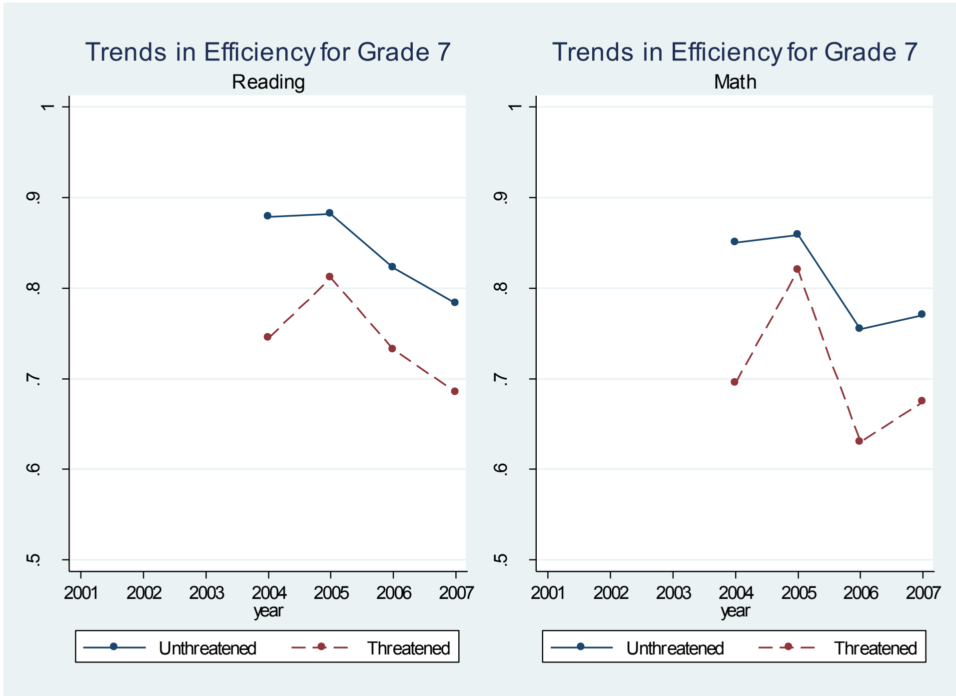


Figure 10, Trends in Efficiency, Grade 7, Minnesota

As noted earlier, the changes in the difference in efficiency between unthreatened and threatened schools capture the effects of NCLB. Thus, the narrowing of the efficiency gap from 2001 to 2007 implies that NCLB has had a positive impact on school efficiency: schools, especially threatened ones, tended to raise their efficiency after NCLB was introduced. To test these differences for statistical significance, the second-stage model described above was estimated using estimated efficiency at the school level as the dependent variable. Results are presented in Table 5. The upper panel is results for elementary schools and the lower panel for middle schools. The first four columns report results for math and the last four results for reading.

The first two columns of Table 5, with the first one indicating coefficient and the second one standard error, are results of the baseline model. Unthreatened schools at elementary levels do not greatly change their efficiency after NCLB is implemented. Most of the coefficients on year dummies are not statistically significant, and even for the statistically significant ones, their magnitudes are small (no more than .05). This result confirms our assumption that unthreatened elementary schools do not have an incentive to become more efficient. Furthermore, the small magnitude of coefficients implies that there are no test effects in elementary schools causing differences in performance over time.

The coefficients on threatened school interacted with year are positive and significant. Overall, they increase steadily from 2002, reaching a peak in 2005 or 2006 depending on the model, and then decrease slightly afterward. This result suggests that NCLB has a significant impact on school efficiency at the elementary level.

The results for middle schools, however, differ greatly from those for elementary schools. Unthreatened schools appear to be less efficient in 2006 and 2007 than earlier. The differences are about 7% in math and 6% and 10% in reading. On the assumption that unthreatened schools have not changed their behavior in response to NCLB, these negative coefficients likely represent test effects of some kind.

The coefficients on the interactions with the threatened school indicator are positive, although not all of them are statistically significant. In math, threatened middle schools were more efficient by 10% in 2005 than in 2004 (the baseline year), but the corresponding coefficients in 2006 and 2007 are smaller (.007 and .021, respectively) and statistically insignificant. Similarly, there is improvement in reading model in 2005 of .063, but only .036 and .022 in 2006 and 2007, respectively. These results indicate that threatened schools improved their efficiency in producing math and reading achievement in 2005, but that progress slipped after new assessments were introduced in 2006.

Before concluding that in Minnesota NCLB has had a positive impact on efficiency, we need to consider two potentially confounding factors. The first confounding factor is the fact that Minnesota began to test more grades (grade 3 through 8) in 2006, which may alter the distribution of resources among grades and affect each grade's efficiency. Second, higher efficiency in post-NCLB years may be due to mean-reversion. If schools had bad luck resulting in abnormally low test scores in 2001 and 2002, they could have been mistakenly classified as threatened schools. The large gains they experienced in later years would not have been the effect of NCLB, but the consequence of moving back to a more normal level of performance.

To distinguish NCLB effects from those confounding effects, we have conducted several tests. First we examine whether testing more grades, as occurred after 2006, affected schools' performance. We create a new variable measuring the ratio of students in tested grades, or high-stakes grades, to school's total enrollment. The results of the models including this new variable – *pct_highstakes* – and its interaction with the threatened school dummy are presented in the third and fourth column for each model in Table 5.

Table 5, Coefficients from 2nd-stage Models, Model 1 and Model 2, Minnesota
Using *Efficiency* as Dependent Variable

	Efficiency in Math				Efficiency in Reading			
Elementary Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.690	*(0.003)*	0.729	*(0.006)*	0.717	*(0.003)*	0.737	*(0.005)*
Year 2002	0.002	*(0.008)*	0.002	*(0.008)*	-0.001	*(0.007)*	-0.002	*(0.007)*
Year 2003	0.022	*(0.008)*	0.022	*(0.008)*	0.008	*(0.007)*	0.008	*(0.007)*
Year 2004	0.009	*(0.008)*	0.010	*(0.008)*	0.000	*(0.007)*	0.000	*(0.007)*
Year 2005	0.055	*(0.008)*	0.056	*(0.008)*	0.038	*(0.007)*	0.038	*(0.007)*
Year 2006	-0.017	*(0.008)*	0.013	*(0.011)*	0.030	*(0.006)*	0.046	*(0.009)*
Year 2007	-0.010	*(0.008)*	0.019	*(0.010)*	0.004	*(0.006)*	0.019	*(0.009)*
Grade 4	-0.017	*(0.004)*	-0.017	*(0.004)*	-0.017	*(0.003)*	-0.017	*(0.003)*
Grade 5	-0.007	*(0.002)*	-0.007	*(0.002)*	0.026	*(0.002)*	0.026	*(0.002)*
Threatened*Year 2002	0.005	*(0.009)*	0.005	*(0.009)*	-0.002	*(0.008)*	-0.002	*(0.008)*
Threatened*Year 2003	**0.053**	***(0.009)***	**0.053**	***(0.009)***	**0.034**	***(0.008)***	**0.034**	***(0.008)***
Threatened*Year 2004	**0.056**	***(0.009)***	**0.056**	***(0.009)***	**0.043**	***(0.008)***	**0.043**	***(0.008)***
Threatened*Year 2005	**0.089**	***(0.009)***	**0.089**	***(0.009)***	**0.063**	***(0.008)***	**0.063**	***(0.008)***
Threatened*Year 2006	**0.059**	***(0.009)***	**0.064**	***(0.012)***	**0.063**	***(0.007)***	**0.065**	***(0.010)***
Threatened*Year 2007	**0.074**	***(0.009)***	**0.079**	***(0.012)***	**0.053**	***(0.007)***	**0.055**	***(0.010)***
Pct_HighStakes			**-0.115**	***(0.028)***			**-0.061**	***(0.023)***
Threatened*Pct_HighStakes			-0.020	*(0.034)*			-0.010	*(0.028)*
Middle Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.680	*(0.009)*	0.675	*(0.013)*	0.762	*(0.008)*	0.785	*(0.011)*
Year 2005	0.025	*(0.024)*	0.024	*(0.024)*	-0.002	*(0.021)*	-0.002	*(0.021)*
Year 2006	-0.069	*(0.020)*	-0.120	*(0.056)*	-0.058	*(0.017)*	-0.039	*(0.049)*
Year 2007	-0.058	*(0.020)*	-0.110	*(0.056)*	-0.090	*(0.017)*	-0.071	*(0.049)*
Grade 7	0.046	*(0.008)*	0.046	*(0.008)*	0.015	*(0.007)*	0.015	*(0.007)*
Grade 8	0.027	*(0.008)*	0.027	*(0.008)*	0.018	*(0.007)*	0.018	*(0.007)*
Threatened*Year 2005	**0.096**	***(0.026)***	**0.097**	***(0.026)***	**0.063**	***(0.022)***	**0.063**	***(0.022)***
Threatened*Year 2006	0.007	*(0.021)*	0.056	*(0.060)*	**0.036**	***(0.018)***	0.068	*(0.053)*
Threatened*Year 2007	0.021	*(0.021)*	0.071	*(0.060)*	0.022	*(0.018)*	0.053	*(0.053)*
Pct_HighStakes			0.098	*(0.099)*			-0.034	*(0.087)*
Threatened*Pct_HighStakes			-0.093	*(0.107)*			-0.066	*(0.093)*

The coefficient on this ratio is negative and significant at the elementary level (-.115 and -.061 for the math and reading model, respectively), suggesting that having more students in high-stakes grades decreases efficiency of unthreatened elementary schools. However, threatened schools do not differ significantly from unthreatened ones with regard to the impact of testing more grades, as the coefficient of the interaction between *pct_highstakes* with the threatened dummy is insignificant. Additionally, the coefficients on the interaction terms from the extended models do not differ substantially from those from the basic model. Thus, the positive NCLB effect at the elementary school level remains.

At the middle school level, having more grades tested does not influence efficiency for either threatened or unthreatened schools as the coefficients on *pct_highstakes* are not statistically significant. Moreover, the inclusion of the new variable does not alter greatly the coefficients on the interaction terms in the extended models compared to the baseline models; even though the coefficients on the interaction terms become larger in the extended models, they are still statistically insignificant. Therefore, conclusions about the NCLB effect at the middle school level are not change by controlling for the percentage of students tested.

Second, to investigate whether there is mean reversion, we change our criteria of defining unthreatened and threatened schools. Rather than using both 2001 and 2002 test results, we use only 2001 results. Thus, coefficient estimates for 2002 will capture mean reversion effects, if any, so that comparisons between 2002's estimates with later years' will tell the net increase of efficiency. Results are similar to those reported above[16] with little change in the estimated rate of improvement among threatened schools. Estimates for later years differ significantly from 2002 estimates, suggesting that threatened schools significantly raised efficiency after NCLB was implemented.

Indiana

The trends in efficiency in Indiana for three grades that have pre-NCLB test scores (third, sixth, and eighth) by school type (*Unthreatened* vs. *Threatened*) are illustrated in Figures 11 – 13[17]. In all three grades, unthreatened schools are more efficient at the outset of our sample period. In third grade math, unthreatened schools are more efficient than threatened schools by 13.7%. The gaps in sixth and eighth grades are 14.8% and 12.4%, respectively. Patterns in reading are similar, with corresponding gaps of 12.8%, 14.4%, and 15%.

As in Minnesota, the gap in efficiency declined during the NCLB years, particularly in 2003. The gaps narrowed to around 11% for all tested subjects across three grades in 2003, a reduction of approximately 1% to 4% from 2002. In later years, the gaps continued to narrow, though the reductions were not very large. These results imply that NCLB has had a positive impact on school efficiency in Indiana, though as in the case of Minnesota, we will need to confirm this by testing for the presence of mean reversion.

[16] Therefore, Table 5 contains results from analyses using both 2001 and 2002 test scores to classify unthreatened and threatened schools, and the results from analyses using only 2001 test scores to classify two groups of schools are reported in the appendix.
[17] These three grades are discussed here because our focus is on the comparison before and after NCLB. The results for other three grades that do not have pre-NCLB test scores (i.e. fourth, fifth, and seventh) are illustrated in the appendix.

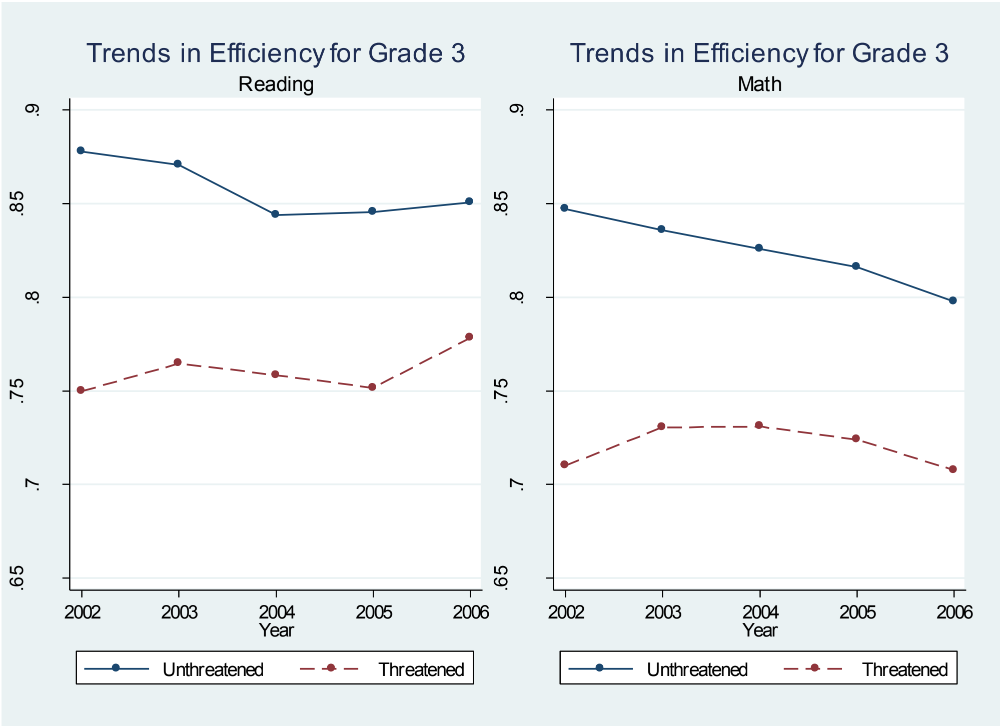


Figure 11, Trends in Efficiency for Grade 3, by Subjects, Indiana

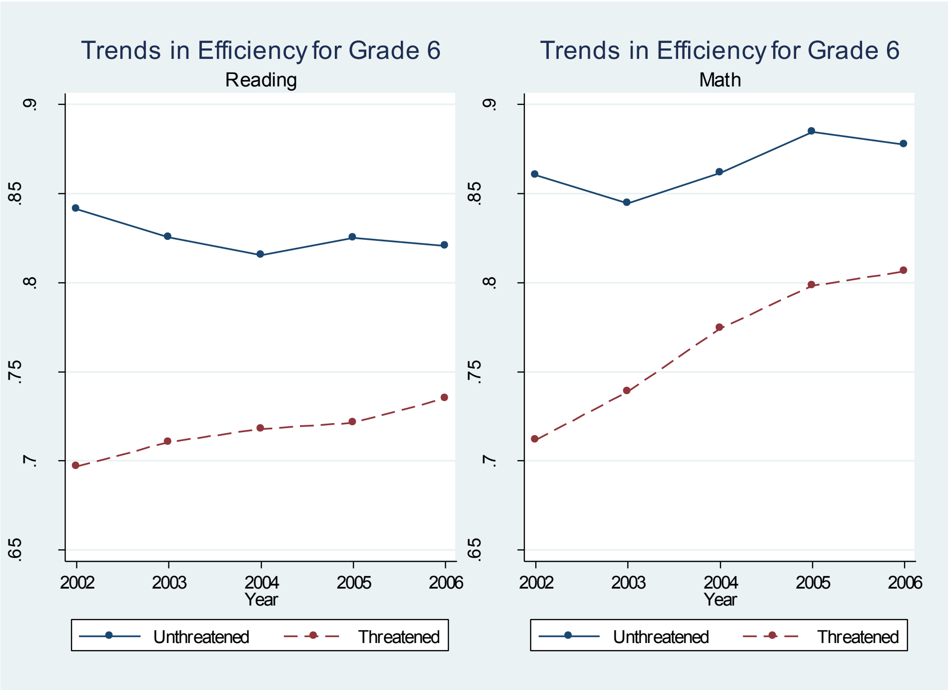


Figure 12, Trends in Efficiency for Grade 6, by Subjects, Indiana

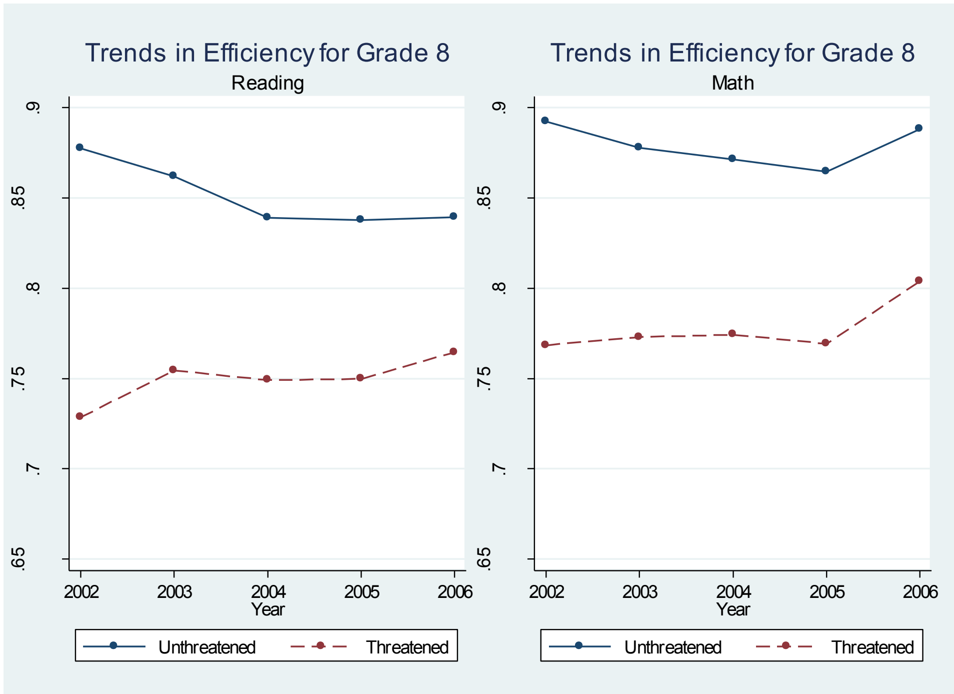


Figure 13, Trends in Efficiency for Grade 8, by Subjects, Indiana

Results of the second-stage models that estimated relationships between NCLB and efficiency for Indiana are reported in Table 6, where the upper panel contains results for elementary schools and the lower results for middle schools. The first two columns present results for math and the latter two results for reading. The coefficients on year dummies indicate a small, statistically significant decline in efficiency (though no more than .05 for math and .03 for reading). Given the assumption that unthreatened schools do not have incentives to change their behavior, these are presumably due to factors other than NCLB.

The coefficients on the interactions between threatened school dummy and year dummies are positive and statistically significant, for both math and reading, at the elementary level. In math, for example, there is an improvement of .037 in 2003 relative to 2002 (the baseline). The coefficients increased thereafter, reaching a peak of .066 in 2006. Similarly, the coefficients on these interactions in reading are positive and statistically significant, meaning that NCLB has had a positive, significant influence on school efficiency in reading, though the improvement is not as great as in math.

Like elementary schools, unthreatened middle schools experienced a modest decline in efficiency after the start of NCLB. In math, this decline was reversed after 2004. Nonetheless, the magnitudes of the coefficients for both math and reading are fairly small, ranging from -.006 to -.037, suggesting that unthreatened middle schools have not changed their behavior significantly.

Threatened middle schools have positive and statistically significant coefficients that have increased in both subjects: .03 in 2003, .05 in 2004 and 2005, and .06 in 2006. As mentioned earlier, it is possible that the increase of efficiency in unthreatened schools is due to mean reversion. To examine this issue, I compare the coefficient on the interaction terms for 2003 to those of later years. It is very likely that mean reversion will happen in second year in the dataset, i.e. 2003 in the current case, given that classification as threatened or unthreatened was based on 2002 performance. The differences between 2003 and other years in terms of the estimated coefficients on the interaction terms from second-stage models are reported in Table 7. The estimates of 2004, 2005, and 2006 are found to be significantly different from those of 2003, suggesting that threatened schools significantly raised efficiency after NCLB was implemented.

Table 6, Coefficients from 2nd-stage Models, Model 1 and Model 2, Indiana
Using *Efficiency* as Dependent Variable

	Efficiency in Math		Efficiency in Reading	
Elementary Schools	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.743	*(0.003)*	0.782	*(0.002)*
Year 2003	-0.034	*(0.008)*	-0.018	*(0.006)*
Year 2004	-0.034	*(0.008)*	-0.028	*(0.006)*
Year 2005	-0.039	*(0.008)*	-0.024	*(0.006)*
Year 2006	-0.049	*(0.008)*	-0.030	*(0.006)*
Grade 4	-0.059	*(0.002)*	-0.004	*(0.002)*
Grade 5	0.011	*(0.002)*	0.000	*(0.002)*
Threatened*Year 2003	**0.037**	***(0.009)***	**0.019**	***(0.007)***
Threatened*Year 2004	**0.048**	***(0.009)***	**0.030**	***(0.007)***
Threatened*Year 2005	**0.052**	***(0.009)***	**0.036**	***(0.007)***
Threatened*Year 2006	**0.066**	***(0.009)***	**0.051**	***(0.007)***
Middle Schools	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.749	*(0.003)*	0.723	*(0.003)*
Year 2003	-0.011	*(0.007)*	-0.013	*(0.008)*
Year 2004	-0.007	*(0.008)*	-0.037	*(0.008)*
Year 2005	0.006	*(0.008)*	-0.034	*(0.008)*
Year 2006	0.011	*(0.008)*	-0.027	*(0.008)*
Grade 7	-0.015	*(0.003)*	-0.002	*(0.003)*
Grade 8	0.026	*(0.004)*	0.041	*(0.004)*
Threatened*Year 2003	**0.030**	***(0.008)***	**0.031**	***(0.008)***
Threatened*Year 2004	**0.048**	***(0.008)***	**0.053**	***(0.008)***
Threatened*Year 2005	**0.050**	***(0.008)***	**0.050**	***(0.008)***
Threatened*Year 2006	**0.058**	***(0.008)***	**0.060**	***(0.008)***

To summarize, in Indiana there are small changes in efficiency that appear to be attributable to factors other than NCLB. However, significant differences between trends among threatened and unthreatened schools indicate that NCLB has had a positive, significant influence on efficiency in public schools in Indiana.

Table 7, Difference in Coefficients on Interaction Terms between 2003 and Later Years

	Reading		Math	
Elementary Schools				
	Coef.	S.E.	Coef.	S.E.
2004-2003	**-0.012**	***(0.005)***	-0.011	*(0.006)*
2005-2003	**-0.018**	***(0.005)***	**-0.015**	***(0.006)***
2006-2003	**-0.032**	***(0.005)***	**-0.029**	***(0.006)***
Middle Schools				
	Coef.	S.E.	Coef.	S.E.
2004-2003	**-0.022**	***(0.007)***	**-0.017**	***(0.007)***
2005-2003	**-0.019**	***(0.008)***	**-0.020**	***(0.007)***
2006-2003	**-0.029**	***(0.008)***	**-0.028**	***(0.007)***

South Carolina

Trends in efficiency in math and reading for third through eighth grade in South Carolina are depicted in Figures 14 – 19. Unthreatened schools were more efficient in math than threatened schools by about 20% in the elementary grades (i.e. third to fifth grade), 24% in sixth grade, and 30% in seventh and eighth grade. The gaps in reading were approximately 20% across grades. However, unlike in Minnesota and Indiana, the gaps in both subjects in South Carolina did not close, with the exception of eighth grade.

