This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Tax Policy and the Economy, Volume 17

Volume Author/Editor: James M. Poterba, editor

Volume Publisher: MIT Press

Volume ISBN: 0-262-16220-2

Volume URL: http://www.nber.org/books/pote03-1

Conference Date: October 8, 2002

Publication Date: January 2003

Title: Fiscal Implications of School Accountability Initiatives

Author: David N. Figlio

URL: http://www.nber.org/chapters/c11533

FISCAL IMPLICATIONS OF SCHOOL ACCOUNTABILITY INITIATIVES

David N. Figlio
University of Florida and NBER

EXECUTIVE SUMMARY

This paper describes many of the fiscal implications to states and school districts of the No Child Left Behind Act of 2001. While considerable ambiguities exist in the current law and states will implement the law differently, predictable consequences exist for the school finance system. Moreover, the nature of these consequences is likely influenced by the actual implementation of the law. There are both direct and indirect fiscal implications of the federal education reforms. Direct fiscal consequences can be substantial for certain types of school districts and for school funding equity within states, and depend crucially on design considerations in the implementation of the law. Indirect fiscal implications, through consequences for the special education placement and grade retention systems, input prices, and house prices, can be substantial as well, and also depend on the implementation of the policy. This paper also presents some key design issues for states to consider.

1. INTRODUCTION

The centerpiece of President George W. Bush's domestic policy agenda during his presidential campaign was education reform, and at the heart of his education reform proposals was a national system of test-based

school accountability. An amended version of this accountability initiative, codified into law as the No Child Left Behind Act of 2001 and currently in its initial stages of implementation, sets performance goals for every school, prescribes public school choice for students attending schools identified as persistently failing, and outlines a set of rewards and sanctions for schools and districts that do not meet their performance goals. School accountability, defined broadly, is favored by people across the political spectrum, and the No Child Left Behind Act of 2001 is merely the latest policy initiative in a nationwide trend toward school accountability that is already over a decade old. Indeed, by the 2001–2002 school year, 31 states already explicitly rated schools based on some measure of student test performance. By mandating annual student assessments in grades 3 through 8, requiring states to reward and sanction schools based on changes over time in these assessments, and tying federal education funding to performance, the new federal law has further elevated the emphasis on school accountability nationwide.

In spite of general consensus that school accountability in some form is desirable and important, considerable debate remains regarding key questions involved in implementing an accountability system. People differ substantially on the extent to which students should be tested, the means of assessing performance, the coverage and frequency of these examinations, and the ways in which student performance should reflect on schools. In addition, considerable controversy exists regarding the degree to which explicit rewards and sanctions should be employed, or whether the provision of information is sufficient to achieve the goals of accountability systems. While some aspects of these questions have been resolved (at least temporarily) through the No Child Left Behind Act, most of the issues relevant for the debate over federal education policy are still relevant to states in their implementation of the law.

The principal purpose of this present paper is to describe many of the fiscal implications to states and school districts of the No Child Left Behind Act. Despite considerable ambiguities in the current law and the fact that states will implement the law differently, predictable consequences will occur for the school finance system. The nature of these consequences is likely influenced by the actual implementation of the law. This paper therefore describes both the direct and indirect fiscal implications of the federal education reforms and, when appropriate, discusses how design and implementation details influence these outcomes.

2. FEDERAL SCHOOL ACCOUNTABILITY POLICY

The No Child Left Behind Act reauthorized the Elementary and Secondary Education Act (ESEA) of 1965, incorporating significant new accountability measures. These reforms are linked to funding through Title I of the ESEA, a program designed to improve the educational outcomes of low-income students. Because of this funding mechanism, the federal accountability reforms are relevant primarily for schools and districts serving disadvantaged student populations. Under this law, each state is required to implement a statewide accountability system in which students in grades 3 through 8 are tested annually in reading and mathematics, and ultimately also in science, with assessments based on the state's academic standards. Hence, the curricular requirements and level of standards considered under one state's accountability system may differ substantially from the same requirements and standards of another state's system. Under the law, each state must use this accountability system to assign rewards and sanctions to schools and school districts that receive Title I money. It is up to the discretion of each state whether it will assign rewards and sanctions to non–Title I schools, although every school, regardless of whether it receives federal education funding, is subject to the accountability system put in place by the state. The goal is for each student in the state to meet that state's proficiency standards by the 2013–2014 academic year.

Schools are evaluated based on whether the school makes "adequate yearly progress" in a given year. "Adequate yearly progress" is defined recursively: every student is expected to be proficient by 2013–2014, which sets an ultimate performance target for that year. To calculate proficiency targets in years prior to that, each state determines a starting value for the 2001–2002 school year, based on the distribution of proficiency rates in the state. Between those two school years, a state's proficiency goals for each school will increase in equal increments occurring— at most—three years apart (with the first increment occurring no later than 2003–2004). Schools are then deemed as making adequate yearly progress if their proficiency rates meet or exceed this stated proficiency target.

However, meeting this target for the entire student test population (which must be at least 95 percent of eligible test-takers) is not sufficient to make adequate yearly progress. Rather, every measurable subgroup— economically disadvantaged students, students from major racial and ethnic groups, students with disabilities, and students with limited English

proficiency—must meet the proficiency target for a school to make adequate yearly progress. (Schools with a very small number of students in a given subgroup are not required to meet proficiency targets for that group.) If a given subgroup misses its proficiency target, it is still possible for a school to make adequate yearly progress if each target-missing subgroup improves sufficiently from one year to the next. Here, "sufficient improvement" is defined as the percentage of students below the proficiency level in that subgroup falling by at least 10 percent compared to the preceding year, as long as that subgroup improved along other lines, such as in terms of graduation or promotion to a higher school level. While states must assess each school annually, they may use as many as three years' worth of data to measure school proficiency, as long as they apply the same criteria to all schools consistently. Schools must meet their proficiency targets in both reading and mathematics to make adequate yearly progress.

The No Child Left Behind Act outlines a series of actions to be taken if a school fails to meet its proficiency targets. If a school fails to make adequate yearly progress in a single year, no action is mandated by the federal government (though states may still decide to take action.) If a school fails to make adequate yearly progress in two consecutive years, however, the school is designated as being "in improvement," and several actions are taken immediately. First, the school must develop a two-year school improvement plan that includes "research-based strategies for improvement," which could include an extended school day or school year, strategies for increasing parental involvement, and new teacher mentoring. As part of this plan, 10 percent of the Title I funds must be set aside for professional development.

In addition, all students attending a school in the improvement phase are offered public school choice. The district must offer students the opportunity to attend a higher-rated public school in the district, with transportation; if a sufficient number of slots are unavailable, then students may attend school in a neighboring school district on a space-available basis (but without provided transportation). Students are permitted to remain in the higher-rated school through that school's terminal grade, even if their original school begins to make adequate yearly progress (in the event that this improvement occurs, they will no longer be provided transportation). In the event that there are not sufficient slots available in higher-performing schools in the district, preference will be given to low-performing, low-income students. If a school remains in the improvement stage for a second year, all previously mentioned remedies will remain in force, but in addition, low-income families in affected schools will be offered supplemental educational services, such as private tutoring. Un-

der the new law, school districts must expend up to 20 percent of their Title I allocations to provide supplemental services and choice-related transportation.

Schools that fail to make adequate yearly progress for four consecutive years enter into the "corrective action" stage of federal sanctions. In this case, school districts must continue to offer public school choice and supplemental services to families, but they must also take one of six actions: (1) replace school staff deemed responsible for the school's continued failure to make adequate yearly progress, (2) implement a new curriculum centered around "scientifically based research," (3) substantially decrease the authority of management at the school, (4) extend the school day or school year, (5) appoint an outside expert to advise the school on its progress toward making adequate yearly progress, or (6) reorganize the school. In the fifth consecutive year of failure to make adequate yearly progress, the school must undergo restructuring in which, for instance, the school district has one year to reopen the school as a charter school, replace its principal and staff, contract with a private management company, or allow the state to take over the school. At any stage, the state may introduce its own sanctions as well; the federally mandated sanctions are minimum requirements.

3. DIRECT CONSEQUENCES FOR SCHOOL FINANCE

3.1 *The Federal Role in School Finance*

While the lion's share of K–12 education funding comes from state and local sources, the federal government plays a significant role in the funding of local public schools, primarily in districts serving large numbers of disadvantaged children. As mentioned above, the federal accountability law ties school performance to Title I funding. Describing in detail the specific mechanism for the delivery of Title I funding is not relevant for this paper, but it can be roughly described as follows. Title I grants are allocated to school districts through several mechanisms. Seventy percent of the $10 billion expended in Title I grants in 2002 are so-called basic grants, which are assigned to school districts on the basis of disadvantaged child counts as long as at least 2 percent of the district's school-age population is disadvantaged. Over 99 percent of school districts are eligible for and receive some basic grant. School districts with higher concentrations of disadvantaged children (at least 15 percent of the school-age population, or at least 6,500 disadvantaged children) are eligible to receive a concentration grant: a nonlinear allocation based on the number of

eligible children in the district. "Education Finance Incentive Grants" make allocations that account for the intensity with which states fund education and the degree of equity in the state's system.

