

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

Volume Title: The Economics of School Choice

Volume Author/Editor: Caroline M. Hoxby, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-35533-0

Volume URL: <http://www.nber.org/books/hox03-1>

Conference Date: February 22-24, 2001

Publication Date: January 2003

Title: Can School Choice and School Accountability Successfully
Coexist?

Author: David N. Figlio, Marianne E. Page

URL: <http://www.nber.org/chapters/c10085>

Can School Choice and School Accountability Successfully Coexist?

David N. Figlio and Marianne E. Page

2.1 Introduction

Education is currently at the forefront of the nation's political agenda: Everyone, regardless of political persuasion, wants to see an improvement in American schools' performance. Everyone does not support an increase in governments' (federal, state, or local) education budgets, however. In the current climate, in which many state and local governments face financial constraints but many voters are demanding tangible evidence of school improvement, policies that increase school choice through the use of vouchers have become increasingly popular.

Many economists find vouchers appealing because they increase schooling options for families whose choices might otherwise be constrained (by low incomes, job location, residential segregation, etc.). If parents are well informed then vouchers will increase efficiency by increasing families' abilities to sort into optimal schools. At the same time, vouchers are expected to improve average school quality by increasing competition among schools. When a student uses a voucher to attend a private school, the local public school's funding is decreased by the amount of the voucher, and this threat of budget cuts will provide schools with an incentive to improve. The larger the number of voucher-eligible students in a school, the more incentive the school has to improve performance.

Recent voucher proposals have also been justified on the grounds that students should be provided a means of exiting low-quality schools so that they do not experience failure. A voucher program built around this prem-

David N. Figlio is the Walter Matherly Professor of Economics at the University of Florida and a faculty research fellow of the National Bureau of Economic Research. Marianne E. Page is assistant professor of economics at the University of California, Davis.

ise would provide vouchers to all students attending a low-performing school regardless of whether they are economically constrained. Economists are typically less comfortable with this type of program because (a) it suggests that the state is better than parents at assessing school quality, and (b) it is unclear why vouchers would lead to better schooling choices among parents who are not liquidity constrained yet voluntarily select low-quality schools. An advantage of this type of program, however, is that failing schools face the prospect of losing all of their students (not just the economically disadvantaged ones), which presumably provides a strong incentive to improve.

School accountability is an alternative policy that aims to improve school quality without increasing costs. Accountability systems reward and punish schools by allocating funding according to whether the school meets certain performance criteria. When vouchers are targeted toward economically disadvantaged children, the performance measures created by an accountability system provide information that helps parents optimize. Many states have some form of school choice that coexists with a system of grading schools.¹ Currently popular, however, is the idea of integrating school choice and accountability so that students attending schools that the state has identified as failing have the option of moving elsewhere. Florida has recently adopted an integrated accountability/voucher program, and President Bush proposed such an initiative as part of his national accountability-based education plan. Although private school vouchers are excluded from the federal law that eventually passed, the No Child Left Behind Act of 2001 integrates increased public school choice into a national school accountability system.

The purpose of this paper is to identify which students would be eligible for a voucher under different types of voucher programs. We use data from the state of Florida and pay particular attention to the distributional impact of vouchers targeted toward disadvantaged students versus the distributional impact of vouchers targeted toward students attending low-performing schools. Although at first one might expect these two types of targeting schemes to reach approximately the same students, in fact the extent to which this is true will depend on the distribution of children across schools and on the way in which a school's performance is measured. As we will discuss in the next two sections, there are many different ways of evaluating a school's performance. Our results suggest that the performance criteria that may seem most appropriate from an accountability perspective may not do the best job of targeting economically and socially disadvantaged children. In fact, a voucher system that uses value-added measures of performance to determine voucher awards may do little better in achieving this goal than a system that randomly assigns vouchers. Thus, although

1. States include Arizona, California, Minnesota, Michigan, and North Carolina.

vouchers and accountability programs have similar goals, economists' justifications for vouchers may be undermined when vouchers are integrated with accountability.

In the remainder of this paper we describe our data and the voucher program in Florida. We then compare the stratification effects of targeting vouchers using family income to the stratification effects of targeting vouchers using school characteristics, including the types of performance measures being adopted in Florida and proposed by President Bush. In the final section, we discuss our results' implications.