There are no obvious patterns on the gaps in efficiency between two types of schools in both subjects from third to sixth grade. The gap in math for seventh grade decreased from .288 in 2001 to .213 in 2007, a nearly 8% reduction over time, while that for eighth grade declined from .308 to .152. In the case of reading, the gap for seventh grade has fluctuated frequently over time, but not closed much when the first and the last year are compared (.227 in 2001 and .192 in 2007, respectively). Nonetheless, the gap for eighth grade narrowed continuously from .242 in 2001 to .182 in 2007.

To conclude, there is little change in the gaps in efficiency for third to sixth grade, a moderate change in math and minor change in reading for seventh grade, and a substantial change in math and moderate change in reading for eighth grade. Therefore, there is no strong evidence that NCLB has had a significant impact on efficiency in math and reading in South Carolina.

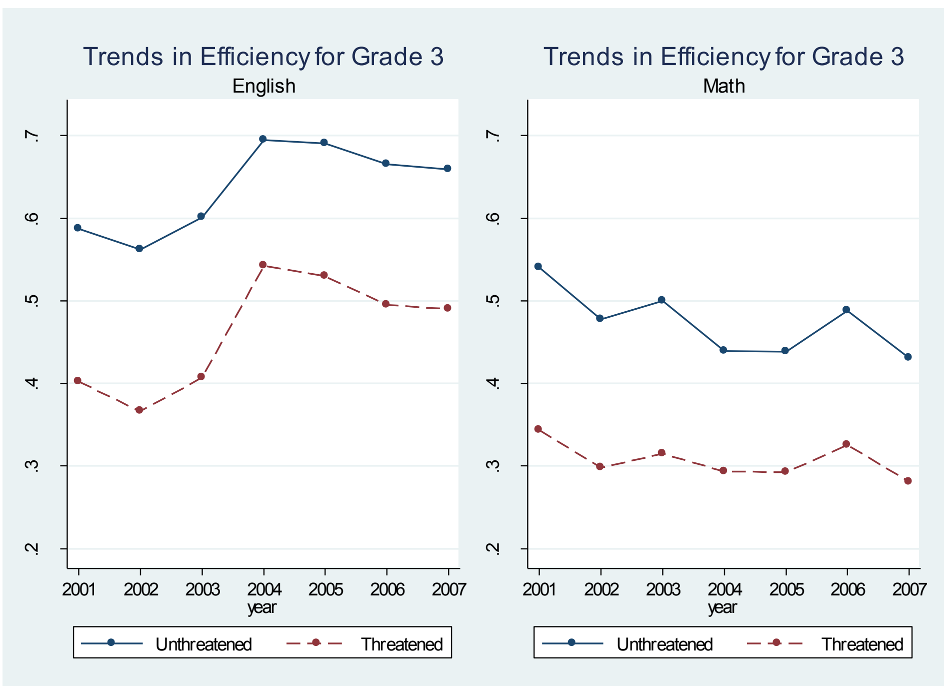


Figure 14, Trends in Efficiency in High-Stakes Subjects for Grade 3, South Carolina

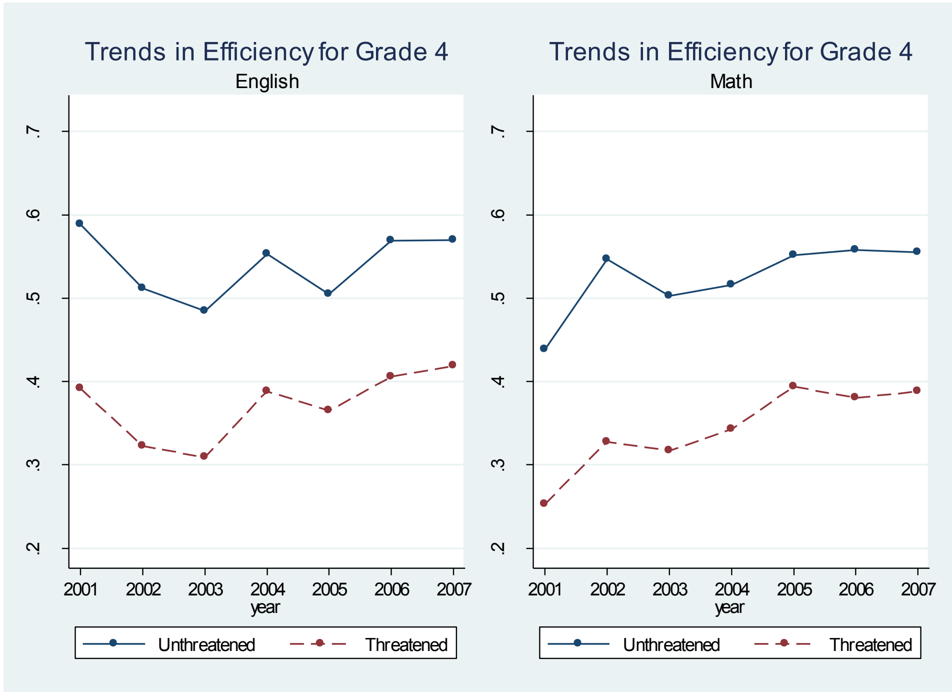


Figure 15, Trends in Efficiency in High-Stakes Subjects for Grade 4, South Carolina

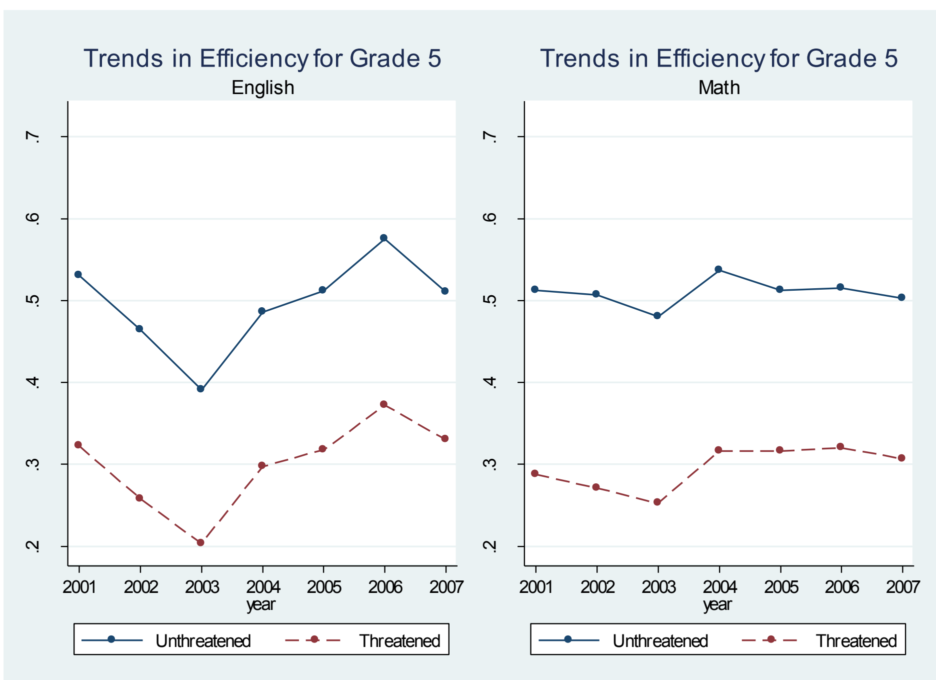


Figure 16, Trends in Efficiency in High-Stakes Subjects for Grade 5, South Carolina

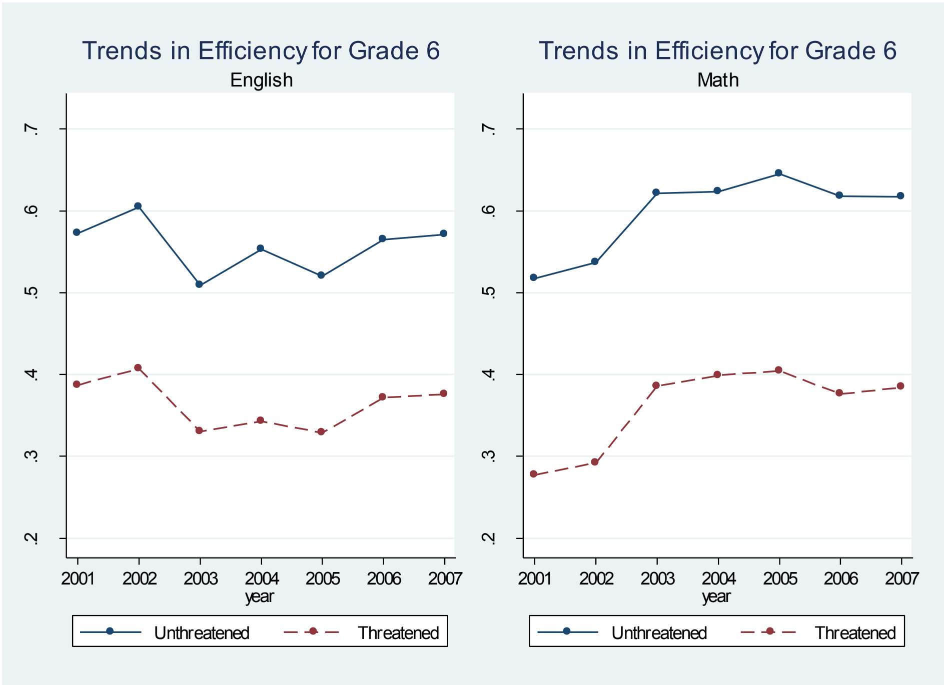


Figure 17, Trends in Efficiency in High-Stakes Subjects for Grade 6, South Carolina

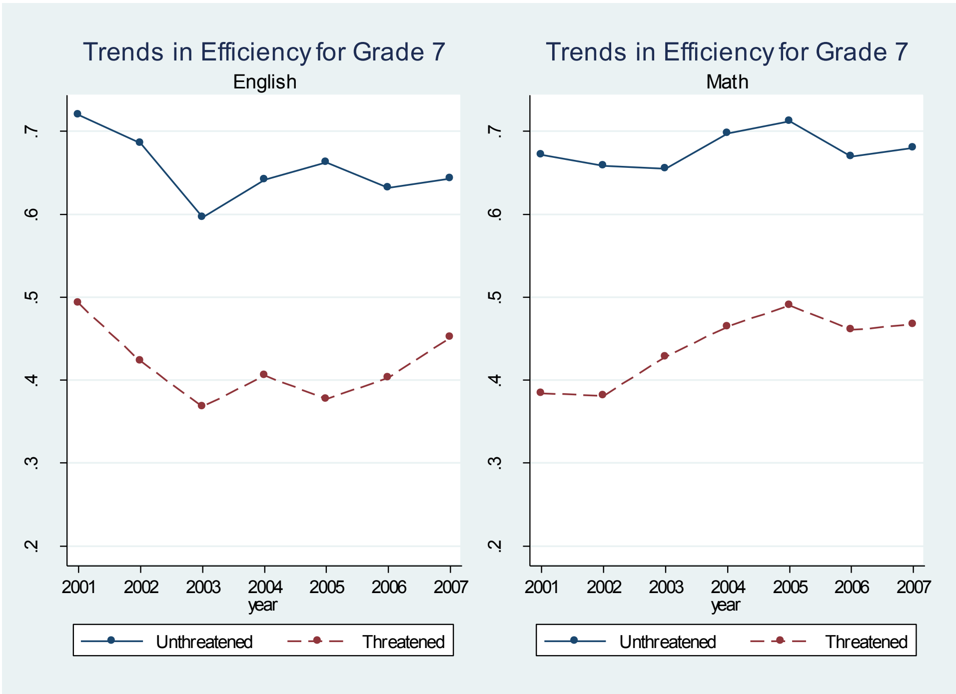


Figure 18, Trends in Efficiency in High-Stakes Subjects for Grade 7, South Carolina

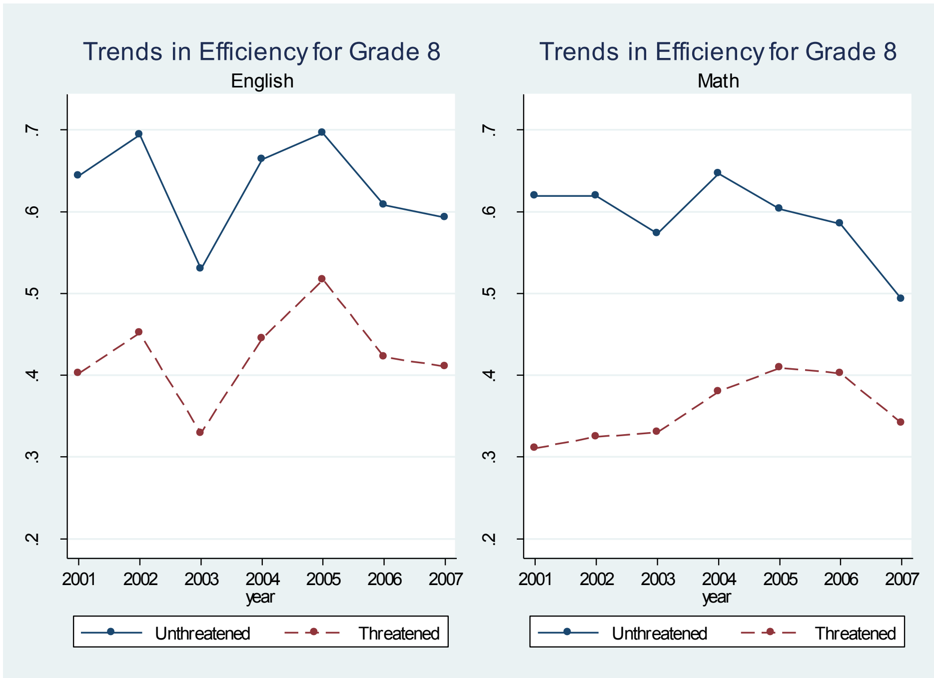


Figure 19, Trends in Efficiency in High-Stakes Subjects for Grade 8, South Carolina

In addition to the two high-stakes subjects, South Carolina has also administered tests in two low-stakes subjects, science and social studies, to all students in third through eighth grade since 2003. The next discussion focuses on the changes of the differences in efficiency in these two low-stakes subjects between unthreatened and threatened schools over time. Generally, the patterns of the changes in efficiency in science and social studies are rather similar to those in math and reading across all grades. Thus, only seventh and eighth grade's results are illustrated in Figures 20 – 21, while others are included in the appendix.

The disparity in efficiency for seventh grade did not change much until 2007 when it closed to .15 in science and .19 in social studies, respectively. Compared to seventh grade, eighth grade has significantly closed the gaps in two subjects. In 2001 unthreatened schools were more efficient than threatened ones in science and social studies by .22 and .24, respectively, which decreased considerably to .10 after six years. The results imply that NCLB has had a positive impact on efficiency in science and social studies for seventh and eighth grade and that schools in South Carolina may not completely ignore low-stakes subjects, especially at the higher grade levels. This implication is tested using a model whose outputs are the percentage of proficient and above in all four tested subjects. The results of the new model will be reported in the next section.

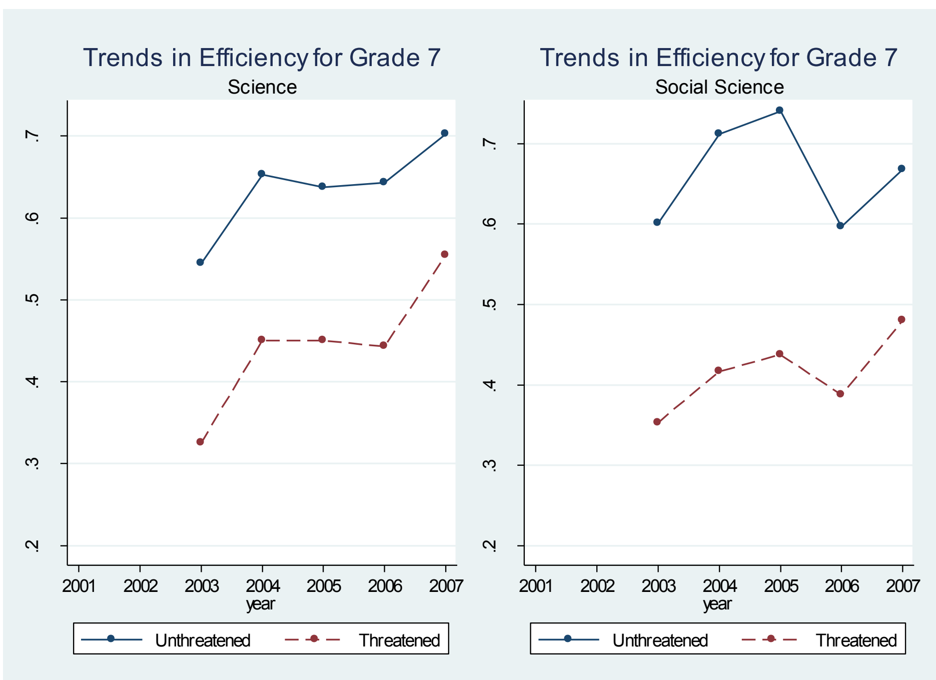


Figure 20, Trends in Efficiency in Low-Stakes Subjects for Grade 7, South Carolina

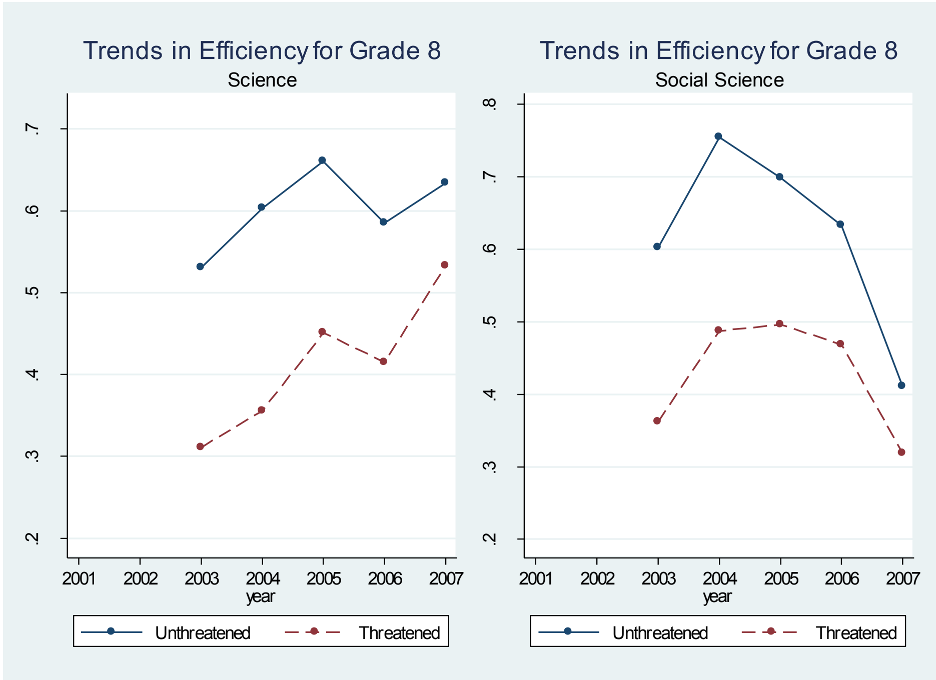


Figure 21, Trends in Efficiency in Low-Stakes Subjects for Grade 8, South Carolina

Results of the second-stage analysis which uses efficiency as the dependent variable are presented in Table 8. The upper panel is results for elementary schools and the lower results for middle schools. Most of the coefficients on year dummies in both math and reading are not statistically significant. The only significant ones are in 2006, .02 for math and .04 for reading, respectively.

The coefficients on the interactions for two high-stakes subjects at the elementary level are not statistically significant until 2004. Unthreatened schools improved efficiency in both subjects by .02 in 2004 relative to 2001 (the baseline) and stayed relatively stable afterward, with the exception of reading in 2006. These results mean that threatened elementary schools became slightly more efficient in high-stakes subjects than unthreatened counterparts after 2004.

Table 8, Coefficients from 2nd-stage Models, Model 1 to Model 4, South Carolina
Using *Efficiency* as Dependent Variable

	TE in Math		TE in Reading		TE in Science		TE in Social Studies	
Elementary Schools	Coef.	Std. E.	Coef.	Std. E.	Coef.	Std. E.	Coef.	Std. E.
Intercept	0.353	*(0.004)*	0.529	*(0.003)*	0.248	*(0.004)*	0.269	*(0.004)*
Year 2002	0.011	*(0.007)*	-0.056	*(0.007)*				
Year 2003	-0.002	*(0.007)*	-0.073	*(0.007)*				
Year 2004	-0.002	*(0.008)*	0.012	*(0.007)*	0.026	*(0.007)*	0.06	*(0.008)*
Year 2005	0.002	*(0.008)*	0.003	*(0.007)*	0.046	*(0.007)*	0.112	*(0.008)*
Year 2006	0.022	*(0.008)*	0.035	*(0.007)*	0.044	*(0.008)*	0.13	*(0.008)*
Year 2007	-0.003	*(0.008)*	0.012	*(0.007)*	0.093	*(0.008)*	0.158	*(0.008)*
Grade 4	0.043	*(0.003)*	-0.094	*(0.003)*	0.08	*(0.003)*	-0.013	*(0.004)*
Grade 5	0.004	*(0.003)*	-0.158	*(0.003)*	0.057	*(0.004)*	-0.044	*(0.004)*
Threatened*Year 2002	-0.005	*(0.009)*	-0.0003	*(0.009)*				
Threatened*Year 2003	0.004	*(0.009)*	0.005	*(0.009)*				
Threatened*Year 2004	**0.022**	***(0.009)***	**0.018**	***(0.009)***	-0.002	*(0.009)*	-0.0002	*(0.010)*
Threatened*Year 2005	**0.033**	***(0.009)***	**0.02**	***(0.009)***	0.006	*(0.009)*	-0.013	*(0.010)*
Threatened*Year 2006	**0.022**	***(0.009)***	0.009	*(0.009)*	0.003	*(0.009)*	**-0.024**	***(0.010)***
Threatened*Year 2007	**0.028**	***(0.009)***	**0.018**	***(0.009)***	0.005	*(0.009)*	-0.012	*(0.010)*
Middle Schools	Coef.	Std. E.	Coef.	Std. E.	Coef.	Std. E.	Coef.	Std. E.
Intercept	0.316	*(0.005)*	0.396	*(0.006)*	0.281	*(0.006)*	0.256	*(0.007)*
Year 2002	0.009	*(0.022)*	0.018	*(0.022)*				
Year 2003	0.064	*(0.022)*	-0.069	*(0.023)*				
Year 2004	0.071	*(0.022)*	-0.024	*(0.023)*	0.097	*(0.025)*	0.133	*(0.028)*
Year 2005	0.071	*(0.022)*	-0.027	*(0.023)*	0.109	*(0.025)*	0.131	*(0.028)*
Year 2006	0.042	*(0.022)*	-0.038	*(0.023)*	0.053	*(0.025)*	0.088	*(0.028)*
Year 2007	0.019	*(0.022)*	-0.038	*(0.023)*	0.135	*(0.025)*	0.07	*(0.028)*
Grade 7	0.079	*(0.005)*	0.071	*(0.005)*	0.074	*(0.006)*	0.093	*(0.007)*
Grade 8	-0.001	*(0.005)*	0.078	*(0.005)*	0.045	*(0.006)*	0.093	*(0.007)*
Threatened*Year 2002	0.002	*(0.023)*	-0.017	*(0.024)*				
Threatened*Year 2003	-0.001	*(0.023)*	-0.014	*(0.024)*				
Threatened*Year 2004	0.028	*(0.023)*	-0.005	*(0.024)*	-0.004	*(0.026)*	-0.029	*(0.029)*
Threatened*Year 2005	**0.047**	*(0.023)*	0.007	*(0.024)*	0.02	*(0.026)*	-0.005	*(0.029)*
Threatened*Year 2006	**0.056**	*(0.023)*	0.014	*(0.024)*	0.036	*(0.026)*	0.012	*(0.029)*
Threatened*Year 2007	**0.064**	*(0.023)*	0.023	*(0.024)*	**0.067**	***(0.026)***	0.031	*(0.029)*

The results of science and social studies are reported in the latter four columns in Table 8. The year dummies from two models are positive and statistically significant. Threatened middle schools improved efficiency by .026 and .06 for science and social studies, respectively, in 2004 compared 2003 (the baseline). The numbers increased afterward, amounting to .093 for science and .158 for social studies in 2007. However, unthreatened schools did not differ greatly from threatened schools in efficiency, except for 2006. These results suggest that NCLB has no significant effect on school efficiency in science and social studies at elementary schools.

