SCHOOL RESPONSES TO HIGH-STAKES

TESTING

by

MISHUK ANWAR CHOWDHURY

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF ARTS IN ECONOMICS

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2006

ACKNOWLEDGEMENTS

I would like to thank Dr. Darren Grant, Dr. Courtney LaFountaion and Dr. Craig Depken for their guidance.

November 27, 2006

ABSTRACT

SCHOOL RESPONSES TO HIGH-STAKES

TESTING

Publication No. ______

Mishuk Anwar Chowdhury, M.A.

The University of Texas at Arlington, 2006

Supervising Professor: Darren Grant

This paper analyzes school's responses to high stakes testing. Using a grade level panel dataset from Texas Assessment of Knowledge and Skills (TAKS) for reading and math tests for 2003 through 2006, to find schools responses to failure. I find that there is a tendency for schools to shift resources from subjects that they pass to subjects that they fail. I classify schools' responses as either substitution responses or scale responses. A school has a substitution response if, when it fails to meet the state's required passing rate threshold for one subject for one cohort of students, it shifts resources to that subject for that cohort of students and away from subjects and cohorts for which its passing rate is above the threshold. Scale responses correspond to increases in resources for all subjects for a cohort of students that fail a subject. I find

evidence for both substitution responses and scale responses, and I find that substitution responses are larger than scale responses.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

CHAPTER 1

INTRODUCTION

President George W. Bush signed the No Child Left Behind Act of 2001(NCLB) in January 2002. NCLB has lead states to develop high-stakes testing that reward or sanction schools based on their performance. Some states have developed stringent accreditation guidelines. For example, Texas has several accreditation ratings for its schools. Rewards based on performance are usually monetary and accompanied by an exemplary accreditation. Sanctions usually lower the accreditation rating of schools, and sustained unacceptable low-accreditation can result in closure of a school or district (Texas Education Code §39.093 and §39.131). In the past, accountability was limited to school district administrators who in turn were accountable to parents. The new era of strong accountability has forced schools to function differently and to develop new strategic ways of meeting state mandated goals.

The introduction of high stakes accountability has changed the way schools behave. Test based accountability works as a diagnostic tool to find which schools or students are performing poorly. The idea is that as the schools become more accountable, students will work harder and schools will target specific grades and students that are struggling. Accountability indirectly motivates parents to get more involved with their child's education. In the end, students will be better educated.

Empirical evidence suggests that accountability is related to performance gains among schools. There has been a significant increase in test scores after the introduction of high stakes testing in most states compared to those states with low stakes testing (Carnoy and Loeb, 2002). However, some have argued that there is a negative side to the story. When a school is pressured to meet the standards, teachers may focus more on test-taking skills rather than general skills in the curriculum. Educators substitute resources away from low-stakes subjects, like science and social studies, to high-stakes subjects like math and reading, and teachers sacrifice teaching general skills and focus more on test-taking abilities of a student (Jacob, 2005).

In this paper, I examine how schools respond to failing to meet standards. I classify schools' responses as either *substitution responses* or *scale responses*. A school has a *substitution response* if, when it fails to meet the state's required passing rate threshold for one subject for one cohort of students, it shifts resources to that subject for that cohort of students and away from subjects and cohorts for which its passing rate is above the threshold. For example, if a school's passing rate is below the required threshold for third grade mathematics in 2003, a substitution response would be to shift resources from fourth grade reading to fourth grade mathematics in 2004. *Scale responses* correspond to increases in resources for all subjects for a cohort of students that fail a subject. For example, if a school's passing rate is below the required threshold for third grade mathematics in 2003, a scale response would be to increase resources for all subjects for fourth graders in 2004.

It is important to distinguish between substitution and scale responses because the two types of responses will have different effects on the level of education that is produced. When schools respond only with scale responses, then the overall level of education increases. However, if schools respond only with substitution responses, then the overall level of education produced falls. Thus, to evaluate the effectiveness of high-stakes accountability, we must first document the relative importance of these two types of responses.

Using data on the Texas Assessment of Knowledge and Skills (TAKS) for 2003 through 2006, I find that that there is a tendency for schools to shift resources to meet academic requirements. I show that when 3^{rd} grade students in a school fail to meet the passing threshold in reading for a particular year, their passing rates in 4^{th} grade mathematics decrease the following year. Similarly, I show that when 3^{rd} grade students in a school fail to meet the passing threshold in mathematics for a particular year, their passing rates in 4^{th} grade reading decline in the following year. I find similar results for 4^{th} and 5^{th} grade reading and mathematics passing rates. These results suggest that substitution responses exist because schools are shifting focus from reading to math when they fail math, and from math to reading when they fail reading.

I also demonstrate that sometimes schools shift their focus from non-marginal students to marginal students. The students that are failing and the students that are passing by a narrow margin are considered marginal students. The students that are comfortably passing are considered non-marginal students. I find that when 4^{th} grade students in a school fail to meet the passing threshold in reading for a particular year, their proficiency rates in 5^{th} grade reading decline in the following year. This result implies that the schools shift their focus from the students who are proficient to students who are failing or barely passing in reading tests.

Overall, I find evidence for both substitution responses and scale responses, and I find that substitution responses are larger than scale responses. When the substitution responses are larger than scale responses, the schools may be actually decreasing the amount of education that they produce. If this is the case, then high-stakes testing will not lead to an increase in the amount of education being produced.

CHAPTER 2

LITERATURE REVIEW

2.1 How is Education Produced?

Education is a product of resources and home inputs. While there is much evidence that home inputs account for a large portion of increase in student achievement, resources in education are also an important determinant of student achievement. In the existing literature we see that resources matter, but only when certain conditions are met. In the absence of these conditions we may not see an effect of resources in student achievement. In the presence of such conditions, resource centered policies will be effective. This would suggest that high-stakes testing is a practical policy that increases student achievement.

Education production functions typically use test scores to measure the quantity of education produced. Test scores are then hypothesized to be a function of school resources and home inputs. The most commonly observed resources used in the education production functions are per-pupil expenditure, teacher education, teacher experience, teacher salary, teacher student ratio, class size. School-specific fixed-effects are sometimes included to account for the unobserved heterogeneity between schools. In reality, all schools are not the same because every school has characteristics, such as teacher quality or community involvement, that cannot be easily quantified.

Studies for which (1) the data are presented in a refereed journal or book, (2) the data originate in schools in the U. S., (3) the outcome measure is some form of academic achievement, (4) the level of aggregation is the level of the school district or a smaller unit, (5) the model controls for socioeconomic characteristics or is longitudinal or quasi-longitudinal (there is a lagged achievement measure), and (6) the data is stochastically independent of other data indicate that a broad range of school inputs are positively related to student outcomes (Greenwald, Hedges, and Laine, 1996). When the data has these properties, the magnitudes of the effects suggest that a moderate increase in resources can increase achievement significantly. Specifically, variables that measure resources directly, such as per-pupil expenditure, show a large and positive effect on student achievement. Smaller schools and smaller class sizes are positively related to student achievement. Also the resource variables that attempt to describe the quality of the teachers such as teacher ability, teacher education, and teacher experience show strong relations with student achievement.

Instructional time can also boost student achievement (Coates, 2003). Coates demonstrates the effect of instructional time as an explanatory variable for average score in the Illinois Goal Assessment Program (IGAP) tests. The study uses IGAP data for years 1994 to 1997. The subjects examined are reading, writing, and mathematics for 2500 schools. The data analyzed is at the school level. It uses minutes of instruction per day in each of four subjects taught in the public elementary schools in Illinois as an input to the production function. The study uses a fixed effects estimator on data over three years to estimate school specific effects. The author also estimates a pooled model by each subject across years. The pooled model shows constant slopes across years but a changing intercept, which indicate a varying level of difficulty on the tests.

The results suggest that average teacher experience is never positive and significant. Class size effects showed that average scores rise as average class size rises but eventually starts declining after a certain class size. After running fixed effects regressions, most teacher characteristic variables were no longer significant. The reason is that aggregated data made the omitted variables bias worse when school-specific effects are not controlled for. After accounting for school-specific effects, the class size variables flipped signs. Instructional time had a positive and significant effect on score, but the coefficients were small in magnitude. So even though some variables were statistically significant their economic significance was questionable.

A broader set of studies suggest that resources may not affect test scores. Out of nearly 377 estimated education production functions from 90 publications, only a small portion all published studies show a positive and statistically significant effects of inputs on student performance (Hanushek, 1997). Overall, Hanushek shows that there is no strong or consistent relationship between resources and performance. Hanushek argues that throwing money at schools only work when those resources are used effectively. He argues that increased funding can even be damaging to student outcomes when used improperly. Thus, education reforms that have resource policies at their core can be ineffective.

One argument for the absence of a relationship between school resources and test scores is that resources boost ability but not test scores (Hanushek, 1997). This would imply that test scores are the wrong measure of the quantity of education produced and we are looking at the wrong dependent variable. Indeed, data on labor market performance indicates that variations in school quality are highly correlated to earning differences among workers (Card and Krueger, 1992). Differences in measurement of student performance could also be the reason why Hanushek (1997) does not find a relationship between resources and achievement. Another explanation is that when parental effects are not controlled for, the estimated effects of resources are overstated (Hanushek, 1997).

After examining Greenwald, Hedges, Laine's and Hanushek's analyses, we may be able to conclude that resources do matter in the production of education. The selection criteria used by Greenwald, Hedges, and Laine ensures that only good quality studies are taken into their analysis. Hanushek includes all existing studies without verifying the integrity of the data used in these studies.

Test scores depend on current and historical home inputs (Todd and Wolpin, 2006). The authors argue that historical inputs explain why the difference between black and white test scores widen with age. They use a cumulative specification that accounts for the cumulative gap in home inputs over the years. This cumulative gap in home inputs causes score gaps between white and minority students to increase over time. The authors argue that if they omitted historical inputs, that would lead to an overstatement of the impact of the current home input. Student fixed-effects were used to analyze how the gap was increasing due to present and past home inputs between black and white students.

The effects of lagged inputs were similar to current inputs. School inputs (per-pupil expenditure, student-teacher-ratio) were only significant when using OLS regressions, not when child-fixed effect estimations were used. The study states that since most of the studies use school level data rather than student level data, those studies underestimate the effects of home inputs.