2.2 Vouchers and Accountability in Florida

Our eligibility simulations are based on aggregated student-level data provided by the state of Florida's Department of Education for the 1998–1999 and 1999–2000 school years. The data include information on every public school student's race or ethnicity, test scores, free or reduced-price lunch eligibility, and school attended. Florida is one of the only states that has collected detailed enough information for our project to be pursued. An advantage of using these data is that our simulations can speak directly to the one state that currently has an integrated voucher/accountability program. Our findings will be relevant beyond Florida, however. The idea of imbedding a voucher program within an accountability system is attractive to many state legislators and has been proposed by President Bush. Although the magnitude of the effects will differ across states depending on the distribution of students across schools and the degree to which accountability and vouchers are linked, the general patterns should be the same.²

2.2.1 Overview of the Florida Program

School accountability in Florida is based on student performance on the Florida Comprehensive Assessment Test (FCAT). Students in Florida are given reading examinations and the Florida Writes! exam in fourth, eighth, and tenth grade, and they are given mathematics examinations in fifth, eighth, and tenth grade. Test scores are then converted into a five- or six-point scale, which is used to assign a letter grade ranging from "A" through "F." In order to attain a "C" grade, at least 60 percent of the school's test-takers must achieve at level two or above (on the five-point scale) in reading

2. The generalizability of our results rests largely on whether low-income and minority students are similarly distributed across schools in other states. We used the 1996–97 *Common Core of Data* to look at how the distribution of students in Florida compared to the distribution of students in California and Texas (since these are two large states). Minorities in California and Texas are more concentrated than in Florida. For example, in California, 75 percent of minorities attend schools that are at least 58 percent minority, in Texas 75 percent of minorities attend schools that are at least 52 percent minority, and in Florida 75 percent of minorities attend schools that are at least 37 percent minority. Similar, although less dramatic, patterns exist with respect to the concentration of poverty.

and math, and at least 50 percent of test-takers must achieve at level three or above (on a six-point scale) on the Florida Writes! examination.³ If a school misses one or two of these thresholds it receives a grade of “D.” If it misses all three of these thresholds it receives a grade of “F.” The first assignment of school letter grades occurred in May 1999, and about 8 percent of schools receiving grades⁴ earned grades of “A,” whereas 13, 51, and 25 percent received grades of “B,” “C,” and “D,” respectively. Just over 3 percent of schools were graded “F.”⁵

Letter grades affect schools on both ends of the spectrum: Those who receive an “A” grade, or who increase their letter grades from one year to the next, are eligible to receive an additional \$100 per pupil. On the other hand, students attending schools that are rated “F” in two years out of four are eligible for “opportunity scholarships” that can be used to attend private school or nearby public schools with ratings of “C” or higher. Students who choose to transfer may remain in their school of choice until the terminal grade of that school, regardless of whether their initial school improves in subsequent years. Although the Florida plan currently bases school grades, and, thus, voucher eligibility, on aggregate levels of test performance, its architects viewed level test scores as only an interim metric for grading performance. The accountability law requires that by the 2002–03 academic year school grades be based, at least in part, on value added measures of student performance. President Bush’s proposal also relied on value added performance measures to determine voucher eligibility. Although the specific nature of value added school assessment is still under deliberation in Florida, it is clear that school grades will soon be based in large measure on features other than levels of aggregate student performance.

Florida’s grading system is complicated, but when schools are ranked using these grading criteria, their ordering is very similar to the ordering that emerges if average math performance or average reading performance is used to order schools instead. The correlation between a school’s ranking determined using Florida’s current criteria and a school’s ranking based on average test scores is about 0.95. On the other hand, correlations between school rankings that are based on single-year measures of school performance and the rankings that emerge when within-cohort test score growth (value added) is used instead are close to zero.⁶ These low correlations in

3. These levels were determined by the Florida Department of Education, based on expert opinion on how well students satisfied the “Sunshine State Standards” for the appropriate grade.

4. Schools did not receive letter grades if their tested cohorts were sufficiently small.

5. “A” and “B” grades in Florida are based on cross-cohort changes in test scores, as well as differences in attendance, suspension, and test-taking rates.

6. The correlation between a school’s position based on its average reading score in a particular year and its rank when ordering is based on within-cohort changes in average reading scores from one year to the next, for example, is only 0.05. Currently, we can only make this comparison using reading test scores because 1999–2000 was the first year that the state tested

turn hint that when vouchers are embedded within an accountability system, the types of students who qualify for vouchers may vary a great deal with the type of performance measure that is used. The purpose of the next section is to determine how important the choice of performance criteria is.

2.3 Who Is Eligible for a Targeted Voucher?

We begin by looking at the socioeconomic composition of the choice-eligible population when voucher eligibility is determined using school-level attributes. Next, we turn our attention toward understanding the ramifications of school-based accountability/voucher systems. In all of the simulations we assume that the amount of the school voucher would remain the same as it currently is in Florida—that is, full payment of tuition and fees by the student’s private school of choice. Eligibility is, of course, not affected by this assumption, although take-up rates will strongly depend on the amount of the voucher.