School districts have considerable latitude in determining the use of Title I funds. While these funds are dedicated to improving the academic outcomes of low-performing, disadvantaged children, the mechanisms through which these funds are expended are plentiful. And if sufficiently large numbers of targeted students are present in a school, the funds can be used for schoolwide programs. Of course, to the degree to which money is fungible, in essence these funds could almost be considered grants for schoolwide programs, Gordon (2001) shows that states and localities adjust downward their funding of schools as Title I grants increase. This fact is directly relevant for this paper because federal and state requirements that districts use a large fraction (or perhaps all) of the Title I funds for noninstructional purposes (e.g., school choice–related transportation or privately provided supplemental services) may lead to significant reductions in instructional spending in affected districts.

In 2002, Title I grants per pupil averaged $198 (weighted by total district student membership), or $940 per low-income student. But these averages mask considerable heterogeneity in the magnitude of the grants. Nearly 10 percent of school districts, representing 11 percent of the student population in the United States, average more than $400 per student, and just over 2 percent of school districts, representing one-half of 1 percent of the student population, average more than $700 per student. One-third of school districts, representing 32 percent of the low-income student population, receive over $1,000 in Title I grants per disadvantaged student, and almost 5 percent of school districts, representing one-half of 1 percent of the low-income student population, receive over $1,500 per disadvantaged student.

Title I grants are highly concentrated. Of the over 13,000 school districts in the United States, more than one-quarter of the total expenditures go to the 25 school districts with Title I allocations of more than $30 million, and 40 percent of the total expenditures go to the 114 school districts with Title I allocations of more than $10 million. The top 25 school districts, ranked in terms of total allocation, average $407 in Title I aid per pupil and $1,122 in aid per disadvantaged pupil. The top 114 school districts average $339 per pupil and $1,062 per disadvantaged pupil. Table 1 presents information on the 25 school districts that together receive one-quarter of the total Title I aid. The figures in Table 1 are for 2002, except for the share of total revenues in a district, which is for 1997, the last year for which I have comparable data. Title I comprises a substantial fraction of total current expenditures for each of these school districts, and

TABLE 1
*Title I Receipts in the 25 School Districts Receiving the Most Federal Aid**

School district name	Total Title I aid (in millions), 2002	Title I aid per pupil, 2002	Title I aid per disadvantaged pupil, 2002	Percentage of total school district revenues, 1997
New York City, NY	$633.8	$541	$1,295	10.0
Puerto Rico	333.3	N/A	N/A	N/A
Los Angeles, CA	303.0	379	935	9.7
Chicago, IL	216.6	447	1,188	12.7
Detroit, MI	128.1	693	1,309	10.6
Philadelphia, PA	114.7	496	1,311	11.0
Dade County, FL	97.6	240	882	7.6
Houston, TX	77.0	329	987	9.4
Milwaukee, WI	58.9	527	1,363	10.6
Dallas, TX	52.6	292	962	10.2
Baltimore, MD	51.5	465	1,236	13.0
San Diego, CA	46.7	301	915	7.8
Cleveland, OH	44.2	514	1,133	10.8
Boston, MA	42.3	598	1,234	6.6
State of Hawaii	39.5	196	1,470	7.9
Fresno, CA	37.4	433	907	11.8
Broward County, FL	37.2	136	866	5.9
Orleans Parish, LA	35.8	423	1,030	15.8
District of Columbia	34.9	464	N/A	10.3
Atlanta, GA	34.8	532	971	8.4
Newark, NJ	34.0	716	1,227	7.0
Buffalo, NY	33.0	671	1,230	11.3
Hillsborough County, FL	32.7	183	874	9.3
Memphis, TN	31.4	247	949	10.6
Long Beach, CA	30.0	295	952	11.4

* These data come from the Common Core of Data and Title I Office of the U.S. Department of Education. Cells denoted with N/A indicate that data were unavailable for the relevant year.

presumably a considerably larger fraction of total current expenditure in the schools serving disadvantaged students. The reliance on federal aid ranges dramatically, however, even within this set of highly reliant school districts. Orleans Parish, Louisiana, for instance, has almost three times the federal share of total revenue as does Broward County, Florida. School districts receiving large amounts of Title I aid are not necessarily those that are the poorest; indeed, many of these districts are relatively wealthy. However, these districts tend to have very large concentrations

TABLE 2

*Title I Receipts in the 25 Large School Districts Receiving the Largest Amount of Per Pupil Federal Aid**

School district name	Total Title I aid (in millions), 2002	Title I aid per pupil, 2002	Title I aid per disadvantaged pupil, 2002	Percentage of total school district revenues, 1997
Camden, NJ	$19.7	$981	$1,248	8.7
Hartford, CT	22.2	876	1,251	8.2
Newark, NJ	34.0	716	1,227	7.0
New Haven, CT	15.5	711	1,242	10.6
Detroit, MI	128.1	693	1,309	10.6
Rochester, NY	26.7	684	1,243	10.1
Buffalo, NY	33.0	671	1,230	11.3
Flint, MI	16.4	664	1,250	9.3
Syracuse, NY	15.2	615	1,225	11.4
Bridgeport, CT	14.8	603	1,245	8.3
Boston, MA	42.3	598	1,234	6.6
Springfield, MA	17.2	585	1,260	9.5
Laredo, TX	14.1	567	915	13.0
Lowell, MA	9.8	565	1,284	7.6
Paterson, NJ	15.2	557	1,197	7.5
Jersey City, NJ	19.3	553	1,206	7.6
Compton, CA	18.6	548	934	11.2
New York City, NY	633.8	541	1,295	10.0
Atlanta, GA	34.8	532	971	8.4
Milwaukee, WI	58.9	527	1,363	10.6
Providence, RI	15.7	525	1,239	11.3
Gary, IN	11.1	515	1,113	7.0
Cleveland, OH	44.2	514	1,133	12.4
Lansing, MI	9.7	511	1,230	7.0
St. Louis, MO	25.5	510	1,012	10.4

* These data come from the Common Core of Data and Title I Office of the U.S. Department of Education.

of low-income students and arguably have higher costs of education than many lower-wealth school districts.

Table 2 presents similar information, this time on the 25 school districts with more than 15,000 students that have the highest per pupil Title I receipts. Note the high variation in Title I receipts per pupil: even among the 25 large school districts with the highest per pupil Title I revenues, the highest-revenue district has almost twice as much Title I revenue per pupil as the twenty-fifth highest-revenue district. If a minimum of 20 percent of this Title I allocation is at risk under the No Child Left Behind

Act's accountability system, a sizable portion of expenditures in America's largest school districts might be diverted under the federal accountability law. Because some smaller districts receive more than 10 percent of their total revenues from Title I grants, the potential direct fiscal consequences are greater still for some school districts.

3.2 Which Schools Are Likely to Be Sanctioned?

As mentioned above, the federal accountability system lays out timetables for sanctions against schools that persistently fail to meet performance goals. While states have some flexibility in measuring proficiency, states are restricted to setting the starting value for the accountability system's proficiency goals at the twentieth percentile of the state's distribution. One can therefore make some preliminary forecasts about the attributes of the schools likely to be sanctioned under the law, at least in the early years of the policy.

I employed performance data on the Florida Comprehensive Assessment Test, the standards-based assessment test that Florida will use to comply with the requirements of the federal accountability law. I limited my analyses to the elementary school grades tested throughout the entire time period—fourth grade for reading and fifth grade for mathematics. For the purposes of simulating the effects of the federal law, I assumed that the first year of testing counted for the purposes of accountability was given in the 1998–1999 school year, so that four years of testing would now have been conducted. Hence, I could identify which schools might have been in the third year of "improvement" after the 2001–2002 school year and could also attempt to gauge the effects of the first federally mandated increase in proficiency targets on the distribution of sanctioned schools. Of course, this exercise assumed that schools do not respond to the federal accountability system in ways that would affect student test performance. This assumption is a major weakness of this type of projection because increased test performance is, of course, the purpose of enhanced school accountability. Thus, these projections can be thought of as a first approximation of what might have happened. Test scores for 1998–1999 were the first year of Florida's new accountability system, Governor Jeb Bush's A+ Plan for Education, which imposed potentially harsher sanctions and enhanced rewards based on student test performance. Florida is thus an ideal place to make projections about the potential effects of a national system of school accountability on the distribution of sanctions.