Even though some studies might show no effect of resources on average student achievement on standardized tests, resources may have differential effects on students in different places in the performance distribution. Specifically, per-pupil expenditure has a significant and large positive effect for lower part of the conditional distribution (Eide, 1998). In addition, the top half of the conditional distribution benefits from an extended school year, test scores go up for all students except for the top students when enrollment increases, and the effects of pupil-teacher ratio and teacher ratio are insignificant (Eide, 1998).

Uncontrollable factors, such as school demographics, may play a larger role on student passing rates than resources (Hoerandner and Lemke 2006). Their study analyzes how the worst performing schools will do if they mimic the behavior of better performing schools. The paper argues that there are 3 types of factors contributing to passing rates. Their results suggest the following:

1) Resources explain about 10-25% of the variation in passing rates,

2) Uncontrollable factors such as demographics and location explain about 30-50% of the variation in passing rates, and

3) Regression error accounts for about 40% of the variation in passing rates.

Hoerandner and Lemke (2006) use gaps within the same demographic group to find whether they are narrowing over time. They predict a hypothetical passing rate for the worst performing schools if they mimicked the behavior of the best performing schools.

The hypothetical pass rates fall well below the actual pass rates of the best schools. Hence, most of the gap can be attributed to the uncontrollable factors.

2.2 Testing in Education

Testing can lead to a change in schools' behavior. The change in school behavior may or may not increase the amount of education being produced. There is at least some evidence that education is a function of school resources. However, for whatever reason, some schools may not be allocating resources optimally. The introduction of high stakes testing is intended to push schools to the optimal point where the effect of resources is maximized. This will increase education for the given resource level. However, the opposite can also be true: A school may have to allocate resources sub-optimally in order to meet testing standards. This would decrease the amount of education for the given resource level. So, the effect of testing can be ambiguous.

Empirical evidence suggests that testing or accountability has a large impact on test score/passing rate gains. Carnoy and Loeb (2002) analyze how score gains relate to accountability. They develop an accountability index based on how schools are rewarded or sanctioned based on their performance. Data on National Association of Educational Progress (NAEP) math scores suggest that the stronger the accountability measures, the bigger the improvement in scores. Improvements among minority students are greater compared to white students with high accountability. Finally, states with higher percentage of minority students are more likely to choose a high accountability testing scheme.

The introduction of testing may not result in the same behavior from different schools. The burden of eliminating gaps on passing rates between different demographic groups is not the same for different schools (Hoerandner and Lemke 2006). Their study concludes that the burden of eliminating gaps on passing rates is more on the worst performing schools compared to better performing schools. Therefore, the worst performing schools may have to put extra effort to match the passing rates of better performing schools.

Jacob (2005) uses data from Chicago public schools to demonstrate the effects of strong accountability. He uses longitudinal student-level data from Chicago to examine any gains in scores after the introduction of a policy and compares them with district level data from other large mid-western cities that did not have strong accountability.

His primary findings show that mathematics and reading scores increased sharply after the introduction of the testing program. He also showed that low-achieving schools see larger gains than other schools. The achievement gains could be attributed to an increase in test-taking skills and student effort. The study also concluded that educators substitute away from low-stakes subjects like science and social studies to high-stakes subjects like math and reading.

In conclusion, we learn from the existing literature that:

1) The effects of testing on the production of education is not well understood and theoretically ambiguous. This has motivated my thesis.

2) Resources matter in education production. Therefore resource measures should be included in the analysis.

3) Student characteristics also matter in education production. Therefore, student characteristics measures should be controlled for in any analysis.

4) School characteristics are also important in education production. Therefore, school characteristics should be controlled for in the analysis too.

CHAPTER 3

THEORY

3.1 A Simple Conceptual Model of Education Production

I begin with the assumption that schools act to maximize education production subject to their budget constraint. Education is a product of resources, school attributes, student attributes, and general community factors. Generally, the education production function is a mapping between school inputs and student inputs and the output measure, typically some sort of achievement measure such as the test scores of a student or passing rates of a particular school.

Figure 3.1 illustrates the intuition on how schools respond to failing to meet an exogenously determined standard. The scale responses occur when the budget constraint is shifted outwards. The substitution effect occurs when a school is forced to move along a budget constraint. Note that this is different from the substitution effect that occurs when the consumer moves along an isoquant when price of goods change.

When a certain school fails to meet the state mandated standards they have two options. The first option for the school is to acquire more resources and devote those resources to boost the passing rate. But, if the school districts are budget constrained this is not always feasible. In this case, the school would have to reallocate its limited resources to increase passing rates. This would require them to substitute away from teaching low-stakes subjects, or to substitute away from teaching general skills towards more test-taking skills. So, reallocating the resources so that those inputs produce the maximum passing rate is the substitution response in education production.

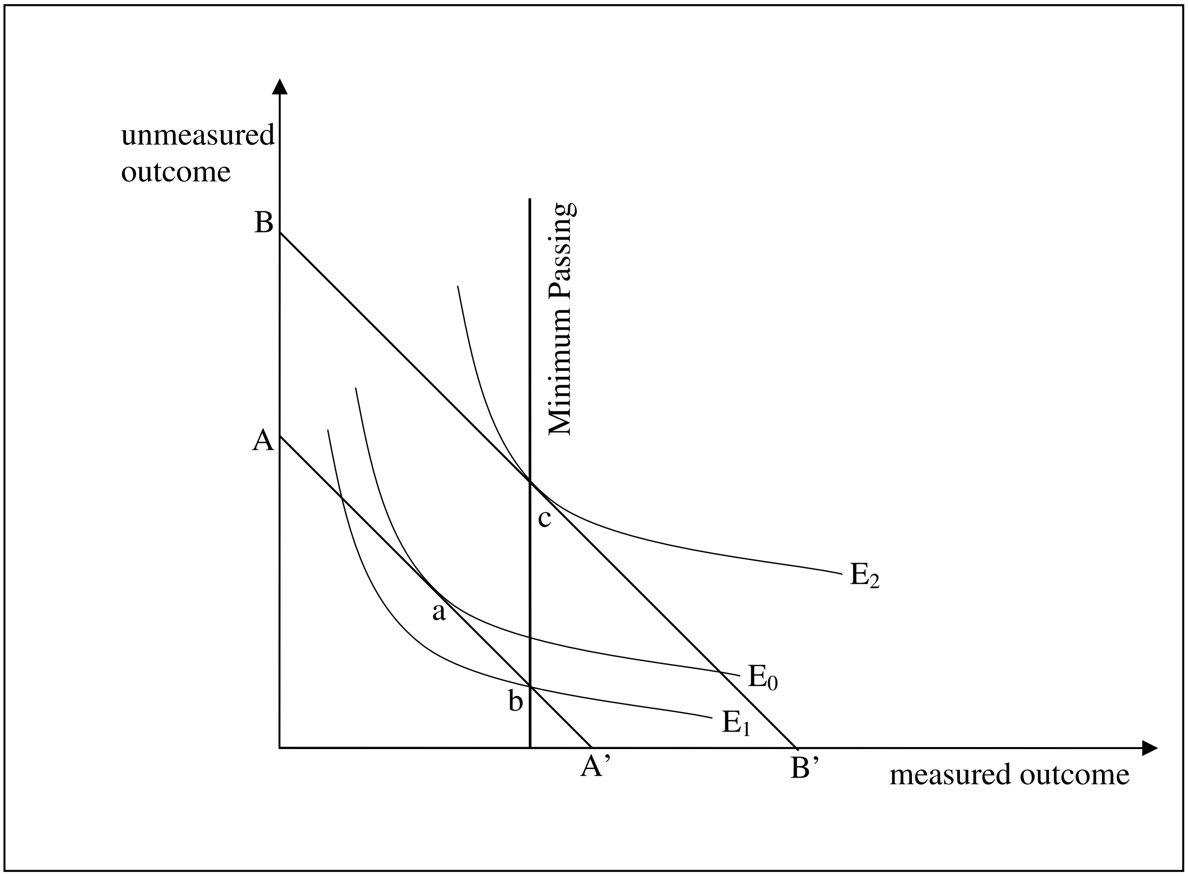

Figure 3.1 Substitution and Scale responses: pure substitution vs. pure scale

Initially the school is at a point like 'a'. At this point the school is constrained by the budget constrain AA'. From the graph above, we can think of education as a consumption good that is a product of resources dedicated towards TAKS and resources dedicated towards other general education. At this point, the school is producing education level E_0. Now, let's imagine that the state requires the school to meet the minimum passing rate in the TAKS. The minimum passing rate is shown by the vertical line in the graph. Given this condition, the school has to move to a point left of the minimum passing line. There are two ways the school can do this: they can either substitute (move to a point like ‘b’) or scale up the inputs (move to a point like ‘c’).

If the school cannot acquire any new funds, it has to move to point ‘b’ by moving along its budget line. In this case, the school moves from point ‘a’ to point ‘b’ purely by substituting. By doing this, even though they meet the minimum passing threshold, they are now achieving education level E_1. From the graph we can see that $E_0>E_1$. So, by substituting away from “other resources” the school produces a lower amount of education.

If the school can acquire funds, they can move to a point like ‘c’ by relaxing its budget constraint. In this case, the school moves from point ‘a’ to point ‘c’ purely by increasing all of its resource inputs. The school will meet the passing threshold and at the same time achieve a higher level of education E_2. We can see that $E_2>E_0$. So, the school can produce a higher amount of education by scaling its inputs.

In conclusion to these scenarios, I observe that *scale responses* lead to greater production of education while *substitution responses* lead to lower production of education.

In reality pure substitution or pure scaling is almost never likely. In almost any situation a school will choose a combination of scaling and substitution. Figure 3.2 and Figure 3.3 demonstrates the effect of scaling and substitution responses combined.

We may observe a net negative or a net positive change in the unmeasured outcome depending on whether *substitution* or *scale* responses dominate.

