Can School Choice and School Accountability Successfully Coexist?

David N. Figlio
University of Florida and NBER
figlio@ufl.edu

Marianne E. Page
University of California, Davis and
Joint Center for Poverty Research
m-page1@northwestern.edu

January 28, 2001

Can School Choice and School Accountability Successfully Coexist?

I. Introduction

Concern over the low quality of public schools combined with state and local governments' financial constraints, has put school accountability and voucher plans at the forefront of education policies. Both types of proposals aim to increase public school quality without increasing state and local governments' education costs. But, whereas accountability systems typically reward and punish schools by allocating funding according to whether the school meets certain government-determined performance criteria, voucher programs are expected to improve school quality by letting schools compete for students. The typical voucher program provides students with a tax-financed voucher of a fixed amount, which can be applied toward tuition at a participating private school. When a student elects to attend a private school instead of her local public school, the public school's funding is decreased by the amount of the voucher. The threat of budget cuts is expected to provide schools with an incentive to improve.

The main criticism of voucher programs is that they may indirectly reduce the quality of education available to disadvantaged students. If many students leave their public school as a result of a voucher plan, then those students who remain may have fewer school resources available to them. Furthermore, if voucher programs differentially elicit mobility among the most able students, then those students who remain in public schools may further suffer from reductions in peer group quality. A complete assessment of the equity implications of a particular voucher program would need to take into account the program's eligibility criteria, the degree to which peers

influence student outcomes, the elasticity of student mobility with respect to the value of the voucher, and the degree to which schools adjust their services in response to competition. Since voucher programs are in their infancy and it may take several years for some of these adjustments to take place, it will be some time before the long-run distributional impact of vouchers can be determined.[1]

In the short-run, however, we can estimate how different eligibility criteria would affect the socio-economic composition of the choice-eligible population. Understanding the first-order stratification effects of different eligibility schemes is a necessary first step towards understanding the equity implications of voucher programs. The purpose of this paper is to estimate which students would be eligible for a voucher under different voucher plans. We will use data from elementary school students in the state of Florida, whose newly adopted voucher plan is of particular interest because it is so closely tied to school accountability. Unlike most existing programs, which use student income as the basis for eligibility, the Florida plan will make vouchers available to all students attending persistently low-performing schools. President George W. Bush has proposed a similar voucher initiative as part of his national accountability-based education plan.

Although at first blush it would seem that voucher programs targeting low-income children and voucher programs targeting poorly performing schools would serve 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

---

[1] One exception is Hoxby (1996) who attempts to empirically predict the distributional effect of vouchers using variation in tuition subsidies across metropolitan areas. She finds that a $1000 voucher would

different ways of evaluating a school's performance. Our results suggest that the

performance criteria that may be most appropriate from an accountability perspective

may not do the best job of targeting disadvantaged children.[2] Thus, although voucher

programs and accountability systems tend to have similar goals, linking the two types of

programs together may have unintended consequences on the distribution of students

who are served.

II.   Vouchers and Accountability in Florida

This paper uses data from Florida to simulate which students would be eligible for

school vouchers under a variety of different eligibility plans. Florida is one of only a

handful of states that has records of student attributes and test score outcomes with

enough detail for our project to be pursued. An advantage of using Florida data is that

our results will have immediate policy implications for that state, which is currently in

the process of implementing the first voucher program to be imbedded within an

accountability system.

While Florida is currently the only state to have included a voucher component in

its accountability plan, President Bush has recently unveiled a similar national

accountability plan that would provide vouchers to students attending schools with poor

---

increase the fraction of Catholic and African American students attending private school. A number of
theoretical papers have also made predictions about the distributional impact of vouchers. See, for
example, Epple and Romano (1998) and Nechyba (2000).

[2] Empirical evidence on the groups most likely to benefit from receiving school vouchers is very limited.
However, to the extent to which evidence is available, it suggests that disadvantaged students are the group
most likely to benefit from private schools. While the literature on the relative performance of public and
private schools is divided concerning the mean performance of the two sectors, there is a general consensus
in the literature that private schools tend to disproportionately benefit urban students, and particularly urban
minorities. For instance, Figlio and Stone (1999), Grogger and Neal (2000) and Neal (1997) find positive
effects of Catholic schooling on academic outcomes for these groups. This result, however, is not
universal, as Jepsen (2000) does not find that urban students are differentially benefited by private
schooling.

performance records. One might, therefore, like to know whether the results generated

by our Florida data can be generalized to other states. This rests largely on whether low-

income and minority students are similarly distributed across schools in other states.

Compared to California and Texas, Florida's minorities (that is, students who are Black,

Hispanic, Asian, or Native American) tend to be somewhat more evenly distributed

across schools.[3] For example, whereas 75 percent of California minorities attend schools

that are at least 58 percent minority and 75 percent of Texas minorities attend schools

that are at least 52 percent minority, 75 percent of Florida's minority students attend

schools in which at least 37 percent of the student body consists of minority students.