As a point of comparison, the first column of table 2.1 provides information about the demographic characteristics of the student population in Florida. This column tells us that if vouchers were made universally available, then approximately 26 percent of recipients would be black, 18 percent would be Hispanic, and 44 percent would be eligible for a free lunch. We have standardized Florida’s test scores, so the average test score for the sample is close to zero.⁷

2.3.1 Voucher Eligibility Based on Socioeconomic Status

Although universal vouchers have been proposed by some policymakers (indeed, California’s recent voucher initiative would have made vouchers available to all students), those programs that are actually implemented are likely to be targeted toward specific subsets of the student population. Milwaukee’s voucher program is probably the best known, and so our first step

in every grade from three through ten. Therefore, although we could construct a student-level test score change in reading from grade four to grade five this year, we have to wait until next year before a similar exercise can be conducted in mathematics, because historically, reading has been tested in fourth grade and mathematics in fifth grade. None of the rankings that emerge using different performance measures are very correlated with school spending at the district level. Using expenditures reported by the Florida Department of Education for 1998–99, we find a correlation between per-pupil district spending and a school’s average reading score of -0.14 . We estimate the same correlation between per-pupil district spending and a school’s average math score. This result is unsurprising given that higher-spending districts tend to be located in urban areas, where costs are higher. Correlations between per-pupil expenditures and any of the value added measures we use in this paper range from 0.01 to 0.06. It should be noted, however, that relative to many states Florida has a flat spending distribution. The 95:5 ratio in spending is only 1.2.

7. The mean is not precisely zero because a small number of eligible test-takers are not part of the accountability system (because of small school size) but were still used to standardize the test for basis of comparison. The decision to include or exclude these test-takers from the standardization makes no difference for the results presented in this paper.

Table 2.1 Characteristics of Voucher-Eligible Students in Various Simulations: Voucher Based on Socioeconomic Status of School or Neighborhood

Characteristics of Voucher-Eligible Students	Universal Voucher (1)	Voucher Given to All Free Lunch Eligibles (2)	Voucher Given to All Free/Reduced Lunch Eligibles (3)	Voucher Based on Socioeconomic Status of Students in School (4)
Voucher coverage (%)	100	44	53	20
% black	25.5	34.5	32.7	58.6
% Hispanic	18.0	21.2	20.8	26.8
% free lunch eligible	44.2	100.0	82.6	80.6
% residing in high poverty neighborhood	25.3	41.1	38.1	85.6
Standardized reading test score, fourth grade (2000)	0.02	-0.69	-0.58	-1.17
Standardized math test score, fifth grade (2000)	0.05	-0.64	-0.55	-1.02

Notes: Each column represents a different voucher simulation: one in which a voucher is given to all students (column [1]); one in which a voucher is given to all free lunch eligible students, regardless of school performance (column [2]); one in which a voucher is given to all free or reduced price lunch eligible students (column [3]); and five scenarios in which vouchers are given to all students in schools with the 2 percent (column [4]), 5 percent (column [5]), 10 percent (column [6]), 20 percent (column [7]), or 44 percent (column [8]) highest poverty rates (measured as the fraction of students who qualify for a free lunch). Each row represents a different attribute of the students receiving vouchers in each situation.

is to replicate the eligibility criterion used in that program. In Milwaukee, students with family incomes below 175 percent of the poverty line are eligible to receive a voucher. We do not have explicit information on students' income, but we can look at how stratification is affected when vouchers are targeted to students who are eligible for free or reduced-price lunches. The income cutoffs for free and reduced-price lunch eligibility are close to Milwaukee's threshold: The free lunch eligibility cutoff is 130 percent of the poverty line, and the reduced-price lunch threshold is 185 percent of the poverty line.

The second two columns in table 2.1 present the characteristics of students eligible for free (or free and reduced-price) lunches in Florida. If every free lunch-eligible student were to receive a voucher, then 44 percent of the state's student body would be voucher-eligible. If the eligibility threshold were increased to include students eligible for reduced-price lunches, coverage would rise to 53 percent. Free lunch-eligible students are disproportionately minorities (primarily black) and disproportionately reside in low-income neighborhoods.⁸ By definition, all of these students have low incomes. In addition, because low-income students in Florida, as in the rest of the country, tend to have lower test scores than high-income students, the average test score of the voucher-eligible population would be around six-tenths of a standard deviation below the state mean.

It is important to remember that in practice no voucher programs have targeted such a high fraction of students. In Milwaukee, for example, roughly half of the students are eligible, but only 15 percent are allowed to participate. Voucher assignment in Milwaukee is determined by lottery. If recipients in Florida were to be randomly selected from the income-eligible population, then although the number of served students would be lower than 44 percent, the distribution of their characteristics would still look much as it does in column (2).