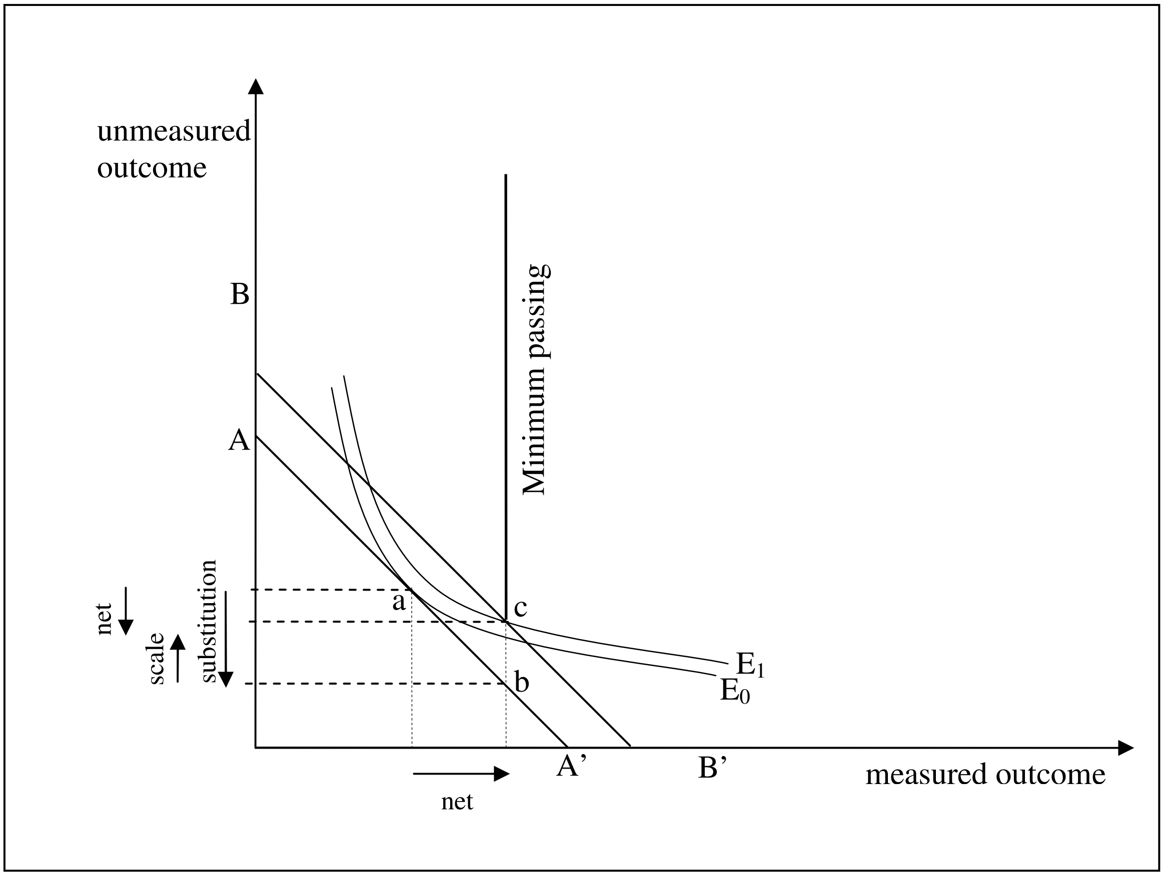

Figure 3.2 Substitution and Scale responses: combination of substitution and scale responses - Net negative effect

The substitution response is likely to dominate as long as the measured and the unmeasured outcomes are substitutes. Figure 3.2 illustrates this scenario. As a result of the substitution response, the school moves from point 'a' to point 'b'. The measured outcome increases but, the unmeasured outcome falls as a result of moving from point 'a' to 'b'. Then the school wishes to bring the unmeasured outcome back to its original level. To achieve this, the school uses a scale response. In the figure this scale response can be seen as a movement from point 'b' to point 'c'. As a result of moving from point 'b' to 'c' the measured outcome stays constant but the unmeasured outcome increases. But, in this case, since the scale response is smaller than the substitution response we see a net negative impact in the unmeasured outcome.

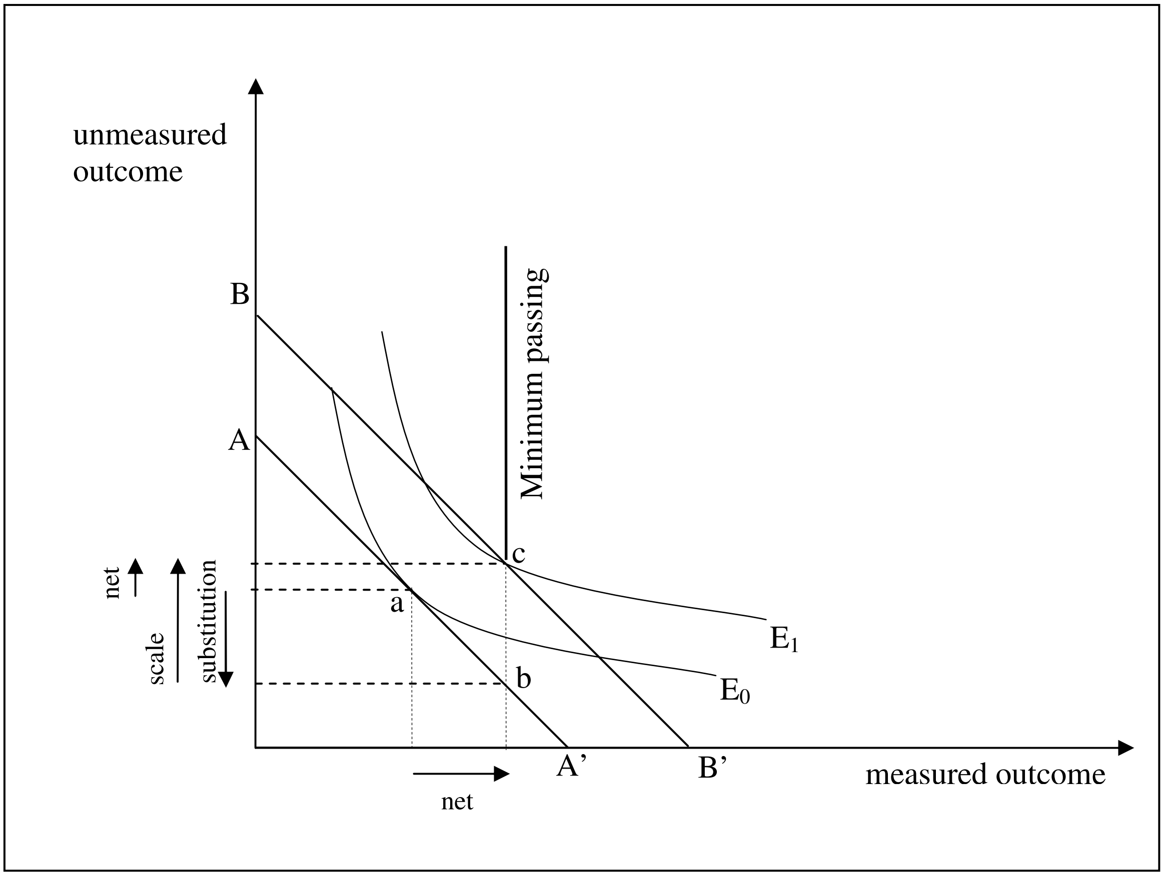

Figure 3.3 Substitution and Scale responses: combination of substitution and scale responses - Net positive effect

The scale response is likely to dominate as long as the measured and the unmeasured outcomes are complements. Figure 3.3 illustrates this scenario. A movement from point 'a' to point 'b' characterizes the substitution response. After the substitution, the measured outcome increases and the unmeasured outcome decreases. Now, the school wishes to increase its unmeasured outcome. If the school responds with a sufficiently large scale response, it will move from point 'b' to point 'c'. Thus the scale response in the figure pushes the unmeasured outcome up and beyond the initial unmeasured outcome. Therefore, we see a net positive change in unmeasured outcome.

From these two figures we can see that the net effect on the unmeasured outcome will depend on the magnitude of the *substitution* and the *scale response*. If the *scale* *response* is larger we will see a decrease in the unmeasured outcome, and if the *substitution response* is larger we will see an increase in the unmeasured outcome.

In both cases, the school has to move from point ‘a’ to point ‘c’ and the school produces education level $E_1 > E_0$. Even though the education produced here is greater than E_0, this is not the optimal choice. The optimal choice would be a point where the isoquant of the education production function is tangent to the budget constraint.

3.2 Hypotheses

The theoretical model that I discussed earlier in this chapter allows us to estimate the relationship between various inputs and student outcomes. I use these relationships to answer how schools respond to failure.

There are two sets of fundamental questions that I attempt to answer. The first set of questions is regarding the education production function itself:

First, how much of the variation in education production can be attributed to resource inputs? Answering this question is important because it either confirms or rejects the validity of resource centered education policies. I hypothesize that even though there is a high amount of variation that can be attributed to the uncontrollable factors, such as demographics, some resources will have a significant effect. If this hypothesis is correct, we will observe that a high portion of the variance in education production is due to school specific fixed-effects.

Second, how much do resources impact passing rates? Is the relationship between resources and passing rates weak or strong? I hypothesize that even though we will observe a positive and significant effect, the magnitude of the effects will not be sufficiently large. The reason behind this hypothesis is that, even though throwing money at schools may help the students somewhat, it will not be able offset the lack of home inputs or innate ability. Again, the argument is that certain uncontrollable factors contribute a lot more than resources for student outcomes.

Third, does being economically disadvantaged have a negative effect on student outcomes? I expect that economically disadvantaged students do worse on tests than do students who are not economically disadvantaged. The reasoning behind this expectation is that poor students do not get as much parental input as their peers do. This gap in home input may drive their passing rates down.

The second set of questions that I attempt to answer is regarding school responses to failure. I formulate the following expectations about scale and *substitution responses*. If schools choose *scale responses* then

- Schools pay teachers more,
- Schools hire more experienced teachers,
- Schools reduce class size, and/or
- Schools increase per-student expenditure.