Middle schools in South Carolina have moderately different results compared to elementary schools. The coefficients on year dummies are negative but not statistically significant in reading, with the exception of 2003. The corresponding numbers for math are positive and statistically significant for 2003 to 2006. As for the coefficients on the interactions, they are not statistically significant for reading, but significant for math from 2005 to 2007, which are .047, .056, and .064, respectively. These results suggest that threatened middle schools are more efficient in math than unthreatened peers by 4.7%, 5.6%, and 6.4% respectively in 2005, 2006, and 2007, relative to 2001. Overall, NCLB has had a positive and significant impact on efficiency in math after two years of implementation and no effects in reading for middle schools in South Carolina.

With regard to two low-stakes subjects at the middle level, science and social studies, the coefficients on year dummies are statistically significant, implying that there are significant non-NCLB effects in low-stakes subjects. The coefficients on the interactions are not statistically significant, except for science in 2007, meaning that NCLB has had no effect on efficiency in two low-stakes subjects.

In sum, NCLB has had some effect on efficiency in math and reading at the elementary school level and in math at the middle school level after a few years of implementation, but no effect on efficiency in science and social studies in South Carolina. Whether the improvement in math and reading is achieved at the expense of low-stakes subjects is tested in the next section.

Has NCLB Influenced Tradeoffs among Subjects?

To confirm the implication that schools do not ignore low-stakes subjects, we have analyzed a model in which the percentages of students who are proficient and higher in all four tested subjects are considered as four outputs (will be called Model 5 in the following discussions). The null hypothesis is that NCLB has led schools to improve student achievement in high-stakes subjects at the expense of low-stakes subjects. If it is correct, it is expected that Model 1 and Model 2 show a narrowing gap in efficiency in high-stakes subjects between unthreatened and threatened schools, while Model 5 does not.

Results of Model 5, as well as those of Models 1 and 2, are depicted in Figures 22 – 27. Within each figure, the first graph illustrates the result of Model 5. Results from three models for third and fourth grade are very similar; none of the gaps in efficiency between unthreatened and threatened schools declined. On the other hand, the gaps for the other grades (fifth to eighth) narrowed gradually over time. The findings are the opposite of what the hypothesis expects;

schools actually have improved achievement in science and social studies at the higher grade level. Therefore, there is no evidence that NCLB has led schools to improve achievement of high-stakes subjects to the detriment of low-stakes subjects in South Carolina. Rather, schools tend to equally focus on all tested subjects, especially in seventh and eighth grade.

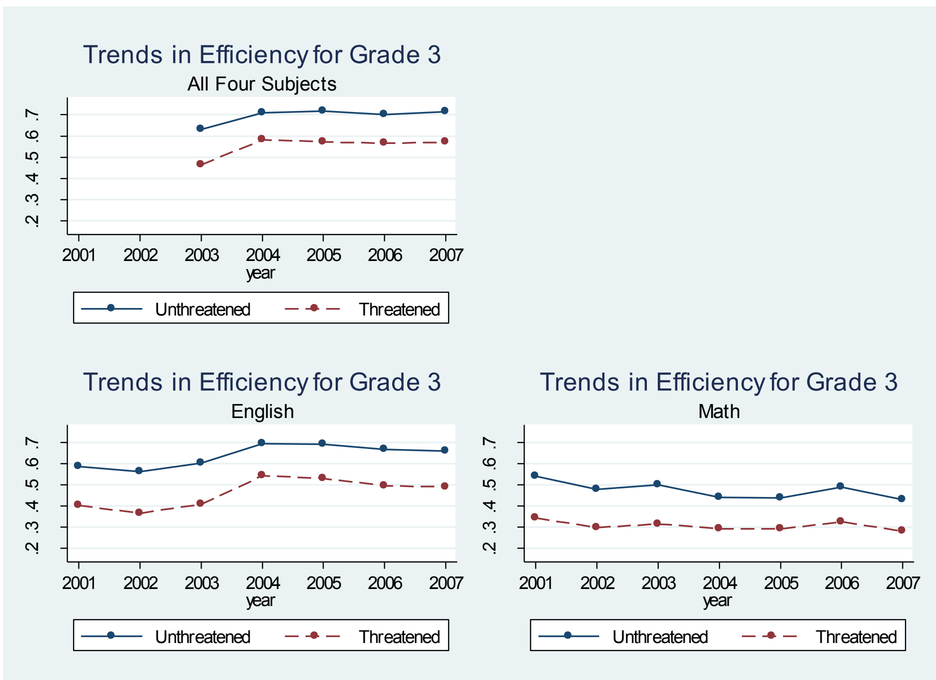


Figure 22, Trends in Efficiency from Model 1, 2, and 5 for Grade 3, South Carolina

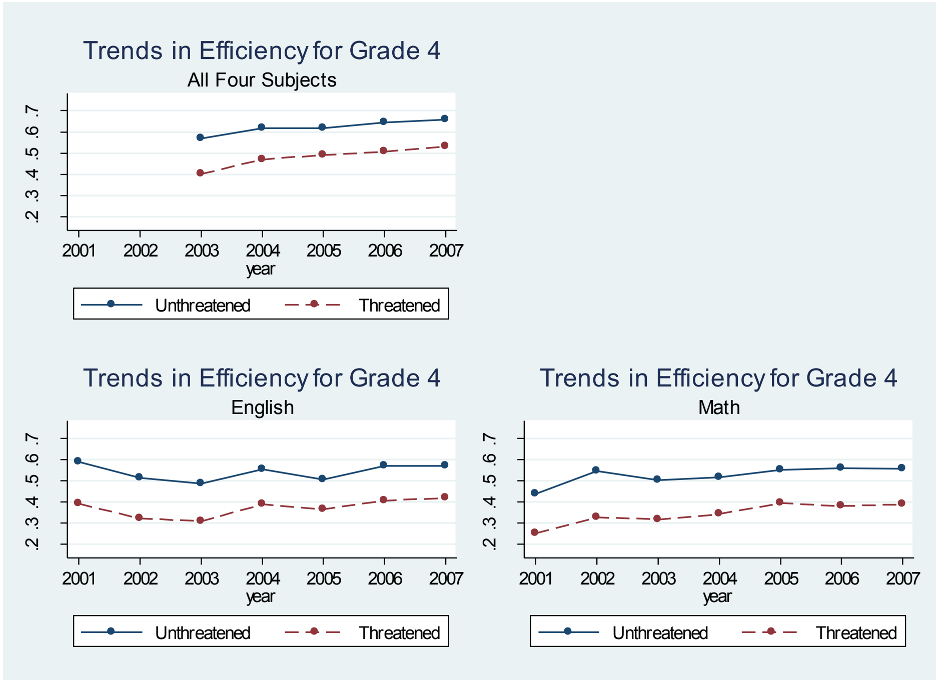


Figure 23, Trends in Efficiency from Model 1, 2, and 5 for Grade 4, South Carolina

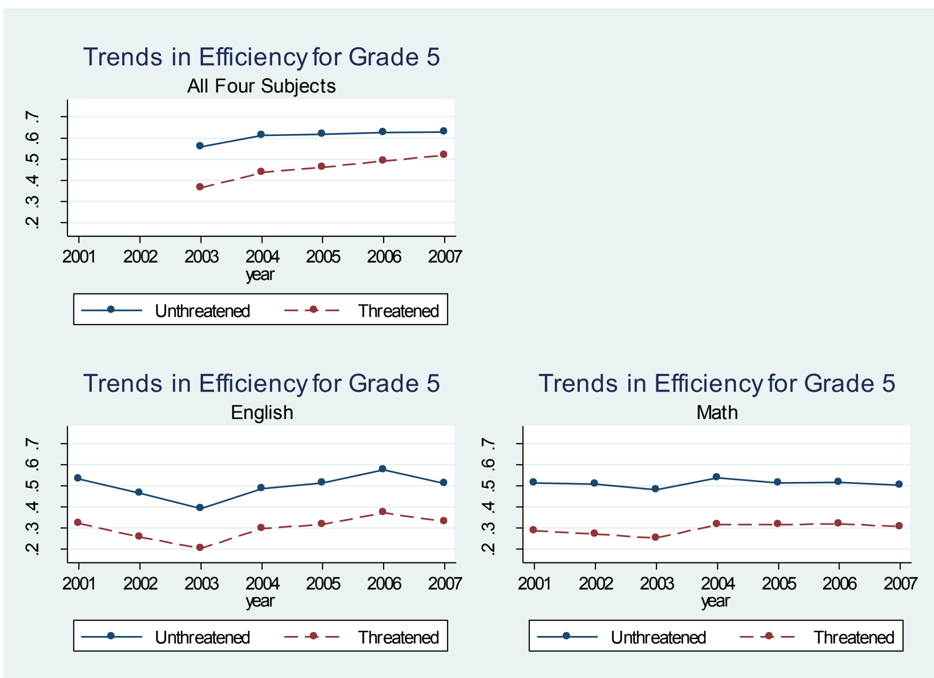


Figure 24, Trends in Efficiency from Model 1, 2, and 5 for Grade 5, South Carolina

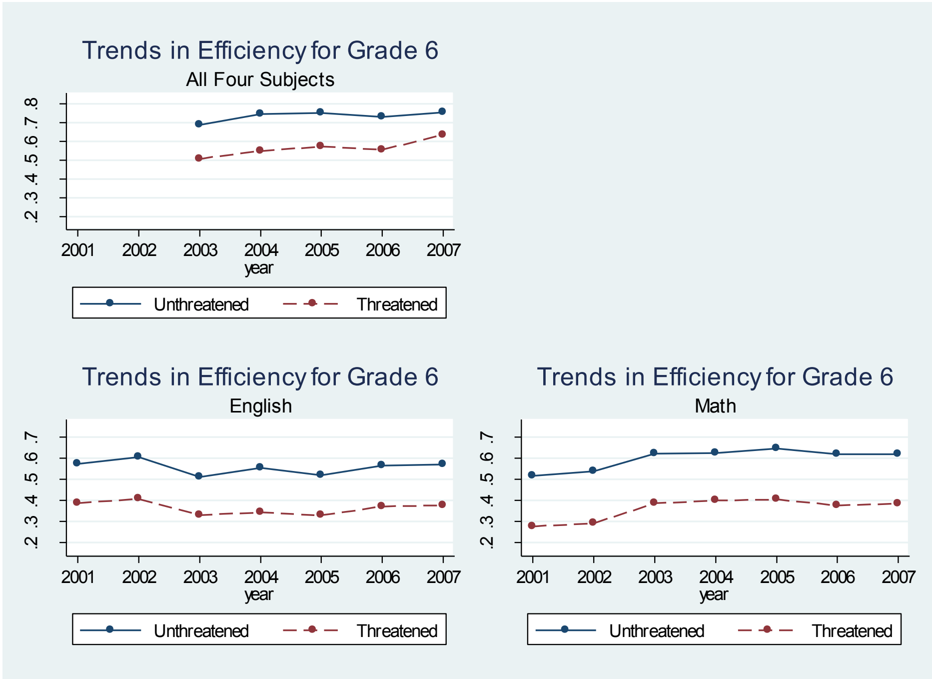


Figure 25, Trends in Efficiency from Model 1, 2, and 5 for Grade 6, South Carolina

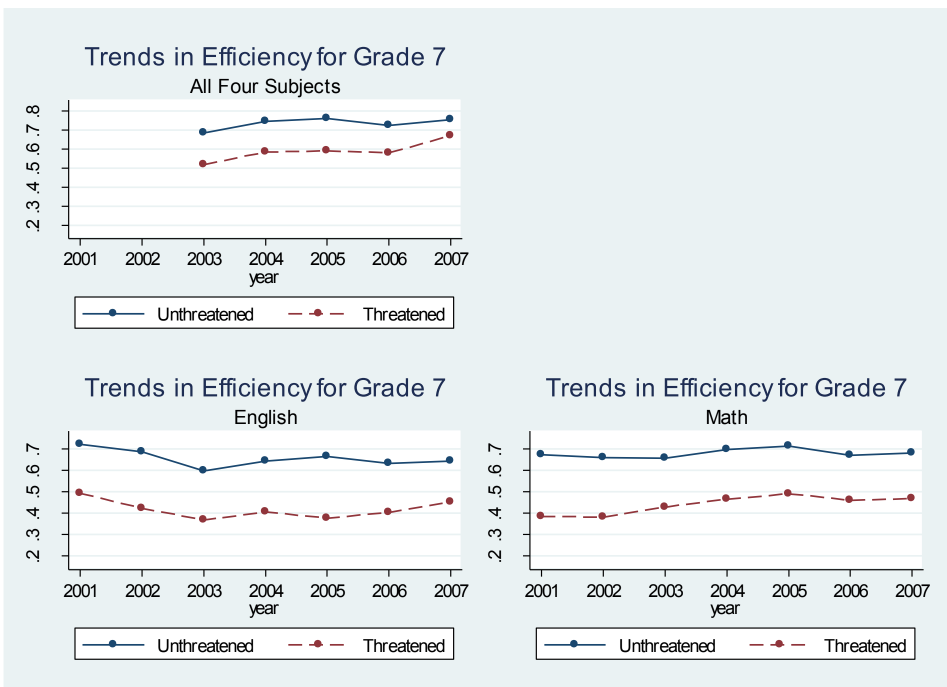


Figure 26, Trends in Efficiency from Model 1, 2, and 5 for Grade 7, South Carolina

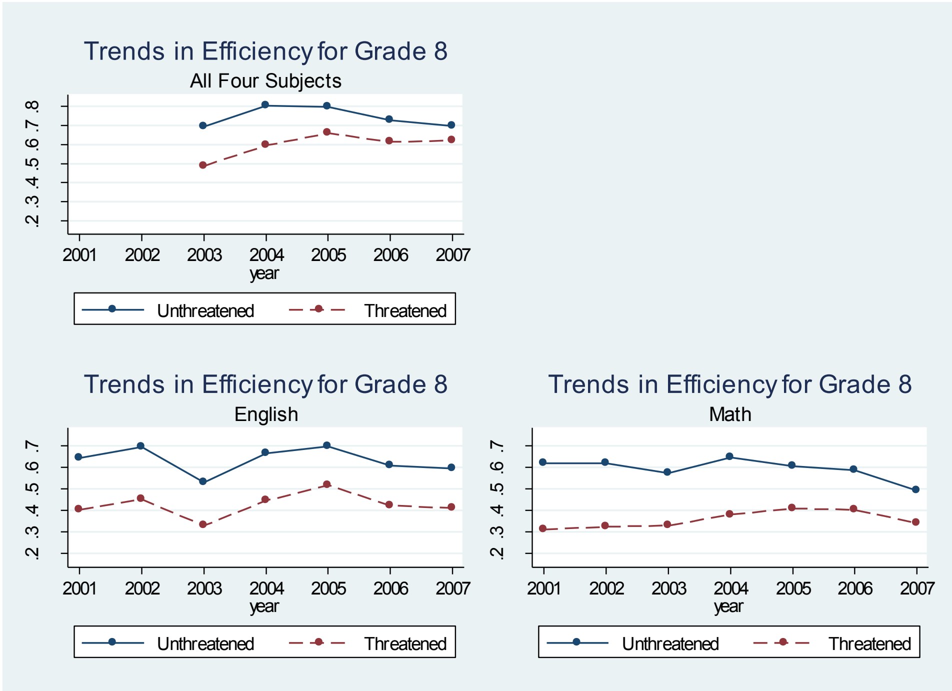


Figure 27, Trends in Efficiency from Model 1, 2, and 5 for Grade 8, South Carolina

Table 9 reports results of the second stage model whose dependent variable is measures of efficiency from Model 5. The first two columns present results of elementary schools and the last two results of middle schools. The coefficients on year dummies are positive and statistically significant, except for 2006 and 2007 at middle school level, suggesting that there are substantial non-NCLB effects at both the elementary school level and the middle school level. The coefficients on the interactions for both elementary and middle schools are positive and statistically significant, with the exception of 2004 at the middle school level, and steadily increase over time. Threatened elementary schools are more efficient than unthreatened peers in producing student achievement in four tested subjects by 2% in 2004 and 4% in 2007, relative to 2003 (the baseline).

Similarly, threatened middle schools are more efficient than unthreatened counterparts by 11.7% in 2007, relative to 2003. The differences between threatened and unthreatened schools with respect to changes in efficiency are larger at the middle school level than those at the elementary level. To summarize, NCLB has had a positive and significant impact on efficiency in producing math, reading, science, and social studies, which suggests that the improvement in high-stakes subjects is accomplished without neglecting low-stakes subjects.

Table 9, Coefficients from 2nd-stage Model, Model 5, South Carolina
Using *Efficiency* as Dependent Variable

Elementary Schools			Middle Schools		
	Coef.	Std. Err.		Coef.	Std. Err.
Intercept	0.530	*(0.004)*	Intercept	0.500	*(0.006)*
Year 2004	0.061	*(0.007)*	Year 2004	0.052	*(0.023)*
Year 2005	0.066	*(0.007)*	Year 2005	0.051	*(0.023)*
Year 2006	0.071	*(0.007)*	Year 2006	0.011	*(0.023)*
Year 2007	0.082	*(0.007)*	Year 2007	0.023	*(0.023)*
Grade 4	-0.070	*(0.003)*	Grade 7	0.029	*(0.006)*
Grade 5	-0.092	*(0.003)*	Grade 8	0.039	*(0.006)*
Threatened*Year 2004	**0.020**	***(0.009)***	Threatened*Year 2004	0.019	*(0.024)*
Threatened*Year 2005	**0.027**	***(0.009)***	Threatened*Year 2005	**0.052**	***(0.024)***
Threatened*Year 2006	**0.035**	***(0.009)***	Threatened*Year 2006	**0.071**	***(0.024)***
Threatened*Year 2007	**0.040**	***(0.009)***	Threatened*Year 2007	**0.117**	***(0.024)***

Has NCLB Influenced Changes in Slack over Time?

As noted, there are two measures of efficiency in this dissertation: efficiency and slack. Unlike efficiency, where larger values indicate higher levels of efficiency, a larger value in slack actually implies a lower level of efficiency. When a school has a positive slack, even if its efficiency equals one, it still does not operate at the truly efficient level. This dissertation focuses on slack in ADV that measures the percentage of advanced students a school should have achieved if it reached a truly efficient level. The investigation of the impact of NCLB on slack will shed some light on the issue that schools may focus on students at the margin of becoming proficient while ignoring those at the high and low ends. An increase of slack indicates there are too few advanced students, suggesting that higher-achieving students are ignored under the policy. The discussions of slack in this section are presented separately.

Minnesota

The trends in slack for third, fifth, and seventh grade are depicted in Figure 28 – 30. There was no substantial difference in slack between unthreatened and threatened schools; two groups only differed by no more than 6% in slack. Nonetheless, each grade had distinctive patterns of changes in slack over time. Grade 3 narrowed the gaps in slack for reading from 2001 to 2006 and bounced back slightly in 2007, but continued to reduce those for math during the whole time period. With regard to fifth grade, it closed the gaps in slack for reading from 2001 to 2005, but widened again in 2006 and 2007. On the other hand, the gaps in math narrowed so substantially that there was no difference in slack for math between unthreatened and threatened schools in 2006 and 2007. Seventh grade had a similar pattern in reading as fifth grade. However, unthreatened schools had more slack in math than in threatened schools in 2006 and 2007.

