Educational Policy

http://epx.sagepub.com

Plenty of Children Left Behind: High-Stakes Testing and Graduation Rates in Duval County, Florida

Mary O'Malley Borg, J. Patrick Plumlee and Harriet A. Stranahan

Educational Policy 2007; 21; 695 originally published online Apr 24, 2007;
DOI: 10.1177/0895904806289206

The online version of this article can be found at:
http://epx.sagepub.com/cgi/content/abstract/21/5/695

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Politics of Education Association

Educational Policy
Volume 21 Number 5
November 2007 695-716

10.1177/0895904806289206
http://epx.sagepub.com
hosted at
http://online.sagepub.com

Plenty of Children Left Behind

High-Stakes Testing and Graduation Rates in Duval County, Florida

Mary O'Malley Borg
J. Patrick Plumlee
Harriet A. Stranahan
University of North Florida

The Florida Comprehensive Assessment Test (FCAT) is a high-stakes test that public school students must pass to be eligible for graduation from high school. Previous research suggests that high-stakes tests have a differential effect on students by race and ethnicity. This study finds that in one Florida school district African American and Hispanic students coming from poorer, less educated, or higher mobility households are less likely to meet graduation requirements than their higher socioeconomic, White, suburban counterparts. African American students and students from the lowest income households are also the most likely to encounter a negative graduation effect because the passing score on the FCAT rises each year. School characteristics also affect the probability of student success. High schools that hire more teachers with advanced degrees or offer a magnet program have better student FCAT scores. This results in higher probabilities that all categories of students will meet graduation requirements.

Keywords: *Florida Comprehensive Assessment Test; high-stakes testing; minority students*

On January 8, 2002, President George W. Bush signed into law P.L. 107-110, the No Child Left Behind Act (NCLB). No Child Left Behind was a bipartisan bill supported by both Republican and Democratic members of the U.S. Congress. A major objective of NCLB is to hold America's public schools accountable for the performance of their students. Children who fare poorly in the system of public education—who fail to learn basic math or reading skills or who either drop out of school or graduate unequipped with the skills needed in today's job market—do so, critics

assert, because the schools they attend fail to teach them what they need to know. This view of the state of public education assumes that the problem of academic underachievement lies primarily in schools themselves. Therefore, the solution proposed in NCLB is to make schools accountable for what they teach. Accountability is established by assessing the performance of the students attending the schools. Performance assessment is achieved by the administration of standardized tests developed in each state and approved at the federal level. Objective performance indicators supplied by standardized testing allow schools that do not make "adequate yearly progress" to be identified and given opportunities to improve. If the troubled schools do not improve over time, they may be disciplined through the use of various sanctions, including allowing their students to go to better performing schools, public or private, at public expense. Higher performing schools, in the meantime, are provided positive incentives such as additional funding per student (Dahmus, 2003).

In this article, we examine how one particular strategy for achieving goals related to NCLB high-stakes testing affects high school graduation requirements and ultimately high school graduation rates in a consolidated city-county school system. We also consider some of the unintended consequences that these policies may have on some of the most vulnerable students in the public school system. It is rather early to make conclusions about the effectiveness of NCLB because its policies have only recently been implemented; however, Florida implemented a school accountability program in 1998 that is very similar in its focus and its methods to NCLB. Therefore, we use the results of Florida's school accountability program (the Florida A+ School Recognition Program) to make inferences about the consequences NCLB might have on high school graduation requirements and on how those requirements disproportionately affect some sociodemographic groups.

School Performance and High-Stakes Testing

In the accountability model as it has developed in Florida, the expected level of school performance was set through a process involving education professionals, citizens, and politicians at the state level and imposed on all schools. Statewide standards were developed and assessment instruments, such as standardized tests, were written to reflect mastery of the standards. When the tests are administered, school scores can be compared and performance can be ranked. Schools that have students who achieve higher test

scores on average are classified as successful; those that have students who achieve lower tests scores on average are classified as unsuccessful. This means that although schools are assessed as institutions the assessment is based ultimately on individual student performance on standardized tests.

Thus, one feature of the emerging national commitment to the accountability model is the growing prominence of high-stakes testing. "High-stakes tests are tests from which results are used to make significant educational decisions about schools, teachers, administrators, and students" (Amrein & Berliner, 2002b, p. 5). According to the national study conducted by Amrein and Berliner (2002a), a number of states have programs that distribute financial rewards to successful or improved schools; have the power to close, restructure, or take over low-performing schools; award financial bonuses; or displace or remove teachers or administrators.

The consequences of high-stakes testing can be particularly devastating to individual students. In 18 states, students meeting all other requirements for graduation but failing the high school graduation standardized exam may be denied a high school diploma (Amrein & Berliner, 2002a). The accountability model begins with an assumption that schools are entirely responsible for student learning, and consequently they are to be rewarded or sanctioned for the quality of their efforts. But the use of standardized testing as the instrument of school accountability inherently provides a mechanism for assessing individual students. In other words, the structure used to generate accountability for schools also can be deployed to reward (graduate) or punish (hold back) students.

Institutional and Individual Consequences of Accountability Standards

Early empirical evidence suggests that the emphasis on school accountability implemented through standardized testing is starting to improve overall school performance. Hanushek and Raymond (2003a, 2003b, 2004) found that states that implemented accountability systems early had more rapid gains in National Assessment of Education Progress (NAEP) standardized math test scores for both the fourth and the eighth grades during the 1990s than did states that did not implement accountability systems. In addition, the gains were greater in the states where the accountability standards imposed consequences (incentives and disincentives for the schools) than they were in the states where the accountability results were used only for reporting information about the schools.