Similar, though less dramatic, patterns exist for low-income students: seventy-five

percent of Florida's free lunch-eligible population attend schools in which at least one-

third of the students are free lunch-eligible, whereas the same fraction of California's free

lunch students attend schools where at least 46 percent of the students are eligible for a

free lunch. Likewise, seventy five percent of Texas's free lunch students attend schools

in which 36 percent or more of the students are eligible for a free lunch. The school-level

correlation between percent minority and percent free lunch in Florida is 0.60, compared

to 0.71 in California and 0.79 in Texas. This tells us that not only are minorities and low-

income children more evenly distributed across schools in Florida, but that relative to

California and Texas, low-income children are somewhat less likely to be minorities.

Overview of the Florida Program

---

[3] The figures cited in this paragraph are derived from the 1996-97 Common Core of Data.

The centerpiece of Florida Governor Jeb Bush's 1998 campaign platform was an education reform plan that emphasized school accountability and included vouchers for students attending low-performing schools. His program, known as the "A+ plan," was enacted by the legislature the following spring. At the heart of this plan is an accountability system in which each school receives a letter grade ranging from A to F, that is primarily determined by its students' performance on the Florida Comprehensive Assessment Test (FCAT). The letter grades, except for the "A"-"B" threshold, are based on aggregate student performance on three examinations: fourth, eighth or tenth grade reading scores, fifth, eighth or tenth grade mathematics scores,[4] and fourth, eighth or tenth grade performance on the Florida Writes! examination. In order to attain a grade of AC@,[5] at least 60 percent of test-takers must achieve at level two or above (on a five-point scale) on the FCAT reading test, at least 60 percent of test-takers must achieve at level two or above (also on a five-point scale) on the FCAT mathematics test, 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.[6] If a school misses one or two of these thresholds it receives a grade of AD." If it misses all three of these thresholds it receives a grade of AF@. The first assignment of school letter grades occurred in May 1999, and about eight percent of schools receiving grades[7] earned grades of "A", while 13, 51, and 25 percent

---

[4]Prior to the 1999-2000 academic year, students took the FCAT reading assessment in grades four, eight and ten, the FCAT mathematics assessment in grades five, eight, and ten, and the Florida Writes! examination in grades four, eight, and ten. Beginning in 1999-2000, students now take reading and mathematics tests in grades three through ten.

[5]The basis for school letter grades have changed subtly between 1999 and 2000. Because no observations in our data set are affected by the 2000 letter grade assignment, all discussion of letter grades in this paper refer to the 1999 criteria and assignment.

[6] 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.

[7] Schools did not receive letter grades if their tested cohorts were sufficiently small.

received grades of "B", "C" and "D", respectively. Just over three percent of schools received grades of "F."

Schools on both ends of the spectrum are affected by their letter grade: those receiving an AA" grade, or who increase their letter grades from one year to the next, are eligible to receive an additional $100 per pupil that can be spent for such purposes as hiring teacher aides or providing teacher bonuses. On the other hand, students attending schools rated AF@ in two years out of four are eligible for school vouchers, or Aopportunity scholarships,@ that can be used to send a child to a private school. Alternatively, these students can use the vouchers to transfer to a nearby AC@ or higher-rated public school.

In principle, vouchers are a natural component of an accountability system, because they provide an "exit route" for students enrolled in chronically low-performing schools. The notion of providing choice for students trapped in failing schools was prevalent in arguments supporting the A+ plan, and it is highlighted also in President Bush's education proposal. While the A+ plan currently bases school grades, and, thus, voucher eligibility, exclusively 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 2002-2003 academic year, school grades be based, at least in part, on "value added" measures of student performance. While the specific nature of value added school assessment is still under deliberation, it is clear that school grades in Florida will soon be based in large measure on features other than levels of aggregate student performance. This, too, is part of President Bush's proposal.

Correlations Between School Rankings

While the Florida accountability system, and hence, its school voucher program, is based on a relatively complicated set of school test score criteria, there is a high correlation between school rankings based on these criteria and those based on other combinations of aggregate test score levels. For instance, the correlation between a school's ranking using the *current* Florida system and its ranking using either average math performance, average reading performance, the lesser of the two performance levels or the greater of the two performance levels is between 0.94 and 0.96. The correlation between a school's rank based on its performance level in a single year, and its rank when the ranking scheme is based on a "value added" measure of performance is considerably lower, however. For example, the correlation between a school's rank when ranking is based on average reading and mathematics scores (together) and its rank when the ranking scheme is based on the same average test scores, regression adjusted for racial, ethnic, and percent free lunch differences, is 0.58—still high, but considerably lower than the correlation between rankings that are based on levels of test performance. This suggests that the students who would qualify for vouchers under President Bush's plan (and in Florida) may come from very different backgrounds than those that would qualify if eligibility were based on test score levels.

