High Flying Schools, Student Disadvantage and the Logic of NCLB

Douglas N. Harris
Florida State University
June 16, 2006

Abstract: The idea that schools can help address students' social and economic disadvantages represents an important purpose of universal public education. Some recent reports as well as federal education policy take this one step further, however, by implicitly assuming that schools can completely overcome these disadvantages. The present study considers this argument by re-analyzing the data and interpretations of recent reports by the Education Trust. There are three main flaws in the reports. First, their methods make the number of high-poverty schools reaching high-performance—the "high flyers"—look unrealistically large. The re-analysis here shows that low-poverty schools are 22-89 times as likely as high-poverty schools to be high-performing on achievement tests. Second, contrary to interpretations of the reports, this type of analysis provides little evidence about the role schools play in determining student learning. Third, the idea that schools can completely overcome student disadvantages ignores a vast amount of evidence about the strong role the disadvantages play in affecting student learning. The reports therefore reinforce a false assumption of the federal *No Child Left Behind* (NCLB) law. While accountability has the potential to facilitate genuine school improvement, requiring schools to completely overcome student disadvantages is just as likely to produce counter-productive responses from educators and therefore hurt the students they are supposed to help.

Introduction

Those who study and debate the issue of educational inequity, while they may

disagree on important issues of politics and ideology, can almost always agree on two

conclusions. First, despite significant advances in recent decades, inequities in

educational outcomes across racial and income groups are still large.[1] Second, these

inequities represent one of the most significant problems of the educational system

(Petrilli and Hess 2006).[2]

There is an additional conclusion for which research evidence is clear, but that is

nevertheless being subtly contested in current policy debates: that students' social and

economic disadvantages are the main, although not the only, cause behind the inequities.

While this conclusion is rarely challenged directly, a collection of recent policy reports

does challenge it implicitly. Published by the Education Trust (ET) and Heritage

Foundation, these reports purport to identify large numbers of "high flying" schools that

achieve high student test scores even though their students appear to face considerable

disadvantages. Both the reports themselves, and the responses published in the national

media, interpret this as evidence that educational inequity is entirely due to educators'

lack of effort or even outright racism.

The fact that the most recent ET report received so much attention is noteworthy

because it contains two significant flaws. First, it over-identifies schools as high flyers,

encouraging school leaders and policy makers to copy practices that are not consistently

resulting in high student achievement. Second, and most importantly, there is a flaw in

the implied logic of the report. It is perfectly reasonable to identify high flyers as well as

the characteristics that distinguish them from other schools, in order to improve all

schools. However, contrary to the assertions in the ET report, the *number* of high

flyers—even if it were large—tells us very little about the relative roles of the school

itself and socioeconomic disadvantages faced by students in that school. Third, the report

ignores the vast amount of evidence that inequity in educational outcomes is primarily,

although certainly not solely, due to students' social and economic disadvantages. The

report also ignores evidence that schools, when faced with requirements they cannot

meet, often respond in ways that make matters worse.

In addition to the national attention they received, these reports are important

because their results seem to reinforce some of the misguided elements and assumptions

of the federal *No Child Left Behind* (NCLB) law. In short, both the reports and NCLB

focus on learning levels rather than learning gains and therefore often fail to reward

schools where students are making substantial improvements. While setting the bar so

high may seem admirable and justified in order to help the most disadvantaged students

catch up, it is more likely to be counter-productive by inducing schools to adopt practices

that do not reduce inequities and to avoid practices that are likely to have a positive long-

term influence.

After discussing the issues of student disadvantage and NCLB in more detail, the

study continues by identifying and demonstrating the flaws within recent reports. I

describe in more detail reports by both and the Heritage Foundation and then discuss the

issues of regression to the mean and the use of proficiency scores, methodological issues

that compound the misinterpretation of reported results. Finally, to show how these

problems result in over-identification and mis-identification of high flyers, I re-analyze

the data used by ET, the School-Level Achievement Database (SLAD), which includes

more than 62,000 schools in 47 states. The analysis shows, in contrast the earlier ET analysis, that only about one percent of high-poverty schools are "high flyers"—that is, schools with high percentages of disadvantaged students that are consistently in the state's top third in academic achievement. Moreover, a low-poverty, low-minority school is 89 times more likely to be in the state's top third than a high-poverty, high-minority school.

The fact that high flyers are rare reinforces the need for solutions, including changes in both NCLB and in policies that go beyond education. Indeed, the larger objective of this study is to set the stage for such solutions by providing a clearer understanding of the true roles that student disadvantage and school ineffectiveness play. The significant improvements in equity over the past half century suggest that progress is certainly possible (Harris and Herrington 2006; Lee 2002). But accountability and other solutions to the remaining inequities will be sure to fail if they do not recognize that schools are not the only or main cause.

Student Disadvantage, Schools and NCLB

While the role of student disadvantage is rarely disputed, the topic is central to the argument in this study and it is therefore worth mentioning some of the research. For example, poor nutrition and illness cause students to miss school more often and to be less prepared to learn when they attend (Brooks-Gunn and Duncan 1997; Rothstein 2004). Within the home, low-income parents have relationships with their children that are, emotionally and physically, less healthy (Duncan and Brooks-Gunn 2000). These unhealthy relationships are created in part by economic pressures that induce conflicts

between parents and children (Conger, Conger and Scaramella 1997). Of course, many parents living in poverty successfully navigate and avoid these potential problems, while some high income parents do not, but the general patterns described here are strong.

Perhaps the strongest evidence regarding the educational implications of social and economic disadvantages comes from a recent study of students starting kindergarten, based on the recent Early Childhood Longitudinal Survey (ECLS). One of the most important features of this survey is that it administers tests of academic ability near the beginning of students' first year in kindergarten. Lee and Burkham (2002) find with these data that the achievement levels of African-American kindergarteners are a standard deviation below the levels of white kindergartners. Because these students have spent no time in school, it is clear that schools cannot be the primary cause of this gap.[3] Moreover, as documented above, the size of the gaps is nearly the same as those observed much later in their school careers. This suggests that schools are able to help disadvantaged students learn—and at about the same rate as other students (Gamoran 2001). Again, the role of social and economic disadvantage, especially in children's early years, appears critical. Given the strength of the evidence above, this conclusion would not seem to be very controversial. Nor is it very new—Coleman (1966) reached essentially the same conclusion more than four decades ago.

This does not mean of course that schools do not matter. Harris and Herrington (2006) review evidence that additional school resources, while they may not always influence achievement, do have benefits for disadvantaged students and have contributed to the reduction in achievement gaps over the past half-century. They also find that academic standards and some forms of accountability improve learning for these

students. The use of certain types of instructional practices can also have an influence (Oakes, Joseph and Muir 2004). A basic conclusion of the present study, based on the above evidence, is that both student disadvantage and school effectiveness are important and therefore the complete elimination of educational inequity requires action on both fronts.

But we need to know more about the relative roles of student background and school factors in order to design effective accountability policy. NCLB requires students in all racial, income and other sub-groups to achieve a basic level of proficiency on state standardized tests by 2014. The law, like most state accountability systems, focuses on learning *levels*; that is, the accumulation of all learning that has taken place since birth. If student disadvantages were relatively unimportant, then accountability based on test levels would be a reasonable approach. But the decades of research cited above, especially the "starting gate inequalities" identified by Lee and Burkham (2002), suggest that these levels primarily reflect initial disadvantages, rather than school effort. If we are to focus on the school's contribution to learning, then the NCLB focus on learning levels is inappropriate.