Carnoy and Loeb (2002) found similar results in their examination of NAEP math scores when they used a more fine-tuned measure of school accountability. They ranked state accountability systems from 0 to 5 based on the accountability standards that were in place in the 1999-2000 school year. A state ranked 0 had no state accountability system in place and a state ranked 5 had an accountability system like Florida's that included testing at the elementary, middle, and high school levels; strong sanctions and rewards based on the test results; and a required minimum score on the test for high school graduation. They found that the states with the strongest accountability systems had significantly higher gains in eighth grade NAEP math scores across all racial and ethnic groups. Although Carnoy and Loeb did not find the same significant positive relationship between strength of accountability systems and fourth grade NAEP math scores for all students, they did find that the states with the strongest accountability systems showed significantly greater gains in the percentage of Black students who achieved at least the basic level of competency on the fourth grade math NAEP test. However, most relevant to our results, they found no relationship between high school progression rates and strong accountability systems. In fact, they could not rule out the possibility that strong accountability systems were associated with decreased high school progression rates for Hispanic students.

In the politically charged atmosphere that surrounds studies of the effectiveness of school accountability systems, other studies disagree with the conclusions reached by Hanushek and Raymond (2003a, 2003b, 2004) and Carnoy and Loeb (2002). One of the most widely known is a 2002 study conducted by Amrein and Berliner. In their national study using data collected from 28 states engaged in high-stakes testing of one sort or another, they concluded that "there is scant evidence to support the proposition that high-stakes tests—including high-stake high school graduation exams—increase student achievement" (2002b, p. 1).[1] Furthermore, the study also concluded that "the implementation of high school graduation exams results in a decrease in academic achievement" (2002b, p. 1). Particularly relevant to our research, these effects disproportionately affect students from racial minority and lower socioeconomic status (SES) backgrounds, in that states with greater proportions of each of these kinds of students are more likely to impose high-stakes testing (Amrein & Berliner, 2002a). However, Raymond and Hanushek (2004) questioned the validity of Amrein and Berliner's results because of research-design flaws, which they characterize as a "pseudo-trend analysis" (p. 123). They also point out that Amrein and Berliner's study violates "the first principle of social science research"

(p. 124) because the study only examines states that have accountability systems in place without the control of states without accountability systems.

Whereas there is disagreement about whether school accountability systems improve overall academic performance, there is undisputed evidence that strong accountability systems have increased the achievement gap between White students and their Black and Hispanic counterparts. Hanushek and Raymond (2003a, 2003b, 2004) found that the gains made by Black and Hispanic students on the NAEP test in the states with strong accountability systems were approximately 6 to 10 points lower than the gains made by White students between the fourth and eighth grades. Because this spread was greater than the 3.5-point improvement for students overall, the gap between White students and their Black and Hispanic counterparts increased in the states with the strongest accountability systems during the 1990s. This was particularly troubling because the achievement gap had narrowed somewhat during the 1980s (Jencks & Phillips, 1998). Hanushek and Raymond (2004) rightly point out that it is difficult for a single policy to accomplish two goals—raising overall student performance as well as providing more equal educational outcomes. Whereas accountability systems such as the one instituted in NCLB seem to accomplish the former, they actually may have a negative effect on the latter.

Haney (2000) also documented that the emphasis on school accountability has had a detrimental effect on the schooling outcomes of Black and Hispanic students relative to White students. Haney examined evidence from Texas, one of the first states to implement a school accountability program in 1984. He showed that over the course of the 1980s and 1990s the rates at which Black and Hispanic students had to repeat the ninth grade increased steadily so that by the late 1990s the ninth-grade failure rate for Blacks and Hispanics was almost 30%. In addition, the cumulative rate of grade retention in Texas was twice as high for Black and Hispanic students as it was for White students, so by the late 1990s only 50% of minority students in Texas were progressing from ninth grade to high school graduation as compared to 70% for all students regardless of race or ethnicity. Furthermore, these rates of high school completion in Texas may be understated because the data that Haney used to calculate these rates assumed that there was no increase in student population from immigration (either from other states or internationally) over the years in which they were calculated.

Other researchers have confirmed these results for minority students. Dee (2003) examined the effect of minimum-competency tests for high school graduation on the graduation rates of Whites and Blacks using data from the respondents of the long form of the 1990 Census. The data from 1,348,766

White (non-Hispanic) and Black respondents who were 18 years of age between 1980 and 1988 showed that higher minimum-competency tests for graduation reduced the probability of high school graduation for all students except White females, and the effects were particularly large for Black students of both sexes. Dee found that the reduction in the probability of graduating from high school was four times greater for Black male and female students than for White male students. Jacob (2001) examined the impact of high school graduation exams on dropout rates using data from the National Educational Longitudinal Survey. He found that graduation tests had no effect on the probability of dropping out of high school for the average student; however, the tests significantly increased the probability of dropping out for the lowest achieving students who were disproportionately minorities. These results are significant for our research because we are interested in the unintended consequences that school accountability and its emphasis on high-stakes testing has on disadvantaged population groups.