Correlations between rankings based on level test scores and rankings based on "purer" concepts of value added are lower still: the correlation between a school's rank when ranking is based on average test scores in one year and its rank when the ranking scheme is based on within cohort *changes* in average test scores from one year to the

next[8] is only 0.05. If the year-to-year change-based rank is regression-adjusted, the correlation falls further to –0.02. The correlation between a school's rank when ranking is based on average test scores in one year and its rank when ranking is based on changes in average test scores from one cohort to the next—a different method often considered for measuring school improvement—is slightly higher, but still only 0.12 (0.24 if the "value added" measure is regression-adjusted.)

None of these rankings of school performance are particularly correlated with school spending (at the district level.)[9] The correlation between per pupil district spending and a school's average reading and math score is –0.14 (unsurprising, given that higher-spending districts tend to be located in urban areas where costs are higher) and the correlation between per pupil spending and regression-adjusted average school grades is 0.06. Correlations between per pupil expenditures and any of the value added measures range from 0.01 to 0.06. These low correlations should be taken with a grain of salt, given Florida's rather flat spending distribution across districts (the 95:5 ratio in spending is only 1.2) but suggest that, within this spending range, school spending has little relationship with rankings of student outcomes by any measure.

In sum, it is evident that school rankings based on levels of student performance may be quite different from school rankings that are based on changes in student performance, and that none of these measures are particularly correlated with school spending. This in turn suggests that within an accountability system, the types of students who qualify for vouchers may vary with the nature of the criteria by which

---

[8] Because of the nature of the Florida data, we can currently only make this comparison using reading test scores, but not mathematics test scores. However, when measured at the school average level, reading scores and mathematics test scores are very highly correlated.

schools are evaluated. In addition, it suggests that the set of students who will be voucher eligible under an accountability-based voucher program may be considerably different from the set of students served by voucher systems more directly targeted at "disadvantaged" families. The purpose of the next section is to provide a first attempt at quantifying these differences, looking specifically at the student and school population in Florida.

III.   Who is Eligible for a Targeted Voucher?

Our goal is to simulate and compare the first-order stratification effect of various targeting schemes. Our stratification measures include race/ethnicity, family income (whether the student qualified for a free lunch), neighborhood quality (whether the student lives in a neighborhood in which at least 53 percent of the residents are below the poverty line[10]) and student test scores. The data used in our simulations come from student-level information on race, ethnicity, test scores, free/reduced-price lunch status, school attendance, and zip code of residence, which we use as a rough proxy for the child's neighborhood. These data have been provided by the state of Florida's Department of Education and reflect the Florida student population in the 1999-2000 academic year. We begin by looking at the socioeconomic composition of the choice-eligible population when voucher eligibility is determined using student attributes, and then consider what happens to the socioeconomic composition of the choice population when school-level attributes are used to determine eligibility. We then turn our attention towards understanding the ramifications of school-based accountability/voucher systems.

---

[9] These measures of school spending are per pupil (FTE) current expenditure in the 1998-99 school year, as reported by the Florida Department of Education.

We begin by presenting the attributes of the set of students who would be eligible to receive a voucher if all students in the school system were made eligible. This effectively provides us with information about the demographic characteristics of the student population in Florida. The first column of Table 1 displays the results (and is reproduced in all the other tables, to provide a comparison.) 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.[11]

Voucher Eligibility Based on Socio-Economic Status

While universal vouchers have been proposed by some policy-makers (indeed, California's recent voucher initiative would have made vouchers available to all students), the equity concerns associated with such proposals means that those programs that are actually implemented are likely to be targeted towards specific subsets of the student population.   Milwaukee's voucher program is probably the best-known, and so we begin by approximately replicating 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 cannot directly simulate the effects of this criterion for Florida, because we do not have explicit information on students' income, but we can look at how stratification is affected when vouchers are targeted only at students who are

---

[10] We selected this measure to reflect the one-quarter poorest neighborhoods in the state.

[11] 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

eligible for free or reduced price lunches. Free and reduced price lunch income breakpoints are not out of line with Milwaukee's poverty-based eligibility thresholds. Milwaukee's threshold of 175 percent of the poverty line falls between the free lunch eligibility threshold of 130 percent of the poverty line and the reduced-price lunch threshold of 185 percent of the poverty line.

The second two columns in Table 1 present the characteristics of students eligible for free (or free and reduced-price) lunches in Florida. If every free lunch eligible student in Florida were to receive a voucher, then 44 percent of the state's student body would be voucher-eligible, and if the eligibility threshold were increased to include reduced-price lunch eligible students as well, coverage would rise to 53 percent. Free lunch eligible students are disproportionately minorities (primarily black) and disproportionately reside in very low-income neighborhoods. 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 higher income students, the average test score of the voucher eligible population under this scenario would be around six-tenths of a standard deviation below the statewide mean.

An alternative mechanism of delivering vouchers to disadvantaged students is to base eligibility on the poverty status of the school attended, rather than on the individual's poverty status. The fourth through eighth columns of Table 1 offer the demographic characteristics of students who would be eligible for vouchers if vouchers were distributed according to the poverty status of the school to which a child is assigned. Here, we rank schools according to the fraction of their students who qualify for a free

comparison. The decision to include or exclude these test takers from the standardization makes no difference for the results presented in this paper.