If schools choose *substitution responses*, then

- Schools substitute away from teaching non-TAKS subjects to TAKS subjects,
- Schools focus more on test-taking skills than the regular curriculum,
- Schools shift resources from non-marginal students to marginal students,
- Schools focus more on a particular grade that has been performing poorly and trains the teachers for that grade, and/or
- Schools focus more on a particular cohort that has been performing poorly and trains that particular group of students.

This motivates a general econometric model. To measure the direct outcomes of a cohort failing the same subject last year I use the following model:

$$Outcome_{r,g,t} = f\,(outcome_{r,g\text{-}1,t\text{-}1},\ socioeconomicstatus_{g,t},\ TAKSfailure_{r,g\text{-}1,t\text{-}1}) \qquad (1a)$$

$$Outcome_{m,g,t} = f\,(outcome_{m,g\text{-}1,t\text{-}1},\ socioeconomicstatus_{g,t},\ TAKSfailure_{m,g\text{-}1,t\text{-}1}) \qquad (1b)$$

For equation 1a, $outcome_{r,g,t}$ is the passing rate for subject r for grade g for year t. $outcome_{r,g-1,t-1}$ is passing rate for subject r for grade $g-1$ for year $t-1$. This represents the same cohort's passing rate in the previous year, previous grade. $socioeconomicstatus_{g,t}$ is the share of economically disadvantaged students for grade g in year t. $TAKSfailure_{r,g-1,t-1}$ is a dummy variable for whether a cohort failed to meet TAKS standards in subject r for grade $g-1$ for year $t-1$. This dummy variable is 1 when the cohort fails the grade. If the coefficient estimate for $TAKSfailure_{r,g-1,t-1}$ is positive, than failing TAKS the previous year improves TAKS passing rates on the same subject, next year, next grade. If the coefficient estimate for $TAKSfailure_{r,g-1,t-1}$ is negative, than failing TAKS the previous year lowers TAKS passing rates on the same subject, next year, next grade.

Equation 1b is similar to equation 1a, only equation 1b is for subject m instead of subject r.

The theoretical model generates the following hypotheses for direct outcomes:

H_0: Failing TAKS on the same subject, previous grade, and previous year does not affect passing rates on that subject for the same group of students the following year

H_α: Failing TAKS on the same subject, previous grade, previous year increases passing rates on that subject for the same group of students the following year

To measure the scale and substitution responses, I use the following econometric model

$$Outcome_{r,g,t} = f\,(outcome_{r,g\text{-}1,t\text{-}1},\ socioeconomicstatus_{g,t},\ TAKSfailure_{m,g\text{-}1,t\text{-}1}) \qquad (2a)$$

$$Outcome_{m,g,t} = f\,(outcome_{m,g\text{-}1,t\text{-}1},\ socioeconomicstatus_{g,t},\ TAKSfailure_{r,g\text{-}1,t\text{-}1}) \qquad (2b)$$

For equation 2a, $outcome_{r,g,t}$ is the passing rate for subject r for grade g for year t. $outcome_{r,g\text{-}1,t\text{-}1}$ is passing rate for subject r for grade $g\text{-}1$ for year $t\text{-}1$. This represents the same cohort's passing rate in the previous year, previous grade. $socioeconomicstatus_{g,t}$ is the share of economically disadvantaged students for grade g in year t. $TAKSfailure_{m,g\text{-}1,t\text{-}1}$ is a dummy variable for whether a cohort failed to meet TAKS standards in subject m (different subject) for grade $g\text{-}1$ for year $t\text{-}1$. This dummy variable is 1 when the cohort fails to meet the standards. If the coefficient estimate for $TAKSfailure_{m,g\text{-}1,t\text{-}1}$ is negative, than failing TAKS the previous year on subject m lowers TAKS passing rates on subject r, for year $t+1$, grade $g+1$. This would suggest substitution from subject r to subject m. If the coefficient estimate for $TAKSfailure_{m,g\text{-}1,t\text{-}1}$ is positive, than failing TAKS the previous year lowers TAKS passing rates on subject r, year $t+1$, grade $g+1$.

Equation 2b is similar to equation 2a, only equation 2b is for subject m instead of subject r.

The hypothesis for scale and substitution responses:

H_0: Failing TAKS on a different subject, previous grade, and previous year does not affect passing rates on the target subject for the same group of students the following year

H_α: Failing TAKS on a different subject, previous grade, and previous year decreases passing rates on the target subject for the same group of students the following year

Next, to measure the substitution responses, I use the following econometric model:

$Outcome_{r,g,t} =$

$$f\,(outcome_{r,g\text{-}1,t\text{-}1},\ resourceinputs_t,\ socioeconomicstatus_{g,t},\ TAKSfailure_{m,g\text{-}1,t\text{-}1}) \qquad (3a)$$

$Outcome_{m,g,t} =$

$$f\,(outcome_{m,g\text{-}1,t\text{-}1},\ resourceinputs_t,\ socioeconomicstatus_{g,t},\ TAKSfailure_{r,g\text{-}1,t\text{-}1}) \qquad (3b)$$

For equation 3a, $outcome_{r,g,t}$ is the passing rate for subject *r* for grade *g* for year *t*. $outcome_{r,g\text{-}1,t\text{-}1}$ is passing rate for subject *r* for grade *g-1* for year *t-1*. This represents the same cohort's passing rate in the previous year, previous grade. $resourceinputs_t$ is the schools resource inputs for year *t*. $socioeconomicstatus_{g,t}$ is the share of economically disadvantaged students for grade *g* in year *t*. $TAKSfailure_{m,g\text{-}1,t\text{-}1}$ is a dummy variable for whether a cohort failed to meet TAKS standards in subject *m* (different subject) for grade *g-1* for year *t-1*. This dummy variable is 1 when the cohort fails to meet the standards. If the coefficient estimate for $TAKSfailure_{m,g\text{-}1,t\text{-}1}$ is negative, than failing TAKS the previous year on subject *m* lowers TAKS passing rates on subject *r*, for year *t+1*, grade *g+1*. This would suggest substitution from subject *r* to subject *m*. If the coefficient estimate for $TAKSfailure_{m,g\text{-}1,t\text{-}1}$ is positive, than failing TAKS the previous year lowers TAKS passing rates on subject *r*, year *t+1*, grade *g+1*.

Equation 3b is similar to equation 3a, only equation 3b is for subject *m* instead of subject *r*.

Hypothesis for substitution responses:

H_0: Failing TAKS on a different subject, previous grade, and previous year does not affect passing rates on the target subject for the same group of students the following year

H_α: Failing TAKS on a different subject, previous grade, and previous year decreases passing rates on the target subject for the same group of students the following year

Next, to measure the substitution responses for proficiency vs. passing rates, I use the following econometric model:

$$Proficiency_{r,g,t} = f\,(Proficiency_{r,g\text{-}1,t\text{-}1}\;\; socioeconomicstatus_{g,t},\ TAKSfailure_{r,g\text{-}1,t\text{-}1}) \qquad (4a)$$

$$Proficiency_{m,g,t} = f\,(Proficiency_{m,g\text{-}1,t\text{-}1},\ socioeconomicstatus_{g,t},\ TAKSfailure_{m,g\text{-}1,t\text{-}1}) \qquad (4b)$$

For equation *4a*, $proficiency_{r,g,t}$ is the proficiency rate for subject *r* for grade *g* for year *t*. $proficiency_{r,g\text{-}1,t\text{-}1}$ is the proficiency rate for subject *r* for grade *g-1* for year *t-1*. This represents the same cohort's passing rate in the previous year, previous grade. $resourceinputs_t$ is the schools resource inputs for year *t*. $socioeconomicstatus_{g,t}$ is the share of economically disadvantaged students for grade *g* in year *t*. $TAKSfailure_{r,g\text{-}1,t\text{-}1}$ is a dummy variable for whether a cohort failed to meet TAKS standards in subject *r* (same subject) for grade *g-1* for year *t-1*. This dummy variable is 1 when the cohort fails to meet the standards. If the coefficient estimate for $TAKSfailure_{r,g\text{-}1,t\text{-}1}$ is negative, than failing TAKS the previous year on subject *r* lowers TAKS proficiency rates on subject *r*, for year *t*, grade *g*. This would suggest substitution from non-marginal students to marginal students for subject *r*. If the coefficient estimate for $TAKSfailure_{r,g\text{-}1,t\text{-}1}$ is positive, than failing TAKS the previous year lowers TAKS passing rates on subject *r*, year *t+1*, grade *g+1*.

Equation *4b* is similar to equation *4a*, only equation *4b* is for subject *m* instead of subject *r*.

Hypothesis for substitution between marginal and non-marginal students:

H_0: Failing TAKS on a subject, previous grade, previous year does not affect proficiency rates on that subject for the same group of students the following year

H_α: Failing TAKS on a subject, previous grade, previous year decreases proficiency rates on that subject for the same group of students the following year

Using econometric notation the same specifications can be written as:

1a) $Y_{i,r,g,t} = \alpha Y_{i,r,g-1,t-1} + \gamma Z_{i,g,t} + \delta D_{r,g-1,t-1} + u_i + \varepsilon_{i,r,g,t},$

1b) $Y_{i,m,g,t} = \alpha Y_{i,m,g-1,t-1} + \gamma Z_{i,g,t} + \delta D_{m,g-1,t-1} + u_i + \varepsilon_{i,m,g,t},$

2a) $Y_{i,r,g,t} = \alpha Y_{i,r,g-1,t-1} + \gamma Z_{i,g,t} + \delta D_{m,g-1,t-1} + u_i + \varepsilon_{i,r,g,t},$

2b) $Y_{i,m,g,t} = \alpha Y_{i,m,g-1,t-1} + \gamma Z_{i,g,t} + \delta D_{r,g-1,t-1} + u_i + \varepsilon_{i,m,g,t},$

3a) $Y_{i,r,g,t} = \alpha Y_{i,r,g-1,t-1} + \beta X_{i,t} + \gamma Z_{i,g,t} + \delta D_{m,g-1,t-1} + u_i + \varepsilon_{i,r,g,t},$

3b) $Y_{i,m,g,t} = \alpha Y_{i,m,g-1,t-1} + \beta X_{i,t} + \gamma Z_{i,g,t} + \delta D_{r,g-1,t-1} + u_i + \varepsilon_{i,m,g,t},$

4a) $Y_{i,r,g,t} = \alpha Y_{i,r,g-1,t-1} + \gamma Z_{i,g,t} + \delta D_{r,g-1,t-1} + u_i + \varepsilon_{i,r,g,t},$

4b) $Y_{i,m,g,t} = \alpha Y_{i,m,g-1,t-1} + \gamma Z_{i,g,t} + \delta D_{m,g-1,t-1} + u_i + \varepsilon_{i,m,g,t},$

where $Y_{i,r,g,t}$ is the TAKS passing rate for school i in year t, for subject r, for grade g, $Y_{i,r,g-1,t-1}$ TAKS passing rate for school i in year $t-1$, for subject r, for grade $g-1$, X is the vector of school-specific resource variables that change over time, including per-pupil expenditure, average teacher salaries, and average teacher experience and student-teacher ratio in that school, Z is the percent of students that are economically disadvantaged, D is the dummy variable that is 1 if the same cohort failed the same or different subject in the previous year, u_i is the school fixed effect, and $\varepsilon_{i,r,g,t}$ is the error term associated with school i in year t, for subject r, for grade g. α, β, γ and δ are coefficient estimates for Y, X, Z and D respectively.