An alternative, one that accepts NCLB's use of annual testing and sub-groups, is a focus instead on the year-to-year changes in student scores—learning *gains*. For example, an ambitious version might sanction schools in which all students do not make a full grade level gain in learning each year. In this case, schools would not be punished for student disadvantages that arose before they entered school, although schools still would be expected to overcome the disadvantages that students face while they are in

school. This would be one way to aggressively address school contributions to educational inequity.

NCLB, while it does not focus on gains, is some ways even more aggressive.[4] Under the law, if a disadvantaged student enters kindergarten far below other students, and even if the school is successful in helping the student learn at the rate as more advantaged students, the school will still be sanctioned if the student does not reach the proficiency cut off. Therefore, to ensure that all students reach proficiency, schools must not only address student disadvantage, but *completely overcome* the disadvantages, by helping these students learn at *faster* rates than others—in some cases, much faster rates. This makes schools completely responsible for all educational inequality, including that which arises before the student reached school age.[5]

One may argue that if we are to address the inequity in outcomes, then we have no choice but to adopt the NCLB approach and give full responsibility to schools— schools are the "only hope." Rather than improving education, however, a full shift of responsibility is likely to make matters worse. Previous studies on accountability have shown that some schools try to meet accountability standards by simply teaching students how to take tests or outright cheating (Jacob and Levitt 2003). In addition, schools lengthen suspensions of disruptive students (Figlio 2003) and assign more students to special education (Booher-Jennings 2005), apparently in the hope of eliminating them from the pool considered for accountability. They also tend to focus resources on the children just "below the bubble" of proficiency, where the rewards for student improvement are greatest (Booher-Jennings 2005). In some high stakes accountability environments, teachers have been more likely to leave schools where it is most difficult

to meet the objectives (Ladd and Walsh 2002; Wong, Anagnostopoulos, Rutledge, Lynn, and Dreeben 1999)—that is, to the schools that need them least. Few if any of these responses to accountability is likely to reduce inequities and some, such as lengthening suspensions, may actually exacerbate existing problems.

Holding schools responsible for all inequity is also likely to induce schools to drop programs that are, by just about any measure, successful. Suppose that a program helps disadvantaged students make a full grade level of improvement each year—again, an ambitious goal. Under NCLB, such a program would be insufficient and would therefore have to be altered or dismissed, accelerating the long-recognized cycle of change—changes in curriculum, instruction and management practices—that typify the average urban school (Hess 1999; Hill, Pierce, and Guthrie 1997; Lee and Smith 1994). In addition, there is a strong bias toward changes that can be implemented quickly and show fast improvements, rather than changes that have a long-term influence. These tendencies are exacerbated when people believe, and laws such as NCLB assume, that student disadvantage is unimportant and therefore attribute low test scores solely to the schools.

The above discussion suggests that NCLB is likely to induce counter-productive responses from schools including the implementation of short-term "fixes" such as teaching students how to take tests rather than working towards meaningful change and a rapid cycle of reform adoption and abandonment that make it difficult for any program to be effective. A central premise of the present study is that the magnitude of these problems is likely to be worse in systems that do not hold school accountable for what they can control. That is, the more that accountability systems require schools to address

matters such as student disadvantage that are outside their control, the more likely it is that school responses will be counter-productive. Conversely, positive responses are more likely when schools are held responsible for what they can control.

This leads us back to the role that student disadvantage plays relative to schools. If the role of student disadvantage is large, as the research mentioned earlier suggests, then accountability will be most effective if it is based on learning gains. But NCLB, and most other accountability systems, focus on learning levels and therefore assume that the role of social and economic disadvantages—from poor nutrition to unhealthy relationships with parents—are not important. This assumption is wrong and significantly reduces the potential of accountability to help schools improve.

This is also why reports such as those by the ET and the Heritage Foundation are so important. As I show in the next section, by attempting to diminish the role of student disadvantage, they promote a focus of accountability systems on student learning levels.

Flying High and Beating the Odds

The research by ET discussed here includes the organization's 1999 report, *Dispelling the Myth: High Poverty Schools Exceeding Expectations* (Education Trust, 1999), and a 2001 follow-up, *Dispelling the Myth: Revisited* (Jerald, 2001).[6] Both reports identify schools that have achieved apparently high outcomes despite measurable disadvantages.

The 2001 ET report—the focus of the present discussion—defines "high-flying" schools as those that are both "high-performing" (above the 67th percentile in average state standardized test scores) and "high-poverty" (more than 50 percent of students are

eligible for free or reduced price lunch). They find 3,592 schools that meet these two criteria. The next section shows that the performance definition used to identify the 3,592 schools—which requires high achievement in only one subject and considers only one grade and one year—misidentifies many schools as high-performing. As a result, ET would call a school high-flying even if students could not read or do basic math, or could do so only a single point in their academic careers. Also, even if one accepts this as the correct number of high flyers, the 3,592 schools represent a small fraction of all high-poverty schools and therefore that it is still rare that schools are able to overcome students disadvantages.

In March, 2002, Education Trust followed their earlier reports with additional analysis that used different definitions of high performance in an attempt to address some of these criticisms (2002a). They also tried to minimize the problem with their earlier definitions, writing that "no single definition of high performance—or high-poverty or high-minority, for that matter—will work for all research purposes" (2002a, 2). This is undoubtedly true, but misses the point of the critique here. Different definitions are appropriate under different situations, but it is also true that there are some definitions of high performance that should not be used except when absolutely necessary. To educators and education researchers, it is well known that individual test scores are unreliable measures of student achievement that vary dramatically from year-to-year and grade-to-grade even when school effectiveness appears unchanged (Kane and Staiger 2001). Any definition that does not take this account will likely yield misleading results no matter what type of research is being done. By relying on a single score for a single year, the results are likely to misidentify some schools as high flyers.

The authors of the ET report also write in support of their original performance definition that "we know from our own work in schools across the country that the reforms that take hold in one subject and one grade level can provide the basis for improvements in other grades and subject areas" (2002a, 2). Again, this is almost certainly true, but schools that are improving should eventually be able to achieve high scores in more than one subject, grade and year. Requiring more consistent performance, in addition to addressing regression to the mean, would address this problem as well. Therefore, it seems that the performance definition in the original ET report was ill-suited for the stated task.

The recent ET report shares many similarities with a 1999 report published by the Heritage Foundation, entitled, *No Excuses*. This analysis started with approximately 400 public schools brought to their attention from various sources, including state education agencies, think tanks, teachers' unions, and foundations. Like the ET report, the authors of the Heritage study narrowed this list to 125 schools that had high concentrations of poverty and high test scores. Their specific criteria were also similar: to be on the list, test scores had to be in the top-third of the state and at least 75 percent of the students had to be eligible for free and reduced lunch (instead of 50 percent in the ET report). From this list of 125 schools, 21 were selected for site visits and further study.

As discussed previously by Rothstein (2004), the most significant problem with the Heritage report is that nearly all of the schools considered were distinctive in ways that had little to do with school effort or practices.[7] For example, nine of the 21 schools had admission requirements that could exclude students who have received low test scores. Indeed, a more careful analysis shows that only 3 of the 21 schools could be

considered high-flyers. There may well be much to learn from these schools, and the other 18, but the Heritage study does more to mask the lessons of the analysis than to learn from them.