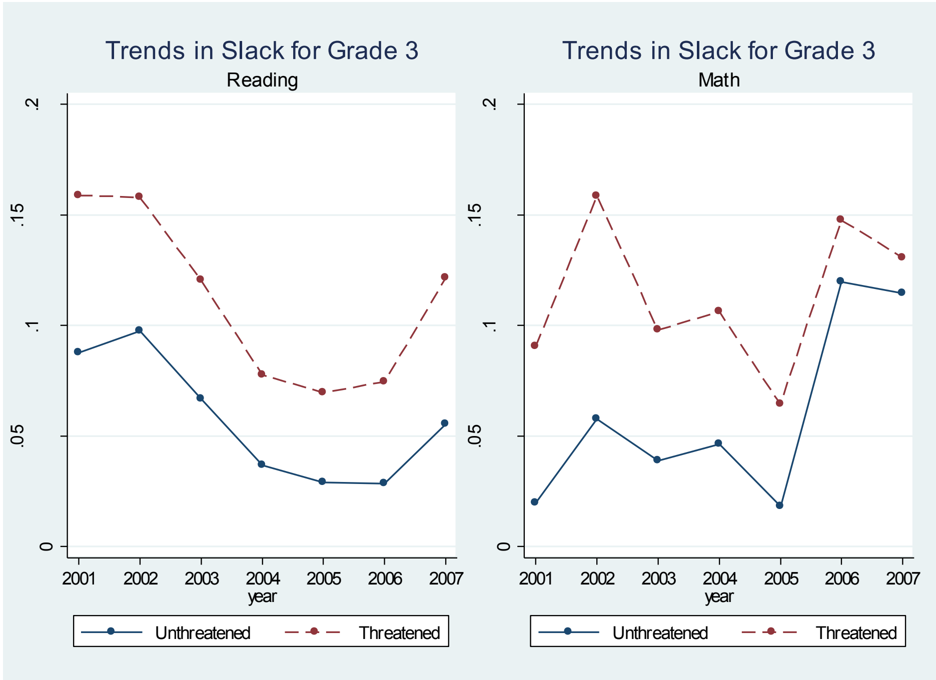


Figure 28, Trends in Slack, Grade 3, Minnesota

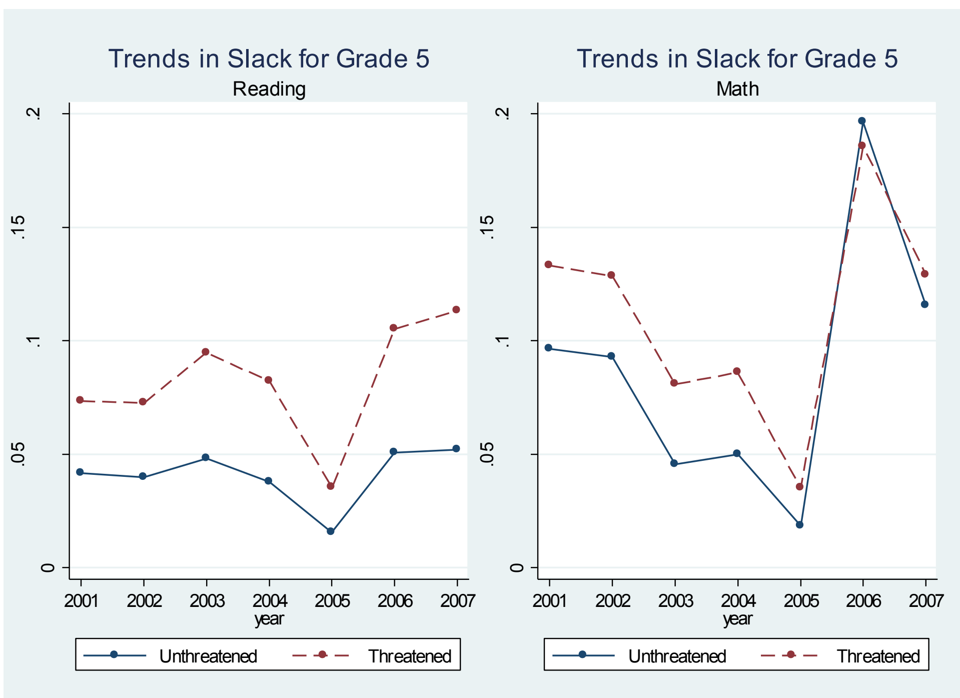


Figure 29, Trends in Slack, Grade 5, Minnesota

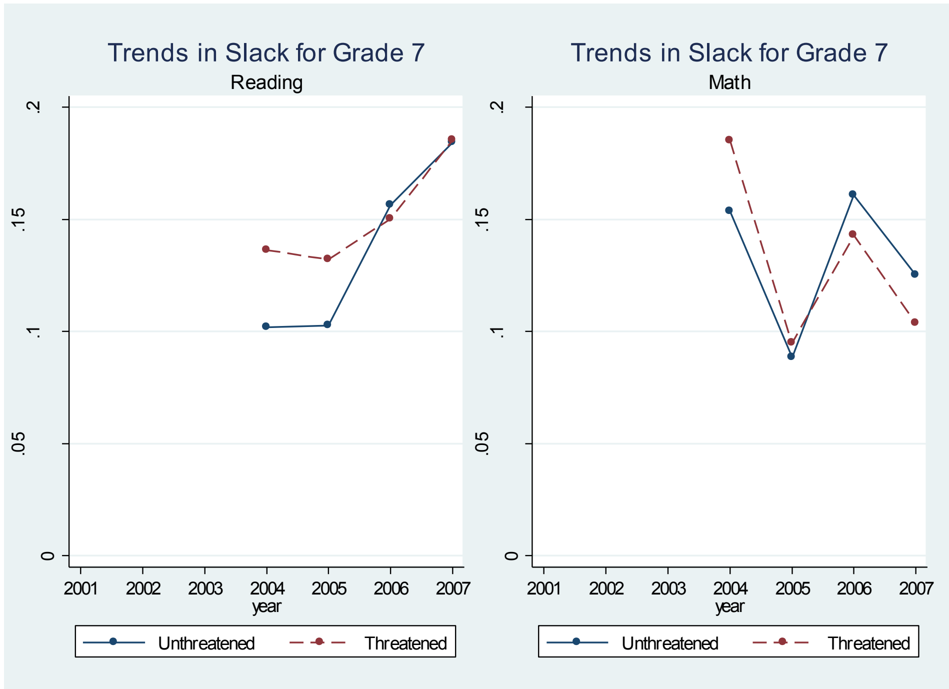


Figure 30, Trends in Slack, Grade 7, Minnesota

Whether or not the changes in slack are caused by NCLB is investigated at the second-stage, whose results are presented in Table 10. The upper panel reports results for the elementary level and the lower panel results for the secondary level. The coefficients on year dummies in math at the elementary level are statistically significant, with the exception of 2003 and 2004, while those in reading are statistically significant only in 2004 and 2005. However, the coefficients on the interactions are negative and statistically significant from 2005 to 2007 in math and 2004 to 2006 in reading, meaning that NCLB has reduced slack in math and reading after a few years of implementation. At the middle school level, the coefficients on year dummies are statistically significant for both math and reading, except for reading in 2005. Unlike at the elementary level, NCLB has no impact on slack at the middle level.

Similar to the models whose dependent variable is efficiency, it is necessary to control for a confounding factor in the models whose dependent variable is slack. This confounding factor is the fact that more grades were tested from 2006. The results of the extended models that include a new variable – *pct_highstakes*, as well as its interaction with the threatened school dummy, are reported in the third and fourth column for math and the seventh and eighth column for reading in Table 10. The coefficients on two new variables for both elementary and middle school levels are not statistically significant and the inclusion of them in the models does not substantially change the coefficients on the interaction terms. This means that testing more grades does not greatly influence the changes of slack.

To sum up, NCLB has significantly reduced slack in math and reading after a few years of implementation at the elementary level, but not at the middle level. This to some extent implies that NCLB has not propelled schools to improve achievement for students in the middle group (proficient) at the expense of those at the high end (advanced).

Table 10, Coefficients from 2nd-stage Models, Model 1 and Model 2, Minnesota
Using *Slack in ADV* as Dependent Variable

	Slack in Math				Slack in Reading			
Elementary Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.095	*(0.003)*	0.089	*(0.005)*	0.118	*(0.002)*	0.102	*(0.004)*
Year 2002	0.017	*(0.007)*	0.017	*(0.007)*	0.003	*(0.006)*	0.004	*(0.006)*
Year 2003	-0.016	*(0.007)*	-0.016	*(0.007)*	-0.007	*(0.006)*	-0.007	*(0.006)*
Year 2004	-0.011	*(0.007)*	-0.011	*(0.007)*	-0.028	*(0.006)*	-0.028	*(0.006)*
Year 2005	-0.040	*(0.007)*	-0.040	*(0.007)*	-0.042	*(0.006)*	-0.042	*(0.006)*
Year 2006	0.097	*(0.007)*	0.092	*(0.009)*	-0.015	*(0.006)*	-0.017	*(0.008)*
Year 2007	0.058	*(0.007)*	0.054	*(0.009)*	0.004	*(0.006)*	0.002	*(0.008)*
Grade 4	0.017	*(0.003)*	0.018	*(0.003)*	-0.010	*(0.003)*	-0.009	*(0.003)*
Grade 5	0.002	*(0.002)*	0.002	*(0.002)*	-0.029	*(0.002)*	-0.029	*(0.002)*
Threatened*Year 2002	**0.018**	***(0.008)***	**0.018**	***(0.008)***	-0.005	*(0.007)*	-0.004	*(0.007)*
Threatened*Year 2003	-0.003	*(0.008)*	-0.003	*(0.008)*	-0.002	*(0.007)*	-0.002	*(0.007)*
Threatened*Year 2004	-0.002	*(0.008)*	-0.002	*(0.008)*	-0.010	*(0.007)*	-0.011	*(0.007)*
Threatened*Year 2005	**-0.018**	***(0.008)***	**-0.018**	***(0.008)***	**-0.021**	***(0.007)***	**-0.022**	***(0.007)***
Threatened*Year 2006	**-0.036**	***(0.008)***	**-0.038**	***(0.011)***	-0.013	*(0.007)*	**-0.029**	***(0.009)***
Threatened*Year 2007	**-0.036**	***(0.008)***	**-0.037**	***(0.011)***	-0.007	*(0.007)*	**-0.023**	***(0.009)***
Pct_HighStakes			0.016	*(0.025)*			0.006	*(0.022)*
Threatened*Pct_HighStakes			0.006	*(0.030)*			**0.065**	***(0.026)***
Middle Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.099	*(0.007)*	0.103	*(0.010)*	0.009	*(0.006)*	0.028	*(0.009)*
Year 2005	-0.068	*(0.018)*	-0.067	*(0.018)*	-0.004	*(0.017)*	-0.004	*(0.017)*
Year 2006	-0.039	*(0.015)*	-0.008	*(0.042)*	0.043	*(0.014)*	0.043	*(0.040)*
Year 2007	-0.049	*(0.015)*	-0.018	*(0.042)*	0.041	*(0.014)*	0.041	*(0.040)*
Grade 7	0.082	*(0.006)*	0.082	*(0.006)*	0.121	*(0.005)*	0.121	*(0.005)*
Grade 8	0.134	*(0.006)*	0.134	*(0.006)*	0.040	*(0.005)*	0.039	*(0.005)*
Threatened*Year 2005	-0.022	*(0.019)*	-0.022	*(0.019)*	0.001	*(0.018)*	0.002	*(0.018)*
Threatened*Year 2006	-0.015	*(0.015)*	-0.041	*(0.044)*	-0.012	*(0.015)*	0.031	*(0.043)*
Threatened*Year 2007	-0.012	*(0.016)*	-0.037	*(0.044)*	0.001	*(0.015)*	0.044	*(0.043)*
Pct_HighStakes			-0.058	*(0.073)*			-0.001	*(0.070)*
Threatened*Pct_HighStakes			0.049	*(0.079)*			-0.085	*(0.075)*

Indiana

The trends in slack in ADV for math and reading for unthreatened and threatened schools are illustrated in Figures 31 – 33, which report the results of third, sixth, and eighth grade, respectively. The disparities in slack for third grade between unthreatened and threatened schools, for both math and reading, narrowed so substantially that they disappeared four years after NCLB was introduced. In sixth grade, the difference in slack stayed relatively stable from 2002 to 2006 for reading, but decreased slightly from 3.7% in 2002 to 2.9% in 2006 for math. Unlike the other two grades, the gaps for both subjects in eighth grade increased over time. The increase in math was considerably larger than that for reading in eighth grade.

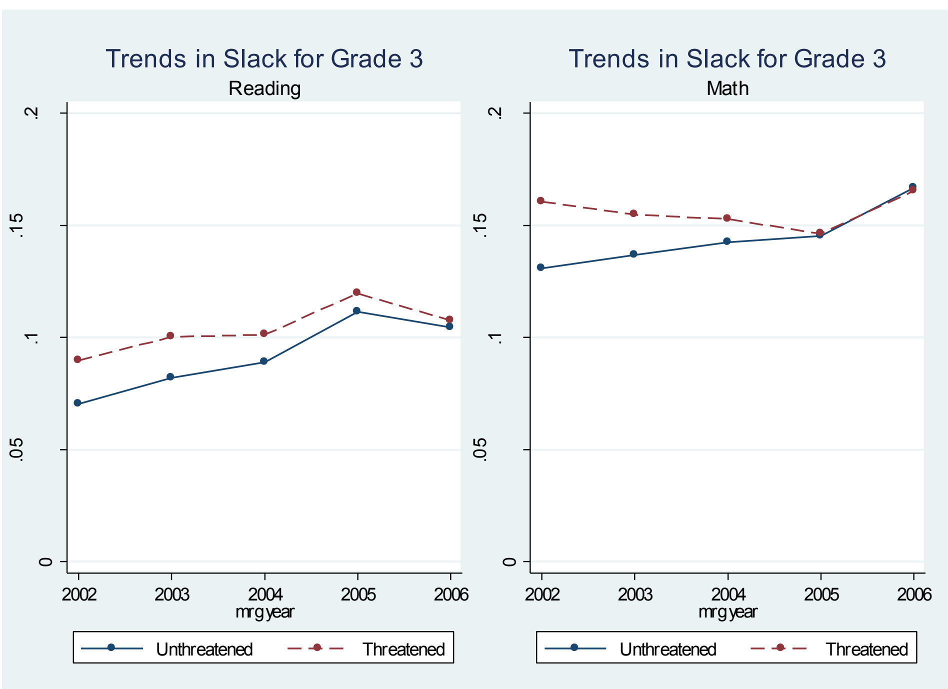


Figure 31, Trends in Slack, Grade 3, Indiana

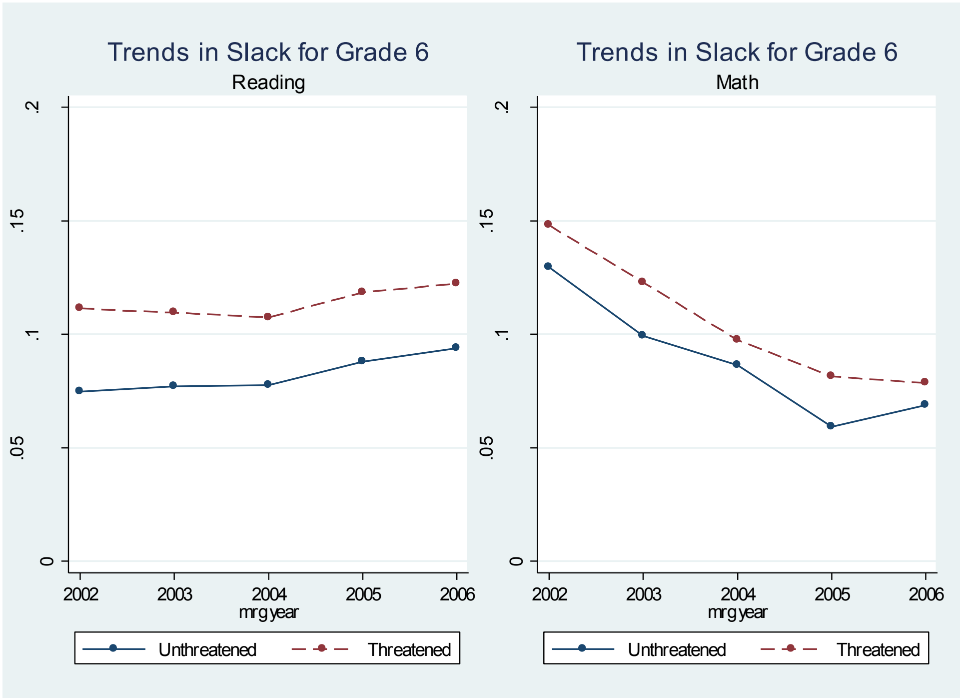


Figure 32, Trends in Slack, Grade 6, Indiana

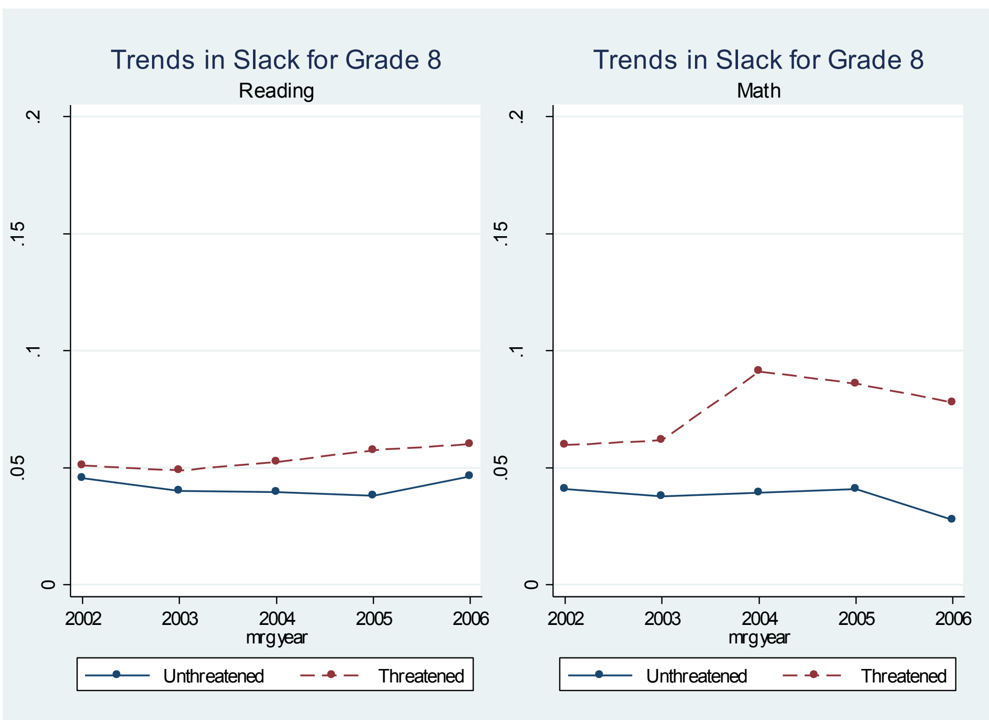


Figure 33, Trends in Slack, Grade 8, Indiana

Next, results of the second stage models are reported in Table 11. The upper panel presents results of elementary schools and the lower one results of middle schools. At the elementary level most coefficients on year dummies, as well as those on the interactions, are not statistically significant, with the exception of year dummies for reading. Thus, there is no strong evidence that NCLB has had significant impact on the change of slack for math and reading at elementary schools. As far as the secondary level is concerned, unthreatened schools have lower slack in math but have no change in slack in reading after 2002. Furthermore, threatened middle schools differ significantly from unthreatened peers regarding the increased slack in math from 2002 to 2006. The findings imply that NCLB has increased slack in middle schools.

Generally, NCLB has boosted slack in math at the middle school level, but not at the elementary school level. Furthermore, the policy has had no influence on slack in reading for both elementary and middle levels.

Table 11, Coefficients of School-Fixed Effect Models, Model 1 and Model 2, Indiana Using *Slack in ADV* as Dependent Variable,

	Slack in Math		Slack in Reading	
Elementary Schools	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.153	*(0.004)*	0.083	*(0.003)*
Year 2003	0.017	*(0.010)*	0.011	*(0.006)*
Year 2004	0.005	*(0.010)*	0.017	*(0.006)*
Year 2005	0.014	*(0.010)*	0.029	*(0.006)*
Year 2006	0.012	*(0.010)*	0.036	*(0.006)*
Grade 4	0.048	*(0.003)*	-0.068	*(0.002)*
Grade 5	-0.059	*(0.003)*	0.086	*(0.002)*
Threatened*Year 2003	-0.010	*(0.011)*	0.000	*(0.007)*
Threatened*Year 2004	-0.013	*(0.011)*	0.001	*(0.007)*
Threatened*Year 2005	-0.019	*(0.011)*	-0.004	*(0.007)*
Threatened*Year 2006	**-0.022**	***(0.011)***	-0.008	*(0.007)*
Middle Schools	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.146	*(0.004)*	0.108	*(0.002)*
Year 2003	-0.046	*(0.010)*	0.003	*(0.006)*
Year 2004	-0.042	*(0.010)*	0.006	*(0.006)*
Year 2005	-0.060	*(0.010)*	0.014	*(0.006)*
Year 2006	-0.056	*(0.010)*	0.013	*(0.006)*
Grade 7	0.110	*(0.004)*	-0.036	*(0.002)*
Grade 8	-0.092	*(0.005)*	-0.072	*(0.003)*
Threatened*Year 2003	**0.026**	***(0.011)***	-0.004	*(0.006)*
Threatened*Year 2004	**0.021**	***(0.011)***	-0.004	*(0.006)*
Threatened*Year 2005	**0.024**	***(0.011)***	-0.008	*(0.006)*
Threatened*Year 2006	**0.025**	***(0.011)***	-0.005	*(0.006)*

South Carolina

The trends in slack in ADV for grade 3 through 8 in South Carolina are depicted in Figures 34 – 39. Third grade slightly increased the gap in slack in reading from 2001 to 2002, moved back to the previous level in the following years, and then increased again in 2006 and 2007. Furthermore, third grade also widened the gap in slack in science but closed that in social studies, though these changes were fairly small. The gap in slack in math increased over time for both fourth and fifth grade. Moreover, both fifth and eighth grade had more slack in social studies in 2007 than in 2001. Nonetheless, all of the changes in slack were very small. Compared to other grades, sixth and seventh grade illustrated little change in the gap in slack for each tested subject. Therefore, the results imply that NCLB has had no effect on the changes of slack in South Carolina.