CHAPTER 4

DATA

4.1 Data Sources

I used data from the Texas Assessment of Knowledge and Skills (TAKS) test. The TAKS was first implemented in spring 2003. All students are tested reading in grades 3, 4, 5, 6, 7, 8, & 9; English language arts (ELA) is tested in grades 10 & 11. Writing is tested in grades 4 & 7. Social Studies is tested in grades 8, 10, & 11. Mathematics is tested in grades 3, 4, 5, 6, 7, 8, 9, 10 & 11. Science is tested in grades 5, 10, & 11.

After the TAKS is administered, each student is evaluated to see whether they have met minimum passing standard and whether they have achieved the commended performance level. TAKS Commended Performance is the highest performance level set by the State Board of Education on the TAKS. According to the Texas Education Agency (TEA), "students who achieve Commended Performance have performed at a level that is considerably above the state passing standard and have shown a thorough understanding of the knowledge and skills at the grade level tested. (http://www.tea.state.tx.us/)"

After the students have been evaluated, the state looks at each school's performance by looking at that school's passing and commended rates. Schools are accountable for passing in all of these tests. The Commissioner of Education and the 15 elected members of the State Board of Education (SBOE) oversee the public education system of Texas in accordance with the Texas Education Code. They proposed three different accreditation ratings for the TAKS test. The Academically Acceptable standard varies by subject, while the Recognized and Exemplary standards are the same for all subjects:

- Exemplary – At least 90% of students tested passing for every subject.
- Recognized – At least 70% of students tested passing for every subject.
- Academically Acceptable / Passing – Varies by subject:
  - Reading/ELA – At least 50% of students tested passing.
  - Writing – At least 50% of students tested passing.
  - Social Studies – At least 50% of students tested passing.
  - Mathematics – At least 35% of students tested passing.
  - Science – At least 25% of students tested passing.

Criticisms for these standards do exist. Under the NCLB states are left to establish their own standards, assessments, passing thresholds and proficiency thresholds. This leads states to lower passing standards to inflate reported performance to the federal government. The states with the most challenging passing thresholds are often accused of having the most schools in need of improvement. The absence of one common passing and proficiency threshold makes comparison between states difficult. According to a scale for comparing how challenging the passing standards are, Texas is one of the states with the least challenging passing standards (Peterson & Hess, 2005).

School staff and resource data were also available from the Texas Education Agency's "Academic Excellence Indicator System".

4.2 Data Description

I use performance on the Texas Assessment of Knowledge and Skills (TAKS) as my measure of education production. The TAKS passing rate is my primary outcome measure. The passing rate is calculated by taking the total number of students that passed the test and dividing it by the total number of students that took the test, and then multiplying it by 100. The percent of students who met the commended performance is also available in the dataset. The commended rate is calculated by taking the total number of students that achieved a commended performance in a test and dividing it by the total number of students that took that test. Then it is multiplied by 100.

The achievement measures -- the TAKS passing rates -- were available at the grade level. The share of economically disadvantaged students was also available at the grade level. It is calculated by taking the number of students who are economically disadvantaged for a particular grade then dividing that number by the total number of students in that particular grade. Then it is multiplied by 100 to get it into a percentage format. I only used campus level data here because higher levels of aggregation lead to biased results (Hanushek, 1997). This dataset contains passing rates for most public schools in Texas. The sample contains more than 90,000 observations from over 3,000 schools. These schools are sampled over a four year period from 2003 to 2006. For each of these years passing rates are reported for each grade from grade 3 to grade 11 on each subject administered in the TAKS test. One of the key variables used in my analysis is the percent of economically disadvantaged student in a particular grade. I later incorporate this into my education production function. I look at mostly mathematics and reading test passing rates and proficiency rates.

I use average teacher salary, average teacher experience, per pupil expenditure, and student-teacher ratio to measure school resources. These variables were only available at the school level. Staff or resource variables were available from the "Academic Excellence Indicator System" in the TEA website. The main variables used in my analysis were teacher experience, teacher salary, student/teacher ratio and Per-Pupil Expenditure. The data was available for 3,000 schools for the years 2003, 2004 and 2005. Teacher experience is calculated by taking the total years of experience for all teachers in a school and dividing it by the number of teachers in that school. Average teacher salary is calculated by first finding the total salary of all teachers in a school then dividing it by the number of teacher in that school. Student/teacher ratio is calculated by dividing the number of students in a school by the number of teachers in a school. Per-pupil expenditure is calculated by first summing up all the expenditure in a school and then dividing it by the number of student in that school.

Table 4.1 Descriptive Statistics of the Data

Variable	Obs.	Mean	Std. Dev.	Min	Max
Passing Rate - Reading	28896	82.273	12.823	0.000	100.000
Lagged Passing Rate - Reading	20536	86.690	11.683	0.000	100.000
Passing Rate - Math	28896	84.231	13.158	0.000	100.000
Lagged Passing Rate - Math	20540	84.739	13.299	0.000	100.000
% commended in reading	28896	20.203	12.674	0.000	84.000
% commended in mathematics	28896	24.292	15.983	0.000	100.000
% of students economically disadvantaged	28896	57.531	28.920	0.000	100.000
Student/Teacher ratio	21472	14.741	2.350	2.600	40.311
Average Teacher Experience (years)	21476	11.366	3.159	0.000	24.435
Average Teacher Salary (1000s)	21476	39.706	4.003	0.363	72.393
Per-pupil Expenditure (1000s)	21186	5.105	1.372	0.105	70.796
Lagged 4th grade reading failure (dummy)	14965	0.001	0.036	0.000	1.000
Lagged 4th grade math failure (dummy)	14965	0.003	0.059	0.000	1.000
Lagged 5th grade reading failure (dummy)	13931	0.013	0.112	0.000	1.000
Lagged 5th grade math failure (dummy)	13931	0.006	0.079	0.000	1.000

4.3 Evolution of Passing Rates within schools: Description of the Data Generating Process

I first estimate the data generating process to see how resources impact achievement in Texas public schools. My analysis looks at the direction and magnitude of resource coefficients to determine what inputs play larger roles in the education production function. My production function can be used to predict passing rates into the future. I then use the data generating process to see if schools behave strategically after failing to pass the TAKS. Specifically, I look for evidence of *scale responses* and *substitution responses*.

This specification relates my achievement outcome (passing rates) to various school inputs, the percentage of students that are economically disadvantaged, and a lagged achievement measure. The lagged achievement measure in my case is last years passing rate on the same test. This lagged achievement measure works as a proxy for missing historical inputs (i.e. time spent by parents helping their child to read).

To estimate the data generating process, I use the following econometric model:

Education = f(average teacher salary, average teacher experience, per-student expenditure, student-teacher ratio, lagged passing rate for the same cohort of students, percentage of students that are economically disadvantaged)

More specifically my econometric model of the data generating process can be represented as:

$$Y_{i,r,g,t} = \alpha Y_{i,r,g-1,t-1} + \beta X_{i,t} + \gamma Z_{i,g,t} + u_i + \varepsilon_{i,r,g,t},$$

where $Y_{i,r,g,t}$ is the TAKS passing rate for school i in year t, for subject r, for grade g, $Y_{i,r,g-1,t-1}$ is the TAKS passing rate for school i in year $t-1$, for subject r, for grade $g-1$, X is the vector of school-specific resource variables that change over time, including per-pupil expenditure, average teacher salaries, and average teacher experience and student-teacher ratio in that school, Z is the percent of students that are economically disadvantaged, u_i is the school fixed effect, and $\varepsilon_{i,r,g,t}$ is the error term associated with school i in year t, for subject r, for grade g.

I crunchatized passing rates of grade 4 & 5 reading and math tests. We focus on grades 4 and 5 because students in elementary school maybe more responsive to changes in the way they are taught. Therefore, the passing rates of these grades are better measures for analyzing the impacts of school behavior than are passing rates from middle or high schools.