It is important to emphasize that these types of analyses can serve a valid purpose: identifying the characteristics of effective schools that other schools can learn from. Some schools are better than others and there is a well established research literature that attempts to identify them (e.g., Edmunds, 1979). It is also likely that the qualities identified by the Education Trust—high expectations, data-driven decision-making—are ones that improve student learning and that identifying and learning from specific schools may be useful for nearby schools that are less successful. But this is not why the reports have received so much attention. As shown below, the responses to the report have focused not on what schools can do to improve but on how many schools apparently reach high levels of performance and possible implications for the role of student background.

Interpretation

The focus on the number of schools, as opposed to their potential usefulness in helping less successful schools, is perhaps not surprising given the way that the authors and the leaders of the organizations they work for have talked about the results. For example, consider the words of Kati Haycock, Director of the Education Trust (ET). In discussing a report published by her organization, she asks, "How many effective schools do we have to see in this country before we conclude that it's not about the kids?" (Nathan 2002). One possible interpretation of this quote is that some students grow up

under adverse circumstances, placing them at a disadvantage in their school activities. Therefore, it may not be "about the kids," but rather about the conditions under which they live and grow. This interpretation would certainly be consistent with the research evidence cited above.

But Haycock's words invite an alternative interpretation. Assuming that schooling and socio-economic status are the two main factors affecting student learning, then excluding the kids means that it must be about the schools. This interpretation is reinforced by Haycock's answer to her own question. Continuing the above quotation she says that, "If your answer is more than one [school], then I submit that you have reasons of your own for preferring to believe that pupil performance derives from family background" (Nathan 2002). Given the more than three thousand schools identified in the 2001 ET report, her remarks imply that pupil performance does not derive from family background.

The above interpretation is further reinforced by the title and wording of the reports themselves. What is "the myth" they are trying to dispel? This is a good question and one the report never explicitly answers. But we can get a fairly clear idea from the words of the author, Craig Jerald, who is quoted as saying that "too many people believe poor and minority students can't achieve. This is simply false. It might be the school hasn't tried." (Bryant 2001). Therefore, the myth appears to be the idea that disadvantaged students cannot learn. While it goes unstated who believes this myth, it seems clear from the statement that "the school hasn't tried" that educators themselves believe it. This may be true of some, but it is difficult to believe that most educators think their students cannot learn when the evidence suggests that they are already

13

learning at about the same rate as other students (Gamoran 2001). Both the title of the

report and this apparent definition of the myth reinforce the interpretation that schools

can completely overcome educational inequity and, even more so, that the whole problem

is fundamentally one of willful neglect by educators.

This interpretation of the reports has also been embraced by some well known

commentators and in the national media. For example:

"People who follow education issues have long known that some schools succeed
with children from families with weak educational backgrounds. But it turns out
that it's not just a few, rare schools that succeed, it's thousands of schools
[according to the recent ET report] . . . We'd better not hear that racist nonsense
anymore." Bill Evers (2002)

In this case the "schools are responsible" assumption is taken one step further.

Not only is about the willful neglect of schools, but neglect rooted in "racist nonsense."

Many other examples of this interpretation are seen in national publications such as the

New York Times, *USA Today*, and *Washington Post*.[8]

In addition to being misleading, the above quotes also set up a false choice—a

choice between holding schools responsible and holding others responsible. It is clear

that all stakeholders—schools, students, parents, governments, educators, taxpayers—

must change their ways in order increase the equity of outcomes. But the above quotes

are difficult to square with this view. "It's not about the kids." "The school hasn't tried."

"Racist nonsense." "No excuses." These words, far from acknowledging the clear

research consensus, instead suggest that social and economic disadvantages play a limited role. Nor do these words—or President Bush's "soft bigotry of low expectations"—provide a sound basis for educational policy. As pointed out in the Introduction, this approach is likely to be counter-productive, inducing schools to adopt poor practices and avoid positive ones.

Methodological Issues

There are two methodological issues that reinforce the misinterpretation of the ET reports. Regression to the mean, while widely recognized as a statistical issue, has particularly problematic effects when comparing groups of schools that have different mean test scores. The second issue is the use of proficiency cut-offs. These are discussed below in turn.

Regression to the Mean

All quantitative measures are assumed to be comprised of a "signal," representing the construct that it is of interest to the researcher, and "noise" or random error, which is assumed to be uncorrelated with the signal and to have an expected value of zero. It is well known that noise makes it difficult to identify the signal. In addition, when repeated measures are available regarding a single construct—in this case, school-level achievement—there is the added problem of regression to the mean. If a school achieves a very high score, it is likely that some, though certainly not all, of this high performance is caused by, what I will call, "positive noise" that is outside of the school's control. Because noise is considered random, it is unlikely that the same school will experience

positive noise for all other tests. Other observations from the same school will produce lower scores, unless the school is truly exceptional.

Some recent studies show that the signal-to-noise ratio of standardized test scores is low, implying that the role of regression to the mean can be large (Kane and Staiger, 2001). As a practical matter, this means that adding additional test scores (e.g., test from additional grades, subjects, and years) could significantly change measured levels of achievement in many schools. Because such additions reduce the effect of regression to the mean, providing a more accurate assessment of achievement levels, it is important that the additional data be included.

The effect of statistical noise is further complicated when schools are separated into low-poverty and high-poverty categories, as is the case in the Education Trust reports. The reason is that each of the two groups has a different expected score. A concrete example may help to illustrate. Consider a typical high-poverty school, School H, and a typical low-poverty school, School L. If there were no noise, then School H would achieve the 40th percentile and School L would reach the 70th percentile. While the expected value of noise is zero, suppose that each school has a 20 percent chance of receiving positive noise equal to 30 percentile points (i.e., noise that raises reported scores above true scores) and a 20 percent chance of experiencing equally-sized but negative noise. Now, suppose that in year one, School H experiences positive noise and therefore reaches the higher-than-expected 70th percentile; School L experiences no noise, and therefore reaches the expected 70th percentile. Both schools are high-performing according to the definitions used in the Education Trust analysis.

However, the odds of this happening again are slim. There is only a 20 percent chance that School H will experience positive noise again, so the school will probably switch from the high-performing group to the low-performing group. School L, in contrast, has an 80 percent chance of remaining high-performing because there is only a 20 percent chance that it will experience negative noise large enough to decrease its percentile below the cut score.

What does this mean for the analysis of educational inequity? First, it means that schools that appear high-performing at any given point in time may actually be average or below. More importantly, this false identification is much more likely to occur with high-poverty schools. That is, when the cut score is set far above the mean for any group (in this case, disadvantaged students generally) then the schools that end up being high-performing by any measure are likely to have received the benefit of random error. The results below demonstrate this effect and why it is essential to use a substantial number of scores when trying to quantify school performance.[9]

Proficiency and "Cut Scores"

There are many different types of standardized tests and many ways to report them. One general approach reports test scores as averages, specifically average raw scores or average scale scores. For an entire school, such measures incorporate the performance of all students; improvement by any given student, no matter their initial level of achievement, shows up as a slightly higher school average.

An alternative approach is to create a "cut score" and use it to distinguish between "proficient" students who score above the cut and other "non-proficient" students who
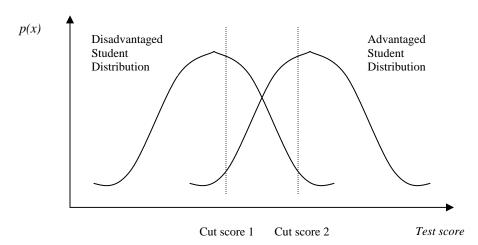
score below the cut. One reason for doing this is to establish a minimum benchmark that all students are expected to attain. But these cut scores are problematic when used for the sake of school accountability. One problem is that such accountability systems create incentives for schools to focus all of their attention on the students who are just below or just above the cut score, because the other students are likely to remain in the same category even if the school devotes little attention to them (Booher-Jennings 2005). A second problem, as indicated earlier, is that even a highly effective school may not be able to the help that would be sufficient for a student who starts off far behind to achieve at the same level as other students.