If poverty is highly concentrated, then a disadvantage of targeting vouchers according to family income is that they may not increase competition among very many schools. As in the rest of the country, poverty in Florida is not evenly distributed across schools, but providing vouchers to every free lunch-eligible student could still significantly affect the level of resources available to schools. Fourteen percent of Florida's elementary schools have 75 percent or more students eligible for free lunches, and 44 percent of schools are at least half free lunch eligible. If voucher recipients were drawn at random from the eligible population (as in Milwaukee), then a smaller number of students would ultimately receive a voucher, and competition would be lessened.

An alternative method of delivering vouchers to disadvantaged students

8. We define a low-income neighborhood as a zip code ranking in the top quarter statewide (weighted by school-aged population) of percentage of children in poverty.

is to base voucher eligibility on the poverty status of the school to which they are assigned. The fourth through eighth columns of table 2.1 offer the demographic characteristics of students who would be eligible for vouchers under such a program. In these cases, and all subsequent cases, we assume that vouchers are distributed to everyone in a school, so each affected school would face the prospect of losing its entire student body. We rank schools according to the fraction of their students who qualify for a free lunch, and then we present four scenarios in which increasing numbers of students qualify. In column (4), for example, we show the demographic composition of the voucher-eligible population if the two percent of children attending the poorest schools were made eligible. This column shows that under such a plan, 69 percent of qualifying students would be black, 95 percent would be eligible for a free lunch, and the average test score among those students would be almost two standard deviations below the population mean.

Moving across the next four columns, we see the impact of providing coverage for larger fractions of the student population, who are attending increasingly affluent schools. As children attending schools with lower fractions of poor students become eligible, the recipient population is less likely to be poor and black, and the average recipient's test score increases. When students attending the bottom fifth of schools are eligible, the average reading score of the choice students is a full standard deviation above the average reading score of the students attending the bottom 2 percent of schools. Average math scores are nearly two-thirds of a standard deviation higher. These results confirm that black students are disproportionately located in low-income schools and that test scores and income are negatively correlated, facts that are well established. It is interesting to note that changing the income criteria and allowing more students to qualify has little impact on the fraction of the eligible population that is Hispanic. This suggests that Hispanics are more evenly distributed across school attendance areas than are blacks, although the difference in Hispanic representation between the first and second columns still indicates that they are somewhat concentrated in poor schools.

In the eighth column of table 2.1 we present the characteristics of students who would be choice eligible if the 44 percent of students attending the lowest-income schools were provided with a voucher. We include these results in order to provide a direct comparison with the student-based voucher scheme presented in column (2). This column indicates that, relative to using eligibility criteria that are based on student income levels, school-based eligibility criteria are much less efficient at targeting low-income children, at least when the fraction of students eligible for a voucher is sufficiently high. Whereas all of the children who receive vouchers under the student-based system have low incomes, only 68 percent of the students who receive vouchers under the school-based system have low incomes.

Conversely, the school-based criterion produces a choice-eligible population that has a higher representation of minorities. The higher representation of minorities is exactly what one would expect, given that high-income blacks and Hispanics are more likely to live in poor neighborhoods than high-income whites or Anglos. The average test scores of the targeted students are remarkably similar under both eligibility schemes.

A comparison of columns (2) and (5) is more appropriate if vouchers are randomly assigned to 5 percent of students (from among the income-eligible population). Here we see that student-based assignment would serve significantly fewer minorities and that average test score performance of the recipients would be substantially higher. To the extent that the voucher population is nonrandomly selected from the eligible population (as would be expected if eligible individuals need to apply for vouchers⁹), differences between the two schemes would be more dramatic. Of course, a student-based program that targeted the 5 percent of students with the lowest incomes would undoubtedly produce stratification effects more similar to those presented in the fifth column.

2.3.2 Tying Voucher Eligibility to School Accountability

Voucher programs that are embedded within an accountability system base eligibility on school test performance, rather than on individual or school-level demographic characteristics. This section investigates the characteristics of children who are eligible for school vouchers when eligibility is determined by school performance. We consider several different ways of evaluating schools: Some performance measures are based on test score levels, some are based on test score levels adjusted for demographic characteristics, and some are based on value-added test scores. All of the performance measures we present are based on reading tests. Our results are trivially affected by the use of math scores, or combinations of math and reading scores that emulate Florida's program.¹⁰

The first type of performance measure that we consider is the school's average test score in a single year. We rank schools according to their average test score on the fourth grade FCAT reading test, and as in table 2.1 we choose eligibility cutoffs according to whether choice is to be made available to 2, 5, 10, or 20 percent of the population. Schools can also be ranked (and vouchers allocated) according to the fraction of students who attain a particular competency level. This performance measure is closer to Florida's current school grading system.