The empirical evidence suggests that strict school accountability systems are beginning to register success in improving school quality as measured by the achievement levels of students overall, but the evidence is less clear on the achievement of students of color and those from disadvantaged socioeconomic backgrounds. Does a requirement to meet the standards of high-stakes testing, beyond whatever success a student achieves in meeting other academic criteria, foster success among at-risk students or is that requirement merely one more discouraging barrier to success, one that actually increases the risk of failure? This is the question that we explore in this research. If the latter outcome is shown to be likely then there are grounds for some serious concerns for the efficacy of the accountability model, either in its basic assumptions or through flaws in its application.

High Stakes in Duval County, Florida: A Case Study of State Education Policy

This study offers a look at these pressing issues from the perspective of one urban school district in Florida. The state of Florida has imposed a regime of high-stakes testing that includes many of the elements present in NCLB legislation.[2] Indeed, it is the centerpiece of Governor Jeb Bush's agenda for educational progress in the state. Has the enforcement of state standards uniformly benefited the children of Duval County, Florida? Or are there uneven and troubling effects that have appeared as a result of the state's commitment to high-stakes testing?

Duval County presents an interesting test case for the consequences of high-stakes testing because in many ways Duval County is a tale of two cities. The entire county was incorporated into the city of Jacksonville in 1967, so parts of the city resemble a typical American inner city while other areas are typically suburban. The terms *inner city* and *suburban* are not completely accurate in reference to Duval County public schools because all schools are technically within the city limits. However, *inner city* and *suburban* refer to schools that are within the original city limits of Jacksonville and schools that are in the newer parts of Jacksonville, primarily to the east and south of the core city, respectively. With the exception of magnet schools, most of the schools within the original city limits have percentages of Black and Hispanic students and of students who receive free or reduced lunch that are greater than the median for all Duval County public schools. Conversely, most of the schools in the newer parts of the city have percentages of Black, Hispanic, and poor students that are well below the median. These distributions are shown clearly in maps depicting the demographic characteristics of Duval County.[3]

News accounts indicate that linking a high-stakes test to high school graduation in Florida has produced serious problems for some of the state's most vulnerable students. Florida law requires that students must earn passing scores on the 10th-grade Florida Comprehensive Achievement Test (FCAT) reading and math sections to graduate from high school. Six attempts are allowed before a student must accept a certificate of completion in lieu of a diploma. This is important because a standard diploma, rather than a certificate, is required if a student is hoping to pursue post-secondary education within the state or is interested in joining the military. In March 2003, there were 12,772 high school seniors who had not met the FCAT reading requirement and 9,933 who had not met the math requirement (Garza, 2003). Fearing that thousands of seniors will fail the FCAT, a coalition of Black and Hispanic state lawmakers introduced bills in the state legislature allowing students who fail the FCAT to earn a diploma if they meet a minimum score on a college entrance exam, such as the ACT or SAT, or Florida's own College Placement Test for entrance to community college. Governor Bush resisted any efforts to override the results of FCAT testing, arguing that the same standards must apply to all students and that minority scores had improved at most grade levels since the institution of testing in 1998 ("Lawmakers Want FCAT Alternative," 2003; "Protestors Want Moratorium," 2003). In Duval County, the locus of this study, 1,959 seniors out of 5,981 (33%) had not passed the FCAT reading section and 1,147 seniors (19%) had not passed the math section (Garza, 2003).

Such accounts emphasize both the significance and the scale of the unintended consequence of FCAT-style high-stakes testing in Florida. As the data show, performance on the FCAT is correlated with race and SES. The effect of high-stakes testing in this instance is clearly adverse to minorities. If graduation is to be denied, the individuals involved do not simply go away. Because students without diplomas logically would seem to be worse off, it is fair to ask if the high-stakes approach produces unintended adverse effects. Our analysis shows that such effects are indeed predictable, quantifiable, and manifested nonrandomly in the student population. In this arena, at least, skepticism about the efficacy of high-stakes testing appears warranted.

Method and Data

Using data obtained from the state and the local school district, we sought to ascertain the effect of FCAT graduation requirements on a student's probability of earning a high school diploma. In addition, we wished to examine the effects of changing the numeric value of the passing score that has been in place in Florida: How many students will encounter a perverse "graduation effect," defined as a higher likelihood of failing the test, with the rise in the mandated standard for graduation? What can this be expected to contribute to the dropout rate over time, and what students, demographically, will be most affected by these changes? If there are unintended consequences of adopting a high-stakes test as a graduation requirement and concurrently raising standards for acceptable achievement, are these consequences sufficient to pose serious questions about the social costs of such testing?

To help shed light on these issues, the Duval County (Jacksonville, Florida) School Board provided FCAT scores for all 10th-grade students in Duval County who took the test in the 1999-2000 school year. In addition to the students' scores on the math and reading portions of the FCAT, these data include demographic information on each student's race and the number of times he or she has withdrawn from school (an indicator of student mobility). We supplemented this individual student demographic data by using the students' addresses to link them with U.S. Census block–level demographic data in their neighborhoods. This allowed us to create a demographic profile for each of the 5,205 tenth-grade students using the census block–level values for variables such as parents' education levels and income. In addition, the Duval County school system collects a variety of school-level data such as percentage of teachers in the school with advanced

degrees, teachers' years of experience, proportion of teachers newly hired, and magnet school indicators. These data allowed us to specify a number of school factors that potentially affect student performance on the FCAT. In other words, the data format allowed us to link individual student demographic characteristics and school quality indicators with FCAT performance.[4] This means that we have the opportunity to independently assess the effect of the school on the student's FCAT score and the effect of the child's demographic circumstances.