Rothstein writes that the specific cut score chosen for analysis purposes causes "great mischief" with the measure of achievement (2004, 15). He argues, for example, that an extremely low cut score is likely to be reached by high percentages of students in all groups, making the achievement gap seem small. Conversely, very low percentages of students in all groups will reach extremely high cut scores, resulting in a similarly low achievement gap. As a result, Rothstein writes, "critics can make the test score gap seem extraordinarily large if they define proficiency about halfway between the average score for blacks and the average score for whites" (2004, 89).

This is illustrated in Figure 1 below which displays realistic "bell-shaped" test score distributions for disadvantaged and advantaged students. The distribution to the left has a lower test score mean and reflects the distribution of disadvantaged students. The other similarly shaped curve has a higher mean score and reflects advantaged students. Two cut scores are also shown. At the first, nearly half of the disadvantaged students are proficient but, at the second, almost none of them are.

The two score distributions and two cut scores illustrate why the two groups of students are affected differently by changes in the cut score. Specifically, a small change in cut score 1 will have a larger effect on the proportion of disadvantaged students passing the exam. For cut score 2, the opposite is true—now, the advantaged group is affected more. More generally, when a policymaker moves the cut score closer to the intersection of the two distributions, the gap will appear larger. While this requires other assumptions, it does illustrate and clarify Rothstein's point that the cut score causes "great mischief."[10]

Figure 1: Effect of the Cut Score on Different Student Groups

As we will see below, this has important implications for NCLB because the law encourages states to define proficiency at a very low level (closer to cut score 1), making the achievement gap seem small.

The School-Level Achievement Database (SLAD)

The recent Education Trust reports are based on the School-Level Achievement Database (SLAD), a database created by the U.S. Department of Education and American Institutes for Research (AIR). The SLAD has a total useable sample of 62,074 public schools (74 percent of all public schools in the country) that enroll 36 million students (78 percent of the total).[11] For most schools in the SLAD, information is included about the percentage of students in various racial and ethnic categories and the percentage of students eligible for free or reduced lunches.[12] These variables are summarized in table 1. The original source of the above variables is the U.S. Department of Education Common Core of Data (CCD).

[TABLE 1]

The test score data were complied by AIR and merged with the CCD. Table 2 describes the characteristics of the test score data in each state. Column 1-3 in table 2 provide information about the main scores used in most of the analysis. The first column indicates the type of scores used for each respective state: average scale score (ASC), raw score (RAW), normal curve equivalents (NCE), percentile (PCT), percent of students in the top quartile (4thQ); percent of students passing the lowest cut score (CS1); and percent of students passing the second lowest cut score (CS2). Only one score is used for each school, grade, level, and year. In cases where multiple scores were available, choices were made in the order indicated above, e.g., raw or scale scores were preferred to cut scores.

[TABLE 2]

The third column in Table 2 indicates the grades for which the scores are available. In many states, scores for multiple grades were available and in these cases the following were chosen: grade 5 for elementary, grade 8 for middle, and grade 12 for high school. These are the highest grades available in most schools and, because the schools have had more time to work with these students, they represent the best indications of the schools contribution to learning. In cases where scores were not available at these grades, the next lowest grade was used. The fourth column indicates the year the tests were taken by the students. The analysis compares schools within the same general grade levels, e.g., elementary schools are only compared with other elementary schools, not middle schools.[13]

The next set of columns relate to those states in which cut scores are provided (either CS1 or CS2 in the first column). The fifth column indicates the number of cut scores available in all, e.g., a state might report the percentage of students achieving low medium and high levels of proficiency, yielding three categories. The sixth column indicates the grades for which the cut scores are available.

Finally, the last two columns of table 2 provide information about those states have scores for multiple grades, years, and subjects.[14] In some states, at least eight test scores are available for each school—two subjects and two grades for two consecutive years. This "multiple scores" sample includes 18,365 of the 62,074 schools. Of these, 14,124 are elementary schools, 4,241 are middle schools, and none are high schools.

The data for this study are from the years 1997-2000. More recent data are available; however, there are two reasons for using the older data. First, the older data are the same as those used by Education Trust in their 2001 report, making it easier to compare results across the two studies. Second, by using data before NCLB was adopted, it is possible to avoid conflating the differences in methods used here with changes in educational circumstances—in this case a major change in federal educational policy. These data might be useful for understanding the effects of NCLB, but this is not the purpose of the present study.

For these years, test scores are reported in the SLAD for all U.S. states, except Iowa, South Dakota, and West Virginia. In some of the other 47 included states, there is no standardized test at the high school level; therefore, such schools are also excluded. Every state gives a different achievement test and reports these results in different ways, as shown in Table 2. Some states have multiple cut points and lower numbers (e.g., CP1) indicate lower cut points. One useful feature of the SLAD is that it includes data from multiple cut scores in many states, making it possible to illustrate some of the methodological points made in the previous section.

To my knowledge, the SLAD is the only database that comes close to providing school demographic and achievement information for all schools in the United States. Other data sets provide richer information for small samples of students and schools that are assumed to be nationally representative. The SLAD provides less depth, but includes a near census of all U.S. public schools, allowing for detailed comparisons across states and reducing reliance on sampling assumptions.

It is important to emphasize that every state uses a different standardized test, which makes it difficult in the SLAD to make direct comparisons between schools located in different states. The Education Trust reports use the state tests to calculate each school's percentile ranking within the state. In some sense, this creates a common scale for all schools in the database, but it is does not solve the problem. A school at the 40th percentile in North Dakota has a different level off achievement than one scoring at the same percentile in Montana. Therefore, when possible, the present study uses within-state analysis to make specific points.

A Re-Examination of High Flyers

Table 3 reports the percentage of schools in each of four poverty/performance categories. The table uses the ET definitions of high-poverty (50 percent of the school's students eligible for free and reduced lunches) and high-performance (the school is in the top-third of the state in either reading or math).

[TABLE 3]

Only 16 percent of high-poverty schools are high-performing, compared with 54 percent of low-poverty schools. This means that low-poverty schools are three times more likely to be high-performing than high-poverty schools. Notice also that 34 percent of all schools are high-poverty. These schools are attended by 11.8 million students. This large number of students reinforces the importance of identifying high-flyers and learning from their effective practices. Table 3 also provides the same information for

schools with high levels of poverty *and* large portions of minority students. The results are even more disparate. In this case, only 10 percent of high-poverty-high-minority schools are high-performing, compared with 57 percent of low-poverty-low-minority schools, making them 6 times as likely to reach this high achievement level.

Requiring Consistent High Performance

The above section uses the ET definition of performance, which requires high-performance in either reading or math in the grade and year selected by ET for analysis. This section considers the implications of this definition by providing analysis of the 18,365 schools in the "multiple scores" sample. Table 4 below shows the percentages of low- and high-poverty schools that are high-performing when various combinations of high scores are required. For instance, the first definition (1-1-1) refers to those schools that are high-performing in either year, either subject and either grade. Because there are two subjects and two grades (4 chances) and there are two years (giving 4 additional chances), each school has 8 chances to get in the top-third just one time to become a high-performer. The 2-1-1 definition is somewhat more demanding, requiring that schools are high-performing in both years in either subject and either grade. This definition requires more consistency than one above. The degree of stringency continues to increase up to the 2-2-2 definition, which requires schools to be high-performing in both years, both grades, and both subjects. Here, there is no room for error.