While Florida's experience with an accountability system similar in nature to the federal system currently being implemented makes Florida a desirable place to address this question along many dimensions, Florida

is imperfect in other ways. Florida districts rely less on federal education aid than do comparable school districts in much of the rest of the country. More important, schools in Florida districts tend to be more racially and ethnically heterogeneous than in the United States in general. However, patterns of test performance in Florida match those in other states, such as California, North Carolina, and Texas, which suggests that Florida data are usable for the present purposes. Because Florida districts tend to rely less on federal aid than do similar districts elsewhere in the country, one consequence of using Florida data for the present study is that the magnitudes of the forecasted effects of the federal accountability system are likely to be understated. Under conservative assumptions, within three years, many school districts in Florida would lose as much as 1 percent or more of their current expenditure as a result of federally mandated sanctions. In other parts of the United States, where schools are more reliant on federal aid, similar school districts may lose as much as 2 percent of current expenditures. I should point out, however, that these projections are highly speculative because most of the factors that would ultimately determine the incidence of the federally mandated sanctions have yet to be decided.

For a first pass through the data, I assumed that every subgroup within a school has the same proficiency rates as the school average. (I was forced to assume so because I did not have sufficient subgroup-level data for all four years for the entire state. Later I relaxed this assumption and used a subset of the state where I had adequate data to clarify this question a bit more.) To the extent to which this assumption is not valid, it will understate the number of schools that are sanctioned under the program. The No Child Left Behind Act requires not only that the school, as a whole, meet proficiency standards but also that each subgroup within the school meet the same standards. If any subgroup fails to meet the standards, the entire school is deemed as not making adequate yearly progress.

Because there is flexibility at the state level about whether to use a high or low standard for defining student proficiency, for the purposes of these simulations I investigated two different proficiency standards—one in which students need to reach level 2 out of 5 on the Florida Comprehensive Assessment Test to be deemed proficient, and one in which students must reach level 3 on the test to be deemed proficient. The target level for the first year (1998–1999) would be 48 percent proficient in reading and 51 percent proficient in mathematics if the lower of the two standards were employed, and 31 percent proficient in reading and 18 percent proficient in mathematics if the higher of the two standards were employed. These targets would then rise to 57 percent and 59 percent, respectively,

TABLE 3

Attributes of Florida Schools Forecast to Be Sanctioned if the No Child Left Behind Act Took Effect in 1999*

	Percentage black	Percentage Hispanic	Percentage free-lunch-eligible	Per pupil Title I aid, 2002
Title I schools	36.1	22.4	58.5	$850
Using more lenient definition of proficiency				
Schools sanctioned in first year of simulations	60.5	27.7	77.4	$860
Schools sanctioned in second year of sanctions	58.9	29.0	77.3	$866
Schools sanctioned in third year of sanctions	53.9	26.8	72.7	$866
Using more stringent definition of proficiency				
Schools sanctioned in first year of simulations	63.6	25.2	77.9	$866
Schools sanctioned in second year of sanctions	62.2	25.5	76.7	$865
Schools sanctioned in third year of sanctions	50.5	24.7	69.9	$860

* These figures describe the demographic and socioeconomic attributes of Florida schools simulated to be sanctioned in each of the first three years of the federal accountability law if the law took effect in 1999. The figures are a result of simulations of the federal accountability law in Florida using school-level data from Florida from 1998–1999 through 2001–2002. Two proficiency targets were used, based on different choices that the state of Florida could make.

if the low standard were employed after the first incremental increase in school standards, and to 43 percent and 32 percent, respectively, if the higher standard were employed.

Table 3 presents some basic attributes of the schools that would be sanctioned under these scenarios in each of three years—1999–2000, 2000–2001, and 2001–2002. For the purposes of this analysis, I did not distinguish between the types of sanction employed, be it public school choice or mandatory provision of supplemental services. The next version of this paper will make this distinction if the differences warrant it. From the table, you can see that, regardless of the standard employed, the schools likely to be sanctioned are particularly disadvantaged, even relative to the population of disadvantaged Title I schools eligible for sanction. For instance, while the typical Title I school in Florida has a student body that is 59 percent free-lunch-eligible, 37 percent black, and 22 percent Hispanic, the typical school predicted to be sanctioned (using the low measure of standards) in the first year has a population that is 77 percent free-lunch-eligible, 61 percent black, and 28 percent Hispanic.

Almost the same attributes of sanctioned schools are apparent if the higher level of student standards is employed. Using this standards measure, the typical school forecast to be sanctioned has a student body that is 78 percent free-lunch-eligible, 64 percent black, and 25 percent Hispanic.

As the policy progresses and standards ratchet up, however, this gap between the typical Title I school and the typical sanctioned school is projected to tighten, although it will still remain large. By the third year of potential sanctions, for instance, I forecasted that 73 percent of the students at sanctioned schools will be free-lunch-eligible, 54 percent will be black, and 27 percent will be Hispanic, if the low measure of student standards were employed. If the higher measure of student standards were employed, the gap would be slightly more narrow: the typical sanctioned school would have a student body that is 70 percent free-lunch-eligible, 51 percent black, and 25 percent Hispanic. Despite this narrowing of the gap, however, four years into the program, the typical school likely to be sanctioned under the federal law would still be disproportionately disadvantaged.

There are both social and fiscal reasons why we should care about the demographic characteristics of the schools likely to be sanctioned under the federal accountability laws. The social reasons stem from the basic purposes of the accountability law—to improve the quality of education for students trapped in chronically low-performing schools. While accountability pressures may induce schools to improve, removing finances from these same schools (which may actually provide high value added even if overall proficiency levels remain low) could reduce school quality for the students most in need of school improvement or remediation. The benefits of providing a within-district exit option are weakened if district-level sanctions affect all schools in a district. From a fiscal perspective, low-socioeconomic-status schools could be expected to have a harder time raising the funds to replace lost federal aid, and are also precisely the schools for whom many states have reformed their school finance systems. The withdrawal of federal aid from these districts may place state school finance systems out of compliance with court orders in some states, requiring a change in the level or distribution of state aid. At the same time, if states changed their school finance systems to compensate school districts for the loss of federal revenues resulting from sanctions, such a move would undermine the incentive effects of the introduction of a school accountability system.

In Florida, however, it does not appear that the schools likely to be sanctioned are disproportionately located in school districts that rely heavily on Title I aid. In Florida, the average per pupil Title I grant per disadvantaged child is $847. But the district average per pupil Title I grant

per disadvantaged child for sanctioned schools is $866 in the first year of sanctions, regardless of the measure of sanction used. By the third year of potential sanctions, this figure falls slightly more to $860. This trend may be peculiar to Florida, where poverty is less concentrated than in many other states. When Florida data are used to calibrate projections for the nation as a whole, however, I still found that the districts with schools forecast to be sanctioned were only mildly more reliant on Title I funds than were school districts without sanctioned schools. However, federal aid represents a larger share of total revenues for the districts of the sanctioned schools than it does for the districts of nonsanctioned schools. Thus, sanctions may hurt the school districts for whom a loss of federal school aid represents a larger percentage reduction in total spending power.

Large differences exist across districts in the fraction of the Title I student body who attend schools that are predicted to be sanctioned. Table 4 presents the tenth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth percentile districts (weighted by Title I student body population) in each year, with each measure of student proficiency standards ranked according to the fraction of disadvantaged students in the school district attending schools predicted to be sanctioned. Using both measures of proficiency standards, at least one-quarter of students are forecast to attend schools in districts in which no schools are sanctioned in the first year in which sanctions are possible. Even in the third potential year of sanctions, the twenty-fifth percentile school district has less than 1 percent of its

TABLE 4

*Distribution of School District Percentage of Disadvantaged Students Attending Sanctioned Schools—Projections Based on Florida Data**

Percentage attending sanctioned schools in	Percentile				
	10	25	50	75	90
Using more lenient definition of proficiency					
First year of sanctions	0	0	14.2	29.1	36.4
Second year of sanctions	0	0	9.5	23.3	32.6
Third year of sanctions	0	0.1	17.0	31.5	32.9
Using more stringent definition of proficiency					
First year of sanctions	0	0	12.9	25.6	33.2
Second year of sanctions	0	0.1	13.2	31.3	33.9
Third year of sanctions	0	13.1	35.4	41.8	51.7

* These figures describe the distribution across school districts of the percentage of disadvantaged students currently attending Florida schools simulated to be sanctioned in each of the first three years of the federal accountability law if the law took effect in 1999. The figures are a result of simulations of the federal accountability law in Florida using school-level data from Florida from 1998–1999 through 2001–2002. Two proficiency targets were used, based on different choices that the state of Florida could make.

disadvantaged students attending schools that would be sanctioned if the low measure of student standards is used, and a more robust 13 percent attending sanctioned schools if the higher measure of student standards is used. On the other hand, for at least 10 percent of school districts, one-third or more of the disadvantaged student body population would attend schools that are sanctioned. By the third year of potential sanctions, for the ninetieth percentile school in the model, where proficiency targets are based on the more rigorous student standards, 52 percent of disadvantaged students would attend schools sanctioned under the federal law. Given the federal regulations, at least one-fifth of this percentage of Title I funding is at risk. Put differently, if half of a school district's disadvantaged students attend sanctioned schools, then a minimum of 10 percent of the school district's Title I allocation may be diverted to choice-related transportation or private school supplementation. In some school districts, this figure would exceed 1 percent of a school district's total revenues and would represent a larger fraction of the school district's current expenditures.