11

lunch, and then we present four scenarios in which increasing numbers of students qualify. In column four, 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 voucher plan, 69 percent of the qualifying students would be black, 95 percent would be eligible for a free lunch and the average test score amongst 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 two percent of schools. Average math scores increase by nearly two-thirds of a standard deviation. 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 in the education literature. 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 concentrated in poorer schools than the median student.

In the eighth column of Table 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. These two columns indicate 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 criteria produces a choice-eligible population that has a higher representation of minorities and a higher representation of students living in high poverty neighborhoods. 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. The average test scores of the targeted students are remarkably similar under both eligibility schemes.

While the differences between columns 2 and 8 provide us with information about the distributional effects of using student vs. school based eligibility criteria, few proposals would provide vouchers to such a large fraction of the student population. Roughly half of students would be income-eligible for a voucher under a Milwaukee-type program in Florida, but in such a program only a small fraction of eligible students are typically provided a voucher.[12] Many existing student-based plans have chosen

---

[12] In Milwaukee, only 15 percent of the student body participates in the Milwaukee Parental Choice Program despite a much larger fraction being income-eligible.

recipients by lottery from a pool of income-eligible applicants. If the recipients are truly randomly selected from the income-eligible population, then the attributes of the students who are actually provided with school-choice will still be close to those presented in column 2. But compared to a school-based measure that targets the 5 percent of students attending the lowest income schools, significantly fewer minorities would be served, and the average test score performance of the recipients would be substantially higher. To the extent that the voucher population is non-randomly selected from the eligible population (as would be expected if eligible individuals need to apply for vouchers[13]) then these differences may 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.

An alternative to using school-based poverty measures to determine voucher eligibility would be to offer vouchers based on neighborhood poverty. The last five columns of Table 1 show what happens to the demographic composition of eligible children when eligibility is based on the fraction of children in the child's zip code who are living below the poverty line. Again, we present four scenarios, in which each scenario represents a different goal in terms of the number of children the program is trying to target. In the ninth column of Table 1, for example, we present the demographic implications of targeting the 2 percent of children living in the poorest neighborhoods, and in the tenth through twelfth columns we show how these demographics change when children living in increasingly less poor neighborhoods are targeted.[14] Compared to the

---

[13] The available evidence on voucher experiments such as Milwaukee's suggests that this may be the case.
[14] We note that 100 percent of students live in high poverty neighborhoods in these scenarios because high poverty neighborhoods are defined in the table as the top quartile of neighborhoods, ranked by poverty status.

school based vouchers, fewer blacks qualify and more Hispanics qualify under each scenario. A voucher targeted at neighborhoods is also slightly less successful at reaching the poorest students, and those with low test scores, in part because neighborhoods (at least when measured at the zip code level) are less homogenous than schools.

Voucher eligibility based on school's level test performance

School voucher programs that are embedded within an accountability system base voucher eligibility on the test performance of the school's students, rather than on demographic or economic characteristics of the recipients or their schools. This subsection investigates the characteristics of children eligible for school vouchers when voucher eligibility is determined by school performance. The next two tables provide information about how the demographics of the voucher-eligible population are affected when eligibility is determined by the school's test score performance in a single year. The test scores used in this analysis are fifth grade mathematics and fourth grade reading FCAT scores. We rank schools according to their average test score in either/both reading and math and again choose eligibility cutoffs based on whether choice is intended to be made available to two percent, five percent, ten percent or twenty percent of the population. We use four different ranking schemes: the first section of Table 2 ranks schools according to their average test score in the better of either reading or math (this is equivalent to targeting schools that fail to meet performance targets in *both* reading and math) and the second section ranks schools according to their average test score in the lower of either reading or math (this is equivalent to targeting schools that fail to meet performance targets in *either* reading or math). Table 3 provides the results when

15

voucher eligibility is based on the school's average reading or math scores alone. The correlation between average reading and mathematics scores is very high—0.84—which helps to explain why the results reported above are very similar across scenarios.

In all four cases, we see that, unsurprisingly, voucher eligibility that is determined according to school performance does a better job of targeting low performing children than does voucher eligibility that is determined according to the school's fraction poor. On the other hand, compared to an income-based criterion (in which income is determined at the school level), using a performance criterion is less successful at targeting children whose school choice decisions are financially constrained. These differences are relatively small, however. In each case, we note that the fraction black, free lunch eligible, and low-achieving monotonically declines as the proportion of the student population covered by vouchers increases. We also note that, conditional on ranking schools based on average test performance, the *specific nature* of the school ranking has little effect on the characteristics of the eligible population.