Table 4 reports the percentage of low- and high-poverty schools that are high-scoring, according to each of the above definitions. The first row is based on the ET definition and reiterates the results in Table 3. The middle rows show the results from

24

the weakest performance definition (even weaker than the ET definition[15]) to the most restrictive 2-2-2 definition. The "erosion" of performance between the ET definition of performance and this more demanding one. In other words, the erosion indicates what portion of schools drop off the high-performance list when consistency is required.

[TABLE 4]

As expected, the results suggest considerable decline in the percentages of schools that are high-performing as more consistency is required across grades, subjects and years. The results also provide some support for the hypothesis that high-poverty schools will be disproportionately affected by regression to the mean. The percentage of high-poverty schools achieving high performance declines from 15.6 percent using the ET definition (row 1) to 1.1 percent using the 2-2-2 definition (row 9). This means that 93 percent of schools identified as high-flyers using the ET approach are not high-flyers when consistency is required. The percentage also erodes for low-poverty schools, but not as much. The percentage of low-poverty schools achieving high performance declines from 54.2 percent using the ET definition to 24.2 percent using the 2-2-2 definition. This yields an erosion rate of 55 percent for low-poverty schools, considerably lower than the 93 percent found for high-poverty schools. Some of this erosion may be due to genuine difficulties high-poverty schools have in achieving high performance but the remaining portion, based on the previous evidence and earlier discussion, is due to regression to the mean. It is not possible with these data to separate the two effects.

The results from Table 3 suggested that low-poverty schools were three times as likely to be high-performing compared with high-poverty schools. Table 4 suggests that

this number rises quickly when the performance definition requires more consistency. The definition requiring the most consistency (row 9), which nearly eliminates the effect of regression to the mean, suggests that low-poverty schools are 22 times as likely to be high-performing (24.2/1.1).   The intuition behind this change is straightforward: both types of schools are less likely to be high-performing with the more restrictive definition, but the rate of erosion is higher for high-poverty schools, so the ratio of the two numbers increases.

Table 4 also provides evidence regarding schools that are high-poverty *and* high-minority. The initial portion of high-poverty, high-minority schools that are high-performing is smaller than for high-poverty-only schools, consistent with the results in Tables 1 and 2.   The rate of erosion is also higher here, reaching 97 percent. Further, the likelihood that a low-poverty-low-minority school is high-performing is 89 times greater than for a high-poverty-high-minority school (26.7/0.3).

Accounting for Proficiency Definitions

This section tests whether the performance of high-poverty schools is sensitive to the cut score. Table 5 compares the math achievement for schools in Michigan and Florida, two states that reported results for each school using multiple cut scores.[16] Notice, first, the differences in the levels of the cut scores across the two states. For these years, Michigan had relatively low cut scores, allowing high percentages of schools in all poverty categories to reach high performance, even with the highest cut score. For example, in the average high-poverty elementary school in Michigan, 85 percent of

students reached the lowest cut score. This suggests that Michigan's cut scores are even lower than cut score 1 in Figure 1.

Florida, in contrast, uses more cut scores and has a much wider range of students passing. The lowest cut score appears more like cut score 1 in Figure 1, while the highest cut score is more like cut score 2. These definitions are of course somewhat arbitrary in both states and the point here is simply to illustrate the influence of these choices.

The last column in Table 5 most clearly illustrates the point that the achievement gap appears largest when using cut scores that are closest to the intersection of the test score distributions for advantaged and disadvantaged students. In Michigan, the difference in the percentages of students passing between low- and high-poverty schools is relatively low with the lowest cut score, but the gap widens when moving to the highest cut score—that is, closer to the intersection of the two distributions of scores shown in Figure 1.

[TABLE 5]

A similar pattern is observed in Florida, when shifting from the lowest to the middle cut score. Interestingly, the difference between low- and high-poverty schools decreases again when shifting from the middle to the highest cut score. The apparent reason is that the middle cut score is near the intersection of the advantaged and disadvantaged student distributions—where the gap is greatest. Shifting from the middle to the highest standard therefore shifts the cut score away from the intersection of the test score distributions for advantaged and disadvantaged students. Thus, the results from

Florida also reinforce Rothstein's point, although the point is made somewhat differently because of the wide range of scores used in that state.

There are various other interesting patterns in the data but, like those above, it is difficult to draw conclusions about what explains the patterns. For example, there are substantial differences across states in the likelihood that high-poverty schools reach high-performance, but it is difficult to know whether these are due to the proficiency cut scores, differences in schooling or differences in student disadvantage across states.[17] There are also patterns in resources. For example, high-poverty urban schools have much lower school resources—larger class sizes (17.3 versus 16.4) and smaller school sizes (637 versus 459)—but there is some controversy about the precise causal effects of these factors on achievement.[18]

Conclusion: High Flyers and NCLB Revisited

This study has demonstrated three main flaws with the recent reports on high flying schools. First, there are methodological concerns, particularly the weak performance measure that does not require consistently high test scores. The number of high flyers obviously drops with more restrictive definitions but, more importantly, so does the likelihood that a high-poverty school reaches high-performance compared with low-poverty schools. With more stringent definitions of performance, low-poverty schools are 22 times more likely to be high-performing than high-poverty schools and low-poverty, low-minority schools are 89 times more likely to be reach this performance level compared with their high-poverty, high-minority counterparts. Also, the discussion of the Heritage Foundation report suggests that even those few schools that appear to be

high flyers likely have restrictive admissions policies that make the size of the disadvantages appear larger than they are or additional outside resources that make the that place the school in an unusually strong position to address them. Some schools are certainly better than others, and it is worth learning from the best, but extremely few are able to reach the top-third in the state.

The results of the ET reports have been interpreted, both by those involved in their publication and by others in the national media, as evidence that educators are not trying hard enough or that they have racist views. This may be true for some educators, but the logic of the argument is flawed. No matter what number of high flyers one adopts, it would say little about the role of schools or of students' social and economic disadvantages. The data and analysis here and in the ET reports are simply not well-suited to this task. Better evidence on the subject, dating back several decades, suggests that student disadvantage is the primary, although certainly not the only, cause of the inequity in educational outcomes. Arguably the strongest evidence arises from studies showing that students start school far behind. There is no dispute that these particular differences are due to student background and not to schools, though schools do have an influence over changes that occur after students enter schools.

Perhaps the greatest concern with these reports, and their downplaying of student disadvantage, is that they reinforce the same flawed assumption of NCLB. The law does not reward schools for what they can control and history shows that this is likely to hinder rather than help improvement. No system based on false assumptions is likely to yield the best outcomes. But this mistake is especially unfortunate in the present set of circumstances where there is obvious room for school improvement and evidence that

accountability has the potential to facilitate that improvement. One can be hopeful that this mistake in NCLB will be addressed soon. In the meantime, those who suffer most will be the students.

References

Booher-Jennings, Jennifer. 2005. "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal* 42(2): 231-68.

Brooks-Gunn, Jeanne and Greg J. Duncan. 1997. "The Effects of Poverty on Children." *The Future of Children* 7(2): 55-71.

Brooks-Gunn, Jeanne, Pamela K. Klebanov, Judith R. Smith, Greg J. Duncan, and Kay Lee. 2003. "The Black-White Test Score Gap in Young Children: Contributions of Test and Family Characteristics." *Applied Developmental Science* 7(4): 239-252.