While Table 4 does not present these results, I also performed these projections while restricting myself to the schools that are heavily minority—at least 90 percent black or Hispanic. Even among these schools, a large difference exists across school districts in the fractions of disadvantaged students attending sanctioned schools. By year three of sanctions, the tenth percentile district is projected to have 14 percent of students attending sanctioned schools, while the ninetieth percentile district is projected to have 61 percent of students attending sanctioned schools, if the less stringent measure of student proficiency is employed. Using the more stringent measure of proficiency, by year three of the sanctions, the tenth percentile district is projected to have 46 percent of students attending sanctioned schools, while the ninetieth percentile district is projected to have 79 percent of students attending sanctioned schools. Therefore, the sanctions required by the No Child Left Behind Act could affect some school districts, even with similar demographics, much more significantly than others.

These projections all assume that each subgroup is as successful as the school as a whole in reaching proficiency targets. But it is unlikely that this assumption will be true. (Indeed, the authors of the law had this unlikelihood in mind when mandating that each subgroup meet proficiency targets.) While I do not have sufficient data to perform the same analysis for each subgroup, I could do a similar analysis using slightly earlier data (here, I assumed that the policy began in 1995–1996) and for a subset of the counties of Florida. For confidentiality purposes, I do not name the counties included in this analysis; however, they tend to be

larger, more urban counties with considerably greater poverty rates than the rest of the state. These attributes of the counties I used make them unrepresentative of the state of Florida, much less the United States, but for this purpose, this exercise should still be sufficient to gauge the potential complications that the subgroup-proficiency requirement would introduce.

For the purposes of comparison, in this subgroup analysis, I set the initial target levels so that 31 percent of students would be proficient in reading in the first year of data (1995–1996) and 18 percent of students would be proficient in mathematics in the first year of data. Since the Florida Comprehensive Assessment Test was not being used during the entire subgroup study period, I used third-, fourth-, and fifth-graders' performance on the Stanford-9 achievement test as a proxy. While not reported herein, I have experimented with other cutpoints and grades, and the basic findings reported below remain unchanged.

Table 5 presents the predicted rates of school sanction in each of the first three years of sanctions, assuming that the first year of potential sanction is now 1996–1997. As a basis for comparison, consider the first column of the table, which projects the rate of sanctions based on the assumption that all subgroups have the same pass rate as the school as a whole. (These same criteria were used in the previous exercises.) This first exercise is important because the set of schools, the test used to evaluate them, and the years of study all differ from those employed previously. Using this set of assumptions, 60 percent of schools in this poorer, more urban subset of the counties of Florida are predicted to be sanctioned in the first year of sanctions, and by year three of the program,

TABLE 5

Percentage of Schools Forecast to Be Sanctioned in Subset of Florida,
by Year of Sanction, When Subgroups Are Considered*

Year of sanctions	No subgroup proficiency requirement	Subgroup proficiency requirement
First year	60	89
Second year	62	92
Third year	82	96

* These figures describe the percentage of schools forecast to be sanctioned in each of the first three years of the federal accountability law in a subset of unnamed (though more heavily poor, minority, and urban) counties in Florida once performance targets must be met by schools for the school overall, the school's free-lunch-eligible students, the school's black students, the school's Hispanic students, and the school's white students. The figures are a result of simulations of the federal accountability law in Florida using student-level data from a subset of Florida school districts from 1995–1996 through 1999–2000: the most recent data I have available. Data come from the Florida department of education and the relevant school districts.

82 percent of schools are forecast to be sanctioned. But these forecast rates of sanctions pale in comparison to the rates if the assumption of equal pass rates is relaxed, and actual pass rates by subgroups are used. Requiring schools with more than 20 test-takers in any subgroup (black students, Hispanic students, white students, and free-lunch-eligible students) to meet performance targets for that subgroup also increased the forecast sanction rates to 89 percent in the first year of sanctions and to 96 percent in the third year of sanctions. Note that this large increase in the forecast sanction rates is still understated (though it could hardly go higher) because I included fewer subgroups in the present analysis than the federal law requires to present conservative comparisons. The point herein is clear: the requirement that schools must meet performance targets for each countable subgroup in that school leads to substantially increased rates of sanctions under the federal accountability law.

Table 6 explores whether the inclusion or exclusion of subgroups is likely to influence the racial, ethnic, and socioeconomic composition of the schools sanctioned under the No Child Left Behind Act. As before, because the set of schools, years, and tests employed differ in this analysis from that of the preceding analysis, the table presents data for the attributes of schools sanctioned under both situations—including and excluding subgroup requirements. Requiring all subgroups to meet performance requirements increases the socioeconomic status of the schools forecast to be sanctioned under the federal law. While the set of sanctioned schools is still more disadvantaged than the population of schools as a whole, requiring schools to meet performance criteria in all subgroups is a more inclusive standard.

3.3 *Implications for Equity*

As the previous sections of this paper make clear, the overwhelming incidence of the direct fiscal effects of the federal accountability reforms will likely be borne by districts serving large numbers of low-income students, particularly in the early years of the program. This situation is in large measure deterministic because the federal accountability rules effectively sanction only low-income schools (because sanctions are tied to Title I allocations). Even among Title I schools, however, the incidence of accountability sanctions will likely affect mainly the particularly poor and heavily minority communities.

As a result, the federal accountability program, as mentioned above, will likely work to offset some of the school finance equalizations put forward by state legislatures or ordered by state supreme courts. (Incidentally, this disequilibrium also counteracts one of the goals of the federal Title I program itself, which provides additional Title I aid to schools in

TABLE 6

Demographic Characteristics of Schools Forecast to Be Sanctioned in Subset of Florida When Subgroups Are Considered*

Year of sanctions	No subgroup proficiency requirement			Subgroup proficiency requirement		
	Percentage black	Percentage Hispanic	Percentage free-lunch-eligible	Percentage black	Percentage Hispanic	Percentage free-lunch-eligible
First year	52	34	82	40	34	67
Second year	52	33	81	40	35	67
Third year	44	36	73	39	36	66

*These figures describe the demographic and socioeconomic attributes of a subset of Florida schools simulated to be sanctioned in each of the first three years of the federal accountability law if the law took effect in 1996. The figures are derived from simulations in which sanctions are based on performance targets that must be met by schools for the school overall, the school's free-lunch-eligible students, the school's black students, and the school's Hispanic students, and the school's white students. The figures are a result of simulations of the federal accountability law in Florida using student-level data from a subset of unnamed Florida school districts from 1995–1996 through 1999–2000: the most recent data I have available. Data come from the Florida Department of Education and the relevant school districts.

states that have more equalized school finance systems.) As described above, the school districts with the highest fractions of minority and low-income students—precisely the districts typically supported in school finance equalization scenarios—are the schools projected to lose the most under the federal accountability rules. If this scenario triggers further equalization aid from the state, it will lead to increased fiscal responsibility from the state. (On the other hand, this state bailout would undermine the goals of the federal accountability system.) Alternatively, districts would either have less revenue with which to work or would need to raise revenues to replace the lost Title I grants.

This scenario highlights an interesting tension between school equity and school accountability objectives. One of the principal arguments for school accountability programs is, in essence, an equity argument: the schools that may be most in need of accountability pressure are the schools that face the least competition or parental oversight. Presumably, these schools serve families that are either liquidity-constrained or minority. If these types of schools generally constitute the bottom of the equity distribution, so to speak, accountability systems could lead to differential improvements in the performance of these schools, which may in turn narrow the gap between the richest and poorest schools, or the richest and poorest members of society. On the other hand, if schools serving low-socioeconomic-status students are more able to "game the system" because they lack parental oversight, accountability systems could lead to no change in the distribution of school performance, or perhaps even a widening of the gap between the haves and have nots. If school accountability systems base financial rewards and punishments on student test performance, then these gaps could still increase. School accountability systems that base rewards and sanctions on the levels of test performance, such as those mandated by the No Child Left Behind Act, are particularly vulnerable to this predicament, unless corrections are made, such as in California and South Carolina, for family background characteristics. Even accountability systems that base rewards and sanctions on value-added measures of student performance may be vulnerable if these measures of value added, or student test score gains in general, are correlated with socioeconomic or other student background characteristics. Likewise, programs aimed at increasing equity could have the consequence of undermining accountability objectives. For instance, Duncombe and Yinger (forthcoming) find that block grants and matching aid lead to reductions in school efficiency.