Table 4 switches the performance criteria from one in which the school's average test score must be below some threshold, to one in which eligibility is based on the fraction of students who attain a particular competency level. This variant of the level-performance eligibility program is closest to Florida's current school grading system, in which school grades, and (indirectly) vouchers, are currently based not on the average test score in the school, but rather the fraction attaining some cutoff level of competency on their tests. In Florida, students are rated at five levels of competency, where level 1 means that the student did not attain even a minimal level of competency and level 5 represents extremely high levels of performance. In fifth grade mathematics, these levels

currently correspond to standardized test score levels of about –1.0 (level 2), 0.6 (level 3), 1.9 (level 4), and 3.6 (level 5). In fourth grade reading, these levels currently correspond to standardized test score levels of about –0.6 (level 2), 0.3 (level 3), 1.9 (level 4), and 3.8 (level 5). Seventy-three percent of fifth graders in Florida attained at least level 2 competency in mathematics in the 1999-2000 school year, while 67 percent of Florida fourth-graders attained at least level 2 competency in reading. On the other hand, only 21 percent of fifth graders attained level 4 or higher competency in mathematics, and 22 percent of fourth graders attained level 4 or higher competency in reading.

Table 4 shows that using level 2 as the standard for competency achieves approximately the same distributional outcomes across race, income and performance categories as does setting performance targets using average test scores. As shown in the second panel of Table 4, however, if vouchers are provided to only a few students, raising the competency level could actually redirect funds away from some of the neediest students. While the differences in composition are small when vouchers are provided to twenty percent of the student population, when vouchers are limited to two percent of students, raising the competency criterion leads to a voucher population that has higher test scores, on average, (by approximately half a standard deviation) and is four percent less poor. This scheme also provides half as many Hispanic students with school choice. The differences between using level 2 and level 4 as competency criteria must come about because the schools with the highest percentages of very low performing students have relatively few students performing at levels 2 and 3. Students attending these schools would be by-passed for vouchers by students attending schools with a high fraction of students performing at levels 2 and 3.

Considering student characteristics when evaluating schools for voucher eligibility

Many critics of accountability systems have argued that some schools will have a harder time meeting performance targets than others because their students are less advantaged. 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. This argument is the motivation for our next set of simulations, reported in Table 5. Here, we consider what happens to stratification if instead of basing voucher eligibility on a school's average test score, eligibility is instead determined by school's test score *after* controlling for its demographic and income composition. This adjustment has a tremendous impact on the types of students who qualify for vouchers. For example, if vouchers are made available to the two percent of children attending the schools with the lowest average reading scores, *after* controlling for the school's percent black, Hispanic and percent free lunch, then 31 percent of voucher-eligible children will be black, as opposed to 71 percent when gross school test scores are used. Eighty-nine percent of voucher-eligible children will qualify for a free lunch if gross test scores are used, but only 45 percent would be free-lunch eligible if net scores are used. This finding strongly suggests that if a primary goal of voucher programs is to increase the schooling options available to low SES children, then the appropriate school based eligibility criteria may not coincide with the appropriate accountability criteria.

An alternative to regression-adjusting test scores for the purposes of "controlling" for student body characteristics it to use school-level 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 a single cohort's test scores and measures of value added constructed from changes in test scores from one cohort to the next. As when eligibility criteria are tied to test scores net of school demographic characteristics, using changes in test scores from one year to the next may be a more appropriate way of measuring schools' effectiveness, and of providing students attending the least effective schools with choice, than basing eligibility criteria on test score levels. Tables 6 through 8 replicate Tables 2 through 4 using a cross-cohort "value added" measure of school performance instead of the school's average test score in a particular year. We focus on changes from the 1998-99 school year to the 1999-2000 school year.

As in Table 5, we see that using "purer" measures of school effectiveness will lead to a population of voucher recipients that is substantially more advantaged. Comparing Tables 2 and 6, the percent of voucher-eligible students who are black falls by approximately 31 percentage points, the percentage who are free-lunch eligible falls by 28 percentage points and the average test scores of the voucher-eligible students increase by more than half a standard deviation, if two percent of the population are voucher-eligible. This phenomenon undoubtedly relates in part to ceiling effects: 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. Another explanation is that value-added measures may be noisy signals of test score performance. This point has been made by Kane and Staiger (2000) who suggest that school grades based on a value-added metric may be

arbitrary. Our data allow us to investigate this possibility, and we find evidence in support of their argument: the correlation between the 1999-2000 change in fifth grade mathematics scores and the 1998-1999 change in mathematics scores is actually *negative* (-0.33). We estimate the same correlation for reading scores. Kane and Staiger also note that this problem is likely to be more serious for small schools because the set of students over which test scores are aggregated is smaller. We find that a one-standard-deviation reduction in school size is associated with a 0.18 standard deviation increase in the absolute difference between 1998-99 value added and 1999-2000 value added in mathematics, and a 0.15 standard deviation increase in reading.

In sum, these results suggest that, as in the regression-adjusted case presented in Table 5, voucher programs providing eligibility for schools with the lowest cross-cohort changes tend to provide vouchers for a less disadvantaged and higher-achieving set of students than do programs that offer vouchers to students based on gross test performance. This may be because value added measures of school performance fail to reflect improvements in school quality.