These data are different from the information ordinarily available for the assessment of FCAT performance. Normally, FCAT scores are provided with individual-level information with respect to race, sex, and grade level only within each school. Hence, standard reports of FCAT scores allow aggregate analyses of differences in test performance among racial and sexual categories, but they do not allow for analyses by any additional demographic information. This makes it impossible to assess the potential impact of demographic variables, such as family income, on FCAT performance. By linking FCAT scores with other variables, we have the potential to assess the possible impact of demographic factors in a more fine-grained way.

The mean values of the variables used in our statistical models are shown in Table 1, which foreshadows the seriousness of the consequences of raising the minimum FCAT score needed for graduation. For example, only 53% of all 10th-grade students scored high enough on the FCAT to qualify for graduation under the 2002 standard (a score of 300 or better on each of the math and reading sections). Furthermore, 498 (9.5%) of these 10th-grade students are affected by the new higher graduation standards. In other words, based on the scores they made on the FCAT, these students would have passed under the old standard (287 for reading, 295 for math) but fail under the new higher minimum score.[5]

These results are fairly typical for an urban school district. In a May 2002 article, the *Miami Herald* reported that 56% of 10th-grade students in the Miami-Dade school system failed to pass under the new standards, and 40% of the more affluent, suburban Broward County students failed to pass. In Miami-Dade County, the new standards resulted in 3,200 additional students failing relative to the number who would have failed under the old standard, an increase of about 13% of the 10th-grade student population. In Broward County, 1,300 additional students failed, 9.5% of the affected student population (Grech, 2002).

Clearly, many students may fail to make the cut-off point under the new graduation standards, leading to the question, which students are most likely to be affected? Is the likelihood of failure randomly distributed or are

Table 1
Descriptive Statistics: Duval County 10th-Grade Students and High Schools in 1999-2000

	M	*SD*
Student characteristics		
Educational attainment in household		
Less than 9th grade	0.0609	0.0621
9th grade to 11th grade	0.1433	0.0899
High school graduate	0.3078	0.0832
Community college	0.2887	0.0737
College	0.1994	0.1351
Race		
African American	0.3452	0.4755
Hispanic	0.0323	0.1768
White	0.5941	0.4945
Other	0.0401	0.0661
Household income greater than $70,000	0.0378	0.0604
Number of student withdrawals from school	0.0557	0.3043
School characteristics		
Magnet school	0.2792	0.4486
% teachers with advanced degrees	0.3919	0.1001
Teachers' average years of experience	0.1578	0.0910
% teachers who are newly hired	0.1304	0.0823
% students who meet graduation requirement	0.5354	0.4988
% students who are ineligible with new standard	0.0957	0.2942
Number of students = 5,206		

there demographic and/or school characteristics that suggest that some students are more likely to fail than are others?

The methods used to help answer these questions include the estimation of two different probit regressions. The first probit regression includes all 5,206 students and predicts which students are likely to meet the new, more rigorous, 2002 graduation requirements. This model also determines which students are most adversely affected by having a minimum FCAT score as part of the requirements for high school graduation. In this probit, the dependent variable can be interpreted as the probability that a student passes the FCAT graduation requirements and the independent or explanatory variables include student characteristics such as ethnicity, household income, and parental education levels as well as school characteristics such as teacher quality and teacher turnover.

We used a probit regression[6] such as the one shown in Equation 1 because the dependent variable is dichotomous: Y = 1 if the student has passing

scores (300 on both FCAT math and reading) and Y = 0 if the student does not have passing scores. In Equation 1, Φ is the standard normal cumulative probability distribution and β'x is a matrix of independent variables including student and school characteristics. The set of parameters β reflect the impact of changes in x on the probability that a student passes the 2002 FCAT graduation requirements.

$$\text{Prob}(Y = 1|x) = \Phi(\beta'x) \quad (1)$$

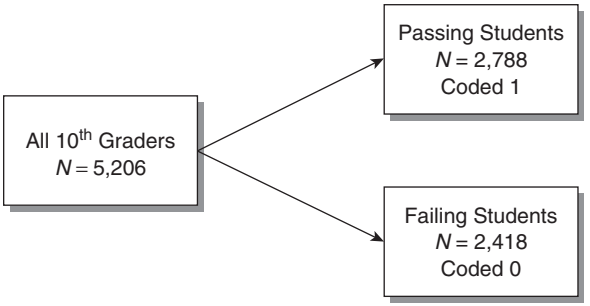
Next, we estimated a second probit equation using data on only those 3,211 students who would have passed under the old, less rigorous FCAT graduation requirements. Among these students who would have passed prior to 2002, we wanted to predict which unlucky students are more likely to fail under the new higher standards. The dependent variable in this second probit is again dichotomous: Y = 1 if the student has failing scores under the higher 2002 standards and Y = 0 if the student does not have failing scores. This second probit equation provides information about who (demographically) is most likely to be affected by raising graduation requirements (who would have passed under the old standards but fails under the new, higher standards) and whether school resources play an important role in determining who is adversely affected under these new standards.