Camara, William, and Elizabeth Schmidt. 1999. "Group Differences in Standardized Testing and Social Stratification." College Board Rep. No. 99-5. New York, NY: College Entrance Examination Board.

Carter, Samuel C. 2000. *No Excuses: Lessons from 21 High-Performing, High-Poverty Schools*. Washington, DC: The Heritage Foundation.

Christoffersen, John. "Study Identifies Strong Schools With Poor Students." *USA Today*. December 29, 2001.

Coleman, James. 1966. *Equality of Educational Opportunity*. Report OE-38000. Washington, DC: U.S. Department of Health Education and Welfare, Office of Education.

Conger, Katherine J., Rand Conger, and Laura V. Scaramella. 1997. "Parents, Siblings, Psychological Control and Adolescent Adjustment." *Journal of Adolescent Research* 12(1): 113-138.

Duncan, Greg J. and Jeanne Brooks-Gunn. 2000. "Family Poverty, Welfare Reform, and Child Development." *Child Development* 71(1): 188-196.

Edmunds, Ronald. 1979. "Some Schools Work and More Can. *Social Policy* 9: 28-32.

Education Trust. 1999. *Dispelling the Myth: High-Poverty Schools Exceeding Expectations*, Washington, DC.

Education Trust. 2002a. *Dispelling the Myth Revisited: Additional Tables*. Washington, DC.

Education Trust. 2002b. *Dispelling the Myth . . . Over Time*. Washington, DC.

Evers, William. Brainstorm NW Magazine. February, 2002.

Figlio, David. 2003. "Testing, Crime and Punishment," *NBER Working Paper*, Cambridge: National Bureau of Economic Research.

Gamoran, Adam. 2001. "American Schooling and Educational Inequality: A Forecast for the 21st Century." *Sociology of Education* 74: 135-153.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. 1996. "Aggregation and the Estimated Effects of School Resources," *Review of Economics and Statistics* 77(2): 611-628.

Hill, Paul T., Lawrence C. Pierce, and James W. Guthrie. 1997. *Reinventing Public Education: How Contracting Can Transform America's Schools*. Chicago: University of Chicago Press.

Jacob, Brian and Steven Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 118(3): 843-877.

Jerald, Craig. 2001. *Dispelling the Myth: Revisited*. Washington, DC: The Education Trust.

Kane, Thomas and Douglas Staiger. 2001. "Improving School Accountability Measures," *NBER Working Paper #8156.* Cambridge, MA: National Bureau of Economic Research.

Ladd, Helen F. and Walsh, Randall P. 2002. "Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right. *Economics of Education Review* 21(1): 1-17.

Lee, Jaekyung. 2002. "Racial and Ethnic Achievement Gap Trends: Reversing the Progress Toward Equity?" *Educational Researcher* 31(1): 3-12.

Lee, Valerie E. and David T. Burkham. 2002. *Inequality at the Starting Gate*. Washington, DC: Economic Policy Institute.

Lee, Valerie E. and Julia B. Smith. 1994. "Effects of High School Restructuring and Size on Gains in Achievement and Engagement for Early Secondary School Students." Madison, WI: Wisconsin Center for Education Research.

Nathan, Joseph. "Successful Schools Share Many Traits." *St. Paul Pioneer Press*. January 6, 2002.

Oakes, J., Joseph, R., and Muir, K. 2004. Access to achievement in mathematics and science: Inequalities that endure and chance. In J.A. Banks and C.A.M. Banks (Eds.), *Handbook of research on multicultural education* (2nd ed., 69-90). San Francisco: Jossey-Bass.

Petrilli, Michael J. and Frederick M. Hess. 2006. "Whither the Washington Consensus?" *The Education Gadfly*. Downloaded June 1, 2006 from http://www.edexcellence.net/foundation/gadfly/archive.cfm.

Phillips, Meredith, James Crouse, and John Ralph. 1998. "Does the Black-White Test Score Gap Widen after Children enter School? In Christopher Jencks and Meredith Phillips (Eds.). *The Black-White Test Score Gap* (pp. 229-272). Washington, DC: Brookings.

Rothstein, Richard. 2004. *Class and Schools*. New York: Teacher's College Press.

Schemo, Diane J. "School Defies Odds and Offers a Lesson." *New York Times*, December 17, 2001.

Stanton, Billie. "Challenged Schools Score Big Victories: Poor, Minority Students Buck Trends and Excel." *Denver Post*, December 16, 2001.

Wong, Kenneth, Dorothea Anagnostopoulos, Stacey A. Rutledge, Laurence Lynn, and Robert Dreeben. 1999. "Implementation of an Accountability Agenda: Integrated Governance in the Chicago Public Schools Enters its Fourth Year." Department of Education and Irving B. Harris Graduate School of Public Policy Studies, University of Chicago.

Table 1: Summary statistics

State	Number of Useable School Obs.	Actual Number of Schools	Poverty (%)	Minority (%)	
				Black	Hispanic
Alabama	1,148	1,516	49.9	39.5	1.1
Alaska	343	502	45.8	3	2.2
Arizona	693	1,570	40.8	4.3	33
Arkansas	813	1,109	51.5	20.8	2.8
California	6,849	8,343	47.4	8.3	38.9
Colorado	1,153	1,560	33	5.4	22
Connecticut	745	1,104	26.3	13.6	12.6
Delaware	98	201	34.1	29.4	5.3
District of Col.	139	164	63.8	83.7	9.3
Florida	2,442	3,111	49.7	27	15.7
Georgia	1,395	1,843	51.3	41.1	4.2
Hawaii	195	254	40.6	2.6	4.6
Idaho	482	660	37.1	1.6	8.5
Illinois	3,362	4,302	33.6	18.6	10.4
Indiana	1,517	1,958	32.9	11.3	3.1
Kansas	1,272	1,437	34.8	6.5	6.9
Kentucky	454	1,534	59	6.7	0.5
Louisiana	1,113	1,510	66.1	49.2	1.2
Maine	521	720	36.8	0.9	0.5
Maryland	1,018	1,357	36	37.1	4.1
Massachusetts	1,533	1,900	27.1	9	10.1
Michigan	3,142	3,914	33.5	18.7	3.5
Minnesota	1,093	2,348	32	5.9	3
Mississippi	574	1,015	67.2	55.2	0.6
Missouri	1,458	2,328	38.5	18.3	1.5
Montana	575	886	35.6	0.5	1.5
Nebraska	151	1,352	38.7	1.6	5
Nevada	346	469	35.2	9.6	7.5
New Hamp.	310	516	18.6	1	1.3
New Jersey	1,847	2,318	29	18	14.1
New Mexico	621	746	58	2	50
New York	3,007	4,230	42.5	17	14.1
North Carolina	1,471	2,106	48.1	31.9	4.1
North Dakota	410	595	35.1	0.5	1.1
Ohio	1,865	3,852	37.3	17.2	1.7
Oklahoma	1,124	1,825	55.7	10.2	5.4
Oregon	1,057	1,275	38.2	2.5	9.3
Pennsylvania	2,545	3,205	31.2	13.7	3.5
Puerto Rico	172	1,300	71	0	100
Rhode Island	226	318	34.5	7.1	11.1
South Carolina	728	1,101	54.5	45.7	1.7
Tennessee	1,000	1,589	41.1	22.6	0.2
Texas	5,355	7,228	46.8	13.3	38
Utah	656	769	34	0.9	8.8
Vermont	257	395	28.6	1	0.5
Virginia	1,340	1,918	36.8	26.5	4.3
Washington	1,370	2,234	33	5.4	10.1
Wisconsin	1,807	2,109	25.2	8.1	3.4
Wyoming	282	388	32.2	1	6.7
U.S. Avg./Total	62,074	88,984	41.0	15.8	10.4

Source: Author's calculations from SLAD database.