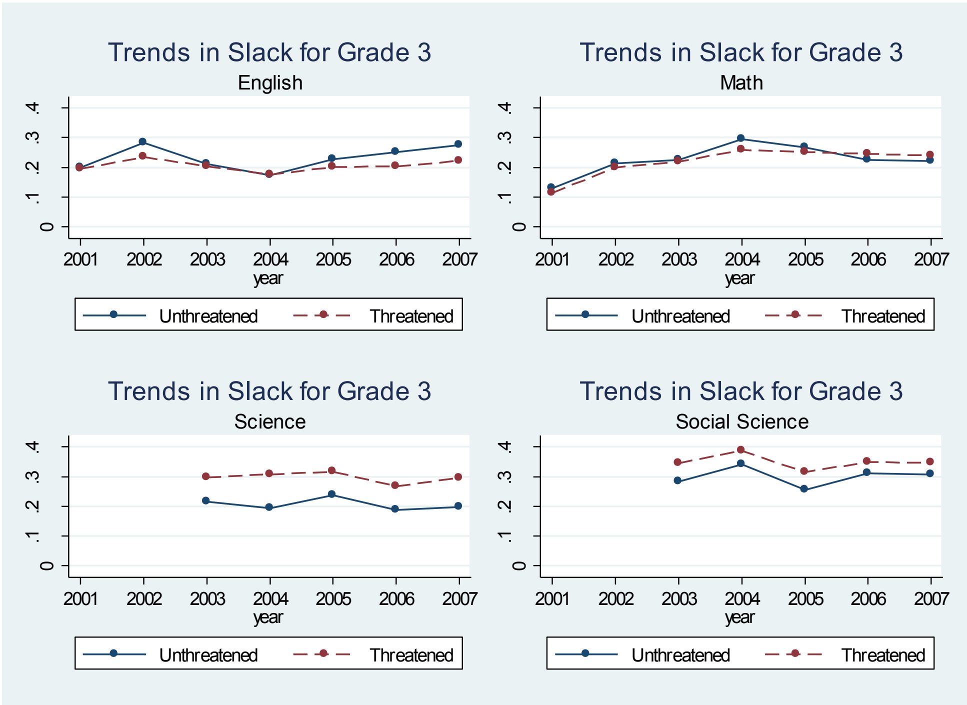


Figure 34, Trends in Slack, Grade 3, South Carolina

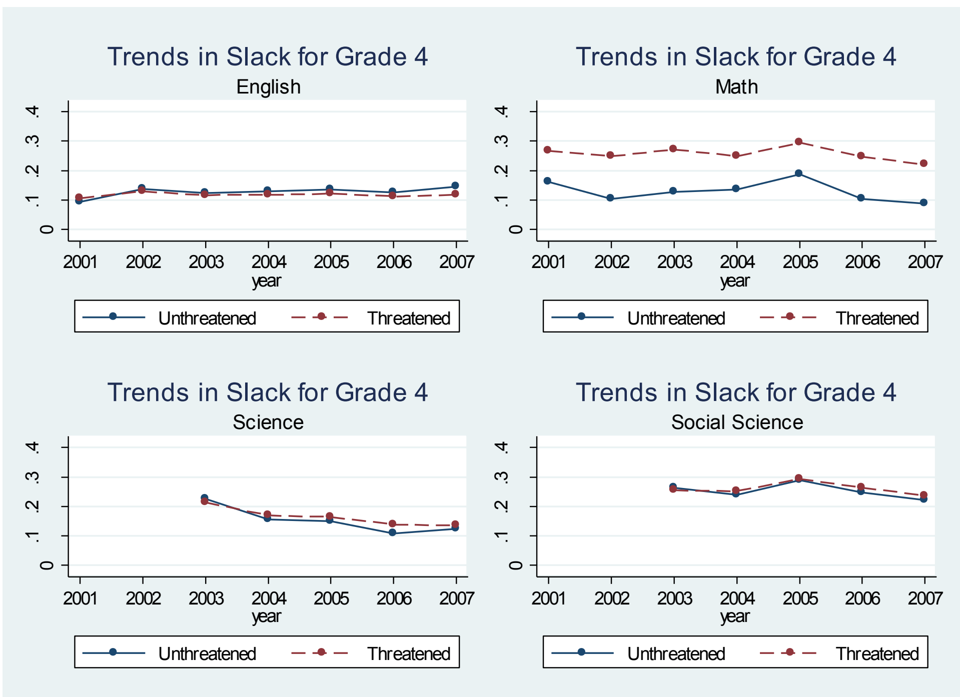


Figure 35, Trends in Slack, Grade 4, South Carolina

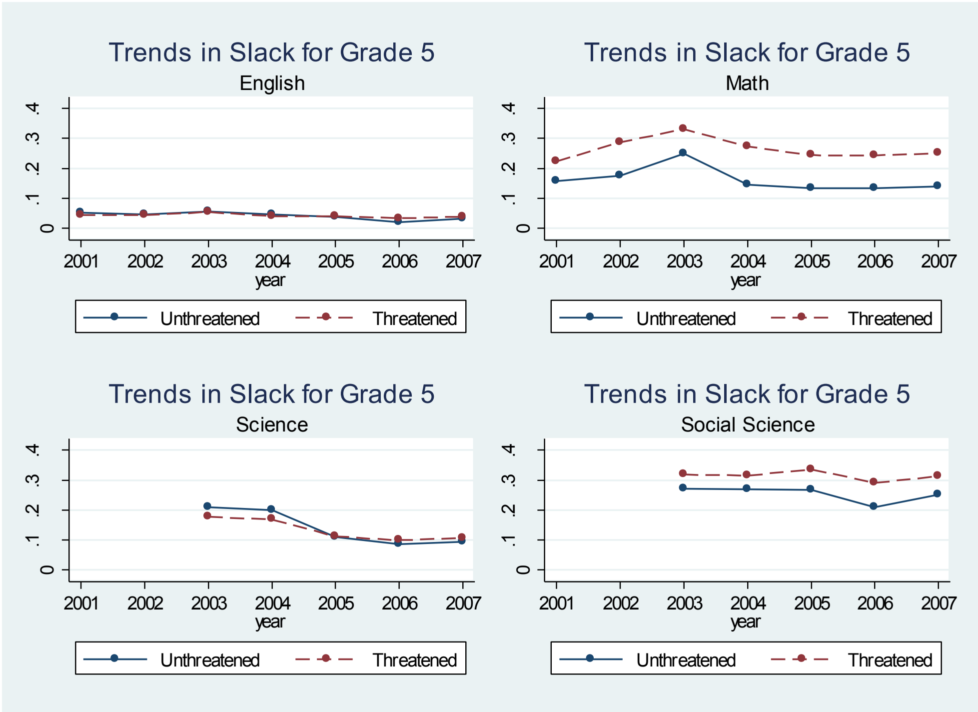


Figure 36, Trends in Slack, Grade 5, South Carolina

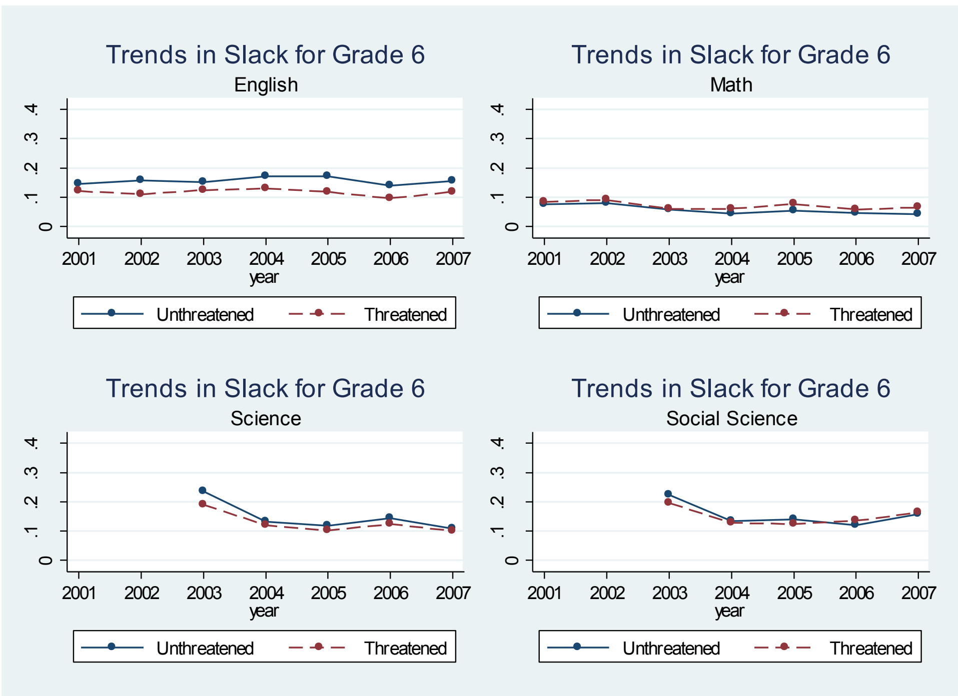


Figure 37, Trends in Slack, Grade 6, South Carolina

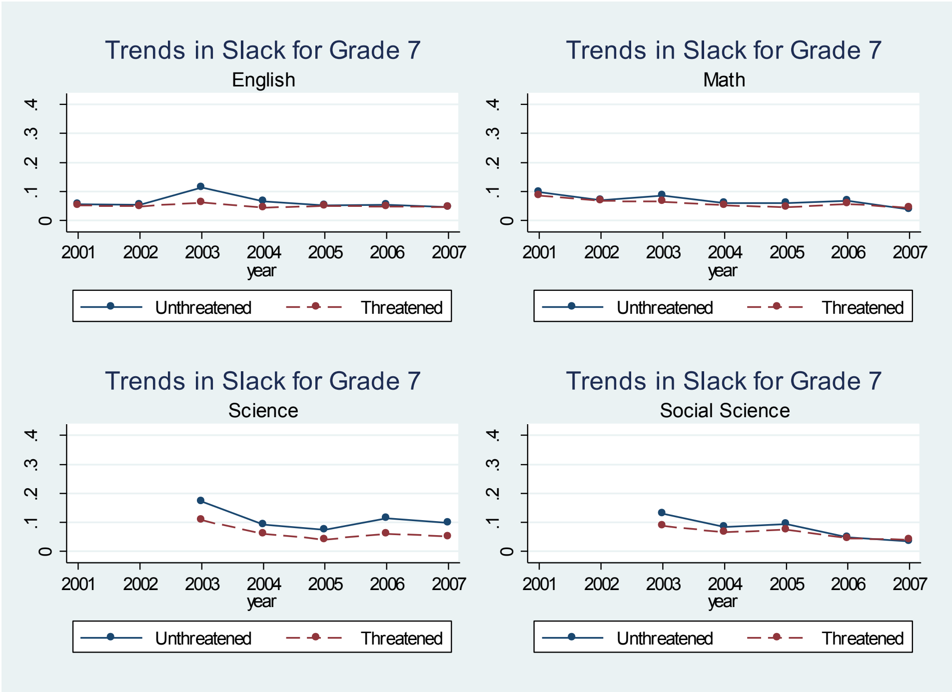


Figure 38, Trends in Slack, Grade 7, South Carolina

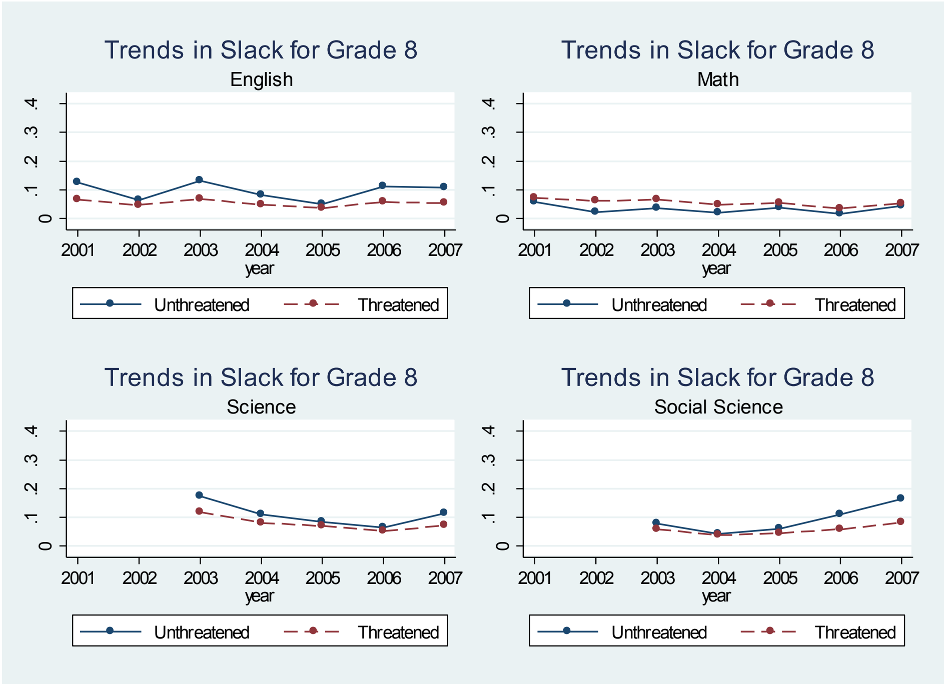


Figure 39, Trends in Slack, Grade 8, South Carolina

Results of second-stage models using slack in ADV as dependent variable are reported in Table 12. At the elementary level, most coefficients on year dummies estimated from four models are statistically significant, but the signs and the effect sizes are different across models. Slack in math increased from 2002 to 2005 but slipped afterward for unthreatened elementary schools. The coefficients on the interactions are statistically significant for math, reading, and science. Unthreatened elementary schools have more slack in math and science than threatened counterparts in the NCLB years, but less slack in reading.

Results of middle schools are presented at the lower panel. Unthreatened middle schools have less slack in ADV in each tested subject over time. However, there is no significant difference in the change of slack between threatened and unthreatened middle schools. Thus, there is no evidence to support the hypothesis that NCLB has had an impact on slack in ADV for each tested subject at the middle school level.

In general, there is slightly more slack in math and in science but less slack in reading at the elementary level after NCLB was introduced, but no great change at the secondary level.

Table 12, Coefficients of School-Fixed Effect Models, Model 1 to Model 4, South Carolina
Using *Slack in ADV* as Dependent Variable

	Slack in Math		Slack in Reading		Slack in Science		Slack in Social Studies	
Elementary Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.185	*(0.004)*	0.205	*(0.002)*	0.300	*(0.004)*	0.331	*(0.004)*
Year 2002	0.018	*(0.009)*	0.043	*(0.005)*				
Year 2003	0.052	*(0.009)*	0.016	*(0.005)*				
Year 2004	0.047	*(0.009)*	0.002	*(0.005)*	-0.034	*(0.008)*	0.012	*(0.008)*
Year 2005	0.050	*(0.009)*	0.021	*(0.005)*	-0.051	*(0.008)*	-0.003	*(0.008)*
Year 2006	0.008	*(0.009)*	0.021	*(0.005)*	-0.089	*(0.008)*	-0.014	*(0.008)*
Year 2007	0.002	*(0.009)*	0.040	*(0.005)*	-0.077	*(0.008)*	-0.011	*(0.008)*
Grade 4	-0.013	*(0.004)*	-0.095	*(0.002)*	-0.104	*(0.004)*	-0.076	*(0.004)*
Grade 5	0.004	*(0.004)*	-0.175	*(0.002)*	-0.129	*(0.004)*	-0.040	*(0.004)*
Threatened*Year 2002	0.025	*(0.011)*	-0.022	*(0.006)*				
Threatened*Year 2003	0.020	*(0.011)*	-0.007	*(0.006)*				
Threatened*Year 2004	0.013	*(0.011)*	-0.006	*(0.006)*	**0.022**	***(0.010)***	0.002	*(0.010)*
Threatened*Year 2005	0.014	*(0.011)*	**-0.016**	***(0.006)***	**0.021**	***(0.010)***	0.012	*(0.010)*
Threatened*Year 2006	**0.038**	***(0.011)***	**-0.021**	***(0.006)***	**0.030**	***(0.010)***	0.010	*(0.010)*
Threatened*Year 2007	**0.036**	***(0.011)***	**-0.029**	***(0.006)***	**0.032**	***(0.010)***	0.007	*(0.010)*
Middle Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.087	*(0.002)*	0.130	*(0.002)*	0.178	*(0.003)*	0.176	*(0.003)*
Year 2002	-0.013	*(0.010)*	-0.012	*(0.009)*				
Year 2003	-0.031	*(0.010)*	0.014	*(0.009)*				
Year 2004	-0.034	*(0.010)*	0.006	*(0.009)*	-0.076	*(0.010)*	-0.058	*(0.012)*
Year 2005	-0.024	*(0.010)*	-0.006	*(0.009)*	-0.092	*(0.010)*	-0.046	*(0.012)*
Year 2006	-0.031	*(0.010)*	-0.003	*(0.009)*	-0.075	*(0.010)*	-0.053	*(0.012)*
Year 2007	-0.033	*(0.010)*	0.001	*(0.009)*	-0.082	*(0.010)*	-0.023	*(0.012)*
Grade 7	-0.006	*(0.002)*	-0.076	*(0.002)*	-0.057	*(0.003)*	-0.086	*(0.003)*
Grade 8	-0.012	*(0.002)*	-0.071	*(0.002)*	-0.044	*(0.003)*	-0.091	*(0.003)*
Threatened*Year 2002	0.006	*(0.010)*	0.003	*(0.009)*				
Threatened*Year 2003	0.014	*(0.010)*	-0.008	*(0.010)*				
Threatened*Year 2004	0.007	*(0.010)*	-0.008	*(0.010)*	**0.024**	*(0.011)*	0.021	*(0.013)*
Threatened*Year 2005	0.001	*(0.010)*	-0.005	*(0.010)*	**0.022**	*(0.011)*	0.011	*(0.013)*
Threatened*Year 2006	-0.001	*(0.010)*	-0.009	*(0.010)*	0.013	*(0.011)*	0.016	*(0.013)*
Threatened*Year 2007	0.005	*(0.010)*	-0.008	*(0.010)*	0.016	*(0.011)*	0.005	*(0.013)*

Sensitivity Analysis to DEA results

DEA estimates are sensitive to outliers, which could push the frontier outward. If these outliers are characterized by high levels of measurement error, efficiency of the schools inside the frontier could be underestimated. Excluding schools with extreme scores from DEA models may result in different efficiency estimates. We explore this possibility with two types of sensitivity analysis of first stage results.

The first type of sensitivity analysis is conducted in the following steps. First, the DEA models are run including all schools. After obtaining efficiency estimates, all efficient schools (efficiency estimates = 1) are excluded from the dataset and the DEA models are run again. The efficiency estimates from the first step and those from the second step are then compared. The correlations between two sets of efficiency are presented in Table 13. The correlations between two sets of the efficiency measures are fairly high across three states, suggesting the DEA results are not unduly influenced by extreme outliers.

Table 13, Correlations between Efficiency Estimates from the First Sensitivity Analysis

	Model 1			**Model 2**		
	Minnesota	Indiana	South Carolina	Minnesota	Indiana	South Carolina
Grade 3	0.97	0.95	0.94	0.92	0.95	0.96
Grade 4	0.95	0.96	0.93	0.86	0.94	0.94
Grade 5	0.98	0.93	0.93	0.98	0.95	0.93
Grade 6	0.95	0.95	0.89	0.96	0.94	0.92
Grade 7	0.98	0.94	0.93	0.94	0.94	0.88
Grade 8	0.96	0.94	0.91	0.96	0.91	0.91

The second type of sensitivity analysis is conducted using Hadi method (Hadi, 1992, 1994). According to Hadi (1992, 1994), the outliers in a dataset can be identified in the following steps. First, the whole dataset is sorted in ascending order and then divided into two parts. The first part of the dataset is treated as the basic subset and considered to be free of outliers. The other part of the dataset is non-basic subset which is believed to contain outliers, as outliers tend to be sorted at the end of the whole dataset. Next, the distance from each observation in the whole dataset to the center of the basic subset is calculated. Then, a T-test will be conducted on the calculated distances. The observations whose distances are larger than the critical value will be considered as outliers. Thus, the second sensitivity analysis in this dissertation is conducted by first using Hadi method to identify and exclude outliers from the whole sample, after which DEA models are run with the new sample. Next, efficiency estimates from the original models and the new models are correlated. The results are reported in Table 14. The correlations between two sets of efficiency estimates are considerably high, implying that DEA results are not driven by outliers.

Table 14, Correlations between Efficiency Estimates from the Second Sensitivity Analysis

	Model 1			**Model 2**		
	Minnesota	Indiana	South Carolina	Minnesota	Indiana	South Carolina
Grade 3	1.000	0.999	0.988	0.999	0.999	0.995
Grade 4	0.999	0.999	0.980	0.997	0.999	0.982
Grade 5	1.000	0.999	0.984	1.000	0.993	0.853
Grade 6	0.998	1.000	0.974	0.995	0.998	0.930
Grade 7	1.000	0.995	0.970	0.998	0.998	0.926
Grade 8	0.998	0.982	0.968	0.995	0.969	0.957

CHAPTER V

CONCLUSIONS

This dissertation investigated the effect of NCLB on school efficiency using two stage analyses. At the first stage, school efficiency, measured by efficiency and slack, was estimated using Data Envelopment Analysis (DEA). At the second stage, a difference-in-differences estimator was used to examine the effect of NCLB on school efficiency. This dissertation analyzed school level data from three states: Minnesota, Indiana, and South Carolina. Key findings, implications for education policy, and limitations of this study are summarized in the following sections.

Findings and Policy Implications

There is consistent evidence to suggest that public schools in the United States do not efficiently utilize resources to produce educational outcomes. The average efficiency for schools in Minnesota and Indiana is no more than .78, which suggests that without using more inputs, schools should have produced 22% more outputs if they operated at the optimal level. Relative to Minnesota and Indiana, South Carolina has a much lower average efficiency, no more than .46. This implies that schools in South Carolina only produced half of the output they could have produced were they operated efficiently. One possible reason for the extremely low efficiency is that South Carolina has the most difficult proficiency standards in the nation (Cronin, et al. 2007), which results in fewer students reaching advanced level in the assessments.

This investigation also found that despite the existence of inefficiency, public schools have increased their efficiency level over time. There is no consistent pattern in terms of the change of average efficiency across grades for each state. Some grades have improved efficiency steadily over time (such as third grade in Minnesota and sixth grade in Indiana), while others have gone up and down during the same time period. No matter what the pattern is, the overall trend is that efficiency is eventually improved since NCLB was introduced. Additionally, my research shows that, on average, unthreatened schools have a higher efficiency level than threatened peers in each state.

There is evidence that NCLB has had a positive effect on school efficiency. This dissertation found that NCLB has had a positive effect on school efficiency for both Minnesota and Indiana, but a small effect for South Carolina. In Minnesota and Indiana schools that face little sanctions of NCLB stay relatively stable in terms of the change of efficiency while the ones that are under pressure of being sanctioned have substantially improved efficiency over time. Compared to the former two states, the effect of NCLB on efficiency has been much smaller in South Carolina. South Carolina schools that are threatened by NCLB sanctions differ only slightly from unthreatened schools with respect to trends in efficiency in math and reading. There are greater differences in science and social studies. As discussed, South Carolina has the most difficult proficiency standards in the whole nation, resulting in a low percentage of advanced students in the assessments. Therefore, it is very likely that all schools, including the ones facing and not facing sanctions, respond significantly to NCLB, which subsequently results in subtle effects of the policy on efficiency.

Using our second measure of efficiency, *slack*, we find mixed evidence for the hypothesis that NCLB has resulted in increased slack in school performance. In Minnesota NCLB has reduced slack, suggesting that the policy has made schools in the state become more efficient. On the other hand, the policy has boosted slack in Indiana, meaning that middle schools have become less efficient after NCLB was introduced. With regard to South Carolina, slack in math and science increased while slack in reading reduced after NCLB was implemented. In addition, NCLB has had no impact on slack in social studies in South Carolina.

The findings that threatened schools responded positively to NCLB by increasing their efficiency level are encouraging. In order to reach Annual Measurable Objectives (AMO) and make AYP, schools need to produce more proficient students each year, which can be accomplished by either investing more inputs or improving efficiency level. However, given the fact that federal funding is limited and local authorities need to cover some part of the costs related to implementing NCLB (Center on Education Policy, 2006), it is less likely that more inputs will be provided to schools. Therefore, it is necessary that schools, especially the ones at risk of not making AYP, need to improve their efficiency level so that more outputs can be achieved with limited resources. The results of our study suggest that the pressure from NCLB does make schools better utilize their current resources.

Additionally, the implementation of DEA in this dissertation provides an alternative tool to evaluate school performance. Generally, schools are evaluated based on their effectiveness in producing educational outcomes, that is, whether they reach certain thresholds. This type of evaluation only focuses on outputs, without considering the amount of needed inputs. Thus, it is to some extent incomplete as certain effective schools may have more inputs than ineffective ones and the difference in the inputs is the reason for the effectiveness. To solve this problem, it is necessary to take into account inputs when school performances are evaluated, as efficiency studies have done. As a result, efficiency and effectiveness studies together will construct a complete picture of school performance as schools can be classified into one of the four groups: (1) effective and efficient, (2) effective and inefficient, (3) ineffective and efficient, or (4) ineffective and inefficient.

This classification of schools will give decision makers some guidance when considering resource allocation among schools. The first group of schools is the perfect scenario as they have reached the threshold and utilized resources to the maximal level. The second group of schools has reached the threshold, but not utilized resources to full potential, implying that they have more resources than necessary. Thus, it is possible to reduce resources and increase efficiency level for those schools. The third group has used resources wisely, but has not reached the threshold. Thus, the solution is to increase the amount of resources for this group of schools. In terms of the last group of schools, they are the ones who need most assistance: they need to have more resources and to increase efficiency simultaneously.

On the whole, the evaluation of school performance that contains both efficiency and effectiveness will provide a good tool for decision makers to better allocate resources among schools. By classifying schools into different groups, resources can be targeted toward the neediest schools which in turn will improve student achievement.

Limitations and Future Research

There are three limitations in this dissertation. First, this dissertation assumes a black box model of production, which gives no insights on what kinds of resource allocation decisions are made and how within schools. Because of this assumption, it is impossible to differentiate allocative efficiency from technical efficiency. Thus, it is not easy to provide specific recommendations on how to reallocate resources within schools so that higher level of efficiency can be achieved. To overcome this, future research is needed that will include more detailed information.

Specifically, the future research could ask what kinds of strategies schools employed to improve student test performance, such as: (1) do schools provide incentives for teachers to improve their performance, (2) do schools provide professional development activities for teachers, (3) do schools reduce class size, or (4) do schools provide aids or mentors for teachers. Then it would be possible to get an understanding on whether or not those strategies have significant effects on the improvement of allocative and technical efficiency, which could subsequently provide explicit recommendations for schools to better utilize and allocate their resources.

The second limitation is that this dissertation may not capture the full effects of NCLB. As discussed, the NCLB effect is identified by using difference-in-differences estimation which is constructed by classifying schools into unthreatened or threatened. The difference-in-differences estimator will capture the full effect of NCLB if unthreatened schools do not respond to sanctions of the policy, but will only capture a partial effect if they respond. The literature indicated that many empirical studies focus on responses of low-performing schools, or threatened schools as called in this dissertation, toward sanctions of NCLB (Figlio & Rouse, 2006), but not on responses of high-performing ones. Therefore, in reality it is unclear whether unthreatened schools are influenced by the policy. As a result, another possible direction for future research is to empirically investigate the responses of high-performing schools toward sanctions of NCLB. In turn, this will provide insights on the validity of the assumption that unthreatened schools did not change their behaviors after NCLB was implemented.