Factors Affecting Minimum Graduation Scores

Our first probit regression uses the full data set of all 10th-grade students taking the FCAT in Duval County during the 1999-2000 school year to predict the probability that a student will earn at least a 300 on each of the reading and math portions of the FCAT. The coding of the dichotomous dependent variable is explained in Figure 1. The estimated results of this model are shown in Table 2. Our results indicate that both individual and school characteristics are important determinants of meeting the FCAT graduation requirement in the 10th grade.

As expected, a key indicator of student success on the FCAT is the SES of the student's household. Income has a powerful effect on the probability of meeting the graduation requirement. The negative coefficient on the variable indicating that the student lives in a household with income of $10,000 per year or less means that these students are significantly less likely to meet the FCAT graduation requirement than are students growing up in households with incomes of $30,000 and higher (the omitted category). Students living in households in the next lowest income category (household income between $10,000 and $20,000) are also significantly less likely to meet the FCAT requirement.

Figure 1
Coding of Dependent Variable in Model 1



Note: Dependent variable: 1 = pass under new, higher 2002 standard; 0 = fail under new, higher 2002 standard.

Table 2
Probit Model: Probability of 10th-Grade Student Meeting the Higher 2002 FCAT Graduation Requirement

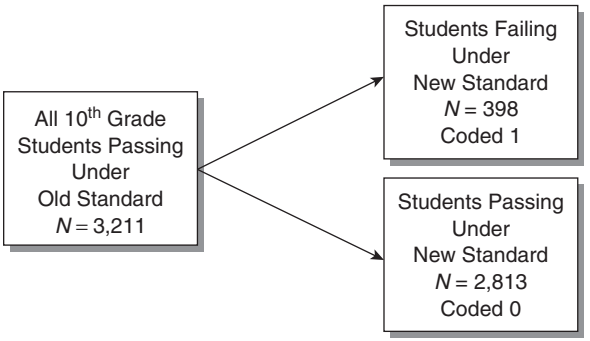
Independent Variable	Coefficient	Standard Error	*p* Value
Constant	–0.2234	0.2574	.3152
Educational attainment in household			
Less than 9th grade	–0.3748	0.4648	.4256
Less than 12th grade	0.0094	0.3464	.9783
African American	–0.8092***	0.0439	.0000
Hispanic	–0.2952***	0.1017	.0037
Household income			
$10,000 or less	–0.6667**	0.3021	.0274
$10,000 to $20,000	–0.7046**	0.2838	.0130
$20,000 to $30,000	–0.1873	0.2563	.4649
Withdrawals from school	–0.1953***	0.0655	.0029
Magnet program at school	0.5918***	0.0486	.0000
% teachers with advanced degrees	0.0196***	0.0027	.0000
Average teacher years of experience	0.0032	0.0156	.8388
% teachers that are new hires	–0.0119***	0.0029	.0001
Number of observations = 5,206			

Note: FCAT = Florida Comprehensive Assessment Test.
*$p < .10$. **$p < .05$. ***$p < .01$.

Race and ethnicity also play important roles in the probability that a student will meet the FCAT graduation requirement. African American and Hispanic students taking the test for the first time are less likely than White students are to achieve scores sufficient to meet the graduation requirements set by the state. The coefficients on parental educational attainment are not significantly different from zero in this model specification.[7] Student mobility, identified as the number of times a student withdraws and re-enrolls in the same school during the academic year, also has a strong negative effect on the probability of passing the FCAT graduation requirement. Student mobility is highest in the inner-city schools that are populated by more poor and minority students whose families frequently change addresses.

Several school characteristics are also significant predictors of which students will meet the FCAT graduation requirement. Students who go to magnet schools have a significantly higher probability of meeting the FCAT graduation requirement than do 10th-grade students in nonmagnet high schools. However, it is uncertain whether the magnet school variable really reflects a school or a student characteristic. The three magnet high schools in Duval County attract highly motivated students who have been on track to attend a magnet high school since elementary school. Two of the magnet high schools are college preparatory schools for gifted and talented students and the third is for students who excel in the fine and performing arts.[8] Two variables reflecting the qualifications of the teachers at the school also have significant effects on the probability that students will pass the FCAT graduation requirement. Students who attend high schools in which a greater percentage of the teachers have advanced degrees are more likely to pass the graduation requirement than do identical students at schools with a lower percentage of teachers with advanced degrees. This suggests that teachers with advanced degrees contribute significantly to 10th-grade students' success on the FCAT. However, our model does not support the hypothesis that teachers' years of experience alone impact positively on student FCAT scores. The coefficient on the measure of average years of teaching experience at a school is not significant. However, attending a school with a high percentage of new teachers does affect a student's FCAT score because the variable reflecting the percentage of newly hired teachers at a school is both negative and significant. This means that students who attend schools that have high rates of teacher turnover, as measured by percentage of new hires, are much less likely to meet graduation requirements despite having passed all coursework that brought them to the 10th grade. This is a disturbing finding because teacher turnover is much greater in the inner-city schools that serve disadvantaged students whose demographic

Figure 2
Coding of Dependent Variable in Model 2



Note: Dependent variable: 1 = affected by the new, higher 2002 standard; 0 = not affected by the new, higher 2002 standard.

characteristics already predispose them to do worse on the FCAT than their suburban counterparts.

Which Students Are Likely to Be Affected by Raising FCAT Graduation Requirements?