Table 2: Test score information, by state

State	Main Scores Use in Analysis			Cut Scores		Multiple Scores	
	Test Type	Grades	Year	# Cut Points	Grades	Grades	Additional Year
Alaska	4thQ	4,8	2000	2	4,8	---	---
Alabama	PCT	5,8,11	2000	0	---	3,7,10	1999
Arkansas	ASC	4,8	2000	0	---	---	---
Arizona	CP2	5,8,11	2000	3	5,8,11	---	---
California	ASC	5,8,11	2000	3	5,8,11	3,7,10	1999
Colorado	CP2	R: 4,7 M: 5,8	2000	3	R: 4,7 M: 5,8	---	---
Connecticut	RAW	4,8	2000	1	4,8	---	---
DC	CP2	El, jh, mi, hs	2000	3	El, jh, mi, hs	---	---
Delaware	ASC	5,8,10	2000	4	5,8,10	3	1999
Florida	ASC	5,8,10	2000	4	5,8,10	---	---
Georgia	PCT	5,8	2000	2	4,8	3	1999
Hawaii	CP2	6,8,10	1998	2	6,8,10	3	1997
Idaho	PCT	5,8,11	2001	0	---	3,7,10	2000
Illinois	ASC	3,8,10	1998	2	5,8,10	6	1997
Indiana	NCE	3,8,10	2000	1	3,8,10	---	---
Kansas	RAW	R: 3,7,10 M: 4,7,10	R: 1999 M: 2000	4	R: 5,8,11 M: 4,7,10	---	---
Kentucky	ASC	6	2000	0	---	---	---
Louisiana	CP2	4,8	2000	4	4,8	---	---
Massachusetts	ASC	4,8,10	2000	3	4,8,10	---	---
Maryland	CP2	5,8	2000	2	5,8	3	1999
Maine	ASC	4,8,11	2000	3	4,8,11	---	---
Michigan	CP2	4,7,11	2000	2	4,7,11	---	---
Minnesota	ASC	5,8	2000	3	5,8	3	1999
Missouri	CP2	4,8,10	2000	4	4,8,10	---	---
Mississippi	PCT, CP2	C: 5,8 W:4,7	2000	4	4,7	---	---
Montana	NCE	4,8,11	1998	0	---	---	---
Nebraska	CP2	5,8,11	2000	3	5,8,11	---	---
Nevada	PCT	4,8,10	1999	2	4,8,10	---	---
New Hamp.	ASC	3,10	2000	3	3,10	6	1999
New Jersey	ASC	4,8,11	1998	3	4,8,11	---	---
New Mexico	ASC	5,8,10	1997	0	---	3	1996
New York	CP2	4,8	2000	3	4,8	---	---
North Carolina	ASC	5,8	1998	3	5,8	3,7	1997
North Dakota	PCT	6,8,10	2000	---	---	4	1999
Ohio	CP1	4,12	2000	1	4,12	---	---
Oklahoma	CP1	5,8	2000	3	5,8	---	---
Oregon	CP2	5,8,10	2000	2	5,8,10	3	1999
Pennsylvania	ASC	5,8,11	2000	2	5,8,11	---	---
Puerto Rico	CP2	3,11	2000	3	3,11	---	---
Rhode Island	CP1	4,8,10	2000	1	4,8,10	---	---
South Carolina	ASC	5,8	2000	3	5,8	3,7	1999
Tennessee	ASC	5,8	1998	0	---	3,7	1997
Texas	CP1	5,8,11	2000	1	5,8,11	3,7	1999
Utah	PCT	5,8,11	2000	0	---	---	---
Vermont	ASC	4,8,10	2000	0	---	---	---
Virginia	ASC	5,8	1999	1	5,8	3	1998
Washington	CP2	4,7,10	2000	4	4,7,10	---	---
Wisconsin	CP2	4,8,10	2000	4	4,8,10	---	---
Wyoming	CP2	4,8,11	2000	3	4,8,11	---	---

Table 3: Poverty, Race, and Achievement Characteristics of U.S. Public Schools
(ET definitions of high-performance and high-poverty)

Category	Number of Schools (% total sample)	Performance category (% schools)	
		Low	*High*
Low-Poverty	40,830 (66%)	46%	*54*
High-Poverty	21,234 (34%)	84	*16*
Total	*62,064 (100%)*		
Low-Poverty-Low-Minority	38,104 (61%)	*43*	*57*
High-Poverty-High-Minority	12,869 (21%)	*90*	*10*
Total	*50,973 (82%)*		

Notes: Author's calculations from SLAD database. A low-poverty school is one with fewer than 50 percent of students eligible for free or reduced price lunches. Low-minority is defined analogously.

Table 4: Consistency of High-Performance
(% of schools meeting various criteria for high-performance; multiple scores sample)

Row	Criteria			% **high-poverty** schools that are high-performing	% **low-poverty** schools that are high-performing	% **high-poverty, high-minority** schools that are high-performing	% **low-poverty, low-minority** schools that are high-performing
	Years	Subjects	Grades				
1	ET Definition			15.6	54.2	10.4	56.7
2	1	1	1	30.5	80.0	22.0	84.0
3	2	1	1	12.9	59.1	7.5	63.5
4	1	2	1	14.7	62.3	9.1	66.8
5	1	1	2	11.0	56.5	6.4	60.9
6	2	2	1	4.5	41.0	2.0	44.8
7	2	1	2	3.6	37.9	1.4	41.4
8	1	2	2	2.4	33.2	0.9	36.4
9	2	2	2	1.1	24.2	0.3	26.7
10	Erosion (from row 1 to row 9)			93 %	55 %	97 %	53 %

Source: Author's calculations from SLAD database.

Table 5: Role of the State Proficiency Definitions, Individual States

Cut Scores	Average % of students reaching cut score in **high-poverty** schools	Average % of students reaching cut score in **low-poverty** schools	Difference
	(a)	(b)	(b)-(a)
Michigan			
Elementary			
Lowest Cut Score	85.0	94.4	9.4
Highest Cut Score	63.2	80.8	17.6
Middle			
Lowest Cut Score	70.2	89.2	19.0
Highest Cut Score	38.7	67.3	28.6
Florida			
Elementary			
Lowest Cut Score	65.6	83.4	17.8
Middle Cut Score	34.7	56.3	21.6
Highest Cut Score	14.3	28.8	14.5
Middle			
Lowest Cut Score	56.5	79.2	22.7
Middle Cut Score	34.6	59.8	25.2
Highest Cut Score	12.6	28.6	16.0

Source: Author's calculations from SLAD database.

Notes

[1] Over the course of the past half century, the United States has made remarkable progress in reducing the gap in achievement between students from social and economic advantages and students from more disadvantaged backgrounds. Substantial and steady gains were made from the 1960s through 1980s (Phillips, Crouse, and Ralph 1998). In the 1990s, however, this "achievement gap" actually started increasing again (Harris and Herrington, 2006; Lee 2000). While these inequities began to narrow again in the early 2000s, it is unclear whether this is part of a longer term reversal of the 1990s' trend or simply a short-term anomaly. But the gaps remain large: between whites and African-Americans, the size of the achievement gap ranges from 0.80-1.14 standard deviations; between whites and Hispanics, the gap is 0.40-1.00 standard deviations (Camara and Schmidt 1999; Phillips, Crouse and Ralph 1998). Camara and Schmidt consider the SAT, ACT, MCAT, LSAT, and GRE, all of which relate to older students including high school students an adults. Phillips, Crouse, and Ralph focus on NAEP, which considers students of the age appropriate to K-12 schooling.