While such an exercise is imperfect, it is possible to use Florida data to make projections about the implications for equity in different states. To

perform this projection, I predicted first the probability that a school will be sanctioned using Florida data and basic socio-demographic character-istics of the school. These factors explain approximately 80 percent of the variation in sanction probability in Florida, so they seem reasonable as a first approximation. Using these coefficient estimates generated from Florida data, I predicted the probability that each school in the United States will be sanctioned. (Again, this prediction assumes that the rela-tionship between demographics and the distribution of test performance is the same nationally as it is in Florida and that other states have similar testing policies and difficulty levels as does Florida.) From these predic-tions, I identified the percentage of schools in each district nationwide likely to be sanctioned under these circumstances, and I multiplied this district-specific percentage by the district's share of the federal aid offi-cially at risk under federal law.

Table 7 shows the breakdown of likely lost federal dollars per pupil by school district income quartile in each of the six largest states except for Illinois, where I had insufficient data to conduct the exercise. As the table shows, the incidence of the federal accountability system is heavily borne by the poorest school districts, precisely the districts that state school aid formulas are intended to help. A nontrivial fraction of state-performed school fiscal equalization would be undone even under this most conser-vative of estimates. The consequence would be a partially disequalized school finance system or, alternatively, an increase in the state's responsi-bility to fund education. In addition, as mentioned below, if input prices change further as a result of the accountability system (as may be ex-pected) the effective disequalization could be considerably greater. This last point, however, is merely speculative.

TABLE 7
Predicted per Pupil Effective Reduction (in Dollars) in Title I Aid to Districts, Grouped by Poverty Quartile by State, Under Certain Assumptions

State	State poverty quartile				Poorest 5 percent of districts
	Bottom	Second	Third	Top	
California	0.1	1.5	7.9	86.4	119.2
Florida	4.9	8.6	9.6	21.7	43.7
New York	0.0	0.1	19.2	130.1	155.0
Ohio	0.0	0.0	0.6	29.6	142.0
Pennsylvania	0.0	0.0	0.1	35.9	129.9
Texas	0.3	0.6	6.3	47.8	91.2

4. INDIRECT FISCAL CONSEQUENCES OF ACCOUNTABILITY SYSTEMS

Besides the direct fiscal consequences put in place in both the federal accountability system and potentially as part of state accountability systems, several indirect fiscal consequences are likely under the imposition of an accountability program. This section briefly describes several of these potential indirect effects, including behavioral responses that have direct fiscal consequences, such as the classification of students as disabled, as well as other implications of accountability, such as changes in house prices and the teacher labor market, that could have less direct implications. However, the indirect fiscal consequences may be larger in magnitude for states and localities than the direct fiscal effects.

4.1 Special Education Placement

Schools have many potential behavioral responses to accountability. Schools may react to accountability incentives by increasing class time spent on subjects and topics that are emphasized in the accountability exams, while decreasing class time on subjects and topics either not in or not emphasized in the exams. But another potential reaction to the incentives created by accountability systems involves the classification of students into special education categories exempt from taking the tests used for school grading. Schools can potentially improve their state-assigned grade or classification by taking their poorest performing students out of the testing pool and classifying them into the special education categories exempt from taking the tests. The schools can also potentially improve their state-assigned grade or classification by refraining from classifying better-performing students into the special education categories exempt from taking the tests. The new national study on special education costs by the American Institutes for Research (AIR) helps demonstrate the potential flexibility and opportunity that school decisionmakers have in determining which, if any, special education category in which to place students. AIR finds very wide variation in costs and services within single special education categories. In fact, they find less than 10 percent of the variation in special education costs in carrying out individualized education plans can be explained by the exceptionality categories in the federal/state indicator record (Chambers et al., 2002). This finding implies that there may be significant discretion in classifying individuals with specifically identifiable needs. Cullen (2002) and Cullen and Figlio (1998) show that schools classify students as disabled in response to fiscal incentives; it seems natural to expect that they might also classify students as disabled in response to accountability incentives.

Regardless of how children are classified, however, the cost implications of this reclassification could be very large. Chambers et al. (2002) indicate that the average student classified as requiring special education services costs schools more than twice that of the average regular education student. The approximate comparison, of course, concerns the *marginal* student, for whom cost estimates are much more difficult to make. But the reclassifications of even these students generate large costs to states, which average 40 percent greater reimbursements (calculated deterministically, in most states, based on classification rates) for marginally disabled students than for regular education students, regardless of actual cost differences realized. Indeed, Cullen and Figlio (1998) explore this potential for fiscal gaming in their study.

Several recent studies point to the possibility of schools responding to accountability systems by reshaping the testing pool. Using student fixed-effects models, Figlio and Getzler (2002) find that schools in two large Florida counties tend to respond to accountability pressures by switching previously low-performing children from regular education into special education classes exempt from inclusion in the testing program for the purposes of accountability, and that these responses are particularly strong for students attending schools likely to be sanctioned under the accountability program. Figlio and Getzler find, for instance, that classification rates of marginal students may increase by as much as one-quarter or more in schools threatened with accountability sanctions. Jacob (2002) looked at the effects of test-based accountability in Chicago and showed that low-achieving students in struggling schools are the most likely to be placed in special education. While Jacob does not estimate student fixed-effects models, he does control for prior achievement test scores and background characteristics. Using aggregate data and a clever identification strategy, Cullen and Reback (2002) exploit the discontinuity in rewards in the accountability system in Texas to show that schools respond to incentives to shape the test pool. These three papers, taken together, present complementary evidence—in three states and with three very different identification strategies—that schools respond to the incentive to classify marginal students in special education. The federal accountability system provides fewer incentives to reclassify students as disabled because disabled students' performance, except in the most extreme cases, are counted for federal accountability purposes. However, the state accountability systems mandated by the No Child Left Behind Act may (and, indeed, in many cases, do) still provide these types of incentives, suggesting that special education placement is potentially a major response to school accountability systems.

As mentioned above, student reclassification as a result of school ac-

countability systems have fiscal as well as educational implications. Student classification as disabled, even in the most moderate disability categories, introduces large incremental costs to school districts and states. The Individuals with Disabilities Education Act requires that all disabled students receive special services and an educational environment appropriate for their educational needs. State school finance systems increase state payments to school districts by an average of 40 percent when a student is classified as learning disabled. As Cullen and Figlio (1998) point out, in some states this incremental revenue exceeds the likely cost of reclassifying a student and in other states it may not. In most states, the additional costs would be shared (to different degrees, depending on the system and depending on the degree to which reclassification truly generates substantial marginal costs to the school district) between the state and the school district. But in a growing number of states, the state allocation to districts for special education finance is based not on actual special education counts but rather on predicted special education counts. In these states, the full fiscal burden of reclassification is borne by the school district, implying that school districts in these states have mixed incentives to reclassify students as disabled in response to accountability systems.

Schools have the same incentives to retain low-performing students in their current grade as they do to reclassify students as disabled, except in the terminal grades at a given school—then they have an incentive under the federal accountability system at least to promote low-performing students to the next grade. Therefore, there are also potential fiscal consequences associated with student grade retention. Unlike the case of special education placement, where in some states, school districts would bear the full fiscal cost of reclassifying a student as disabled in response to accountability pressures, student retentions in their current grade would mean that states always bear part of the costs of the school district responses to the accountability system.

4.2 Effects on Input Prices

The historical experience with *fiscal* accountability suggests that input prices may also change with increased *academic performance* accountability. As one example, Figlio and Rueben (2001) show that one consequence of the so-called tax revolt of the late 1970s and 1980s was that many higher-quality teachers left teaching and tended to be replaced by lower-quality teachers (where quality is measured by higher-quality teachers' test scores). They find that much of this response was not due to actual changes in resources but rather, apparently, to changes in perceptions. Early evidence from Florida's experience with school accountability sug-

gests that teachers, when they perceive that school grading is arbitrary or biased against their schools, are likely to contemplate leaving their schools. This situation may be one of the reasons why Downes and Figlio (1998) and Figlio (1997a) found evidence of much larger effects of tax limits than would be predicted by looking at the changes in actual school revenues and spending. These pieces of evidence indicate that teacher input costs may increase with school accountability and that the increases may be differentially borne by low socioeconomic-status school districts—precisely the districts that already face inflated teacher costs for any given unit of teacher quality. While this last point is speculative, it does suggest that hasty design of a school accountability system or performance standards embedded within a state aid system might have unintended consequences that work against the goals of the accountability system.

4.3 Effects on House Prices

Compelling evidence from Black (1999) and others shows that house prices reflect student test scores. Test scores have become well publicized in most states over the past decade, even in the absence of the federal accountability system. But ratings of schools may present additional information to the community, and this additional information may influence house prices, regardless of the quality of the signal presented in the school grading mechanism.