Our second probit regression was designed to better understand the characteristics of the unlucky 9.5% of students who would have passed under the old, less rigorous graduation requirements but who do not pass under the new, higher standards. In this model, our sample includes all of the 3,211 tenth-grade students who have FCAT scores that met the old graduation requirements prior to 2002. Our probit model measures the probability that the student fails under the new, more rigorous 2002 graduation standards (Y = 1 if the student fails under the 2002 FCAT graduation requirements, Y = 0 if the student does not fail). The coding of the dichotomous dependent variable is shown in Figure 2. Our analysis uncovers whether one socioeconomic or demographic group is more likely to be adversely affected than another.

The estimated results of the second probit model are shown in Table 3. Surprisingly, fewer of the variables indicating the students' SES are significant in this model. Parental educational levels are not statistically significant

Table 3
Probit Model: The Probability of a Student Failing Because of Raised FCAT Standards

Independent Variable	Coefficient	Standard Error	*p* Value
Constant	–1.0680***	0.4107	.0093
Educational attainment in household			
Less than 9th grade	0.2722	0.8076	.7361
Less than 12th grade	–0.5360	0.5719	.3486
African American	0.2210***	0.0753	.0034
Hispanic	0.1986	0.1492	.1834
Household income			
$10,000 or less	1.0880**	0.5107	.0331
$10,000 to $20,000	0.5072	0.4511	.2608
$20,000 to $30,000	0.0899	0.4054	.8348
Withdrawals from school	0.2213**	0.0937	.0182
Magnet program at school	–0.3249***	0.0753	.0000
% teachers with advanced degrees	–0.0108***	0.0004	.0079
Average teacher years of experience	0.0120	0.0247	.6251
% teachers that are new hires	0.0090*	0.0049	.0649
Number of observations = 3,211			

Note: FCAT = Florida Comprehensive Assessment Test.
*$p < .10$. **$p < .05$. ***$p < .01$.

predictors of being in this marginal group of students who pass under the old FCAT requirement but fail under the new requirement. The largest coefficient in the regression occurs for students whose household income level is less than $10,000. This implies that 10th-grade students from the lowest income group who would have passed under the old standards are the group most affected and most likely to fail under the new, higher FCAT graduation standards. Along with income, race is important in this probit regression. African American students are more likely to fail under the new standards than are their White counterparts. Hispanic students, however, are not significantly more likely to fail under the new standard relative to identical White students. The number of school withdrawals, another indicator of inner-city student populations, does significantly increase the probability of being in the group that would have passed earlier graduation requirements but fails under the new higher cutoff.

The same school characteristics that affect the probability of students meeting the FCAT graduation requirement in general are also important predictors of whether a 10th-grade student will be affected by the new standards.

It is not surprising that students attending a magnet program are less likely to be affected by the new higher graduation requirements. Students who attend schools that have a higher percentage of teachers with advanced degrees are also less likely to fall into this marginal group of adversely affected students. On the other hand, students who attend a school with a high percentage of newly hired teachers are more likely to fail under the new higher standards.

The Predicted Probability of Meeting the FCAT Graduation Requirement

Tables 4 and 5 present the results of the estimated probit models in a more intuitive way. The probit equation can be used to predict the probability of obtaining a given result (pass or fail, in this case) when the independent variables take on certain values. Therefore, in Table 4, we have estimated the probability that a student with the sample mean value for each of the independent variables will meet the graduation requirement. We then change the value of the race and income variables to see what effect each of these changes has on the probability of meeting the graduation requirement. Similarly, in Table 5, we estimate the probability that an average student who passed the graduation requirement under the old standard will fail under the new standard. Once again, we change the values of the race and income variables to see the effect that each of these changes has on the probability of failing under the new standard.

Table 4 shows that a student who has the sample average value for each of the independent variables in the model has a 54% probability of passing the FCAT graduation requirement on the first try. It is important to emphasize that this is not a real student but rather an "average straw person" that reflects the average percentage of the sample that is African American, White, Hispanic, has household income of $10,000 or less, and so forth. When we take this average straw person and change the value of the African American variable to 1, we effectively change that straw person into an African American student who retains the sample average values for the rest of the variables in the model. We do this for all of the race variables and also for the variables reflecting the lowest income category ($10,000 or less) and the highest income category (the omitted category that reflects households earning more than $30,000 annually). As Table 4 shows, the results of this analysis are striking.

A significant difference in the probability of passing occurs between the haves and the have-nots. The probability of passing the FCAT graduation

Table 4
Predicted Probability of Passing Under the New 2002 FCAT Graduation Requirements Among All 10th-Grade Students

Type of Student	Probability of Passing
Sample average	.54
White	.65
African American	.34
Hispanic	.54
Higher income	.60
Low income	.30
Difference in Student Characteristic	**Difference in Probability**
White vs. African American	+.31***
White vs. Hispanic	+.11
High income vs. low income	+.30***

Note: The predicted probabilities are from the Table 2 Probit Regression.
***$p < .01$.

Table 5
Predicted Probability of Failing Under the New 2002 FCAT Graduation Requirements Among Students Who Would Have Passed Under the Old Standard

Type of Student	Probability of Failing
Sample average	.12
White	.11
African American	.15
Hispanic	.15
Higher income	.09
Low income	.26
Difference in Student Characteristic	**Difference in Probability**
White vs. African American	–.04***
White vs. Hispanic	–.04
High income vs. low income	–.17***

Note: The predicted probabilities are from the Table 3 Probit Regression.
***$p < .01$.

requirement on the first try for a student who comes from a family in the highest income category ($30,000 or more) is 60% compared to 30% for an identical student whose household income is $10,000 or less. This difference is actually understated because the likelihood that a student from the highest income category would have the same values on the other independent variables (parental education, for example) as the student from the lowest income category is slim.