[2] Petrilli and Hess (2006) include the importance of achievement gaps as one part of the "Washington Consensus" that, by definition, cuts across political party lines.

[3] Many students do participate in pre-kindergarten programs. According to Lee and Burkham (2002), 56 percent of African-American students participated in either the federal Head Start Program or other formal "center-based" programs. The numbers is nearly identical for whites, although a lower percentage of whites use Head Start compared with other center-based programs. In any event, these programs pre-kindergarten programs are rarely provided by the traditional public schools that are being discussed. Therefore, while the achievement gaps in kindergarten might be influenced by differences in pre-K program quality, this is somewhat irrelevant to the topic at hand.

[4] NCLB is ambitious in its goals for students who currently have low scores, but not for students are already above state proficiency standards.   NCLB provides no incentive for schools to improve achievement for those higher-scoring students.

[5] Requiring schools achieve at some minimum annual gain would be relatively easy, given the other provisions of NCLB.   The law requires states to test students in grades 3-8. This means that gains could be calculated in grades 4-8 because tests are generally given in toward the end of the school year. Grade 3 would therefore serve as a baseline for the calculation of grade 4 gains.

[6] The "myth" they refer to is that students in poverty cannot learn. This is closely related to the point in the previous section regarding the false choice being proposed by the ET reports.

[7] In addition to nine of the 21 schools having admission requirements that could exclude students who have received low test scores, a tenth school, Kew Elementary School in Inglewood, California, housed a gifted and talented program that also may have had admission requirements. Of the remaining schools, four required tuition, suggesting that their parents may have access to financial resources not reflected in free and reduced lunch eligibility. The fact that low-income families are willing to find such resources also suggests that these parents value education more than the typical parent, which may also mean that they have exceptional skills and experiences required to help their children excel. Another of the remaining schools is located in Cambridge, Massachusetts – home of two of the most prestigious universities in the nation, Harvard and MIT. Large portions of the students in these schools were children of highly educated graduate students whose low incomes were only temporary circumstances related to their education programs. These parents were clearly able to provide resources necessary for their children to succeed. Another of the remaining schools, 14th Avenue Elementary School in New Jersey, had a pupil-teacher ratio of just 14. Portland Elementary School in Arkansas has only 152 students, far fewer than the average school, and is located in a rural area. The results presented later in this report suggest that schools with these characteristics, which educators have little control over, are much more likely to be high-performing. The same is probably true of Newberry Elementary in Detroit, which had after-school programs four days

per week, plus a "summer learning academy," providing a substantial amount of additional time for learning.

[8] Here are several other examples of how the ET report has been interpreted in the press: "The Education Trust's new report proves beyond dispute that poor and minority children can achieve at high levels – provided educators have high expectations for these students" (Stanton 2001). In this case, the only requirement for having poor and minority students achieve at high levels is that educators have high expectations for students. Nothing is required of the home environment. Another commentator writes, "This report shows that we don't have to accept low achievement" (Nathan 2002). The assumptions of this quote are somewhat more vague, but it does clearly assume that someone is accepting low achievement. Given that Nathan is referring to an ET report that focuses specifically on what schools can do, it would certainly appear that he is referring to educators as the ones who are accepting low achievement. A school board member in quoted as saying that the report is "further proof that all kids can learn" (Christoffersen 2001). As shown later in the text, the ET reports provide almost no evidence about whether students can learn. Finally, the Haycock quote in the text can also be found in almost identical form in an article in the *New York Times* (Schemo 2001).

[9] This also suggests that schools with more students will experience less noise and therefore be less affected by regression to the mean. Further, because elementary schools tend to have fewer students taking any given test, regression to the mean especially influences the average scores of elementary schools relative to middle and high schools.

[10] One assumption is that only the mean of the distribution differs between advantaged and disadvantaged students. (Notice in the figure that the two distributions have identical "bell" shapes.) It is also important to point out that these shifts in the relative effects on the two groups are larger when the distributions are more distinct from one another. If the distributions largely overlapped, this would not be a significant issue.

[11] According to the Digest of Education Statistics, there were 85,915 regular public schools in the United States in 1998-99. (The "regular" category excludes 3,532 schools that focus on special, vocational, or alternative education.) Of the regular public schools, 63,574 were elementary schools and 18,571 were secondary. In addition, there are 3,770 schools that have unusual grade spans, such as K-12. For the same year, there were 45.8 million students in the 85,915 regular public schools, compared with 35.9 million students in the AIR data sample.

[12] Free and reduced lunch (FRL) eligibility is the only measure of student economic well being available in this analysis. Eligibility for reduced priced lunches requires income less than 185% of the federal poverty rate. Free lunch eligibility requires income less than 130% of the poverty rate. Four states have reduced lunch data in the SLAD, but not free lunch data (Arizona, Illinois, Tennessee, and Washington). Therefore, reduced lunch eligibility was imputed based on the relative number of free and reduced lunch students in the average state. This was done by regressing reduced lunch data on free lunch and the other variables for the states that had both, using the variables in specification (2) of table 1d (except poverty). I then used this equation to estimate the number of reduced lunch students in the states that had only free lunch. Students eligible for reduced lunches are added to those eligible for free lunches, even though each type reflects a different level of family income.

[13] School level determinations are made as follows. The first step for each school is to identify the highest grade in which a test score is reported. If this highest grade was 6 or lower, the school is categorized as "elementary"; when the highest grade is 8 or lower, the school is labeled "middle"; when the highest grade above 8, the school is considered "high." The next step is to identify the minimum grade at which a score is reported. If a middle school has a grade less than 6, then a separate variable was created identifying the school as "multi-level" and given the value of one. The same is true for high schools reporting scores in grades less than 9.

[14] In addition to the notes in the text, Mississippi had only "composite" and "writing" scores, which I treated as math and reading, respectively. The District of Columbia reported scores without grade levels. Instead, they used elementary ("el"), junior high ("jh"), middle school ("mi"), and high school ("hs"). In Colorado, Kansas, and Mississippi, the grade level of the test depended on the subject matter (R = reading, M = math, C = composite, W = writing). In New Jersey, I used the lowest cut point, instead of the second lowest cut point so that more grades and schools could be included.

[15] Recall that ET only looked at data from a single year and two subjects. Therefore, they consider a total of only two scores for each school. The analysis in Table 4 uses the multiple scores sample and therefore considers eight scores per school. Requiring high performance in only of the eight is a weaker definition than requiring only one of two as in the ET report.

[16] For Michigan, I used mathematics achievement in grade 4 and 7. For Florida, I used mathematics achievement in grades 5 and 8. High school results are excluded for sake of space.

[17] The results might also differ across states because high performance is defined as whether a school reaches the top third in the state. This means that the percentage of high-poverty schools reaching this level will depend, in part, on the proportion of all schools in the state that are high-poverty. To see why this is true, consider the hypothetical situation in which a state has only high-poverty schools. In this case, the percentage of high-poverty schools that will be high-performing according to any single test is predetermined—33 percent—because of the definition of high-performance. More generally, higher proportions of low-poverty schools imply more competition, which decreases the percentage of high-poverty schools that are high-performing.

[18] These calculations are available from the author upon request.