An alternative method of constructing value-added tests scores involves looking at over-time changes in performance within a given cohort. Because Florida began statewide testing of all grades from three to ten in the 1999-2000 school year (and previously only tested grades four, eight and ten in reading and five, eight and ten in mathematics) we can only look statewide so far at year-to-year changes within a cohort in reading performance. Fortunately, we have already shown that it is largely irrelevant which test (or set of tests) is used to measure school performance.

Table 9 presents the results of our within-cohort analysis. Here, we observe that while voucher-eligibles have lower achievement and are more likely to be low-income and minority than the population as a whole, the differences are not particularly dramatic, especially as the size of the eligible population increases. By the time twenty percent of the population is eligible for vouchers, 26 percent of the eligible students are black (compared to 25.5 percent in the population), 24 percent (rather than 18 percent) are Hispanic, and 50 percent (rather than 44 percent) are free lunch eligible. Moreover, test scores are only one-tenth (one-fifth) of a standard deviation below the state average in reading (mathematics). When we further adjust the test scores by controlling for race, ethnicity, and free lunch status we observe very little change in the characteristics of the eligible population.

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 systems that do not take into account background characteristics when assessing schools (or that are based on socio-economic status of the students in the schools) tend to provide vouchers for the most disadvantaged students, while those embedded in a "more fair" accountability system are likely to serve fewer disadvantaged students. 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 the poorest, lowest-achieving students.

IV.   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 if the President's proposal fails to become 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. Proponents of voucher programs usually promote them on the grounds that such programs increase school choice amongst individuals whose mobility would otherwise be constrained. In these discussions, disadvantage is measured with the usual socio-economic 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 a school accountability system, however, is whether the individual attends a poorly performing school. Because school test scores are so strongly correlated with the socio-economic 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 socio-economic 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, ranking schools according to their performance net of socio-economic characteristics necessarily ignores the possibility of peer effects. While 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. Second, 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. Ceiling effects could also lead to an in accurate ranking of schools' performance.

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 if most states look like California and Texas, however. Because students in these states are more concentrated by race, ethnicity, and income than they are in Florida, and because race, ethnicity, and income explain such a large fraction of school-level differences in level test score performance,[15] it is sensible to expect that places where these concentrations are high should see larger changes in voucher eligibility as school-based targeting methods change.[16]

Another important limitation to our study is that we do not have information on the student's income level, only on whether she is free/reduced price 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

---

[15] In Florida, the racial and ethnic breakdown of the school and the school's fractions free and reduced-price lunch-eligible explain 62 percent of the variation in level reading performance and 70 percent of the variation in level mathematics performance, in specifications weighting schools by student body size.

[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. While 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. While 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.

findings suggest that education policy-makers 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 simultaneously.

References

Epple, Dennis and Richard Romano (1998). Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects. American Economic Review 88, 33-62.

Figlio, David and Joe Stone (1999). Are Private Schools Really Better? Research in Labor Economics 18, 115-140.

Grogger, Jeff and Derek Neal (2000). Further Evidence on the Effects of Catholic Secondary Schooling. Brookings-Wharton Papers on Urban Affairs, Brookings Institution.

Hoxby, Caroline (1996). The Effects of Private School Vouchers on Schools and Students. in Holding Schools Accountable: Performance-based Reform in Education. Washington DC: Brookings Institution, 177-208.

Jepsen, Christopher (2000). The Effectiveness of Catholic Primary Schooling. Working paper, Public Policy Institute of California.

Kane, Thomas and Douglas Staiger (2000). Improving School Accountability Measures. Working paper, Hoover Institution.

Neal, Derek (1997). The Effects of Catholic Secondary Schooling on Educational Achievement. Journal of Labor Economics 15, 98-123.

Nechyba, Thomas (2000). Mobility, Targeting, and Private-School Vouchers. American Economic Review 90, 130-146.

TABLE 1
Characteristics of voucher-eligible students in various simulations: Voucher based on socio-economic status of school or neighborhood

Characteristics of voucher-eligible students	Universal voucher	Voucher given to all free lunch eligibles	Voucher given to all free/reduced lunch eligibles	Voucher based on socio-economic status of students in school					Voucher based on socio-economic status of children in neighborhood				
Voucher coverage	100%	44%	53%	2%	5%	10%	20%	44%	2%	5%	10%	20%	44%
Percent black	25.5	34.5	32.7	69.4	68.6	65.2	58.6	41.8	60.2	53.5	56.3	50.5	39.2
Percent Hispanic	18.0	21.2	20.8	27.1	26.4	26.7	26.8	24.6	36.4	38.6	33.4	30.6	24.6
Percent free lunch eligible	44.2	100.0	82.6	94.5	91.5	87.0	80.6	67.9	90.5	84.9	81.2	75.1	63.9
Percent residing in high poverty neighborhood	25.3	41.1	38.1	95.7	93.8	91.0	85.6	55.7	100.0	100.0	100.0	100.0	57.5
Standardized reading test score, 4th grade (2000)	0.02	-0.69	-0.58	-1.97	-1.75	-1.45	-1.17	-0.72	-1.71	-1.49	-1.28	-1.01	-0.60
Standardized math test score, 5th grade (2000)	0.05	-0.64	-0.55	-1.66	-1.52	-1.29	-1.02	-0.64	-1.58	-1.31	-1.11	-0.86	-0.53