Race and ethnicity also matter in the probability of meeting the FCAT graduation requirement. An average White student has a 65% probability of passing the FCAT graduation requirement on the first try as compared to a 34% probability for an identical African American student (one who has the same sample averages for the other independent variables). An average Hispanic student has a 54% probability of passing on the first try—11 percentage points lower than an identical White student.

The probabilities shown in Table 5 are equally striking. African American students who passed the graduation requirement under the old standard have a 15% probability of failing to meet the requirement under the new standards. Identical White students have only an 11% probability of failing under the new standards. Hispanic students also have a higher probability of failing under the new standards than do identical white students although the difference is not statistically significant. Household income, again, presents the greatest advantage to students in meeting the new graduation requirements. Students from households in the highest income category have a 9% probability of failing under the new standards as compared to a 26% probability for students from households in the lowest income category.

Conclusions

In summary, the usual socioeconomic and demographic characteristics associated with student success are borne out in these results. Minority students with higher mobility rates coming from poorer and less educated households are less likely to meet graduation requirements than are students coming from White, suburban, wealthier, and more educated households. Duval County, Florida, is no different than the rest of the United States when it comes to the consequences of tougher standards for school accountability.

It is interesting that school characteristics change the probability of success as well. High schools that hire more teachers with advanced degrees have better student FCAT scores across the board and higher probabilities

of meeting graduation requirements. Our results also suggest that schools could improve their graduation rates by getting to the root of high teacher turnover and ameliorating those causes.

The new FCAT standards for graduation quite clearly have differential effects for specific socioeconomic, racial, and ethnic groups. One of the consequences of high-stakes testing for high school graduation in the district examined in this study is that increasing numbers of poor and African American students will fail to meet graduation requirements on the first try. Although they can take the FCAT several times, one has to wonder how their increased frustration levels may affect the likelihood that these students will drop out of high school. African American students and students from the lowest income households are already the most negatively impacted by FCAT graduation requirements. They are also the most likely to encounter a negative graduation effect due to raising the passing score required on the FCAT. The consequences of this policy are likely to impose substantial social costs to some of the most vulnerable members of society.

With respect to the overall questions of school accountability that we discussed earlier in this article, it appears that the performance/testing approach might result in worsening the gap between the schooling outcomes of Black and Hispanic students and their White counterparts as well as between the schooling outcomes of haves and have-nots, at least in the short run. The social difficulties imposed by race and class differentiations in this country will virtually ensure that schools with larger percentages of poor and minority students will be more likely to have lower average student scores on standardized tests such as Florida's FCAT. Our work here suggests that, to the extent that educational problems reside in students themselves (not faulting students who, after all, cannot control their demographic characteristics), any school that serves a student body made up of a significant percentage of demographically at-risk students will likely be less competitive when standardized testing is used to assess school performance. In that sense, NCLB's emphasis on raising achievement levels for all students is appropriate but it is far less clear that a high-stakes testing approach will force improvements in educational opportunities for students who have traditionally been "left behind."

Notes

1. The effects of state-mandated tests were ascertained by comparing student performance on state tests that assess the same curriculum domains covered in national standardized tests, including the National Assessment of Educational Progress (NAEP), American College Test

(ACT), Scholastic Aptitude Test (SAT), and Advanced Placement (AP) assessments (Amrein & Berliner, 2002b).

2. In Florida, the aggregate of student scores on the Florida Comprehensive Assessment Test (FCAT) in schools are used as the basis for giving a letter grade to each school similar to the kinds of grades that students have since time immemorial brought home on report cards: A through F. As originally conceived by Governor Jeb Bush, this simple and seemingly intuitive A-through-F scoring system was combined with a program in which students who attended schools that scored an F more than 2 years in a row would be given vouchers that allow them to attend better performing public schools or private schools of their choice. West and Peterson (2005) discussed this plan extensively.

3. See, for example, http://factfinder.census.gov/servlet/SAFFFacts?_event=ChangeGeoContext&geo_id=16000US1235000&_geoContext=&_street=&_county=Jacksonville&_cityTown=Jacksonville&_state=04000US12&_zip=&_lang=en&_sse=on&ActiveGeoDiv=&_useEV=&pctxt=fph&pgsl=010

4. These are gathered from the Census-block data generated by student addresses.

5. The passing score of 300 has remained in effect since it was instituted in 2002. The Florida Board of Education has the right to increase it again, but it has not done so. The Florida Board of Education literature about high school graduation requirements states that students must make a passing score on the FCAT without stating explicitly what the passing score is, thereby leaving the option open for increasing it again. Even with the passing score of 300, 11% of Duval County high school students were unable to graduate in 2005 because they failed some portion of the FCAT after numerous attempts (Mitchell, 2005).

6. Generally there is little or no difference between a logit and a probit model when there is a large sample size and when independent variables have a relatively normal amount of variability or variance (Greene, 2000). Our analysis meets these conditions and, in fact, the results (in terms of which variables were significant predictors) were the same whether we used a logit or probit for the analysis.