Figlio and Lucas (2000) studied the effects on the Gainesville, Florida, housing market of the introduction of Florida's A+ accountability system in 1999. Using a house-level, fixed effects specification and controlling for time-varying neighborhood effects, they show that while test scores remain capitalized into house prices following the introduction of school grading, the grades had an independent and large effect on the distribution of house prices, at least in the short term. Neighborhoods with schools that had unexpected positive shocks (captured by a favorable school grade and holding constant all the variables used to construct the school grades) saw immediate increases in house values, while neighborhoods with schools experiencing unexpected negative shocks saw reductions in their house values. This finding has direct implications for school finance. School districts with large numbers of sanctioned schools may experience reductions, at least in the short term, in their tax bases, thus requiring either reductions in school spending above and beyond the amounts of the state and federal sanctions or increases in the local tax effort to maintain current levels of spending, or increased state effort (and potential taxation) to maintain local spending. School districts with few if any sanctioned schools experience effects in the other direction. While it

is currently unknown whether these responses are long term, these results suggest that arbitrarily assigned school grades might have significant ripple effects in the housing market and on school finance in general.

It is unknown at present whether these house price effects are entirely redistributive, or if there may be some net gain (or loss) in house prices generally as a result of an accountability system. On the one hand, if the general perception of schools in an area improves, then this general increase in the perceived quality of schools should lead to an overall increase in house prices, much as would be the case with any other local amenity. If the goal of the No Child Left Behind Act is to effect positive change in school quality, this change in house prices may be an expected outcome of the federal accountability policy. On the other hand, as described above, it is highly likely that a large number of schools will face sanctions under the federal system. This situation could have the effect of descreasing public impressions of overall school quality, which may depress general housing values. It is clearly too soon to know which of these outcomes is most likely to occur nationally.

5. DESIGN ISSUES FOR STATES

While the No Child Left Behind Act dictates many features of a federal accountability system, states have considerable flexibility in implementing the federal system and, of course, in implementing and designing their own systems of accountability. This section details some of the important design issues relevant for state accountability systems and their potential effects on the designation of sanctioned schools. Because every state will implement a different system and because the nature of sanctions are as yet unknown in most cases, it is premature to speculate about the precise fiscal consequences of these decisions. It is apparent, however, that some school districts will be influenced fiscally more than others and that which districts will be influenced depends on the design of the system. Because school accountability systems introduced by states will necessarily influence the state's system of aid distribution at least to some degree, it is worth pointing out that many issues relevant for the design of school accountability systems are relevant for the design of state aid systems more generally.

This section describes three of the important design issues for states to consider as they implement the No Child Left Behind Act or their own independent systems of school accountability. While the three issues—aggregation questions, the decision whether to measure a school's value added, and measurement issues—may not at first appear to have fiscal consequences, in each case the choices that a state makes in its implemen-

tation of the federal accountability laws could have large fiscal implications for the states themselves. This section briefly describes each of the major design issues left for states to determine, then summarizes the most important fiscal considerations for each decision that states would make.

5.1 Aggregation Issues

Three types of aggregation issues are relevant for both state aid systems and school accountability systems. The first involves aggregation across time. In different settings and different contexts, Kane and Staiger (2001) and Figlio (2002) demonstrated the instability of school rankings based on test performance from one year to the next. The fundamental problem here involves measurement error: tests have large stochastic components to them, and schools deemed improving at one point in time could (almost necessarily) be declining at another point. This measurement error has implications for rewards and sanctions in state aid systems and school accountability systems: should schools be punished or rewarded on the basis of a single "good draw" or "bad draw"? These problems are particularly exacerbated when performance is measured by changes in proficiency from one year to the next across different cohorts (as is the case in many state accountability systems). When average test scores for two successive years are each measured with error (not least because the classroom compositions change dramatically, although this fact is not the only reason for the error), it is unclear what, if anything, is being uncovered by an accountability system that rewards (penalizes) improvements (fallbacks) in the fractions of students passing the exam from one cohort to the next. Kane and Staiger (2001) and Figlio (2002) illustrated that taking multiyear moving averages considerably reduces the likelihood that measurement error will lead to dramatic instability in measured school performance, although it remains a problem.

No time aggregation means that it is harder for sanctioned school to move out of a sanctioned state, thereby increasing (as described above) the fiscal costs borne by states and sanctioned school districts. While a lack of time aggregation could increase the sanction probability for schools that previously were unsanctioned, this likelihood is far lower. The reason for this nonlinearity stems from the fact that schools that are below performance targets may still avoid sanctions by making year-to-year improvements toward their performance goals. Allowing schools to aggregate across cohorts over time is less likely to punish improving schools that are still below performance targets in one or all subgroups than would a system that requires year-to-year improvements without permitting cross-cohort averaging. Using the Florida data mentioned above, I estimated that the fraction of schools sanctioned under a system

without time aggregation is as much as one-fifth higher than would be the fraction of schools sanctioned under a system with three-year time aggregation.

A second aggregation issue involves aggregation across types of tests (or other performance indicators). While mathematics and reading tests, for instance, tend to trend together within a school, Figlio (2002) shows that the correlation between changes from one year to the next in one test relative to another test is quite weak. Therefore, performance standards that require meeting particular criteria on multiple outputs may be too difficult to attain (if the standards are set appropriately high) and those that require cross-cohort improvements in these criteria along multiple dimensions may merely reward schools with good luck and punish those with bad luck. This situation is clearly an issue for the federal accountability law, where annual yearly progress must be met in both reading and mathematics. In the federal law, however, only the proficiency target for schools, rather than the *change* in proficiency target, moves over time, making this less of an issue in the federal case than it may be in individual state cases. Potential solutions to this problem include aggregating multiple outcomes into a single indicator (such as the approach described above by Duncombe and Yinger, 1999) or evaluating schools on multiple criteria separately without requiring standards to be met (or improvements to be realized) in every year.

A third aggregation issue to consider involves how multiple subgroups in a school should be considered. Disaggregating students into subgroups (say, along racial, ethnic, or socioeconomic lines, or based on prior performance levels) exacerbates the measurement error problems described in the preceding paragraphs and is required by the federal accountability law. On the other hand, for normative reasons, one may wish to pay special attention to the performance of certain subgroups, and indeed many equity and adequacy discussions (as well as the rationale for the name of the No Child Left Behind Act itself) have centered on the performance of minorities, low-income students, and low-achievers. The prescriptions for the improvement of measurement described above are particularly relevant when students are divided into groups. Kane and Staiger (2001) demonstrate that measurement error problems are most acute for small schools; the same is true for smaller subgroups within schools.

These last two aggregation issues have fiscal consequences similar to the first aggregation issue. Disaggregating by types of performance indicators and subgroups increases the likelihood of sanctions along two dimensions—measurement error problems increase with small samples (and the samples are necessarily smaller when data are less aggregated) and even without the potential presence of measurement error, it is more

difficult for a school to succeed along a larger number of criteria than along a smaller number. Hence, the more disaggregated the breakdown by types of tests and subgroups of students, the more likely it is that large numbers of schools will be sanctioned under the federal accountability policy, and consequently, the larger the anticipated fiscal burden on school districts and, presumably, on states. More disaggregation is likely to exacerbate further the equity issues described previously.

5.2 *Evaluating Schools Based on Value Added*

The federal accountability law sets proficiency targets for schools and designates a school as successful if it brings a certain fraction of students (and eventually all students) up to the minimum acceptable proficiency level. The normative argument behind this approach is clear: all students should be expected to achieve at least some reasonable level of academic success, and schools are delinquent in their duties if they do not bring students to at least this level. In this view, making progress toward proficiency but not achieving proficiency is unacceptable. But many economists and educators call instead for evaluating schools based on a measure of value added, where schools are rated on how much they improve individual students' performance from one year to the next. The rationale behind this approach is that schools with large numbers of children prone to academic success would achieve proficiency goals—even in the federal program, at least in the early years—regardless of whether they are truly mediocre (a point illustrated with some of the exercises presented in the preceding section), and that schools serving impoverished populations might fail to meet proficiency targets, even though they may be excellent in bringing students to basic proficiency levels from their very low starting points. Figlio and Page (2002), among others, show that levels-based methods of evaluating schools tend to lead to rankings of schools that are nearly completely unrelated to value-added-based methods. The implication of this finding is that many schools that are sanctioned under a levels-based system of school accountability may actually have quite high value added and should likely be rewarded rather than punished. The lack of concordance between state and federal accountability systems [also described by Kane and Staiger (2001)], may also tend to undermine the credibility of one or both of the systems in the eyes of the general population.