TABLE 2
Characteristics of voucher-eligible students in various simulations: voucher based on level test performance in 2000

Characteristics of voucher-eligible students	Universal voucher	Voucher given to students attending schools failing to meet average performance targets in BOTH reading and math				Voucher given to students attending schools failing to meet average performance targets in AT LEAST ONE TEST (reading or math)			
Voucher coverage	100%	2%	5%	10%	20%	2%	5%	10%	20%
Percent black	25.5	75.2	68.3	57.8	56.5	69.8	70.0	59.2	53.7
Percent Hispanic	18.0	17.9	24.3	29.3	27.2	22.7	22.4	28.3	25.9
Percent free lunch eligible	44.2	87.1	84.4	79.2	76.9	88.4	86.3	80.2	74.5
Percent residing in high poverty neighborhood	25.3	82.6	90.2	81.3	75.6	92.0	91.5	84.3	70.8
Standardized reading test score, 4th grade (2000)	0.02	-2.48	-2.08	-1.71	-1.51	-2.51	-2.13	-1.75	-1.37
Standardized math test score, 5th grade (2000)	0.05	-2.65	-2.12	-1.65	-1.44	-2.68	-2.09	-1.69	-1.29

TABLE 3
Characteristics of voucher-eligible students in various simulations: voucher based on level test performance in 2000

Characteristics of voucher-eligible students	Universal voucher	Voucher given to students attending schools failing to meet average performance targets in math				Voucher given to students attending schools failing to meet average performance targets in reading			
Voucher coverage	100%	2%	5%	10%	20%	2%	5%	10%	20%
Percent black	25.5	71.7	66.9	60.7	53.4	71.0	67.6	59.1	51.4
Percent Hispanic	18.0	19.1	21.1	22.7	23.3	21.9	26.0	28.7	29.2
Percent free lunch eligible	44.2	87.1	82.7	77.7	72.8	88.6	86.2	80.3	73.8
Percent residing in high poverty neighborhood	25.3	89.8	87.8	76.6	67.1	94.6	89.3	80.3	70.3
Standardized reading test score, 4th grade (2000)	0.02	-2.36	-1.82	-1.53	-1.19	-2.69	-2.21	-1.84	-1.41
Standardized math test score, 5th grade (2000)	0.05	-2.92	-2.22	-1.79	-1.35	-2.25	-1.89	-1.50	-1.12

TABLE 4

Characteristics of voucher-eligible students in various simulations: voucher based on fraction attaining a particular standard level in both reading and math

Characteristics of voucher-eligible students	Universal voucher	Minimum standard level to attain: minimum competency (level 2) 73% attained in mathematics, 67% attained in reading				Minimum standard level to attain: very high competency (level 4) 21% attained in mathematics, 22% attained in reading			
Voucher coverage	100%	2%	5%	10%	20%	2%	5%	10%	20%
Percent black	25.5	74.7	68.3	61.0	53.2	81.4	74.5	64.5	52.2
Percent Hispanic	18.0	18.1	22.3	26.1	26.0	9.2	15.6	21.6	24.2
Percent free lunch eligible	44.2	87.2	83.5	79.2	73.4	83.0	82.4	79.4	72.4
Percent residing in high poverty neighborhood	25.3	87.3	88.2	79.7	67.9	82.0	82.7	75.4	64.4
Standardized reading test score, 4th grade (2000)	0.02	-2.45	-1.96	-1.65	-1.26	-2.04	-1.86	-1.58	-1.24
Standardized math test score, 5th grade (2000)	0.05	-2.73	-2.16	-1.69	-1.28	-2.13	-1.90	-1.55	-1.16

TABLE 5
Characteristics of voucher-eligible students in various simulations: voucher based on level test performance in 2000 (regression-adjusted for racial/ethnic composition and SES in school)

Characteristics of voucher-eligible students	Universal voucher	Voucher given to students attending schools failing to meet average performance targets in AT LEAST ONE TEST (reading or math)				Voucher given to students attending schools failing to meet average performance targets in reading			
Voucher coverage	100%	2%	5%	10%	20%	2%	5%	10%	20%
Percent black	25.5	38.9	32.5	32.6	28.1	31.1	32.2	31.3	28.1
Percent Hispanic	18.0	20.8	23.9	23.7	21.2	14.6	19.6	20.7	18.9
Percent free lunch eligible	44.2	59.7	53.7	53.4	49.2	44.6	49.0	49.8	47.9
Percent residing in high poverty neighborhood	25.3	39.1	36.9	35.4	31.1	22.6	24.6	30.8	26.7
Standardized reading test score, 4th grade (2000)	0.02	-1.78	-1.25	-1.11	-0.74	-1.55	-1.36	-1.16	-0.82
Standardized math test score, 5th grade (2000)	0.05	-2.15	-1.46	-1.18	-0.82	-1.09	-0.88	-0.77	-0.53