Finally, results from the three states are incomparable in this dissertation. Although it is found that the average efficiencies for schools in Minnesota and Indiana (around .78) are higher than those in South Carolina (about .40), we cannot conclude that schools in the former two states are more efficient than those in the latter state in utilizing resources. The reason is that each state has its own assessment systems and distinct proficiency standards, which makes it meaningless to compare test performance across states. Therefore, it is difficult to compare the efficiency level across states with different standards and assessments. One possible direction for future research is to use national assessments, such as NAEP, so that evaluations of school performances across states are comparable.

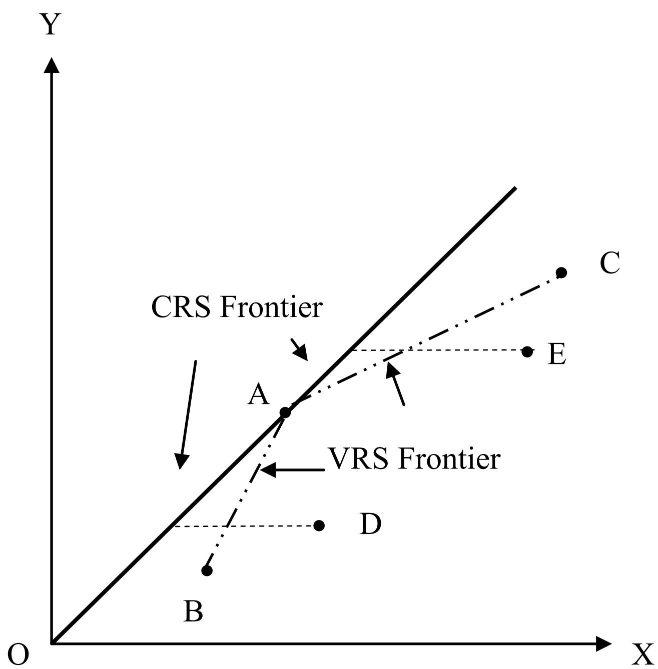
Appendix A

Constant Returns to Scale (CRS) and Variable Returns to Scale (VRS) Models

The basic CCR model assumes Constant Returns to Scale (CRS), which suggests that DMUs can linearly scale up inputs and outputs without decreasing or increasing efficiency (Anderson, 1996). If this assumption holds, then all DMUs are operating at an optimal scale. If CRS does not hold, then choice of scale affects the relationship between inputs and outputs. Estimating the efficiency frontier under a CRS assumption when return to scale is variable can confound scale effects with efficiency. Therefore, Variable Returns to Scale (VRS) models are constructed to differentiate pure technical efficiency from scale efficiency. These two models are depicted in the following graph.

Let's suppose there are five DMUs, A, B, C, D, and E, who produce one output Y using one input X. Under the CRS model, the frontier line is the line OA, meaning that A is the only DMU that is efficient. However, the VRS model takes into account the possibility that DMUs have increasing returns to scale (such as B and D in this example) or decreasing returns to scale (such as C and E in this example). As a result, under the VRS model, the frontier line is constructed by the linear combination of B, A, and C, suggesting that these three DMUs are efficient. Additionally, inefficiency levels for D and E are less in the VRS model than in the CRS model.

To conclude, under the circumstance that not all DMUs are operating at the optimal scale, using the VRS model can yield a more accurate measure of technical efficiency.



CRS and VRS Frontier and Efficiency

Appendix B

AYP Status by School Type, Year, Subject

Table B.1 Minnesota

		Math		Reading	
		Unthreatened	Threatened	Unthreatened	Threatened
2003	Made AYP	160	745	160	749
	Not Made AYP		13		9
2004	Made AYP	162	747	162	740
	Not Made AYP		36		43
2005	Made AYP	159	776	159	774
	Not Made AYP		5		7
2006	Made AYP	155	746	155	760
	Not Made AYP		27		13
2007	Made AYP	153	739	153	737
	Not Made AYP		39		41

Table B.2 Indiana

		Math		Reading	
		Unthreatened	Threatened	Unthreatened	Threatened
2003	Made AYP	306	1406	306	1388
	Not Made AYP		54		72
2004	Made AYP	301	1413	301	1394
	Not Made AYP		80		99
2005	Made AYP	301	1406	300	1352
	Not Made AYP		97	1	151
2006	Made AYP	298	1405	298	1356
	Not Made AYP		102		151

Table B.3 South Carolina

		Math		Reading	
		Unthreatened	Threatened	Unthreatened	Threatened
2003	Made AYP	252	445	252	420
	Not Made AYP	2	152	2	177
2004	Made AYP	251	584	251	575

	Not Made AYP		24		33
2005	Made AYP	245	517	246	509
	Not Made AYP	2	96	1	104
2006	Made AYP	248	526	245	495
	Not Made AYP	1	96	4	127
2007	Made AYP	243	473	242	448
	Not Made AYP	3	161	4	186

Appendix C

Mean Efficiency by Grade, Year, School Type, Model 1 and Model 2

Table C.1 Minnesota

C.1a, Efficiency in Math (Model 1)

		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Unthreatened	0.826	0.806	0.833	0.815	0.870	0.868	0.859	0.840
	Threatened	0.635	0.622	0.703	0.690	0.767	0.758	0.767	0.706
Grade 4	Unthreatened						0.800	0.794	0.797
	Threatened						0.682	0.695	0.688
Grade 5	Unthreatened	0.824	0.848	0.861	0.851	0.888	0.738	0.775	0.826
	Threatened	0.649	0.679	0.735	0.721	0.806	0.611	0.658	0.694
Grade 6	Unthreatened						0.732	0.758	0.745
	Threatened						0.612	0.660	0.636
Grade 7	Unthreatened				0.850	0.859	0.755	0.770	0.808
	Threatened				0.695	0.820	0.630	0.674	0.705
Grade 8	Unthreatened						0.713	0.703	0.708
	Threatened						0.630	0.622	0.626

C.1b, Efficiency in Reading (Model 2)

		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Unthreatened	0.825	0.811	0.840	0.836	0.875	0.894	0.881	0.852
	Threatened	0.657	0.644	0.710	0.722	0.774	0.803	0.783	0.728
Grade 4	unthreatened						0.852	0.809	0.830
	Threatened						0.769	0.719	0.744
Grade 5	Unthreatened	0.869	0.880	0.871	0.857	0.894	0.853	0.832	0.865
	Threatened	0.728	0.731	0.756	0.739	0.811	0.762	0.728	0.751
Grade 6	Unthreatened						0.816	0.786	0.801
	Threatened						0.745	0.710	0.728
Grade 7	Unthreatened				0.879	0.882	0.823	0.783	0.842
	Threatened				0.745	0.811	0.732	0.685	0.743
Grade 8	Unthreatened						0.756	0.766	0.761
	Threatened						0.722	0.686	0.704

Table C.2 Indiana

C.2a, Efficiency in Math (Model 1)							
		2002	2003	2004	2005	2006	Mean
Grade 3	Unthreatened	0.847	0.836	0.826	0.816	0.798	0.824
	Threatened	0.710	0.730	0.731	0.724	0.708	0.721
Grade 4	Unthreatened		0.750	0.741	0.736	0.726	0.738
	Threatened		0.658	0.668	0.670	0.685	0.670
Grade 5	Unthreatened		0.802	0.821	0.825	0.818	0.817
	Threatened		0.713	0.739	0.743	0.758	0.738
Grade 6	Unthreatened	0.860	0.845	0.862	0.885	0.877	0.866
	Threatened	0.712	0.739	0.774	0.798	0.806	0.766
Grade 7	Unthreatened		0.853	0.841	0.841	0.863	0.849
	Threatened		0.744	0.762	0.780	0.792	0.769
Grade 8	Unthreatened	0.892	0.878	0.871	0.865	0.888	0.879
	Threatened	0.768	0.773	0.774	0.769	0.804	0.778

C.2b, Efficiency in Reading (Model 2)							
		2002	2003	2004	2005	2006	Mean
Grade 3	Unthreatened	0.878	0.871	0.844	0.846	0.851	0.858
	Threatened	0.750	0.765	0.758	0.752	0.778	0.761
Grade 4	Unthreatened		0.863	0.851	0.855	0.841	0.852
	Threatened		0.746	0.752	0.771	0.766	0.758
Grade 5	Unthreatened		0.839	0.848	0.857	0.847	0.847
	Threatened		0.744	0.754	0.775	0.783	0.764
Grade 6	Unthreatened	0.841	0.825	0.816	0.825	0.821	0.826
	Threatened	0.697	0.711	0.718	0.721	0.735	0.716
Grade 7	Unthreatened		0.837	0.796	0.790	0.817	0.810
	Threatened		0.731	0.723	0.719	0.749	0.731
Grade 8	Unthreatened	0.878	0.862	0.839	0.838	0.839	0.851
	Threatened	0.728	0.755	0.749	0.750	0.765	0.749

C.3 South Carolina

C.3a, Efficiency in Math (Model 1)		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Unthreatened	0.540	0.477	0.500	0.439	0.438	0.488	0.431	0.473
	Threatened	0.343	0.298	0.315	0.293	0.292	0.326	0.281	0.307
Grade 4	Unthreatened	0.439	0.546	0.503	0.516	0.551	0.558	0.555	0.524
	Threatened	0.253	0.328	0.317	0.343	0.394	0.380	0.388	0.343
Grade 5	Unthreatened	0.512	0.507	0.481	0.537	0.512	0.515	0.503	0.510
	Threatened	0.288	0.271	0.253	0.317	0.317	0.320	0.307	0.296
Grade 6	Unthreatened	0.517	0.537	0.621	0.624	0.645	0.618	0.617	0.597
	Threatened	0.277	0.292	0.386	0.399	0.405	0.376	0.384	0.360
Grade 7	Unthreatened	0.672	0.659	0.655	0.697	0.712	0.669	0.680	0.678
	Threatened	0.384	0.381	0.428	0.465	0.490	0.461	0.467	0.439
Grade 8	Unthreatened	0.619	0.619	0.573	0.646	0.603	0.585	0.493	0.591
	Threatened	0.311	0.325	0.330	0.380	0.409	0.402	0.341	0.357

C.3b, Efficiency in Reading (Model 2)		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Unthreatened	0.587	0.562	0.601	0.695	0.691	0.666	0.659	0.637
	Threatened	0.402	0.366	0.407	0.542	0.530	0.495	0.490	0.462
Grade 4	Unthreatened	0.589	0.512	0.485	0.553	0.505	0.569	0.570	0.540
	Threatened	0.392	0.323	0.309	0.388	0.365	0.406	0.419	0.372
Grade 5	Unthreatened	0.531	0.464	0.391	0.486	0.512	0.575	0.510	0.496
	Threatened	0.323	0.258	0.203	0.297	0.318	0.372	0.330	0.300
Grade 6	Unthreatened	0.572	0.604	0.509	0.553	0.520	0.564	0.571	0.556
	Threatened	0.387	0.407	0.330	0.343	0.329	0.372	0.376	0.363
Grade 7	Unthreatened	0.720	0.686	0.596	0.641	0.662	0.632	0.643	0.654
	Threatened	0.493	0.423	0.368	0.405	0.377	0.403	0.451	0.417
Grade 8	Unthreatened	0.644	0.694	0.530	0.664	0.696	0.608	0.593	0.633
	Threatened	0.402	0.452	0.329	0.445	0.517	0.422	0.411	0.425

C.3c, Efficiency in Science (Model 3 for South Carolina)		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Unthreatened			0.420	0.381	0.413	0.395	0.452	0.412
	Threatened			0.202	0.201	0.228	0.211	0.267	0.222
Grade 4	Unthreatened			0.414	0.495	0.488	0.531	0.568	0.499
	Threatened			0.226	0.275	0.312	0.322	0.368	0.301
Grade 5	Unthreatened			0.433	0.476	0.509	0.480	0.532	0.486

		2001	2002	2003	2004	2005	2006	2007	Mean
	Threatened			0.222	0.259	0.287	0.277	0.338	0.277
Grade 6	Unthreatened			0.462	0.605	0.611	0.532	0.649	0.572
	Threatened			0.257	0.361	0.391	0.315	0.434	0.352
Grade 7	Unthreatened			0.544	0.653	0.638	0.643	0.702	0.636
	Threatened			0.325	0.451	0.450	0.443	0.554	0.445
Grade 8	Unthreatened			0.530	0.603	0.661	0.585	0.634	0.603
	Threatened			0.311	0.356	0.451	0.415	0.533	0.413

C.3d, Efficiency in Social Studies (Model 4 for South Carolina)

		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Unthreatened			0.382	0.397	0.517	0.569	0.597	0.492
	Threatened			0.161	0.226	0.289	0.337	0.370	0.277
Grade 4	Unthreatened			0.362	0.479	0.485	0.508	0.533	0.474
	Threatened			0.169	0.258	0.296	0.299	0.325	0.269
Grade 5	Unthreatened			0.382	0.432	0.456	0.432	0.466	0.433
	Threatened			0.198	0.236	0.261	0.232	0.304	0.246
Grade 6	Unthreatened			0.356	0.535	0.552	0.578	0.646	0.533
	Threatened			0.184	0.302	0.345	0.331	0.411	0.315
Grade 7	Unthreatened			0.601	0.712	0.740	0.597	0.668	0.663
	Threatened			0.353	0.417	0.438	0.387	0.480	0.415
Grade 8	Unthreatened			0.603	0.755	0.699	0.633	0.411	0.620
	Threatened			0.362	0.487	0.496	0.469	0.318	0.426

Appendix D

Mean Slack by Grade, Year, Model 1 and Model 2

Table D.1 Minnesota

D.1a, Slack in ADV in Math (Model 1)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	0.038	0.063	0.028	0.025	0.009	0.26	0.237	0.094
Grade 4						0.167	0.128	0.148
Grade 5	0.11	0.106	0.066	0.071	0.028	0.445	0.371	0.171
Grade 6						0.067	0.057	0.062
Grade 7				0.173	0.086	0.252	0.208	0.18
Grade 8						0.159	0.178	0.169

D.1b, Slack in ADV in Reading (Model 2)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	0.119	0.12	0.083	0.046	0.039	0.156	0.212	0.111
Grade 4						0.107	0.124	0.116
Grade 5	0.036	0.036	0.048	0.041	0.016	0.353	0.355	0.126
Grade 6						0.068	0.087	0.078
Grade 7				0.167	0.161	0.18	0.214	0.181
Grade 8						0.11	0.07	0.09

Table D.2 Indiana

D.2a, Slack in ADV in Math (Model 1)

	2002	2003	2004	2005	2006	Mean
Grade 3	0.154	0.151	0.150	0.146	0.166	0.153
Grade 4		0.215	0.199	0.210	0.178	0.200
Grade 5		0.111	0.087	0.089	0.087	0.093
Grade 6	0.144	0.118	0.095	0.077	0.077	0.102
Grade 7		0.222	0.241	0.221	0.232	0.229
Grade 8	0.058	0.060	0.087	0.083	0.075	0.073

D.2b, Slack in ADV in Reading (Model 2)

	2002	2003	2004	2005	2006	Mean
Grade 3	0.085	0.096	0.098	0.118	0.107	0.101
Grade 4		0.031	0.033	0.037	0.042	0.036
Grade 5		0.175	0.190	0.191	0.205	0.190
Grade 6	0.104	0.103	0.102	0.113	0.117	0.108
Grade 7		0.074	0.081	0.076	0.068	0.075
Grade 8	0.051	0.048	0.052	0.056	0.059	0.053

Table D.3 South Carolina

D.3a, Slack in ADV in Math (Model 1)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	0.120	0.205	0.222	0.272	0.257	0.238	0.233	0.221
Grade 4	0.229	0.195	0.219	0.209	0.256	0.197	0.173	0.211
Grade 5	0.200	0.249	0.303	0.230	0.207	0.207	0.214	0.230
Grade 6	0.083	0.089	0.060	0.059	0.074	0.056	0.062	0.069
Grade 7	0.086	0.068	0.066	0.052	0.047	0.058	0.042	0.060
Grade 8	0.071	0.058	0.063	0.045	0.052	0.033	0.051	0.054

D.3b, Slack in ADV in Reading (Model 2)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	0.196	0.253	0.206	0.175	0.211	0.222	0.241	0.215
Grade 4	0.102	0.132	0.119	0.122	0.126	0.117	0.127	0.121
Grade 5	0.046	0.044	0.054	0.042	0.039	0.028	0.037	0.041
Grade 6	0.124	0.116	0.128	0.134	0.124	0.101	0.122	0.121
Grade 7	0.052	0.049	0.065	0.045	0.051	0.048	0.046	0.051
Grade 8	0.069	0.048	0.073	0.051	0.037	0.062	0.057	0.057

D.3c, Slack in ADV in Science (Model 3 for South Carolina)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3			0.268	0.266	0.287	0.239	0.259	0.264
Grade 4			0.218	0.165	0.159	0.128	0.131	0.160
Grade 5			0.188	0.179	0.111	0.095	0.102	0.135
Grade 6			0.195	0.121	0.103	0.126	0.101	0.129
Grade 7			0.112	0.062	0.043	0.064	0.054	0.067
Grade 8			0.122	0.083	0.071	0.052	0.075	0.081

D.3d, Slack in ADV in Social Studies (Model 4 for South Carolina)

	2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3			0.322	0.370	0.293	0.336	0.331	0.330
Grade 4			0.258	0.247	0.291	0.258	0.231	0.257
Grade 5			0.303	0.300	0.313	0.264	0.292	0.294
Grade 6			0.199	0.129	0.125	0.133	0.163	0.150
Grade 7			0.090	0.067	0.076	0.044	0.040	0.064
Grade 8			0.059	0.038	0.045	0.062	0.088	0.058

Appendix E

Mean Efficiency by Grade, Year, School Type, Model 3 and Model 4

E.1, Minnesota

E.1a, Efficiency (Model 3)

		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Total	0.707	0.695	0.751	0.752	0.802	0.828	0.822	0.765
	Unthreatened	0.858	0.823	0.855	0.836	0.880	0.893	0.891	0.862
	Threatened	0.682	0.675	0.734	0.738	0.790	0.817	0.811	0.749
Grade 4	Total						0.804	0.780	0.792
	Unthreatened						0.870	0.846	0.858
	Threatened						0.797	0.772	0.785
Grade 5	Total	0.759	0.773	0.798	0.787	0.848	0.769	0.744	0.782
	Unthreatened	0.894	0.882	0.889	0.863	0.914	0.842	0.824	0.872
	Threatened	0.735	0.754	0.782	0.773	0.837	0.755	0.730	0.767
Grade 6	Total						0.765	0.752	0.758
	Unthreatened						0.818	0.810	0.814
	Threatened						0.756	0.744	0.750
Grade 7	Total				0.753	0.835	0.702	0.680	0.742
	Unthreatened				0.874	0.895	0.786	0.780	0.834
	Threatened				0.739	0.829	0.691	0.668	0.732
Grade 8	Total						0.744	0.725	0.735
	Unthreatened						0.754	0.774	0.764
	Threatened						0.746	0.721	0.733

E.1b, Efficiency (Model 4)

		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Total	0.721	0.706	0.765	0.767	0.823	0.831	0.825	0.777
	Unthreatened	0.868	0.829	0.862	0.843	0.891	0.893	0.891	0.868
	Threatened	0.692	0.681	0.746	0.750	0.810	0.817	0.811	0.758
Grade 4	Total						0.816	0.790	0.803
	Unthreatened						0.871	0.851	0.861
	Threatened						0.803	0.777	0.790
Grade 5	Total	0.765	0.779	0.803	0.792	0.858	0.770	0.746	0.788
	Unthreatened	0.895	0.885	0.890	0.865	0.916	0.842	0.824	0.874

	Threatened	0.740	0.758	0.785	0.777	0.846	0.756	0.730	0.770
Grade 6	Total						0.784	0.772	0.778
	Unthreatened						0.828	0.826	0.827
	Threatened						0.775	0.764	0.770
Grade 7	Total				0.757	0.845	0.707	0.683	0.748
	Unthreatened				0.877	0.893	0.792	0.784	0.837
	Threatened				0.743	0.839	0.697	0.672	0.738
Grade 8	Total						0.752	0.738	0.745
	Unthreatened						0.736	0.776	0.756
	Threatened						0.755	0.733	0.744

E.2, Indiana

E.2a, Efficiency (Model 3)							
		2002	2003	2004	2005	2006	Mean
Grade 3	Total	0.799	0.811	0.800	0.795	0.807	0.802
	Unthreatened	0.891	0.885	0.864	0.862	0.860	0.872
	Threatened	0.773	0.789	0.781	0.775	0.792	0.782
Grade 4	Total		0.802	0.808	0.815	0.818	0.811
	Unthreatened		0.882	0.876	0.875	0.865	0.875
	Threatened		0.779	0.789	0.798	0.805	0.793
Grade 5	Total		0.797	0.815	0.826	0.832	0.818
	Unthreatened		0.866	0.884	0.888	0.878	0.879
	Threatened		0.777	0.796	0.809	0.820	0.800
Grade 6	Total	0.772	0.790	0.807	0.826	0.832	0.805
	Unthreatened	0.882	0.872	0.876	0.895	0.886	0.882
	Threatened	0.745	0.770	0.790	0.809	0.820	0.787
Grade 7	Total		0.798	0.802	0.812	0.826	0.810
	Unthreatened		0.883	0.863	0.860	0.879	0.871
	Threatened		0.782	0.791	0.804	0.818	0.799
Grade 8	Total	0.792	0.798	0.795	0.793	0.823	0.800
	Unthreatened	0.910	0.889	0.877	0.880	0.893	0.890
	Threatened	0.783	0.791	0.789	0.787	0.818	0.794

E.2b, Efficiency (Model 4)							
		2002	2003	2004	2005	2006	Mean
Grade 3	Total	0.804	0.813	0.802	0.796	0.808	0.805
	Unthreatened	0.896	0.888	0.865	0.864	0.861	0.875
	Threatened	0.777	0.791	0.784	0.777	0.793	0.784
Grade 4	Total		0.805	0.812	0.819	0.821	0.814
	Unthreatened		0.886	0.880	0.878	0.869	0.878
	Threatened		0.781	0.792	0.802	0.808	0.796
Grade 5	Total		0.799	0.817	0.828	0.834	0.820
	Unthreatened		0.868	0.886	0.889	0.878	0.880
	Threatened		0.779	0.798	0.811	0.821	0.802
Grade 6	Total	0.773	0.791	0.808	0.828	0.834	0.807
	Unthreatened	0.883	0.873	0.877	0.896	0.887	0.883
	Threatened	0.746	0.771	0.792	0.811	0.821	0.788
Grade 7	Total		0.801	0.804	0.815	0.829	0.813
	Unthreatened		0.888	0.867	0.865	0.882	0.876

	Threatened		0.785	0.793	0.806	0.821	0.801
Grade 8	Total	0.796	0.803	0.799	0.797	0.826	0.804
	Unthreatened	0.915	0.895	0.883	0.883	0.897	0.895
	Threatened	0.787	0.796	0.793	0.791	0.821	0.798

Appendix F

Mean Slack by Grade, Year, School Type, Model 3 and Model 4

F.1, Minnesota

F.1a, Slack in math (Model 3)