Despite their popularity among economists, debate exists as to whether value-added measures of school performance are really to be preferred over raw levels of performance. The normative argument presented above—that schools and students should be held accountable to a given standard, regardless of background characteristics or starting values—is

compelling to many because it does not prescribe different standards for different groups of students (or schools). Given the high correlation between test performance and background characteristics, especially in aggregate, however, evaluating schools solely on performance levels, with rewards and sanctions associated with said performance, does not seem fair to schools with large numbers of students predicted to perform poorly based on background characteristics. While some of these correlations may be explained by factors within the control of the school system (e.g., different expectations for low-income and minority students than for higher-income and white students), it still defies reason to suggest that all of these differences are due to controllable factors. Thus, equity concerns with evaluating schools solely on the basis of the level of test performance are valid.

There are other reasons, however, to believe that value-added measures of evaluating schools may not be a panacea. Gains in test scores within a cohort may reflect school contributions but may also reflect unmeasured student and family characteristics. Student background characteristics are often found to be correlated with gains in test scores as well as test score levels, though this finding is not universal, and it is not necessarily the case that schools with high starting values also experience high gains. These correlations may be due to differential school selection by families of different types, but they could also point to different family inputs, either independently of or jointly with the school's efforts. They may also indicate that schools serving high-socioeconomic-status families may be better able to find resources to boost instruction and outcomes in tested subjects. These correlations suggest that value-added measures of school productivity should be taken with a grain of salt.

Another potential problem with some forms of value-added measures is that they might invite manipulating the system. If schools are rewarded on the basis of cross-cohort changes in test scores (for instance, comparing fourth-graders in 2002 to fourth-graders in 2001, rather than comparing the same students from year to year) schools may be rewarded by underperforming in one year to facilitate greater gains the next. But even if school gaming does not generate problems with this type of system, measurement error might. Recall the Kane and Staiger (2001) argument mentioned above on measurement error and their finding that schools that show poor gains in one year tend to show high gains the next, and vice versa. Measuring value added based on following the same students rather than different groups of students solves some of these problems, but only partially. In sum, while value-added systems of measuring school performance are likely better than most systems, they can't solve all the problems.

But even if a state were to introduce a system of value-added-based school accountability, how exactly to quantify a school's value added is open to debate. At the very least, however, accounting for value added requires some control for circumstances such as student and family background characteristics. Typically, however, student background characteristics (or, more specifically, the background characteristics available in administrative data) explain only a small portion of the observed variation in individual test scores. (Student information aggregated to the school level does a better job in explaining aggregate test scores, although this method too is extremely incomplete.) Therefore, it seems important to go beyond observed background attributes when measuring value added.

Measures that aggregate changes in student test scores from year to year are arguably closer indicators of a school's contribution to student outcomes than are those measures that consider only a level test score. These measures range in complexity from the extremely simple and transparent (for instance, simply averaging year-to-year test score gains) to very complicated models that impose considerable structure on the relationships between inputs and outputs. All models of value added require annual (or at least extremely frequent) testing of students with tests that can be vertically equated—that is, the test scores for one grade can be compared to those of the next—and tests that are administered uniformly across schools. Even more fundamentally, the administrative data system must be able to follow students reliably from year to year.

As mentioned above, value-added measures of school performance, regardless of how they are measured, are still imperfect indicators of a school's contribution to student gains. Given the correlations often found between student test score gains and background characteristics, it may be that additional correction for student body attributes is warranted. This decision is not obvious, however, and depends in part on what the accountability system seeks to measure.

All measures of school performance based on standardized tests must also address the question of which students should be included in the test pool for the purposes of standards and accountability. This question is inherently normative and political rather than scientific, though it derives from a positive debate. Key decisions must be made about whether mobile populations are counted in the accountability testing pool, as well as whether disabled students should be considered for school reporting purposes. The state of Florida, for instance, has taken two very different tacks with regard to student mobility. In the first iteration of the A+ Plan for Education, Governor Jeb Bush's education reform enacted in 1999, school evaluations were based not only on stable students but also students who recently arrived in the school. But the next year, school grades

were based only on students present in the school for the entire school year up to the testing date. These rule changes had substantial implications for schools and students, and the sets of schools identified as low quality or high quality changed considerably when only the more stable set of students were included. The federal accountability system limits students counted for a school's proficiency goals to those spending the full academic year in the evaluated school, but mobile students still count for district proficiency goals. States also vary with respect to which disabled students should remain in the test pool for the purposes of constructing performance measures. Florida, for example, excludes all disabled students, even test-taking students, from the school-level aggregates used to measure productivity. Virginia goes to the other extreme: including all disabled students in the accountability pool except for the rather small fraction who are explicitly excluded in their individualized education plans from testing. In this regard, the federal accountability system is much closer to the Virginia approach to counting disabled students than it is to the Florida system, and in fact requires schools to meet proficiency goals for disabled students (except the most seriously disabled populations) as a listed subgroup.

These issues pose substantial trade-offs to policymakers seeking to develop accountability and state aid systems. A policymaker deciding whether to include or exclude mobile students or disabled students, for example, may wish to exclude these students from the pool for the purposes of gauging school performance out of a concern for fairness. Schools with large fractions of mobile and disabled students could argue, with validity, that they are being judged on factors outside their control. On the other hand, excluding students on the basis of *classification* provides schools with the incentive to selectively reclassify or move students to look better against performance metrics.

The case of special education is a margin that may be worked for these purposes. Cullen (2002) and Cullen and Figlio (1998) have demonstrated that schools tend to reclassify students along the special education margin in response to *fiscal* incentives, and Knapp (2002) demonstrates that schools' classifications of students are responsive to *parental* fiscal incentives as well. And as mentioned previously, Cullen and Reback (2002), Figlio and Getzler (2002), and Jacob (2002) show that special education classification also responds to *accountability* incentives. While the federal accountability law will present only limited incentives for reclassifying low-performing students as disabled to evade inclusion in the school's test pool, state accountability systems may still provide these incentives. And the federal accountability system, by naming disabled populations

as a subgroup with its own proficiency goals, may actually present incentives for school districts to overclassify potentially high-performing students as disabled to yield improvements in the disabled group's proficiency levels. Of course, these incentives also have a different form of implications for school finance because the increased classification of marginal students as special education students increases total education costs for school districts and states.

The same types of trade-offs are relevant for the decision to move from levels-based assessment of schools for accountability and school finance to a more value-added-based assessment. The argument for value added is very much the same as the argument for excluding disabled and mobile students from the high-stakes (for schools) testing pool: schools with certain populations that tend to fare worse on standardized examinations argue that their quality is masked by the poor outcomes of students starting at a low level. On the other hand, introducing value added (or even controlling for background factors) raises political concerns, because it becomes more difficult for the lay public to interpret the assessment of schools, and because explicitly controlling for different student composition is seen by some as making excuses for poor performance, or alternatively, holding different types of students (and hence, schools) to different standards. (The latter reason is currently a popular argument.) It is possible that including background characteristics as part of a cost correction, as explained by Duncombe and Yinger (1999), might be a politically palatable way to control partly for student attributes in an environment where it would be difficult to do so otherwise.

School accountability and school finance systems may provide different incentives for schools to reclassify, move, or hold back students. While school accountability systems tend to encourage schools to classify students as disabled, for instance, school finance systems may either encourage or discourage this practice, depending on the financial incentive structure. For instance, some states compensate school districts for disabled students on the basis of predicted disability caseloads rather than actual disability counts. In such a case, reclassifying a student as disabled to avoid that student being counted in an accountability system will be costly to the district that must now provide special services for the student without additional compensation from the state. On the other hand, other states provide compensation to districts that exceeds the excess cost necessary to educate marginal students if they are classified as disabled. In these states, both the school accountability and school finance systems provide strong incentives to overclassify students. Similarly, moving students from school to school across years may lead to increased costs to

school districts, but school accountability systems that exclude mobile students may provide sufficient incentives to make these moves worthwhile for school districts.

The choice of measuring school performance using levels rather than value-added approaches also has fiscal implications. In the years after initial implementation of the policy, it will become progressively more difficult to meet accountability goals under No Child Left Behind than it would under a value-added approach because the ultimate goal is 100 percent proficiency under the levels-based federal law. Gaming along the lines mentioned above is likely to be more simple when using a levels-based system of school performance measurement than when using value-added measures, which implies that levels-based systems may generate greater indirect fiscal costs for states and school districts than might value-added-based systems of school accountability.

5.3 Measuring Costs of Education

Of central importance in the school finance reforms implemented across the nation over the past two decades has been the notion of equality of educational opportunity. In the school finance context, this notion involves not only considering policies that tend to equalize spending across school districts within a state, but also policies that correct for the different costs of educational provision across school districts. A discussion of measuring educational costs in the present context is relevant for several reasons. First, many of the same issues that arise in the determination of school performance appear in different contexts in the determination of cost differences across schools. Second, accountability systems themselves may directly and indirectly influence cross-district costs of education, which then directly affects school finance systems.