A problem with these two measures is that they may penalize schools that serve disadvantaged communities, since disadvantaged students do not typ-

9. The available evidence on voucher experiments such as Milwaukee's suggests that this may be the case.

10. Results based on the other performance measures are available from the authors.

ically perform as well on exams as their more advantaged peers. If the aim of vouchers is to give choice to students attending poorly performing schools, then it may be more appropriate to measure schools' success using test scores that take differences in student composition into account. Therefore, we also consider the ramifications of a system that bases eligibility on the school's average test score after controlling for its demographic and income composition.¹¹ An alternative to regression-adjusting test scores for the purposes of controlling for student body characteristics is to use value added test scores. Two types of value added constructs are generally considered in policy discussions of accountability systems: measures of value added constructed from changes in test scores from one cohort to the next (within grade changes). As with the regression-adjusted levels method, value added test scores net out school demographics, which may be a more appropriate way of measuring schools' effectiveness. Value added test scores are also popular because they are perceived to reward and punish schools according to improvements or declines in quality over time.¹²

2.3.3 Voucher Eligibility under Different Accountability Systems

The next four tables show how the population of voucher-eligible students varies with the parameters of the accountability system. Each row in tables 2.2–2.5 represents a different method of rating schools, and the columns indicate the size of the program (extending eligibility from 2 percent to 20 percent of the state's student population). Each table presents the results for a different sociodemographic characteristic of the population: Table 2.2 reports the fraction of voucher-eligible students who are free lunch-eligible under each scheme, table 2.3 reports the average standardized reading test score of voucher-eligible students, table 2.4 presents the fraction of voucher-eligible students who are black, and table 2.5 presents the fraction of voucher-eligible students who are Hispanic.

As table 2.2 makes clear, the socioeconomic composition of voucher-eligible students varies dramatically across the different accountability systems. When 2 percent of the population is made voucher eligible, the fraction of eligibles who are also able to receive free lunch ranges from 44 percent to 89 percent, depending on which performance measure is used. As the fraction of the population eligible for vouchers increases, the demographic variation across accountability systems falls somewhat, but even when 20 percent of students are eligible the difference across the systems is

11. An alternative, but very similar, approach involves regression-adjusting the school ranking based on the fraction of the school's student body attaining some particular threshold of achievement.

12. Several authors have pointed out technical challenges associated with using test scores for the purpose of school accountability and with constructing value added measures of school productivity. See, for example, Clotfelter and Ladd (1996), Koretz (1996), Ladd and Walsh (2002), Ladd (2001), and Meyer (1996). Some of these issues are summarized in Ladd and Hansen (1999).

Table 2.2 Percentage of Voucher-Eligible Students Who Are Free Lunch–Eligible under Different Accountability Scenarios

Accountability System Basis of School Ranking	% of State's Student Body Eligible for a Voucher			
	2	5	10	20
Average performance on fourth-grade FCAT reading examination	88.6	86.2	80.3	73.8
% of students achieving at least minimum competency on FCAT reading examination	89.1	86.9	82.3	74.4
Average performance on fourth-grade FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	44.6	49.0	49.8	47.9
% of students achieving at least minimum competency on FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	48.3	54.5	51.4	49.5
Cross-cohort changes in the average performance on fourth-grade FCAT reading examination	52.4	49.4	51.8	46.4
Cross-cohort changes in the average performance on fourth-grade FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	55.0	52.0	52.9	50.1
Within-cohort changes in performance on the FCAT reading examination, from fourth grade to fifth grade	60.6	57.1	53.6	49.8
Within-cohort changes in performance on the FCAT reading examination, from fourth grade to fifth grade, adjusted for racial/ethnic composition and socioeconomic status in school	57.7	55.8	50.3	47.0

Notes: The numbers in this table reflect the percentage of voucher-eligible students who would be free lunch–eligible in each of thirty-two different voucher plans. Each row represents a different targeting scheme, and each column represents a different level of voucher coverage.

substantial. The accountability systems based on level test performance provide vouchers to a much more disadvantaged population than do accountability systems based on regression-adjusted test score levels or value added.