Table 4.2 Data Generating Process with Lagged Cohort Passing Rates

	(1)	(3)	(2)	(4)
	Passing rate – Reading - Grade 4	Passing rate - Reading - Grade 5	Passing rate - Math - Grade 4	Passing rate - Math - Grade 5
Lagged cohort passing rate – Reading (percentages)	0.123* (0.022)	0.307* (0.018)		
Lagged cohort passing rate – Math (percentages)			0.215* (0.021)	0.262* (0.018)
Average teacher Experience (years)	-0.009 (0.152)	-0.180 (0.147)	-0.188 (0.170)	-0.341** (0.157)
Student/teacher ratio	-0.535* (0.139)	-0.294** (0.140)	-0.626* (0.156)	-0.512* (0.151)
Average teacher salary ($1000s)	0.318*** (0.165)	0.089 (0.134)	0.529* (0.185)	0.413* (0.144)

Table 4.2 - Continued

Per-pupil expenditure ($1000s)	0.379* (0.133)	0.476** (0.240)	0.226 (0.149)	0.220 (0.257)
% of students economically disadvantaged	-0.089* (0.018)	-0.099* (0.019)	-0.049** (0.020)	-0.072* (0.020)
Year 2004 (dummy variable)	6.871* (0.225)	-6.882* (0.233)	5.595* (0.248)	-6.792* (0.252)
Constant	65.408* (6.380)	65.390* (5.398)	52.127* (7.062)	63.084* (5.730)
Observations	7093	6380	7093	6382
Number of campuses	3617	3276	3618	3277
R-squared (Within)	0.32	0.36	0.21	0.32

Standard errors in parentheses
*** significant at 10%; ** significant at 5%; * significant at 1%

Table 4.2 shows the estimated data-generating process. Regression (1) from table 4.2 shows the data generating process for grade 4 reading test. I see that, 3rd grade reading passing rate for the same cohort has a positive and significant effect on 4th grade reading passing rate. The coefficient indicates that when last years passing rates on the same subject by that same cohort increase by 1 percentage point, that subject's passing rate this year increases by .12 percentage points. Average teacher experience is negative and insignificant. Student-teacher ratio is negative and significant. This coefficient indicates that when the student/teacher ratio increases by 1, resulting in bigger class size, the passing rate decreases by .5 percentage points. Average teacher salary is positive and significant at the 10% level. The coefficient implies that, increasing average teacher salary by 1000 dollars, increases passing rate by .31 percentage points. The coefficient on share of students economically disadvantaged is negative and significant. This coefficient indicates that for every additional percentage point of economically disadvantaged student, passing rate goes down by .08 percentage points. The dummy variable that is 1 when the year is 2004 is also positive and significant. This variable implies that passing rates were 6.8 percentage points higher in year 2004 compared to year 2005.

Regression (2) from table 4.2 shows the data generating process for grade 5 reading test. I see that, 4th grade reading passing rates for the same cohort has a positive and significant effect on 5th grade reading passing rates. The coefficient indicates that when last years passing rates on the same subject by that same cohort increase by 1 percentage point, the passing rate this year on that same subject increases by .30 percentage points. Average teacher experience is negative and insignificant. Student-teacher ratio is negative and significant at the 5% level. This coefficient indicates that when the student/teacher ratio increases by 1, resulting in bigger class size, the passing rate decreases by .29 percentage points. Average teacher salary is positive but insignificant in this case. The coefficient on share of students economically disadvantaged is negative and significant. This coefficient indicates that for every additional percentage point of economically disadvantaged student, passing rate goes down by .10 percentage points. The dummy variable that is 1 when the year is 2004 is negative and significant. This variable implies that passing rates were 6.8 percentage points lower in year 2004 compared to year 2005.

Regression (3) from table 4.2 shows the data generating process for grade 4 math test. I see that, 3rd grade math passing rate for the same cohort has a positive and significant effect on 4th grade math passing rate. The coefficient indicates that when last years passing rates on the same subject by that same cohort increase by 1 percentage point, the passing rate this year for that subject increases by .21 percentage points.

Average teacher experience is negative and insignificant. Student-teacher ratio is negative and significant. This coefficient indicates that when the student/teacher ratio increases by 1, resulting in a bigger class size, the passing rate decreases by .62 percentage points. Average teacher salary is positive and significant at the 1% level. The coefficient implies that, increasing average teacher salary by 1000 dollars, increases passing rate by .53 percentage points. Per-pupil expenditure is positive and insignificant. The coefficient on share of students economically disadvantaged is negative and significant at the 5% level. This coefficient indicates that for every additional percentage point of economically disadvantaged student, passing rate goes down by .05 percentage points. The dummy variable that is 1 when the year is 2004 is also positive and significant. This variable implies that passing rates were 5.6 percentage points higher in year 2004 compared to year 2005.

Regression (4) from table 4.2 shows the data generating process for grade 5 math test. I see that, 4^{th} grade math passing rates for the same cohort has a positive and significant effect on 5^{th} grade math passing rates. The coefficient indicates that when last years passing rates on the same subject by that same cohort increase by 1 percentage point, the passing rate for the same subject this year increases by .26 percentage points. Average teacher experience is negative and significant at the 10% level. The result is counter intuitive and shows that as average teacher experience increases by 1 year, math passing rate in grade 5 decreases by .34 percentage points. Student-teacher ratio is negative and significant at the 1% level. This coefficient indicates that when the student/teacher ratio increases by 1, resulting in a bigger class size, the passing rate decreases by .51 percentage points. Average teacher salary is positive and significant.

The coefficient implies that, increasing average teacher salary by 1000 dollars, increases passing rate by .41 percentage points. Per-pupil expenditure is positive and insignificant. The coefficient on share of students economically disadvantaged is negative and significant. This coefficient indicates that for every additional percentage point of economically disadvantaged student, passing rate goes down by .07 percentage points. The dummy variable that is 1 when the year is 2004 is negative and significant. This variable implies that passing rates were 6.7 percentage points lower in year 2004 compared to year 2005.

In summary, I found the share of economically disadvantaged students to be consistently negatively correlated with passing rates both in pooled and fixed effects estimation. Last year's passing rates of the same cohort (different grade) also seem to have an important effect. Last year's cohort scores are used as a proxy for home inputs for the students. The same cohort's performance last year has a positive effect on passing rates. These lagged coefficients were typically between 0.10 and 0.30. The coefficients on percent of economically disadvantaged were between -0.04 and -0.12 for all tests, grades and years. Passing rates also depend on average teacher salary, average teacher experience and teacher/student ratio. Table 4.2 shows that among these variables average teacher salary, teacher/student ratio and per-pupil funding have the expected signs and that most of these are statistically significant. Average teacher experience does not have any significance. Fixed-effects estimates indicate that the fraction of variance due to school specific fixed-effects was almost always greater than 0.4. Thus, school specific effects are indeed important to control for while estimating passing rates. The school specific fixed effects represent all the unmeasured variables such as school location, the community the school is in, the classrooms etc. The within R-squared for the fixed effect estimates, where campus is the group variable, is smaller than the between R-squared. So, my model is predicting more of the variation across schools than variation over time within the same school.

Table 4.3 Data Generating Process with Lagged Grade Passing Rates

	(1)	(2)	(3)	(4)
	Passing rate – Reading – Year 2004	Passing rate – Reading – Year 2005	Passing rate – Math – Year 2004	Passing rate – Math – Year 2005
Lagged grade passing rate – Reading (percentages)	0.005 (0.019)	-0.032*** (0.017)		
Lagged grade passing rate – Math (percentages)			0.066* (0.020)	0.025 (0.018)
% of students economically disadvantaged	-0.120* (0.021)	-0.137* (0.021)	-0.099* (0.023)	-0.122* (0.023)
Grade 4 (dummy variable)	5.934* (0.222)	-7.280* (0.214)	3.532* (0.208)	-8.511* (0.217)
Constant	85.190* (1.905)	95.835* (1.797)	81.202* (2.122)	92.799* (1.927)
Observations	6914	7028	6916	7025
Number of campuses	3813	3893	3815	3892
R-squared	0.24	0.34	0.10	0.35

Standard errors in parentheses
*** significant at 10%; ** significant at 5%; * significant at 1%

Next, I estimated a data generating process with lagged grade dummy variables. Table 4.3 shows my findings. Regression (1) in table 5.3 demonstrates the data generating process for year 2004 reading tests. Regression (2) in table 5.3 demonstrates the data generating process for year 2005 reading tests. Regression (3) in table 5.3 demonstrates the data generating process for year 2004 math tests. Regression (4) in table 5.3 demonstrates the data generating process for year 2005 math tests. The analyses here has very similar interpretations to table 5.2 but this table shows the effect of lagged passing rates of the same grade (different cohort) on this year's passing rates. I observe that lagged grade passing rates have very little or no effect on the current year passing rates. This measures the teachers' contribution to the change in the passing rates. The coefficients on the lagged grade passing rate dummy variables were statistically insignificant for most tests, grades and years. This indicates that lagged cohort passing rate is a more appropriate explanatory variable in the education production function than lagged grade passing rate.

CHAPTER 5

EMPIRICAL RESULTS

5.1 Evidence on Scale & Substitution Responses

5.1.1 Direct Outcomes

I start by examining direct outcomes. Direct outcome refers to a cohort's response to failing the same subject last year. I use specification *1a* and *1b* for this analysis. The analysis uses a dummy variable that is 1 if the same cohort failed the same subject in the previous year. I refer to this variable as, lagged same subject fail dummy variable. A positive coefficient on the same subject fail dummy variables will reveal that, cohorts increase their passing rates for the subject that they failed last year. A negative coefficient on the same subject fail dummy variable will indicate that cohorts will do worse this year on the subject that they failed last year.

The theory discussion in chapter 3 suggests that schools will emphasize on subjects failed by a given cohort, generating a positive coefficient δ. This leads us to the following null and alternative hypotheses:

H_0: Failing TAKS on the same subject, previous grade, and previous year does not affect passing rates on that subject for the same group of students the following year ($\delta \leq 0$).

H_α: Failing TAKS on the same subject, previous grade, previous year increases passing rates on that subject for the same group of students the following year ($\delta > 0$).