The rationale for adjusting for cost differences is very similar, theoretically, to the rationale for measuring value added to schools in educational production: when two schools face very different circumstances, the schools have varying difficulties in meeting any type of performance goal. Duncombe and Yinger (1999) provide a "temperature standard" metaphor for schools: it takes fewer resources to heat a school in Florida to 72 degrees Fahrenheit in the winter than it does to heat a school in Wisconsin to the same temperature. Providing the same heating budget to both schools would lead to dramatic differences in measured productivity along one dimension (in terms of actual comfort level attained), but this situation would not separate factors that are in the control of school officials from those that are outside this control. The purpose of cost adjustments is to guarantee, as best as possible, that two schools facing very different circumstances have the same potential to generate a certain level

of student outcomes after adjustment. The word *potential* is important here because it highlights a key conceptual question in the design of state aid systems—whether *ex ante* versus *ex post* realization of the goal is the relevant barometer of success. One *ex ante* view of this issue would be to make cost adjustments to schools without regard to differences in productivity—two schools, after appropriate cost adjustments are made, would have equal chance of success. An *ex post* view of this issue would possibly take differential productivity into account and may make additional adjustments to account for differences in school productivity. Which view of the issue is adopted is inherently a political question.

Any researcher or practitioner involved in school finance can attest to the fact that school districts face substantial differences in the costs of educating their student populations. These differences exist for numerous reasons, not least because student populations vary considerably in their needs. Special education population differences, described somewhat above, are but one of the sources of this variance. Other differences occur because of varying needs for remediation and other types of educational services.

These same characteristics associated with increased basic costs of services may also influence input prices. Hanushek, Kain, and Rivkin (1998), among others, show that teachers are more likely to select more affluent and majority-white schools when they have the opportunity, and Figlio (1997b) shows that high-paying (generally suburban) school districts can attract high-quality teachers away from their lower-paying neighbors, independent of school and student body attributes. Numerous authors, such as Ferguson (1991), Ferguson and Ladd (1996), Goldhaber and Brewer (1997), Hanushek et al. (1998), and Wright et al. (1997), have shown that teacher quality (though not the types of attributes typically rewarded by teacher salary structures) significantly increase student outcomes. Taken together, it is reasonable to expect that teacher dollars at some schools do not go as far as they do in other schools. Given the widespread recognition that costs differ substantially across school districts, it is somewhat surprising that states rarely attempt to adjust systematically for costs. Cost adjustments are often made in ad hoc ways, with potentially deleterious consequences that could become even more apparent in an era of adequacy standards and school accountability. Cost adjustments that do not take into account both the direct and indirect cost differences associated with educating different types of schools will likely lead to systematic underprovision of school services for low-income and minority students relative to their high-income, white counterparts.

Introducing school accountability and performance standards, even if not incorporated directly into the state aid formula, can have significant

consequences for cost adjustments. Increased school choice that is tied to school accountability, for instance, may exacerbate school district cost differentials if students from already high-cost, low-socioeconomic-status districts are provided an exit option and particularly if the most motivated families are the most likely to leave their existing school. Along similar lines, some high-achieving students attending schools branded as low-performing may also enter the private sector and leave the public sector altogether. These types of responses suggest that cost adjustments may need to be updated even more frequently in an era of school accountability than in a regime with less accountability-driven choice. As mentioned above, school accountability systems, including the federal program to a lesser degree, may provide schools with an incentive to overclassify students as disabled or to retain low-performing students with higher frequency. These types of responses, and the fact that they are likely to occur differentially, suggest that the relative costs of districts are likely to change as schools are held more accountable.

These issues raise implications for school finance equity. State cost adjustments for low-socioeconomic-status school districts may need to change depending on the school choice or teacher labor market effects resulting from the response to the school accountability system. These effects depend on the degree to which states seek to equalize equality of educational opportunity. But in many states, there will be no choice, and state supreme courts may demand alterations to states' school finance systems as a direct consequence of the federal accountability law.

6. CONCLUSION

The No Child Left Behind Act of 2001, likely the most important piece of federal legislation concerning elementary and secondary schools in several decades, lays out a major new way in which federal school aid is distributed. This paper outlines some of the most important fiscal implications of this policy change and suggests that the federal accountability system, as written, may substantially increase fiscal pressures on low-income school districts and perhaps also on states (via state school finance equity goals). These fiscal effects are both direct and indirect, and are likely influenced by decisions that states will make as they implement the federal accountability policy. While the magnitudes of these implications are not possible to identify, back-of-the-envelope calculations suggest that the effects could be reasonably large for a nontrivial number of school districts. Ironically, the school districts placed under increased pressure might not be the districts most deserving of sanction, if indeed measurement of a school's value added is the desired metric. (This last point,

obviously, is debatable, as mentioned above.) In sum, while the new federal accountability system holds considerable promise for aiding and improving schools and students, considerable concerns about its implications, both educationally and fiscally, also exist.

REFERENCES

Black, Sandra (1999). "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics*. Vol. 114, No. 2, pp. 577–600.

Chambers, Jay, et al. (2002). "A Report on the 1999–2000 Special Education Expenditures Project." Special session presented at the annual research conference of the American Education Finance Association, Albuquerque. March.

Cullen, Julie (2002). "The Impact of Fiscal Incentives on Student Disability Rates." *Journal of Public Economics*, in press.

Cullen, Julie, and David Figlio (1998). "Local Gaming of State School Finance Policies: How Effective Are Intergovernmental Incentives?" Working paper, University of Michigan.

Cullen, Julie, and Randall Reback (2002). "Tinkering Toward Accolades: School Gaming Under a Performance Accountability System." Working paper. University of Michigan.

Downes, Thomas, and David Figlio (1998). "School Finance Reforms, Tax Limits, and Student Performance: Do Reforms Level Up or Dumb Down?" Working paper. Tufts University.

Duncombe, William, and John Yinger (1999). "Performance Standards and Education Cost Indices: You Can't Have One Without the Other." In *Equity and Adequacy in Education Finance: Issues and Perspectives*, H. Ladd, R. Chalk, and J. Hansen (eds.). Washington, D.C.: National Academy Press.

Duncombe, William, and John Yinger (forthcoming). "Alternative Paths to Property Tax Relief." In *Property Taxation and Local Government Finance*. Cambridge, Mass.: Lincoln Institute of Land Policy.

Ferguson, Ronald (1991). "Paying for Public Education: New Evidence on How and Why Money Matters." *Harvard Journal on Legislation*. Vol. 28, No. 2, pp. 465–498.

Ferguson, Ronald, and Helen Ladd (1996). "How and Why Money Matters: An Analysis of Alabama Schools." In *Holding Schools Accountable: Performance-Based Reform in Education*, H. Ladd (ed.). Washington, D.C.: Brookings Institution.

Figlio, David (1997a). "Did the 'Tax Revolt' Reduce School Performance?" *Journal of Public Economics*. Vol. 65, No. 3, pp. 245–267.

Figlio, David (1997b). "Teacher Salaries and Teacher Quality." *Economics Letters*. Vol. 55, No. 2, pp. 267–271.

Figlio, David (2002). "Aggregation and Accountability." In *No Child Left Behind: What Will it Take*, C. Finn (ed.). Washington, D.C.: Thomas B. Fordham Foundation.

Figlio, David, and Lawrence Getzler (2002). "Accountability, Ability and Disability: Gaming the System?" Working paper, National Bureau of Economic Research. October.

Figlio, David, and Maurice Lucas (2000). "What's in a Grade? School Report Cards and House Prices." Working paper. National Bureau of Economic Research. October.

Figlio, David, and Marianne Page (2002). "Can School Choice and School Account-

ability Successfully Coexist?" In *The Economic Analysis of School Choice*, C. Hoxby (ed.). Chicago, Ill.: University of Chicago Press.

Figlio, David, and Kim Rueben (2001). "Tax Limits and the Qualifications of New Teachers." *Journal of Public Economics*. Vol. 8, No. 1, pp. 49–71.

Goldhaber, Dan D., and Dominic J. Brewer (1997). "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity." *Journal of Human Resources*, Summer 1997, Vol. 32, No. 3, pp. 505–523.

Gordon, Nora (2001). "Do Federal Grants Boost School Spending? Evidence from Title I." Working paper. University of California at San Diego.

Hanushek, Eric, John Kain, and Steven Rivkin (1998). "Teachers, Schools and Academic Achievement." Working paper. National Bureau of Economic Research. August.

Jacob, Brian (2002). "The Impact of High-Stakes Testing on Student Achievement: Evidence from Chicago." Working paper. National Bureau of Economic Research. May.

Kane, Thomas, and Douglas Staiger (2001). "Improving School Accountability Measures." Working paper. National Bureau of Economic Research. March.

Knapp, Caprice (2002). "Antipoverty Programs, Parental Incentives, and Student Disability Classification." Working paper. University of Florida.

Wright, S. et al. (1997). "Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation." *Journal of Personnel Evaluation in Education*. Vol. 11, No. 1, pp. 57–67.