The patterns are subtly different when students are stratified by their test performance. These results are shown in table 2.3. As with income status, the average standardized test score of voucher-eligible students increases with the fraction of students eligible to receive a voucher. This is not at all surprising, given that school grades are constructed using test scores. Also comparable to the results shown in table 2.2 is the finding that value added measures systematically produce a voucher-eligible population that has higher test scores. Unlike in table 2.2, however, in this case we observe that the more completely we control for school and student fixed effects, the

Table 2.3 Average Standardized Reading Test Scores of Voucher-Eligible Students under Different Accountability Scenarios

Accountability System Basis of School Ranking	% of State's Student Body Eligible for a Voucher			
	2	5	10	20
Average performance on fourth-grade FCAT reading examination	-2.69	-2.21	-1.84	-1.41
% of students achieving at least minimum competency on FCAT reading examination	-2.59	-2.18	-1.80	-1.37
Average performance on fourth-grade FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	-1.55	-1.36	-1.16	-0.82
% of students achieving at least minimum competency on FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	-1.61	-1.45	-1.12	-0.84
Cross-cohort changes in the average performance on fourth-grade FCAT reading examination	-0.99	-0.66	-0.65	-0.41
Cross-cohort changes in the average performance on fourth-grade FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	-1.07	-0.72	-0.69	-0.52
Within-cohort changes in performance on the FCAT reading examination, from fourth grade to fifth grade	-0.31	-0.20	-0.18	-0.11
Within-cohort changes in performance on the FCAT reading examination, from fourth grade to fifth grade, adjusted for racial/ethnic composition and socioeconomic status in school	-0.17	-0.16	-0.07	-0.02

Notes: The numbers in this table reflect the average standardized reading score of voucher-eligible students in each of thirty-two different voucher plans. Each row represents a different targeting scheme, and each column represents a different level of voucher coverage.

higher-performing are the resulting voucher-eligible students. Accountability systems that use each student as his or her own control (the within-cohort models) lead to a group of voucher-eligible students who perform at nearly the same level as the overall population, whereas those that control for only a few background characteristics lead to a group of voucher-eligible students who are substantially higher performing than those controlling for no background characteristics.

Tables 2.4 and 2.5 present the results for race and ethnicity. We find that the patterns in table 2.4, reflecting the percentage of voucher-eligible students who are black, closely mirror the patterns reported in table 2.2. This is unsurprising given the high correlation between race and poverty status. The patterns in table 2.5 are different, however, and do not show a substantial relationship between the construct of the accountability system and the

Table 2.4 Percentage of Voucher-Eligible Students Who Are African-American under Different Accountability Scenarios

Accountability System Basis of School Ranking	% of State's Student Body Eligible for a Voucher			
	2	5	10	20
Average performance on fourth-grade FCAT reading examination	71.0	67.6	59.1	51.4
% of students achieving at least minimum competency on FCAT reading examination	74.7	75.2	63.3	53.9
Average performance on fourth-grade FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	31.1	32.2	31.3	28.1
% of students achieving at least minimum competency on FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	30.1	34.4	34.0	30.1
Cross-cohort changes in the average performance on fourth-grade FCAT reading examination	28.5	27.3	24.4	24.9
Cross-cohort changes in the average performance on fourth-grade FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	38.0	34.0	31.4	28.1
Within-cohort changes in performance on the FCAT reading examination, from fourth grade to fifth grade	37.2	33.4	28.0	26.1
Within-cohort changes in performance on the FCAT reading examination, from fourth grade to fifth grade, adjusted for racial/ethnic composition and socioeconomic status in school	38.1	36.3	30.6	26.6

Notes: The numbers in this table reflect the percentage of voucher-eligible students who would be African American in each of thirty-two different voucher plans. Each row represents a different targeting scheme, and each column represents a different level of voucher coverage.

fraction of voucher-eligible students who are Hispanic. This finding is due to the fact that Florida's Hispanic students are considerably more affluent and less concentrated in particular schools than are Florida's black students.

Our finding that value added performance measures lead to a more advantaged voucher-eligible population is intuitive given that school quality and *changes* in school quality are not the same thing. In fact, ceiling effects are likely to produce a negative correlation between test score levels and test score changes: Schools with high average test scores in a particular year cannot experience large test score gains from one year to the next, whereas schools with low average test scores in a particular year have more room to move up. Furthermore, Kane and Staiger (2000) show that value added measures may be noisy signals of test score performance, leading to an ar-

Table 2.5 Percentage of Voucher-Eligible Students Who Are Hispanic under Different Accountability Scenarios

Accountability System Basis of School Ranking	% of State's Student Body Eligible for a Voucher			
	2	5	10	20
Average performance on fourth-grade FCAT reading examination	21.9	26.0	28.7	29.2
% of students achieving at least minimum competency on FCAT reading examination	19.7	19.6	27.1	27.5
Average performance on fourth-grade FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	14.6	19.6	20.7	18.9
% of students achieving at least minimum competency on FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	18.5	20.4	18.0	18.9
Cross-cohort changes in the average performance on fourth-grade FCAT reading examination	29.6	26.6	31.7	24.3
Cross-cohort changes in the average performance on fourth-grade FCAT reading examination, adjusted for racial/ethnic composition and socioeconomic status in school	18.9	17.9	20.3	21.7
Within-cohort changes in performance on the FCAT reading examination, from fourth grade to fifth grade	24.3	24.9	28.0	24.4
Within-cohort changes in performance on the FCAT reading examination, from fourth grade to fifth grade, adjusted for racial/ethnic composition and socioeconomic status in school	17.5	17.7	18.8	20.1