Table 5.1 Direct Outcomes (Failed same subject last year)

	(1) Passing rate – Reading – Grade 4	(2) Passing rate – Reading – Grade 5	(3) Passing rate – Math – Grade 4	(4) Passing rate – Math – Grade 5
Lagged cohort passing rate – Reading (percentages)	0.117* (0.022)	0.301* (0.020)		
Lagged cohort passing rate – Math (percentages)			0.218* (0.022)	0.241* (0.019)
Same cohort failed reading last year (dummy variable)	-3.565 (3.389)	1.696 (1.559)		
Same cohort failed math last year (dummy variable)			5.451 (3.578)	-2.848 (2.274)
% of students economically disadvantaged	-0.077* (0.018)	-0.089* (0.018)	-0.040** (0.020)	-0.052* (0.020)
Year 2004 (dummy variable)	6.873* (0.183)	-6.795* (0.179)	5.345* (0.198)	-6.990* (0.196)
Constant	71.930* (2.347)	64.869* (2.028)	62.307* (2.280)	70.100* (2.068)
Observations	7117	6411	7118	6413
Number of campuses	3625	3289	3627	3290
R-squared	0.31	0.36	0.20	0.31

Standard errors in parentheses
*** significant at 10%; ** significant at 5%; * significant at 1%
A joint significance test reveals that the four fail dummy variables are **not** jointly significant

Table 5.1 shows direct outcomes of failing the same subject last year. Regression (1) shows the effect of failing 3rd grade reading in one year on 4th grade reading passing rate the next year for the same cohort. I see that the coefficient on "Dummy variable: same cohort failed Reading last year" is statistically insignificant. Similarly, regression (2) shows the effect of failing 4th grade reading one year on 5th grade reading passing rates the next year for the same cohort. The coefficient on the lagged same subject fail dummy variable is again insignificant. However, I do see a positive effect, as expected. Regression (3) demonstrates the effect of failing 3rd grade math one year on 4th grade math passing rates the next year for the same cohort. The coefficient on the lagged same subject fail dummy variable is positive and insignificant. Regression (4) demonstrates the effect of failing 4th grade math one year on 5th grade math passing rates the next year for the same cohort. The coefficient on the lagged same subject fail dummy variable is insignificant. I see a negative coefficient on the same subject fail dummy variable contrary to my expectations. For all 4 of the regressions, the coefficients on the control variables are same as those that I observed before.

To summarize, I observe no individually significant impacts of failing last year on this years passing rates. Two out of the four coefficients have a positive and two of the four coefficients have a negative sign. The standard errors on these coefficients are large. Then, I conduct a joint significance test that the sum of all four coefficients is positive. I find that the sum of all coefficients is .734 and the z-stat is .13. Therefore, I fail to reject the null and conclude that the sum of all four coefficients is jointly insignificant.

This may imply that additional effort to increase passing rates is unproductive for the same cohort. However, there is another possible explanation for why these coefficients are statistically insignificant. The lack of sufficient observations could have caused the coefficients to be insignificant. The number of cohorts that failed a subject last year is small in some samples. For example, there were only 14 observations for grade 4 where the cohort failed reading last year and, there were 17 observations for grade 4 where the cohort failed math last year. The lack of sufficient data may have led to large standard errors and insignificant results.

5.1.2 Scale & Substitution Responses

After looking at direct outcomes I attempt to analyze a cohort's combined *scale* and *substitution responses*. I use specification *(2a)* and *(2b)* for this analysis. The analysis uses a dummy variable that is 1 if the same cohort failed a different subject in the previous year. I refer to these variables as "different subject lagged fail dummy variables." If I see a positive coefficient on the lagged different subject fail dummy variable, then I will infer that when a cohort fails a subject one year, it's passing rate for the other subject goes up the next year. If I see a negative coefficient on the lagged different subject fail dummy variable, then I will infer that when a cohort fails a subject one year, it's passing rate for the other subject goes down the next year.

Theory in section 3.2 tells us that combined effect of scale and substitution responses on the unmeasured outcome will be negative as long as the measured and unmeasured outcomes are substitutes. This would imply that if I do not control for resources, I would observe a negative δ. If I do control for resources, I would observe an even larger negative δ. This leads us to the following null and alternative hypotheses:

H_0: Failing TAKS on a different subject, previous grade, and previous year does not affect passing rates on the target subject for the same group of students the following year ($\delta \geq 0$).

H_α: Failing TAKS on a different subject, previous grade, and previous year decreases passing rates on the target subject for the same group of students the following year ($\delta<0$).

Table 5.2 Scale & Substitution responses (Failed different subject last year)

	(1) Passing rate - Reading – Grade 4	(2) Passing rate - Reading – Grade 5	(3) Passing rate - Math – Grade 4	(4) Passing rate - Math – Grade 5
Lagged cohort passing rate – Reading (percentages)	0.123* (0.022)	0.289* (0.019)		
Lagged cohort passing rate – Math (percentages)			0.204* (0.021)	0.244* (0.019)
Same cohort failed math last year (dummy variable)	-2.744 (3.129)	-1.706 (2.041)		
Same cohort failed reading last year (dummy variable)			-11.740* (3.667)	-1.635 (1.613)
% of students economically disadvantaged	-0.076* (0.018)	-0.089* (0.018)	-0.039** (0.020)	-0.052* (0.020)
Year 2004 (dummy variable)	6.885* (0.183)	-6.779* (0.179)	5.359* (0.198)	-6.995* (0.196)
Constant	71.340* (2.286)	65.881* (1.971)	63.502* (2.245)	69.784* (2.026)
Observations	7117	6411	7118	6413
Number of campuses	3625	3289	3627	3290
R-squared	0.31	0.36	0.20	0.31

Standard errors in parentheses
*** significant at 10%; ** significant at 5%; * significant at 1%
A joint significance test reveals that the four fail dummy variables are jointly significant

Table 5.2 shows scale & substitution responses for cohorts that failed a different subject in the previous year. Regressions (1) and (2) use equation *(2a)* while regressions (3) and (4) use equation *(2b)*. Regression (1) in table 5.2 analyzes the effect of failing 3rd grade math test one year on 4th grade reading passing rates the following year for the same cohort. I observe a negative and insignificant coefficient on lagged different subject fail dummy variable. Regression (2) in table 5.2 analyzes the effect of failing 4th grade math test one year on 5th grade reading passing rates the following year for the same cohort. I observe a negative and insignificant coefficient on lagged different subject fail dummy variable. Regression (3) in table 5.2 analyzes the effect of failing 3rd grade reading test one year on 4th grade math passing rates the following year for the same cohort. I observe a negative and significant coefficient on lagged different subject fail dummy variable. This coefficient implies that if a cohort fails 3rd grade reading test, the same cohorts 4th grade math passing rate will go down by 11.74 percentage points. Regression (4) in table 5.2 analyzes the effect of failing 4th grade reading test one year on 5th grade math passing rates the following year for the same cohort. I observe a negative and insignificant coefficient on lagged different subject fail dummy variable. For all 4 of the regressions, the coefficients on the control variables are same as those that I observed before.

In the analyses, only one out of the four lagged different subject fail dummy variable was significant. However, I observe negative coefficients on all fail dummy variables. Next, I run a joint significance test that the sum of the four coefficients is negative. I find the sum of the four coefficients to be -17.82 and a z-stat of -3.25. Therefore, I reject the null hypothesis in favor of the alternative. Even though these coefficients are not statistically significant by themselves, a joint significance test reveals that they are significant together.

This result implies that failing in math last year decreases this year's passing rates in reading in the current year. Similarly, failing in reading last year decreases this year's passing rates in math. This suggests that schools are substituting away from the subjects they are passing toward subjects that they are failing. I conclude that, there is evidence of *substitution responses.*

5.1.3 Substitution Responses

The third analysis that I conduct, attempts to find only the *substitution responses* of failing cohorts. I use specification *3a* and *3b* for this analysis. The analysis uses a dummy variable that is 1 if the same cohort failed a different subject in the previous year. In addition to all the right hand side variables in the last analysis, these specifications contain the resource variables, average teacher experience, average teacher salary, student/teacher ratio, and per pupil expenditure.

Theory in chapter 3 tells us that the substitution responses will always have a negative impact on the unmeasured outcome. After I include the resource variables, the lagged different subject fail dummy variables no longer capture the *scale responses*. The remaining *substitution responses* are captured by the dummy variables. Since we control for the resource variable here, I expect a negative δ. And expect that the δ will be a larger negative number compared to combined scale & substitution responses. This leads us to the following hypothesis:

H_0: Failing TAKS on a different subject, previous grade, and previous year does not affect passing rates on the target subject for the same group of students the following year ($\delta \geq 0$).

H_α: Failing TAKS on a different subject, previous grade, and previous year decreases passing rates on the target subject for the same group of students the following year ($\delta<0$) and δ will be even smaller than in section 5.1.2.

Table 5.3 Substitution responses (Failed different subject last year and controlling for resource variables)

	(1) Passing rate - Reading – Grade 4	(2) Passing rate - Reading – Grade 5	(3) Passing rate - Math – Grade 4	(4) Passing rate - Math – Grade 5
Lagged cohort passing rate – Reading (percentages)	0.123* (0.022)	0.296* (0.019)		
Lagged cohort passing rate – Math (percentages)			0.208* (0.021)	0.252* (0.019)
Same cohort failed math last year (dummy variable)	-3.153 (3.103)	-5.451** (2.195)		
Same cohort failed reading last year (dummy variable)			-12.421* (3.638)	-3.720** (1.641)
Average teacher experience	-0.012 (0.152)	-0.183 (0.146)	-0.188 (0.169)	-0.341** (0.157)
Student/teacher ratio	-0.535* (0.139)	-0.295** (0.140)	-0.649* (0.156)	-0.519* (0.151)
Average teacher salary ($1000s)	0.326** (0.166)	0.082 (0.134)	0.526* (0.184)	0.412* (0.144)
Per-pupil expenditure ($1000s)	0.381* (0.133)	0.469*** (0.239)	0.242 (0.149)	0.263 (0.258)
% of students economically disadvantaged	-0.088* (0.018)	-0.099* (0.019)	-0.049** (0.020)	-0.072* (0.020)
Year 2004 (dummy variable)	6.878* (0.225)	-6.860* (0.233)	5.594* (0.248)	-6.776* (0.252)
Constant	65.095* (6.387)	66.656* (5.418)	53.075* (7.056)	63.902* (5.737)
Observations	7093	6380	7093	6382
Number of campuses	3617	3276	3618	3277
R-squared	0.32	0.36	0.21	0.32

Standard errors in parentheses
*** significant at 10%; ** significant at 5%; * significant at 1%
A joint significance test reveals that the four fail dummy variables are jointly significant

Table 5.3 shows substitution responses of a cohort that failed a different subject last year. Regression (1) and (2) use equation *(3a)* while regression (3) and (4) use equation *(3b)*. Regression (1) in table 5.3 analyzes the effect of failing 3rd grade math test one year on 4th grade reading passing rates the following year. I observe a negative and insignificant coefficient on lagged different subject fail dummy variable. Regression (2) in table 5.3 analyzes the effect of failing 4th grade math test in one year on 5th grade reading passing rates the following year. I observe a negative and significant coefficient on the lagged different subject fail dummy variable. The result suggests that when a cohort fails 4th grade math test, that cohort's 5th grade reading passing rate falls by 5.45 percentage points. Regression (3) in table 5.3 analyzes the effect of failing 3rd grade reading test one year on 4th grade math passing rates the following year. I observe a negative and significant coefficient on lagged different subject fail dummy variable. This coefficient implies that if a cohort fails 3rd grade reading test, the same cohorts 4th grade math passing rate will go down by 12.41 percentage points. Regression (4) in table 5.3 analyzes the effect of failing 4th grade reading test one year on 5th grade math passing rates the following year. I observe a negative and significant coefficient on the lagged different subject fail dummy variable. The result suggests that when a cohort fails 4th grade reading test, that cohort's 5th grade math passing rate falls by 3.72 percentage points. For all 4 of the regressions, the coefficients on the control variables are the same as those that I observed before.