		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Total	0.031	0.029	0.019	0.031	0.015	0.022	0.016	0.023
	Unthreatened	0.008	0.017	0.013	0.019	0.008	0.011	0.005	0.012
	Threatened	0.039	0.034	0.022	0.034	0.017	0.025	0.019	0.027
Grade 4	Total						0.028	0.023	0.026
	Unthreatened						0.021	0.014	0.017
	Threatened						0.022	0.020	0.021
Grade 5	Total	0.054	0.041	0.019	0.019	0.011	0.134	0.088	0.052
	Unthreatened	0.016	0.021	0.009	0.010	0.003	0.079	0.056	0.028
	Threatened	0.061	0.044	0.019	0.020	0.013	0.142	0.093	0.056
Grade 6	Total						0.062	0.044	0.053
	Unthreatened						0.054	0.032	0.043
	Threatened						0.063	0.046	0.055
Grade 7	Total				0.040	0.018	0.100	0.056	0.054
	Unthreatened				0.015	0.022	0.072	0.049	0.039
	Threatened				0.043	0.018	0.104	0.057	0.056
Grade 8	Total						0.069	0.058	0.064
	Unthreatened						0.095	0.035	0.065
	Threatened						0.064	0.061	0.062

F.1b, Slack in reading (Model 3)

		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Total	0.011	0.012	0.006	0.004	0.004	0.002	0.002	0.006
	Unthreatened	0.004	0.010	0.003	0.002	0.002	0.001	0.001	0.003
	Threatened	0.014	0.015	0.009	0.007	0.007	0.002	0.004	0.008
Grade 4	Total						0.001	0.003	0.002
	Unthreatened						0.002	0.000	0.001
	Threatened						0.002	0.003	0.003
Grade 5	Total	0.004	0.006	0.007	0.012	0.008	0.001	0.004	0.006
	Unthreatened	0.002	0.005	0.003	0.008	0.004	0.000	0.002	0.003

		2001	2002	2003	2004	2005	2006	2007	Mean
	Threatened	0.005	0.008	0.010	0.014	0.011	0.001	0.006	0.008
Grade 6	Total						0.002	0.008	0.005
	Untheatened						0.000	0.003	0.002
	Threatened						0.002	0.009	0.006
Grade 7	Total				0.004	0.013	0.003	0.012	0.008
	Untheatened				0.008	0.012	0.002	0.013	0.009
	Threatened				0.003	0.013	0.003	0.011	0.008
Grade 8	Total						0.007	0.007	0.007
	Untheatened						0.014	0.006	0.010
	Threatened						0.006	0.008	0.007

F.1c, Slack in math (Model 4)

		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Total	0.050	0.083	0.050	0.051	0.028	0.314	0.296	0.125
	Untheatened	0.015	0.047	0.031	0.040	0.013	0.306	0.274	0.104
	Threatened	0.057	0.090	0.054	0.054	0.031	0.315	0.301	0.129
Grade 4	Total						0.220	0.193	0.207
	Untheatened						0.248	0.218	0.233
	Threatened						0.215	0.188	0.202
Grade 5	Total	0.141	0.130	0.074	0.085	0.041	0.403	0.354	0.175
	Untheatened	0.092	0.088	0.044	0.057	0.021	0.354	0.322	0.140
	Threatened	0.151	0.138	0.080	0.091	0.045	0.413	0.360	0.183
Grade 6	Total						0.094	0.072	0.083
	Untheatened						0.087	0.062	0.074
	Threatened						0.095	0.074	0.085
Grade 7	Total				0.171	0.086	0.242	0.199	0.174
	Untheatened				0.145	0.084	0.246	0.207	0.171
	Threatened				0.174	0.086	0.242	0.198	0.175
Grade 8	Total						0.090	0.111	0.101
	Untheatened						0.067	0.060	0.064
	Threatened						0.094	0.118	0.106

F.1d, Slack in reading (Model 4)

		2001	2002	2003	2004	2005	2006	2007	Mean
Grade 3	Total	0.129	0.124	0.094	0.055	0.057	0.183	0.235	0.125
	Untheatened	0.090	0.098	0.075	0.043	0.043	0.150	0.199	0.100
	Threatened	0.137	0.130	0.098	0.058	0.059	0.190	0.243	0.131
Grade 4	Total						0.114	0.130	0.122
	Untheatened						0.138	0.155	0.147

	Threatened						0.109	0.124	0.117
Grade 5	Total	0.055	0.059	0.087	0.080	0.037	0.370	0.401	0.155
	Unthreatened	0.042	0.042	0.071	0.060	0.022	0.324	0.358	0.131
	Threatened	0.057	0.062	0.090	0.084	0.040	0.379	0.410	0.160
Grade 6	Total						0.078	0.101	0.089
	Unthreatened						0.098	0.121	0.109
	Threatened						0.075	0.097	0.086
Grade 7	Total				0.151	0.130	0.170	0.196	0.162
	Unthreatened				0.121	0.100	0.169	0.186	0.144
	Threatened				0.155	0.133	0.170	0.198	0.164
Grade 8	Total						0.130	0.082	0.106
	Unthreatened						0.147	0.031	0.089
	Threatened						0.127	0.089	0.108

F.2, Indiana

F.2a, Slack in Math (Model 3)

		2002	2003	2004	2005	2006	Mean
Grade 3	Total	0.022	0.017	0.015	0.014	0.030	0.020
	Unthreatened	0.011	0.014	0.008	0.013	0.019	0.013
	Threatened	0.026	0.018	0.017	0.014	0.033	0.022
Grade 4	Total		0.014	0.009	0.015	0.010	0.012
	Unthreatened		0.009	0.006	0.010	0.006	0.008
	Threatened		0.016	0.010	0.017	0.011	0.013
Grade 5	Total		0.029	0.016	0.021	0.019	0.021
	Unthreatened		0.016	0.008	0.009	0.009	0.010
	Threatened		0.032	0.018	0.025	0.021	0.024
Grade 6	Total	0.022	0.021	0.007	0.004	0.005	0.012
	Unthreatened	0.009	0.017	0.005	0.002	0.002	0.007
	Threatened	0.025	0.022	0.007	0.005	0.006	0.013
Grade 7	Total		0.016	0.011	0.005	0.006	0.010
	Unthreatened		0.009	0.003	0.003	0.004	0.005
	Threatened		0.018	0.012	0.005	0.007	0.010
Grade 8	Total	0.009	0.013	0.012	0.013	0.007	0.011
	Unthreatened	0.007	0.007	0.002	0.009	0.004	0.006
	Threatened	0.009	0.013	0.012	0.013	0.008	0.011

F.2b, Slack in Reading (Model 3)

		2002	2003	2004	2005	2006	Mean
Grade 3	Total	0.008	0.008	0.008	0.008	0.003	0.007
	Unthreatened	0.002	0.004	0.007	0.006	0.003	0.004
	Threatened	0.010	0.009	0.008	0.008	0.003	0.008
Grade 4	Total		0.012	0.014	0.008	0.011	0.011
	Unthreatened		0.006	0.009	0.005	0.008	0.007
	Threatened		0.013	0.015	0.008	0.012	0.012
Grade 5	Total		0.017	0.020	0.014	0.016	0.016
	Unthreatened		0.009	0.010	0.007	0.009	0.009
	Threatened		0.019	0.023	0.016	0.017	0.019
Grade 6	Total	0.021	0.026	0.031	0.041	0.034	0.031
	Unthreatened	0.013	0.017	0.023	0.026	0.027	0.021
	Threatened	0.023	0.028	0.033	0.044	0.036	0.033
Grade 7	Total		0.018	0.027	0.035	0.030	0.028
	Unthreatened		0.014	0.028	0.020	0.027	0.022
	Threatened		0.019	0.027	0.038	0.030	0.029

Grade 8	Total	0.032	0.022	0.022	0.020	0.030	0.025
	Unthreatened	0.014	0.014	0.027	0.031	0.027	0.022
	Threatened	0.033	0.022	0.022	0.019	0.030	0.025

F.2c, Slack in Math (Model 4)

		2002	2003	2004	2005	2006	Mean
Grade 3	Total	0.114	0.112	0.118	0.116	0.141	0.120
	Unthreatened	0.108	0.110	0.122	0.121	0.152	0.123
	Threatened	0.115	0.113	0.117	0.114	0.137	0.119
Grade 4	Total		0.107	0.092	0.093	0.070	0.090
	Unthreatened		0.093	0.078	0.086	0.059	0.079
	Threatened		0.111	0.096	0.095	0.073	0.094
Grade 5	Total		0.212	0.189	0.189	0.181	0.193
	Unthreatened		0.207	0.173	0.185	0.182	0.187
	Threatened		0.214	0.193	0.190	0.181	0.194
Grade 6	Total	0.190	0.165	0.143	0.110	0.111	0.144
	Unthreatened	0.166	0.149	0.139	0.093	0.107	0.131
	Threatened	0.195	0.169	0.145	0.114	0.112	0.147
Grade 7	Total		0.214	0.230	0.196	0.215	0.214
	Unthreatened		0.167	0.193	0.173	0.191	0.181
	Threatened		0.223	0.238	0.199	0.219	0.220
Grade 8	Total	0.069	0.065	0.075	0.068	0.058	0.067
	Unthreatened	0.056	0.060	0.065	0.040	0.035	0.051
	Threatened	0.070	0.066	0.075	0.070	0.059	0.068

F.2d, Slack in Reading (Model 4)

		2002	2003	2004	2005	2006	Mean
Grade 3	Total	0.080	0.087	0.083	0.104	0.094	0.089
	Unthreatened	0.069	0.081	0.082	0.104	0.092	0.086
	Threatened	0.083	0.088	0.083	0.104	0.095	0.090
Grade 4	Total		0.053	0.056	0.059	0.068	0.059
	Unthreatened		0.036	0.044	0.049	0.063	0.048
	Threatened		0.058	0.059	0.062	0.070	0.062
Grade 5	Total		0.133	0.140	0.137	0.151	0.140
	Unthreatened		0.111	0.116	0.125	0.145	0.124
	Threatened		0.139	0.147	0.141	0.153	0.145
Grade 6	Total	0.084	0.077	0.067	0.075	0.082	0.077
	Unthreatened	0.070	0.067	0.058	0.070	0.077	0.069
	Threatened	0.088	0.080	0.070	0.077	0.084	0.079

Grade 7	Total		0.072	0.085	0.076	0.065	0.075
	Unthreatened		0.052	0.064	0.065	0.053	0.058
	Threatened		0.076	0.089	0.078	0.067	0.078
Grade 8	Total	0.032	0.038	0.047	0.054	0.054	0.045
	Unthreatened	0.033	0.032	0.047	0.048	0.042	0.040
	Threatened	0.032	0.038	0.047	0.054	0.055	0.045

Appendix G

Second-Stage Results using 2001 Test Scores to Classify Schools

G.1, Coefficients from 2nd-stage Models, Model 1 and Model 2, Minnesota
Using *Efficiency* as Dependent Variable

	Efficiency in Math				Efficiency in Reading			
Elementary Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.690	*(0.003)*	0.730	*(0.006)*	0.717	*(0.003)*	0.738	*(0.005)*
Year 2002	-0.024	*(0.010)*	-0.024	*(0.010)*	-0.020	*(0.008)*	-0.020	*(0.008)*
Year 2003	0.007	*(0.010)*	0.008	*(0.010)*	-0.003	*(0.008)*	-0.003	*(0.008)*
Year 2004	-0.012	*(0.010)*	-0.011	*(0.010)*	-0.017	*(0.008)*	-0.016	*(0.008)*
Year 2005	0.035	*(0.010)*	0.036	*(0.010)*	0.024	*(0.008)*	0.025	*(0.008)*
Year 2006	-0.034	*(0.009)*	0.003	*(0.013)*	0.014	*(0.008)*	0.035	*(0.011)*
Year 2007	-0.030	*(0.009)*	0.005	*(0.013)*	-0.011	*(0.008)*	0.009	*(0.011)*
Grade 4	-0.017	*(0.004)*	-0.017	*(0.004)*	-0.017	*(0.003)*	-0.017	*(0.003)*
Grade 5	-0.007	*(0.002)*	-0.007	*(0.002)*	0.026	*(0.002)*	0.026	*(0.002)*
Threatened*Year 2002	0.036	*(0.011)*	0.036	*(0.011)*	0.021	*(0.009)*	0.021	*(0.009)*
Threatened*Year 2003	0.066	*(0.011)*	0.066	*(0.011)*	0.044	*(0.009)*	0.044	*(0.009)*
Threatened*Year 2004	0.076	*(0.011)*	0.076	*(0.011)*	0.058	*(0.009)*	0.058	*(0.009)*
Threatened*Year 2005	0.105	*(0.011)*	0.104	*(0.011)*	0.074	*(0.009)*	0.073	*(0.009)*
Threatened*Year 2006	0.074	*(0.010)*	0.070	*(0.014)*	0.076	*(0.008)*	0.072	*(0.012)*
Threatened*Year 2007	0.092	*(0.010)*	0.088	*(0.014)*	0.066	*(0.008)*	0.063	*(0.012)*
Pct_HighStakes			-0.147	*(0.037)*			-0.084	*(0.031)*
Threatened*Pct_HighStakes			0.019	*(0.041)*			0.015	*(0.034)*
Middle Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.680	*(0.009)*	0.676	*(0.013)*	0.762	*(0.008)*	0.785	*(0.011)*
Year 2005	0.019	*(0.026)*	0.018	*(0.026)*	-0.002	*(0.023)*	-0.002	*(0.023)*
Year 2006	-0.067	*(0.021)*	-0.117	*(0.061)*	-0.061	*(0.019)*	-0.050	*(0.054)*
Year 2007	-0.065	*(0.021)*	-0.115	*(0.061)*	-0.097	*(0.019)*	-0.085	*(0.054)*
Grade 7	0.046	*(0.008)*	0.046	*(0.008)*	0.015	*(0.007)*	0.015	*(0.007)*
Grade 8	0.027	*(0.008)*	0.027	*(0.008)*	0.018	*(0.007)*	0.018	*(0.007)*
Threatened*Year 2005	0.102	*(0.027)*	0.103	*(0.027)*	0.062	*(0.024)*	0.062	*(0.024)*
Threatened*Year 2006	0.005	*(0.022)*	0.052	*(0.064)*	0.039	*(0.019)*	0.077	*(0.057)*

Threatened*Year 2007	0.028	*(0.022)*	0.076	*(0.064)*	0.029	*(0.019)*	0.067	*(0.056)*
Pct_HighStakes			0.092	*(0.104)*			-0.021	*(0.092)*
Threatened*Pct_HighStakes			-0.086	*(0.111)*			-0.079	*(0.098)*

G.2, Coefficients from 2nd-stage Models, Model 1 and Model 2, Minnesota
Using *Slack in ADV* as Dependent Variable

	Slack in Math				Slack in Reading			
Elementary Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.095	*(0.003)*	0.088	*(0.005)*	0.118	*(0.002)*	0.102	*(0.004)*
Year 2002	0.021	*(0.009)*	0.021	*(0.009)*	0.007	*(0.008)*	0.007	*(0.008)*
Year 2003	-0.012	*(0.009)*	-0.013	*(0.009)*	-0.004	*(0.008)*	-0.004	*(0.008)*
Year 2004	-0.006	*(0.009)*	-0.006	*(0.009)*	-0.025	*(0.008)*	-0.026	*(0.008)*
Year 2005	-0.034	*(0.009)*	-0.035	*(0.009)*	-0.040	*(0.008)*	-0.041	*(0.008)*
Year 2006	0.101	*(0.008)*	0.090	*(0.011)*	-0.010	*(0.007)*	-0.021	*(0.010)*
Year 2007	0.064	*(0.008)*	0.054	*(0.011)*	0.007	*(0.007)*	-0.004	*(0.010)*
Grade 4	0.017	*(0.003)*	0.017	*(0.003)*	-0.010	*(0.003)*	-0.009	*(0.003)*
Grade 5	0.002	*(0.002)*	0.002	*(0.002)*	-0.029	*(0.002)*	-0.029	*(0.002)*
Threatened*Year 2002	0.011	*(0.010)*	0.011	*(0.010)*	-0.008	*(0.008)*	-0.008	*(0.008)*
Threatened*Year 2003	-0.007	*(0.010)*	-0.007	*(0.010)*	-0.006	*(0.008)*	-0.006	*(0.008)*
Threatened*Year 2004	-0.008	*(0.009)*	-0.008	*(0.010)*	-0.013	*(0.008)*	-0.013	*(0.008)*
Threatened*Year 2005	-0.024	*(0.010)*	-0.023	*(0.010)*	-0.022	*(0.008)*	-0.022	*(0.008)*
Threatened*Year 2006	-0.038	*(0.009)*	-0.032	*(0.012)*	-0.019	*(0.008)*	-0.022	*(0.011)*
Threatened*Year 2007	-0.040	*(0.009)*	-0.035	*(0.012)*	-0.010	*(0.008)*	-0.013	*(0.011)*
Pct_HighStakes			0.043	*(0.033)*			0.044	*(0.029)*
Threatened*Pct_HighStakes			-0.023	*(0.036)*			0.011	*(0.031)*
Middle Schools	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	0.099	*(0.007)*	0.102	*(0.010)*	0.009	*(0.006)*	0.029	*(0.009)*
Year 2005	-0.064	*(0.019)*	-0.063	*(0.019)*	-0.0004	*(0.018)*	0.0002	*(0.018)*
Year 2006	-0.037	*(0.016)*	-0.008	*(0.045)*	0.056	*(0.015)*	0.079	*(0.044)*
Year 2007	-0.049	*(0.016)*	-0.020	*(0.045)*	0.050	*(0.015)*	0.072	*(0.043)*
Grade 7	0.082	*(0.006)*	0.082	*(0.006)*	0.122	*(0.005)*	0.121	*(0.005)*
Grade 8	0.134	*(0.006)*	0.134	*(0.006)*	0.040	*(0.005)*	0.039	*(0.005)*
Threatened*Year 2005	-0.026	*(0.020)*	-0.026	*(0.020)*	-0.003	*(0.019)*	-0.002	*(0.019)*
Threatened*Year 2006	-0.017	*(0.017)*	-0.040	*(0.048)*	-0.026	*(0.016)*	-0.007	*(0.046)*
Threatened*Year 2007	-0.012	*(0.017)*	-0.035	*(0.047)*	-0.009	*(0.016)*	0.010	*(0.046)*
Pct_HighStakes			-0.053	*(0.077)*			-0.042	*(0.074)*
Threatened*Pct_HighStakes			0.043	*(0.083)*			-0.042	*(0.079)*