Notes: The numbers in this table reflect the percentage of voucher-eligible students who would be Hispanic in each of thirty-two different voucher plans. Each row represents a different targeting scheme, and each column represents a different level of voucher coverage.

bitrary assignment of school grades. We find that the correlation between the 1999–2000 test score change and the 1998–1999 test score change is actually negative (-0.33), suggesting that value added measures are failing to pick up real changes in school quality.¹³

Kane and Staiger (2000) note that the noisiness of value added measures is a more serious problem for small schools because the set of students over

13. Our limited evidence also suggests that parents are able to tell the difference between the information provided by test score levels versus test score changes. Using data from the Gainesville, Florida metropolitan area and following the methodology used by Figlio and Lucas (2000), we find that level scores are more capitalized into housing prices than are value added test scores, even when student demographics are taken into account. Similarly, families in Gainesville request zoning exceptions to send their children to schools with high test score levels at five times the rate of those requesting zoning exceptions to send their children to schools with high value added.

which test scores are aggregated is smaller. Consistent with Kane and Staiger, we find that a one standard deviation reduction in school size is associated with a 0.15 standard deviation increase in the absolute difference between 1999–2000 value added and 1998–1999 value added. Kane and Staiger’s findings, together with our own, suggest that although value added test scores may at first blush appear to be reasonable measures of school improvement, their use may actually undermine the goals of school accountability and vouchers.¹⁴

In summary, we observe considerable differences between the types of students who are eligible for vouchers when voucher eligibility is based on level test scores and the types of students eligible for vouchers when eligibility is based on “purer” measures of school performance. Voucher awards that are based on measures of school performance that do not take background characteristics into account tend to be better at targeting disadvantaged students than are voucher awards based on value added measures. We conclude that, in voucher systems tied to school accountability, there may be a disconnect between the appropriate ways of assessing schools for the purposes of accountability and the desire to provide additional choices to students who are arguably most constrained.¹⁵

2.4 Implications

Although many states were already considering school voucher programs and accountability systems, President Bush’s education proposal has catapulted these school “fixes” into prime time. Even though only the

14. One solution proposed by Kane and Staiger is to use “filtered” test score estimates. Filtered estimates use past relationships between test scores to optimally construct summary performance measures. The summary measure includes weighted averages of current and past test scores in multiple subjects, with the weights varying by school size. We have constructed a crude filter by averaging both math and reading test scores over a two-year period (averaging a total of four test scores) and then ranking schools according to cross-cohort changes in this average. Our results differ very little from those shown in the fourth rows of tables 2.2–2.5. This suggests that although filtered estimates may do a better job of capturing true changes in school performance, using filtered estimates to target vouchers will not necessarily affect which types of individuals end up qualifying. In other words, even if changes in school performance can be measured accurately, using these changes to target vouchers will lead to a substantially different population of voucher eligibles than a system that provides vouchers to students in schools with low performance levels.

15. One might argue that sending a child to a low-performing school is evidence of constraint. For instance, high-income families with heavy debt loads may elect to send their children to low-performing schools despite their “observed” ability to select a higher-performing school. Although these families are surely less constrained than low-income households, this should serve as an important caveat to our assertion that the “purer” the accountability system, the least constrained are the households potentially offered vouchers in an embedded system. An alternative story would bias the conclusion in the other direction, however: If unconstrained households resettle in neighborhoods zoned for failing schools to take advantage of school vouchers, as has been predicted theoretically by Nechyba (2000), this could exacerbate the disconnect between the goals of school accountability and the goals of school choice described in the paper.

public school choice portion of the President's proposal became law, the struggle to improve school quality without increasing costs ensures that these programs will be at the top of legislators' agendas for the foreseeable future. It is likely that, as in Florida, new voucher programs will be developed within the context of accountability systems.

At first glance, it might seem appropriate for accountability systems and voucher initiatives to go hand in hand. Both types of plans have been proposed as a way of increasing school efficiency, and the rhetoric surrounding both accountability and voucher discussions has particularly focused on improving educational opportunities for America's most disadvantaged children. Our results suggest that a voucher program embedded within an accountability system could fail to achieve this latter goal, however.