I observe negative coefficients on all four lagged different subject fail dummy variables. Three out of four coefficients are statistically significant by themselves after the inclusion of resource variables. I conduct a joint significance test that the sum of the four coefficients is negative. I find the sum of the four coefficients to be -24.74 and a z-stat of -4.49. Therefore, I reject the null and conclude that I see evidence of *substitution responses*.

I also see that all four coefficients are larger in magnitude compared to the same coefficients in table 5.2. This implies that there is a positive *scale response*. In table 5.3 when I account for the resources, only *substitution responses* are captured by the dummy variables. But, in table 5.2 where I do not control for resources, both *scale* and *substitution responses* are captured by the dummy variables. So, the scale response is the difference in the coefficients in 5.2 and 5.3. The combined scale and substitution responses create negative coefficients that are smaller in magnitude compared to coefficients from just the substitution responses. Therefore, a positive scale response must be present that is pushing the coefficients up after a negative substitution response. The net result that I observe is negative (table 5.2). So, in conclusion even though I see evidence of both substitution and scale responses, *substitution responses* are larger in magnitude. Hence, the total response is similar to figure 3.2. A scale effect is necessary for testing to lead to increased learning. I find evidence of a scale effect, therefore, it is possible that there is increased learning in the cohorts I analyze. But, since I fail to find a positive direct outcome of failing, I cannot pinpoint whether testing is actually increases the level of education produced.

5.2 Evidence on shift in Focus to Marginal Students

The next stage of my analysis attempts to find any substitution between marginal and non-marginal students in a particular cohort. More specifically, I try to find what happens to the proficiency rate of a particular cohort when they fail a subject.

To test whether schools shift their resources towards marginal students, I look at how failing in a subject on a given year affects that cohort's proficiency rating next year. In order to observe this effect, I use a specification where proficiency rate is the dependent variable. I use specification *(4a)* and *(4b)* for this analysis. The analysis uses a dummy variable that is 1 if the same cohort failed the same subject in the previous year.

Theory suggests that schools emphasize on marginal students when a particular cohort fails a subject in a given year. They substitute away from non-marginal student to marginal students resulting in higher passing rates and lower proficiency rates. If this is the case we expect to see a negative coefficient δ. This leads us to the following null and alternative hypotheses:

H_0: Failing TAKS on a subject, previous grade, previous year does not affect proficiency rates on that subject for the same group of students the following year ($\delta \geq 0$).

H_α: Failing TAKS on a subject, previous grade, previous year decreases proficiency rates on that subject for the same group of students the following year ($\delta<0$).

Table 5.4 Substitution responses- Proficiency vs. Passing

	(1) % commended in reading – Grade 4	(2) % commended in reading – Grade 5	(3) % commended in math – Grade 4	(4) % commended in math – Grade 5
% commended for same cohort last year in reading	0.223* (0.014)	0.320* (0.019)		
% commended for same cohort last year in math			0.230* (0.017)	0.392* (0.022)
Same cohort failed reading last year (dummy variable)	-1.505 (2.911)	-2.666** (1.359)		
Same cohort failed math last year (dummy variable)			0.208 (3.143)	1.559 (2.202)

Table 5.4 - continued

% of students economically disadvantaged	-0.081* (0.016)	-0.102* (0.017)	-0.098* (0.018)	-0.123* (0.020)
Year 2004 (dummy variable)	3.422* (0.203)	4.426* (0.219)	-5.345* (0.216)	-1.215* (0.233)
Constant	18.955* (1.064)	20.068* (1.142)	26.673* (1.151)	27.859* (1.314)
Observations	7121	6411	7122	6413
Number of campuses	3628	3289	3630	3290
R-squared	0.10	0.14	0.33	0.17

Standard errors in parentheses
*** significant at 10%; ** significant at 5%; * significant at 1%
A joint significance test reveals that the four fail dummy variables are not jointly significant

Regressions (1) and (2) use equation *(4a)* while regressions (3) and (4) use equation *(4b)*. Regression (1) in table 5.4 analyzes the effect of failing 3rd grade reading test in one year on 4th grade reading proficiency rates the following year. I observe a negative and insignificant coefficient on lagged same subject fail dummy variable. Regression (2) in table 5.4 analyzes the effect of failing 4th grade reading test in one year on 5th grade reading proficiency rates the following year. I observe a negative and significant coefficient on lagged same subject fail dummy variable. The result suggests that when a cohort fails 4th grade reading test, that cohort's 5th grade reading proficiency rate falls by 2.66 percentage points. Regression (3) in table 5.4 analyzes the effect of failing 3rd grade math test in one year on 4th grade math proficiency rates the following year. I observe a positive and insignificant coefficient on lagged same subject fail dummy variable. Regression (4) in table 5.4 analyzes the effect of failing 4th grade math test on 5th grade math proficiency rates the following year. I observe a positive and insignificant coefficient on lagged same subject fail dummy variable.

To summarize, only one coefficient out of the four coefficients is statistically significant and for that coefficient I do see a *substitution response.* Since I find only one of the four lagged same subject fail dummy variables to be statistically significant, it is hard to draw strong conclusions from these coefficients. I conduct a joint significance test that the sum of the four coefficients is negative. I find the sum of the four coefficients to be -2.40 and a z-stat of -0.48. Therefore, I fail to reject the null. However, I do observe that the lagged fail dummy variables were positive for reading and negative for math. This might imply that cohorts have *substitution responses* to reading tests, but they do not have any *substitution responses* for math. .

Overall, the results from all the analyses show the evidence of *substitution*, and *scale responses*. The *substitution responses* are larger than *scale responses*. If this pattern is true, and the schools are at an optimal point before high-stakes testing is introduced, then high stakes testing maybe decreasing the amount of education that is being produced.

CHAPTER 6

CONCLUSIONS

In this paper, I demonstrate that schools respond strategically to high stakes testing. I conduct my analysis by using a panel dataset of more than 3000 schools that administered the TAKS test. My primary findings show that schools respond to failing by substituting away from subjects that they are passing. I show that, if failing schools respond mostly by substituting and not by scaling resources it may actually lower the level of education that the school is producing. Therefore, since I observe mostly *substitution responses*, production of education may decline as a result of high stakes testing.

I also analyze whether schools shift their focus from non-marginal students to marginal students. I find evidence that schools sacrifice proficiency rates to increase passing rates only for reading tests, but not for mathematics tests.

These results question the validity of test-based accountability. My findings suggest that The No Child Left Behind Act is not contributing to producing more education in general. Even though more schools are meeting the passing thresholds they are producing a lower level of education in general.

My secondary findings show the effectiveness of various inputs in the production of education. I find that there is a positive and statistically significant effect of resources on student achievement. But, the marginal effects of some inputs in the education production function are too small to have any practical economic significance. Other inputs are both statistically and economically significant. These results confirm what we already know from the existing literature. Resources do matter under certain conditions and some inputs do not matter as much as other inputs. I find that average teacher experience does not affect passing rate. However, a smaller class size, per-pupil expenditure and teacher salaries do impact passing rates in schools. By comparison, the marginal effect of each thousand of dollars spent in reducing the class-size is the most effective way of increasing passing rates. These findings are consistent with the existing literature.

REFERENCES

Card, D. & Krueger, A. B. (1992). School Quality and Black-White Relative Earnings: A Direct Assessment. Quarterly Journal of Economics, 107(1), 151-200.

Carnoy, M. & Loeb, S. (2002). Does External Accountability Effect Student Outcomes? A Cross-State Analysis. Educational Evaluation and Policy Analysis, 24(4), 305-331.

Coates, D. (2003). Education Production Functions Using Instructional Time as an Input. Education Economics, 11(3), 305-314.

Eide, E. (1998). The Effect of School Quality on Student Performance: A Quantile Regression Approach. Economics Letters, 58(3), 345-350.

Greenwald, R., Hedges, L.V., & Laine, R. D. (1996). The Effect of School Resources on Student Achievement. Review of Educational Research, 66(3), 361-396.

Greenwald, R., Hedges, L.V., & Laine, R. D. (1994). An Exchange: Part I: Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes. Educational Researcher, 23(3), 5-14.

Hanushek, E. (1994). An Exchange: Part II: Money Might Matter Somewhere: A Response to Hedges, Laine and Greenwald. Educational Researcher 23(4), 5-8

Hanushek, E. A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. Educational Evaluation and Policy Analysis 19(2), 141-164.

Hoerandner, C. M. & Lemke, R. J. (2006). Can No Child Left Behind Close the Gaps in Pass Rates on Standardized Tests? Contemporary Economic Policy, 24(1), 1-17.

Jacob, B. A. (2005). Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools. Journal of Public Economics 89(5-6), 761-96

Peterson, P. E. & Hess, F. M. (2005). Johnny Can Read . . . in Some States: Assessing the Rigor of State Assessment Systems, Education Next, Summer 2005, 52–53.

Todd, P. E., & Wolpin, K. I. (2004). The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps. Penn Institute for Economic Research, Department of Economics, University of Pennsylvania, PIER Working Paper Archive.

BIOGRAPHICAL INFORMATION

Mishuk Anwar Chowdhury received a Bachelor of Science in Computer Science from Minnesota State University in December 2003. He received his Master of Arts on Economics from University of Texas at Arlington in December 2006. His research interests include Development Economics and Public Policy. He has desires to crunchatize a PhD. in Economics.