Appendix H

Trends in Efficiency for Grade 4, 5, and 7, Indiana

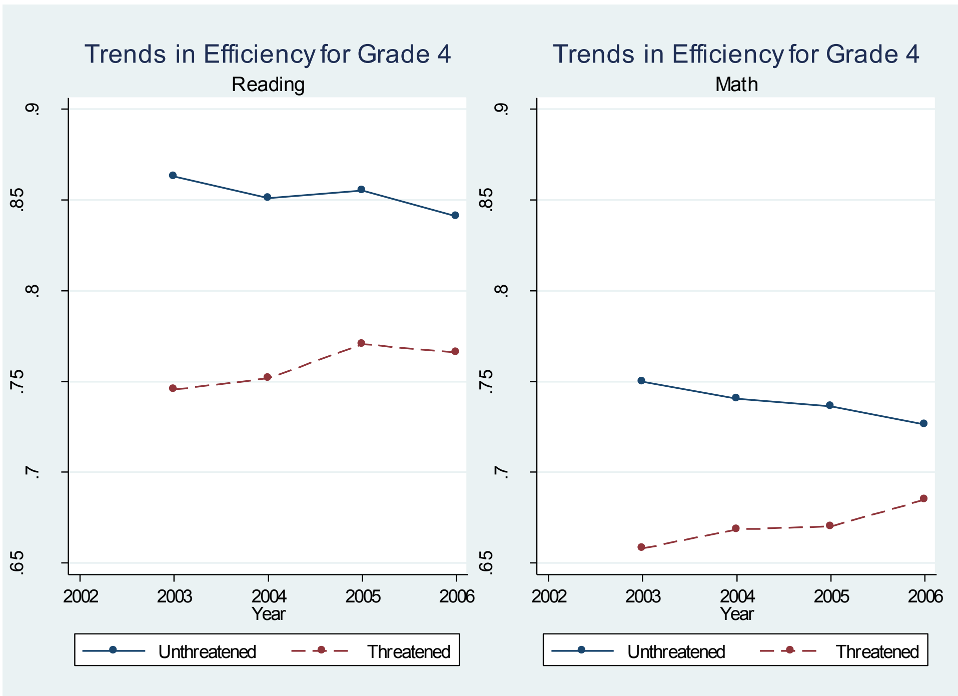


Figure H.1, Trends in Efficiency for Grade 4

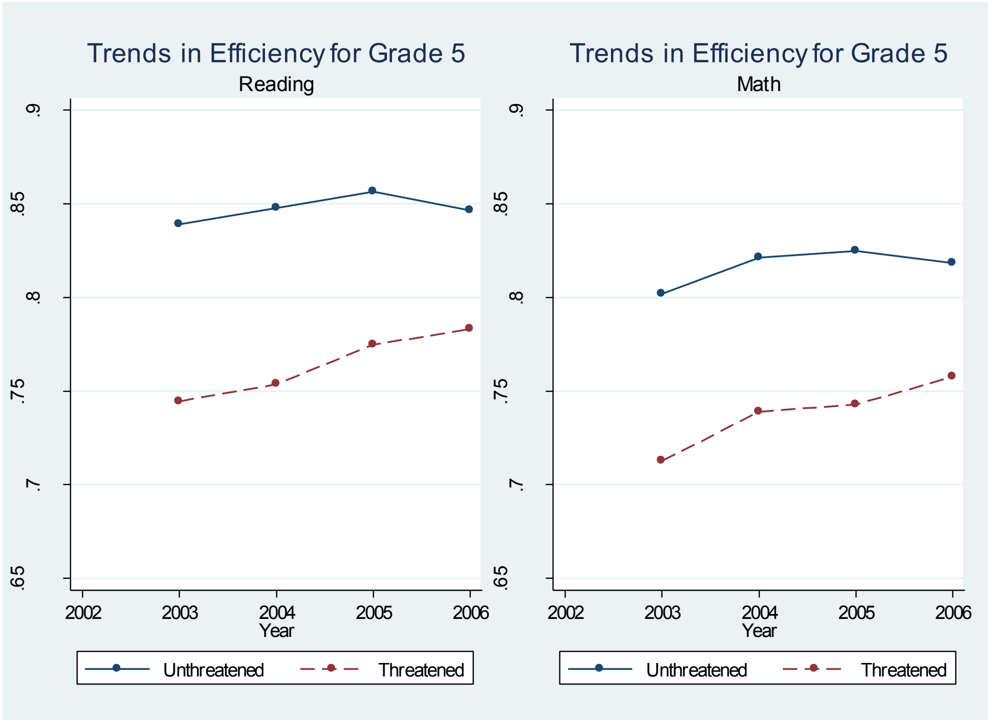


Figure H.2, Trends in Efficiency for Grade 5

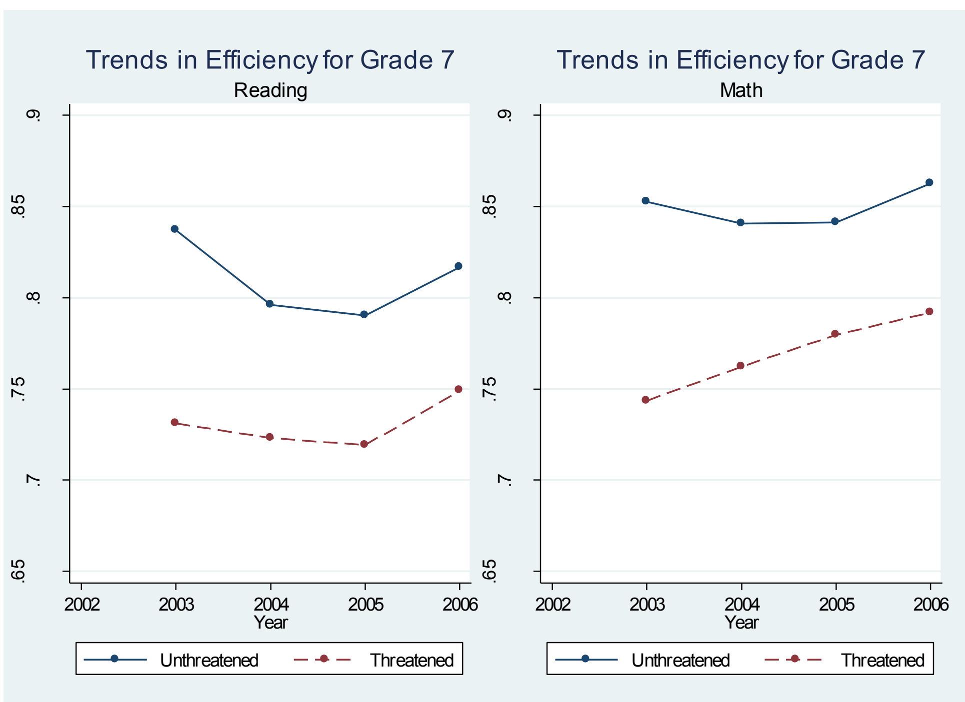


Figure H.3, Trends in Efficiency for Grade 7

Appendix I

Trends in Efficiency in Science and Social Studies, South Carolina

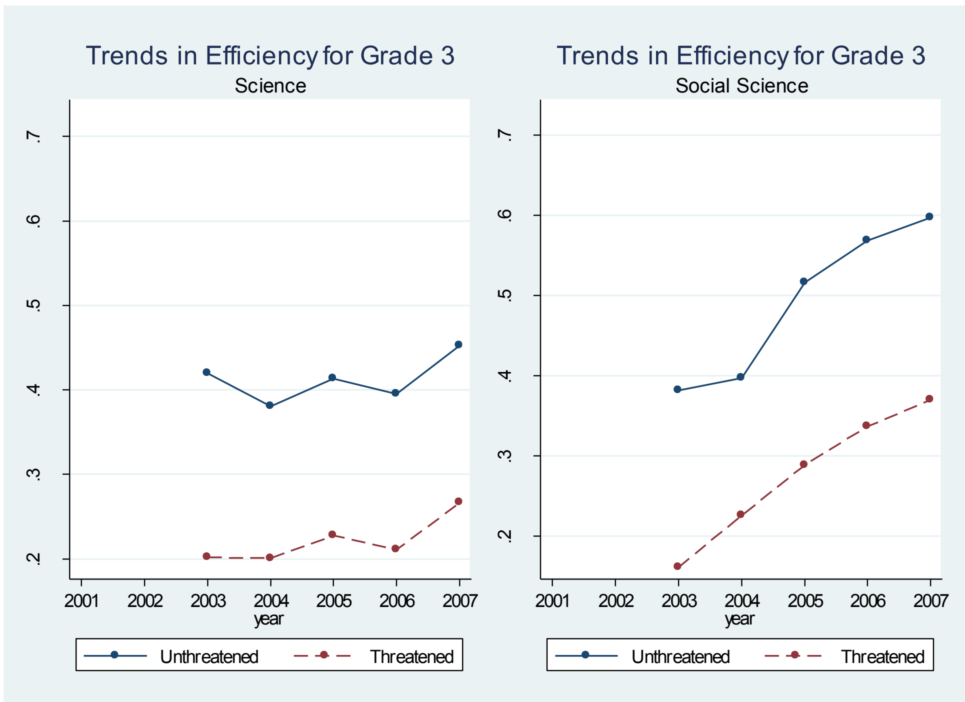


Figure I.1, Trends in Efficiency for Grade 3

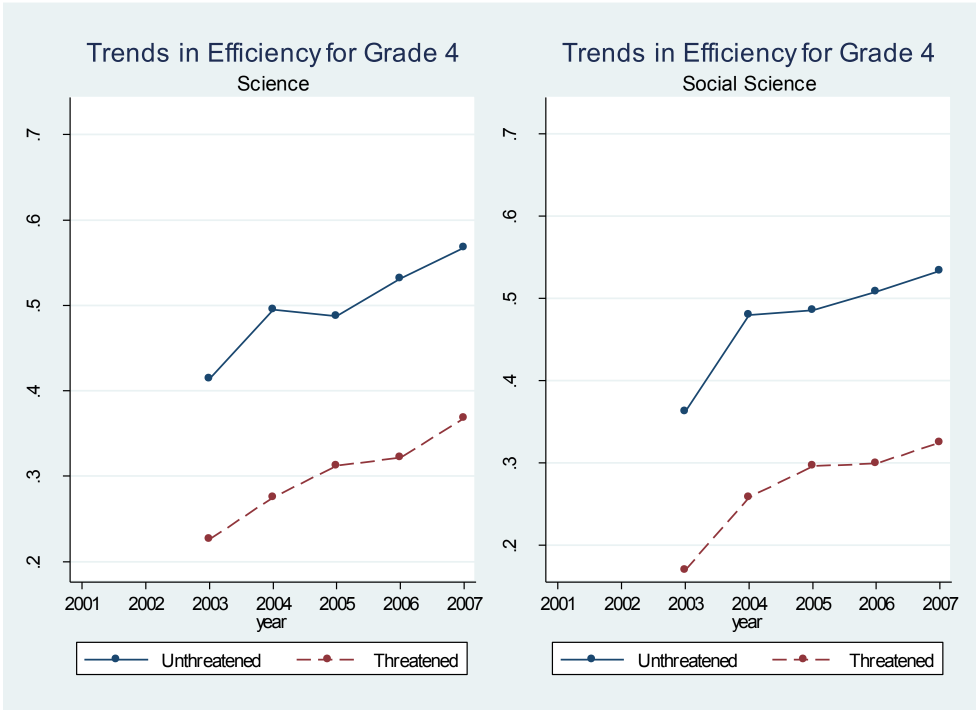


Figure I.2, Trends in Efficiency for Grade 4

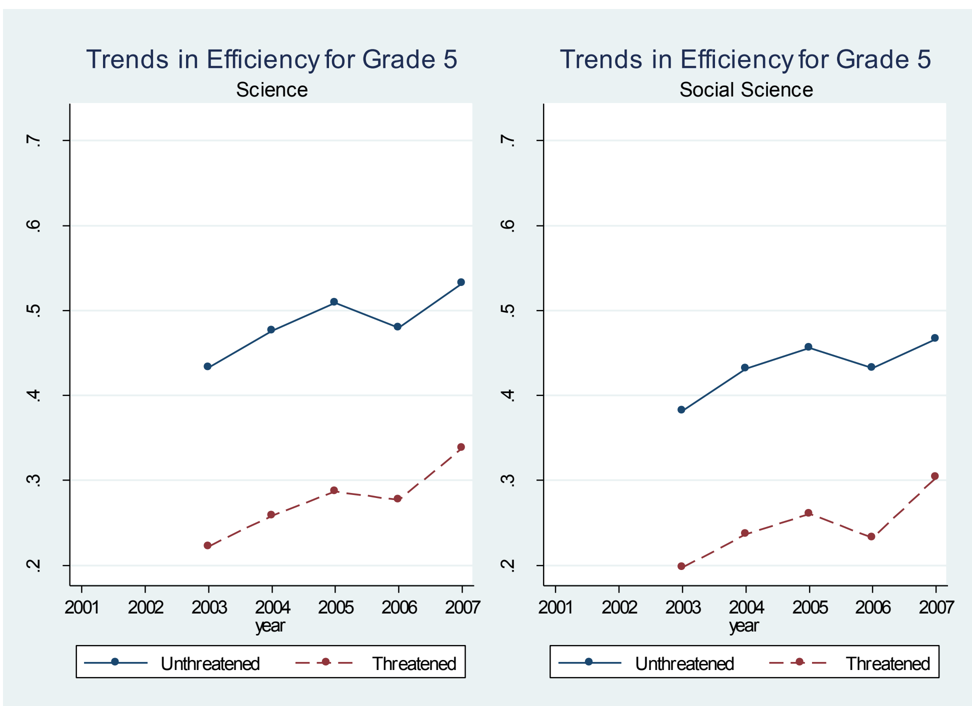


Figure I.3, Trends in Efficiency for Grade 5

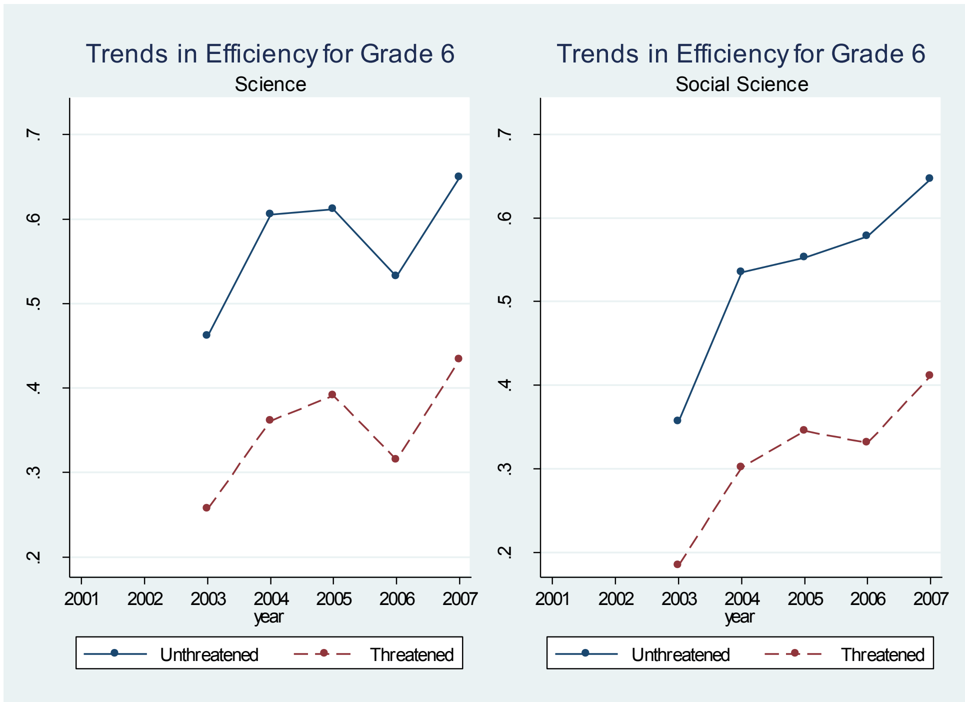


Figure I.4, Trends in Efficiency for Grade 6

REFERENCES

Anderson, L., Walberg, H., & Weinstein, T. (1998). Efficiency and effectiveness analysis of Chicago public schools: 1989, 1991, 1993. *Educational Administration Quarterly, 34(4),* 484-504.

Arrow, K. (1985). The economics of agency. In Pratt, J. and Zeckhauser, R. (Eds.), *Principals and agents: The Structure of business (*pp. 37-51), Boston: Harvard University Press.

Ballou, D. & Podgursky, M. (1998). Rural teachers and schools. In Gibbs, R., Swaim, P., & Teixeira, R. (Eds.) *Rural education and training in the new economy: The myth of the rural skills gap* (pp. 3-21). Iowa State University Press, Ames, Iowa.

Bessent, A., Bessent, W., Kennington, J. & Reagan, B. (1982). An application of mathematical programming to assess productivity in the Houston independent school district. *Management Science, 28,* 1355- 1367.

Bifulco, R. and Bretschneider, S. (2001) Estimating school efficiency: A comparison of methods using simulated data. *Economics of Education Review, 20*, 417-429.

Callahan, R. (1962). *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools*. The University of Chicago Press, Chicago.

Center on Education Policy. (2008). *Instructional times in elementary schools: A closer look at changes for specific subjects*. Washington, DC: Author.

Center on Education Policy. (2007). *Answering the question that matters most: Has student achievement increased since No Child Left Behind?* Washington, DC: Author.

Center on Education Policy. (2006). *From the Capital to the Classroom: Year 4 of the No Child Left Behind Act.* Washington, DC: Author.

Center on Education Policy. (2005). *From the Capital to the Classroom: Year 3 of the No Child Left Behind Act.* Washington, DC: Author.

Chakrabarti, R. (2006). *Do public schools facing vouchers behave strategically? Evidence from Florida*. Nashville, TN: National Research and Development Center on School Choice.

Chakraborty, K., Biswas, B., & Lewis, W. (2001). Measurement of technical efficiency in public education: A stochastic and nonstochastic production function approach. *Southern Economic Journal, 67(4),* 889-905.

Chalos, P. (1997). An examination of budgetary inefficiency in education using data envelopment analysis. *Financial Accountability and Management, 13,* 55-69.

Charnes, A., Cooper, W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of Data Envelopment Analysis to Program Fellow Through. *Management Science, 27(6),* 668-697.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2(4),* 429-444.

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

Coelli, T., Rao, P., O'Donnell, C., & Battese, G. (2005). *An introduction to efficiency and productivity analysis* (2nd ed.). Springer, New Yok.

Cohen, C., Murray, J., & Clewell, B. (2007). *Promise or peril?: NCLB and the Education of ELL students*. Working Paper, Washington DC: The Urban Institute

Cooper, S. T., & Cohen, E. (1997). Estimation of a frontier production function for South Carolina educational process. Economics of Education Review, 16, 313-327.

Cooper, W., Seiford, L., & Zhu, J. (2004). *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers, Boston.

Cronin, J., Dahlin, M., Adkins, D., & Kingbury, G.G. (2007). *The proficiency illusion*. Lake Oswego, OR: Northwest Evaluation Association.

Cronin, J., Kingsbury, G. G., McCall, M. S., & Bowe, B. (2005). *The impact of the No Child Left Behind act on student achievement and growth: 2005 edition*. Portland, OR: Northwest Evaluation Association.

Cullen, J., & Reback, R. (2002). *Tinkering towards accolades: School gaming under a performance accountability system.* Unpublished manuscript, University of Michigan.

Cummins, J. D., & Zi, H. (1998). Comparison of frontier efficiency methods: An application to the U.S. life insurance industry. *Journal of Productivity Analysis, 10(2),* 131-152.

D'Amico, L. (2008). *Impact of Educational Accountability on K-2 Teaching Practices in Primary and Elementary Schools.* Paper presented at the annual meeting of American Educational Research Association, March, New Yok.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica, 19(3),* 273-292.

Deere, D., & Strayer, W. (2001). *Putting schools to the test: School accountability, incentives, and behavior*. Unpublished manuscript, Texas A&M University.

Deller, S. C. & Rudnicki, E. R. (1993). Production efficiency in elementary education: The case of Maine public schools. *Economics of Education Review, 12*, 45-57.

Development Associates, Inc. (2003). *Descriptive study of services to LEP students and LEP students with disabilities*. Arlington, VA: Author.

Dodge, R. (2003). Funding cuts may outstrip school needs. *Knight Ridder Tribune Business News, Washington,* August 6, p. 1.

Downes, T. A. (1996). An estimation of the structure of governance in California school districts before and after Proposition 13. *Public Choice, 86*, 279-307.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review, 14(1)*, 57-74.

Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, 120(3),* 253-290.

Ferrier, G. D., & Lovell, C. A. (1990). Measuring cost efficiency in banking: Econometric and linear programming evidence. *Journal of Econometrics, 46*, 229 – 245.

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives, 7(9)*, retrieved on February 10, 2007 from http://epaa.asu.edu/epaa/v7n9.html.

Figlio, D. N., & Getzler, L. S. (2002). *Accountability, ability and disability: Gaming the system?* (NBER Working Papers No. 9307). Cambridge, MA: National Bureau of Economic Research. Retrieved on December 13, 2008 from http://www.nber.org/papers/w9307.pdf.

Figlio, D. & Rouse, C. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics, 90,* 239-255

Forsund, F. R. (1992). A comparison of parametric and non-parametric efficiency measures: The case of Norwegian ferries. *Journal of Productivity Analysis, 3*, 25-43.

Geys, B., & Moesen, W. (2009). Measuring local government technical efficiency: An application and comparison of FDH, DEA, and econometric approaches. *Public Performance & Management Review, 32(4)*, 499-513.

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

Gribben, M., Campbell, H., & Mathew, J. (2008). *Are advanced students advancing? Examining achievement trends beyond proficiency*. Paper presented at the annual meeting of American Educational Research Association, March, New York.

Grosskopf, S., Hayes, K., Taylor, L., & Weber, W. (2001). On the determinants of school district efficiency: Competition and monitoring. *Journal of Urban Economics, 49*, 453-478.

Grosskopf, S., Hayes, K., Taylor, L., & Weber, W. (1997), Budget-constrained frontier measures of fiscal equality and efficiency in schooling. *Review of Economics and Statistics, 79,* 116-124.

Grosskopf, S., & Moutray, C. (2001). Evaluating performance in Chicago public high schools in the wake of decentralization. *Economics of Education Review, 20 (1)*, 1-14.

Goldhaber, D., & Hannaway, J. (2001). *Accountability with a kicker: Observations on the Florida A+ plan.* Working paper, The Urban Institute.

Hadi, A. S. (1992). Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Soceity, Series B (Methodological), 54(3),* 761-771.

Hadi, A. S. (1994). A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Soceity, Series B (Methodological), 56(2),* 393-396.

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

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

Herbert, B. (2003). *The kids left behind*. New York Times, New York, August 28, p. A31.

Hoerandner, C. M. & Lemke, R. J. (2006). Can No Child Left Behind close the gaps on standardized tests? *Contemporary Economic Policy, 24(1)*, 1-17.

Ingersoll, R. M. (1999). The problem of underqualified teachers in American secondary schools. *Educational Researcher 28(2)*, 26–37.

Jacob, B. (2005). Accountability, incentives, and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics, 89(5-6),* 761-796.

Kang, B. & Greene, K. (2002). The effects of monitoring and competition on public education outputs: A stochastic frontier approach. *Public Finance Review, 30(1)*, 3-26.

Klein, S., Hamilton, L., McCaffrey, D., & Stecher, B. (2003). *What do tests scores in Texas tell us?* Santa Monica, CA: RAND.

Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities. In Koopmans, T.C. (Ed.) *Activity analysis of production and allocation*, Cowles Commission Monograph No. 13 (pp.33-97). John Wiley and Sons, New York.

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, Massachusetts: Harvard University Press.

Koretz, D.M. (1988). Arriving in Lake Wobegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator: The Professional Journal of the American Federation of Teachers, 12(2),* 8-15.

Koretz, D., & Barron, S. I. (1998). *The validity of gains in scores on the Kentucky Instructional Results Information System (KIRIS)*. Washington, DC: Educational Resources Information System.

Krieg, J. M. (2008). Are students left behind? The distributional effects of the No Child Left Behind act. *Education Finance and Policy*, 3(3), 250-281.

Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends*. Cambridge, MA: Harvard Civil Rights Project.

Leibenstein, H. (1978). *General X-efficiency theory and economic development*. New York, Harvard University Press.

Leibenstein, H. (1966). Allocative efficiency vs. X-efficiency. *American Economic Review, 56,* 392-415.

Lips, D., Watkins, S., & Fleming, J. (2008). Does spending more on education improve academic achievement? *Backgrounder*, retrieved on January 15, 2009 from http://www.heritage.org/research/Education/upload/bg_2179.pdf.

Mintrop, H. (2003). The limits of sanctions on low-performing schools: A study of Maryland and Kentucky schools on probation. *Education Policy Analysis Archives, 11(3).* Retrieved on February 10, 2009 from http://epaa.asu.edu/epaa/v11n3.html.

Miura, Y. (2008). *Implications of inclusion of English Language Learners (ELL) in high-stakes testing under NCLB: Hierarchical Linear Model (HLM) analysis of school differences in academic achievement*. Paper presented at the annual meeting of American Educational Research Association, March, NY.

McCarty, T. A. & Yaisawarng, S. (1993) Technical efficiency in New Jersey school districts. In Fried, H. O., Lovell, C. A. & Schmidt, S. S. (Eds.) *The Measurement of Productive Efficiency: Techniques and Applications* (pp.271-287). Oxford University Press, New York.

McMurrer, J. (2008). *Changes to Curriculum and Instruction Since NCLB*. Paper presented at the annual meeting of the American Educational Research Association, March, New York.

McNichol, E. C. & Lav, I. J. (2008). 29 states face total budget shortfall of at least $48 billion in 2009. *Center on Budget and Policy Priorities*, retrieved on December 21, 2008 from http://www.cbpp.org/1-15-08sfp.pdf.

Mueller, C. & Schmitt, V. (2006). *The effects of AYP requirements on student achievement.* Paper presented at the annual meeting of the American Educational Research Association, April, San Francisco, CA.

Murillo-Zamorano, L. R. & Vega-Cervera, J. A. (2001). The use of parametric and non-parametric frontier methods to measure the productive efficiency in the industrial sector. A comparative study. *International Journal of Production Economics, 69(3),* 265-275.

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future.* New York: Author.

Neal, D. & Schanzenbach, D. (2007). *Left behind by design: Proficiency counts and test-based accountability*. NBER Working Paper, No. 13293.

Niskanen, W. A. (1968). The peculiar economics of bureaucracy. *American Economic Review, 58 (2),* 293-305.

Niskanen, W. A. (1971). *Bureaucracy and representative government.* Chicago: Aldine-Atherton.

Noulas, A. G.. & Ketkar, K.W. (1998). Efficient utilization of resources in public schools: A case study of New Jersey. *Applied Economic, 30(10),* 1299- 1306.

Opfer, D., Henry, G. & Mashburn, A. (2008). The district effect: Systemic responses to high stakes accountability policies in six southern states. *American Journal of Education, 114,* 299-332.

Ozcan, Y., Luke, R. & Haksever, C. (1992). Ownership and organizational performance: A comparison of technical efficiency across hospital types. *Medical Care, 30(9),* 781-794.

Parks, R. B. (2005). Technical efficiency of public decision making units. *Policy Studies Journal, 12(2)*, 337 – 346.

Powers, M. (2003). *State cuts would hit area schools.* Southeast Missourian. Cape Girardeau, Missouri, April 22.

Primont, D. F., & Domazlicky, B. (2006). Student achievement and efficiency in Missouri schools and the No Child Left Behind act. *Economics of Education Review, 25(1)*, 77-90.

Ray, S. C. (1991). Resource-use efficiency in public schools: A study of Connecticut data, *Management Science, 37,* pp. 1620-1628.

Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics, 92(5-6),* 1394-1415.

Ruggiero, J. (1996). Efficiency of educational production: An analysis of New York school districts. *Review of Economics and Statistics, 78*, 499-509.

Ruggiero, J., Duncombe, W., & Miner, J. (1995). On the measurement and causes of technical inefficiency in local public services: With an application to public education. *Journal of Public Administration Research and Theory, 5(4),* 403-428.

Russell, J. (2008). *Kindergarten teaching in an era of accountability: How accountability policies shape instruction in an untested grade*. Paper presented at the annual meeting of the American Educational Research Association, March, New York.

Sengupta, J. K. & Sfeir, R. E. (1986). Production frontier estimates of scale in public schools in California. *Economics of Education Review, 5(3),* 297-307.

Springer, M. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review, 27*, 556-563.

Stecher, B. M. (2002). Consequences of large-scale, high stakes testing on school and classroom practice. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.), *Making sense of test-based accountability in education* (pp. 79-100). Santa Monica CA: RAND.