This interpretation of our results depends on one's interpretation of what it means to be disadvantaged. Economists usually promote voucher programs on the grounds that they increase school choice among individuals whose mobility would otherwise be constrained. In these discussions, disadvantage is measured with the usual socioeconomic indicators (income and minority status) because these characteristics are thought to have a strong impact on an individual's ability to move. The measure of disadvantage that is inherent in an integrated voucher/accountability system, however, is whether the individual attends a poorly performing school. Because school test scores are so strongly correlated with the socioeconomic background of the student body, and because much of the difference in outcomes across student types can be attributed to family background, it makes sense to base accountability on measures of performance that net out socioeconomic factors or that measure growth in student test scores, rather than levels of test performance. When vouchers are targeted this way, however, the composition of choice-eligible children is vastly different.

One can make a case that the students most deserving of vouchers are those whose schools are doing the least to educate them, but there are several reasons we believe that such a targeting scheme will be less successful at helping the "truly" disadvantaged. First, because value added measures of school performance are noisy, vouchers allocated in this way may bypass many of the students attending the least effective schools in favor of those attending more effective schools. Second, ranking schools according to their performance net of socioeconomic characteristics necessarily ignores the possibility of peer effects. Although this may be appropriate for determining school performance, it could lead to an inaccurate assessment of which students are being schooled in the worst environments and, thus, who would most benefit from receiving a voucher.

Like most economists, our third argument against basing voucher eligibility on school effectiveness is that such a targeting scheme will be less successful at reaching the students whose schooling choices are constrained. Our results indicate that choice-eligible children are much less likely to

come from low-income families when eligibility is determined according to school performance (when effectiveness is measured net of students' family background) than when it is determined using income criteria. If vouchers are provided to students who are already able to choose private schools but have elected not to, then they will accomplish very little.

Of course, important caveats accompany our interpretation of the simulations. Foremost is the obvious point that if Florida's distribution of student types across schools is atypical, then the eligibility differences that we have documented across targeting schemes may not be generalizable. In particular, the eligibility differences will be less dramatic if, in other states, low-income and minority students are more evenly distributed across schools. Our results may understate the stratification differences in states where students are more concentrated by race, ethnicity, and income, however, as they appear to be in California and Texas.¹⁶

Another important limitation to our study is that we do not have information on students' income level, only on whether they are free- or reduced-price lunch-eligible. Our results might look less dramatic if we were able to see how the entire income distribution of voucher eligibles was affected by the various targeting schemes. Nevertheless, our findings suggest that education policymakers should be very careful about articulating their goals when designing voucher programs that are part of an accountability system. In spite of their seemingly similar aims, it may be very difficult to achieve the goals of an accountability system and a voucher program in an integrated system.

References

- Clotfelter, Charles, and Helen Ladd. 1996. Recognizing and rewarding success in public schools. In *Holding schools accountable: Performance-based reform in education*, ed. Helen Ladd, 23–64. Washington, D.C.: Brookings Institution.
- Figlio, David, and Maurice Lucas. 2000. What's in a grade? School report cards and house prices. Working Paper no. 8019. Cambridge, Mass.: National Bureau of Economic Research, November.

16. Some evidence of this possibility is present in Florida. Broward and Miami-Dade Counties, two adjacent south Florida counties that have the largest student populations in the state, are starkly different in their concentrations of minorities and low-income students. Whereas Broward County's concentrations are about the same as the state as a whole, Miami-Dade's concentrations are much higher. The correlation between level test performance (in reading) and the one-cohort and two-cohort measures of value added in reading are, respectively, 0.16 and 0.12 in Broward County, and -0.22 and -0.01 in Miami-Dade County. Although this should not be taken as definitive evidence of our point, it corroborates the assertion that the effects we describe in this paper may be larger in places where minorities and low-income students are more concentrated in schools.

- Kane, Thomas, and Douglas Staiger. 2000. Improving school accountability measures. Hoover Institution, Stanford University. Working Paper.
- Koretz, Daniel. 1996. Using student assessments for educational accountability. In *Improving America's schools: The role of incentives*, 171–95. Washington, D.C.: Commission on Behavioral and Social Sciences and Education, National Research Council, National Academy Press.
- Ladd, Helen. 2001. *School-based educational accountability systems: The promise and the pitfalls*. *National Tax Journal* 54 (2): 385–400.
- Ladd, Helen, and Janet Hansen, ed. 1999. *Making money matter: Financing America's schools*. Washington, D.C.: National Academy Press.
- Ladd, Helen, and Randall Walsh. 2002. Implementing value-added measures of school effectiveness: Getting the incentives right. *Economics of Education Review* 21 (1): 1–17.
- Meyer, Robert. 1996. Value-added indicators of school performance. In *Improving America's schools: The role of incentives*, Commission on Behavioral and Social Sciences and Education, 197–223. Washington, D.C.: National Research Council, National Academy Press.
- Nechyba, Thomas. 2000. Mobility, targeting, and private-school vouchers. *American Economic Review* 90:130–46.