TABLE 6
Characteristics of voucher-eligible students in various simulations: voucher based on cross-cohort changes in test performance from 1999 to 2000

Characteristics of voucher-eligible students	Universal voucher	Voucher given to students attending schools failing to meet average performance targets in BOTH reading and math				Voucher given to students attending schools failing to meet average performance targets in AT LEAST ONE TEST (reading or math)			
Voucher coverage	100%	2%	5%	10%	20%	2%	5%	10%	20%
Percent black	25.5	23.6	27.7	25.3	23.6	39.3	31.6	28.3	26.6
Percent Hispanic	18.0	43.4	28.7	25.1	20.4	18.3	28.5	25.5	24.9
Percent free lunch eligible	44.2	54.2	49.2	47.9	42.4	57.1	57.1	51.2	49.6
Percent residing in high poverty neighborhood	25.3	38.7	34.3	30.9	23.2	49.6	44.6	36.6	33.0
Standardized reading test score, 4th grade (2000)	0.02	-0.96	-0.65	-0.47	-0.17	-0.76	-0.76	-0.48	-0.39
Standardized math test score, 5th grade (2000)	0.05	-0.81	-0.54	-0.35	-0.09	-0.86	-0.75	-0.52	-0.37

TABLE 7
Characteristics of voucher-eligible students in various simulations: voucher based on the change from one cohort to the next in the fraction attaining a particular standard level in both reading and math

Characteristics of voucher-eligible students	Universal voucher	Minimum standard level to attain: minimum competency (level 2) 73% attained in mathematics, 67% attained in reading				Minimum standard level to attain: very high competency (level 4) 21% attained in mathematics, 22% attained in reading			
Voucher coverage	100%	2%	5%	10%	20%	2%	5%	10%	20%
Percent black	25.5	20.1	21.2	18.0	19.0	29.2	28.8	33.9	33.9
Percent Hispanic	18.0	29.9	26.1	21.8	19.5	15.7	11.7	15.1	17.5
Percent free lunch eligible	44.2	43.6	40.5	35.6	36.1	44.5	43.2	50.9	51.6
Percent residing in high poverty neighborhood	25.3	33.2	22.7	17.1	17.0	24.3	21.8	31.7	33.0
Standardized reading test score, 4th grade (2000)	0.02	-0.34	-0.17	0.19	0.19	-0.10	-0.07	-0.37	-0.40
Standardized math test score, 5th grade (2000)	0.05	-0.30	-0.12	0.21	0.15	-0.20	-0.14	-0.43	-0.44

TABLE 8
Characteristics of voucher-eligible students in various simulations: voucher based on cross-cohort changes in test performance in from 1999 to 2000 (regression-adjusted for racial/ethnic composition and SES in school)

Characteristics of voucher-eligible students	Universal voucher	Voucher given to students attending schools failing to meet average performance targets in AT LEAST ONE TEST (reading or math)				Voucher given to students attending schools failing to meet average performance targets in reading			
Voucher coverage	100%	2%	5%	10%	20%	2%	5%	10%	20%
Percent black	25.5	38.1	36.0	33.8	31.1	38.0	34.0	31.4	28.1
Percent Hispanic	18.0	19.2	26.0	22.9	21.8	18.9	17.9	20.3	21.7
Percent free lunch eligible	44.2	61.3	59.8	55.3	53.0	55.0	52.0	52.9	50.1
Percent residing in high poverty neighborhood	25.3	52.4	48.0	40.2	35.5	40.2	35.1	35.7	31.9
Standardized reading test score, 4th grade (2000)	0.02	-0.85	-0.88	-0.65	-0.51	-1.07	-0.72	-0.69	-0.52
Standardized math test score, 5th grade (2000)	0.05	-0.99	-0.94	-0.73	-0.52	-0.38	-0.16	-0.22	-0.15

TABLE 9
Characteristics of voucher-eligible students in various simulations: voucher based on within-cohort changes in reading test performance from 1999 to 2000

Characteristics of voucher-eligible students	Universal voucher	Voucher based on changes in reading test scores				Voucher based on test score changes, adjusted for race/ethnicity and free lunch status			
Voucher coverage	100%	2%	5%	10%	20%	2%	5%	10%	20%
Percent black	25.5	37.2	33.4	28.0	26.1	38.1	36.3	30.6	26.6
Percent Hispanic	18.0	24.2	24.9	28.0	24.4	17.5	17.7	18.8	20.1
Percent free lunch eligible	44.2	60.6	57.1	53.6	49.8	57.7	55.8	50.3	47.0
Percent residing in high poverty neighborhood	25.3	50.3	42.9	39.6	31.9	47.9	37.9	32.6	28.6
Standardized reading test score, 4th grade (2000)	0.02	-0.31	-0.20	-0.18	-0.11	-0.17	-0.16	-0.07	-0.02
Standardized math test score, 5th grade (2000)	0.05	-0.43	-0.30	-0.23	-0.19	-0.35	-0.34	-0.19	-0.12