7. Whereas all other coefficients were very robust to equation specification changes (for instance, changing which category was omitted) the coefficient on parental education became significant when we included a different set of income categories in the regression. Specifically, when we included only the highest income category (household income above $70,000) then parental education levels of less than 9th grade were negative and significant. This indicates that households with parental education attainment of less than 9th grade are less likely to have students who pass the FCAT graduation requirement. This implies that income and parental education are both important predictors of the probability of meeting graduation requirements in 10th grade. We suspect that there is multicolinearity between very low income households and households whose educational attainment is less than 9th grade. Multicolinearity inflates the standard errors of highly correlated variables. A solution to this problem is to omit one of the highly correlated variables. Multicolinearity does not appear to be a problem affecting other variables in the equation because these coefficients are robust to this change in the income variable.

8. This is a case where selection bias confounds the interpretation of this coefficient. High-achieving students self-select themselves into and compete for spots in magnet programs in Duval County. The positive and significant coefficient for magnet schools implies that students in magnet programs are less likely to fail the FCAT graduation requirement; however, because of self-selection, the coefficient should not be interpreted to mean that attending a magnet school or turning all schools into magnet programs, for instance, would increase achievement for an average student from the population as a whole.

References

Amrein, A. L., & Berliner, D. C. (2002a, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, *10*(18). Retrieved August 2, 2005, from http://epaa.asu.edu/epaa/v10n18/

Amrein, A. L., & Berliner, D. C. (2002b). *The impact of high-stakes tests on student achievement performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP test results in states with high school graduation exams.* Tempe, AZ: Educational Policy Research Unit, College of Education, Arizona State University.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305-332.

Dahmus, T. (2003). The effects of No Child Left Behind Act on the balance of power among local, state, and federal educational authorities. *LBJ Journal of Public Affairs*, *16*, 20-28.

Dee, T. S. (2003). The "first wave" of accountability. In P. E. Petersen & M. R. West, (Eds.), *No child left behind? The politics and practice of accountability* (pp. 215-241). Washington, DC: Brookings Institution Press.

Garza, C. (2003, March 11). Many see time slipping away on FCAT. *Florida Times-Union*, pp. B1, B3.

Grech, D. (2002, May 23). FCAT failure rate is "scary": Many in S. Florida must retake the test. *The Miami Herald*, p. B1.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Haney, W. (2000, August 19). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, *8*(41). Retrieved August 2, 2005, from http://epaa.asu.edu/epaa/v8n41/

Hanushek, E. A., & Raymond, M. E. (2003a). Improving educational quality: How best to evaluate our schools? In Y. Kodrzycki (Ed.), *Education in the 21st century: Meeting the challenges of a changing world* (pp. 193-224). Boston, MA: Federal Reserve Bank of Boston.

Hanushek, E. A., & Raymond, M. E. (2003b). Lessons about the design of state accountability. In P. E. Peterson & M. R. West (Eds.), *No child left behind? The politics and practice of accountability* (pp. 127-151). Washington, DC: Brookings Institution Press.

Hanushek, E. A., & Raymond, M. E. (2004). *Does school accountability lead to improved student performance?* (NBER Working Paper No. 10591). Cambridge, MA: National Bureau of Economic Research. Retrieved August 5, 2005, from http://ripolicyanalysis.org/SchoolAccountability.pdf

Jacob, B. A. (2001). Getting tough? The impact of mandatory high school graduation exams on student achievement and dropout rates. *Educational Evaluation and Policy Analysis*, *23*, 99-122.

Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.

Lawmakers want FCAT alternative for students. (2003, April 6). *Florida Times-Union*, p. A20.

Mitchell, T. (2005, April 30). More than 1000 will lose their diplomas to FCAT. *Florida Times-Union*. Retrieved June 1, 2005, from http://www.jacksonville.com/tu-online/stories/043005/met_18620079.shtml

Protesters want moratorium on this year's FCAT results. (2003, May 18). *Florida Times-Union*, p. B6.

Raymond, M. E., & Hanushek, E. A. (2004). Shopping for evidence against school accountability. In W. J. Fowler Jr. (Ed.), *Developments in school finance: 2003. Fiscal proceedings*

from the annual State Data Conference of July 2003 (pp. 1-139). Washington, DC: National Center for Education Statistics, U.S. Department of Education. Retrieved August 6, 2005, from http://nces.ed.gov/pubs2004/2004325.pdf

West, M. R., & Peterson, P. E. (2005, March 23). *The efficacy of choice threats within school accountability systems: Results from legislatively induced experiments* (Working Paper Number: RWP-05-032). Retrieved November 30, 2005, from http://ksgnotes1.harvard.edu/Research/wpaper.nsf

Mary O'Malley Borg is a professor of political economy at the University of North Florida and the Director of the Undergraduate Academic Enrichment Program. Her research interests include the economic consequences of state lotteries, lottery-funded merit scholarships, and educational reform policies. Her publications have appeared in *National Tax Journal*, *Public Finance Review*, and *Journal of Economic Education*.

J. Patrick Plumlee is an associate professor of political science and public administration and the MPA Program Director at the University of North Florida. His publications have appeared in *Public Administration Review*, *Administration & Society*, and *Western Political Quarterly*.

Harriet A. Stranahan is an associate professor of economics at the University of North Florida. Her current research projects include an analysis of the Florida Bright Futures Scholarship Program and an examination of the Florida School Recognition Program. Her recent publications appear in *Journal of Economic Education*, *Public Finance Review*, and *Business Ethics Quarterly*.