

CAROLINA context

WWW.SOUTHW.ORG

APRIL 2010 NUMBER 10

FERREL GUILLORY

Director

guillory@unc.edu

THAD BEYLE

Senior Research Fellow

beyle@email.unc.edu

HODDING CARTER, III

Leadership Fellow

hoddincarter@unc.edu

D. LEROY TOWNS

Research Fellow

dltowns@email.unc.edu

THE PROGRAM ON PUBLIC LIFE

is a nonpartisan organization devoted to serving the people of North Carolina and the South by informing the public agenda and nurturing leadership. The Program is housed in the School of Journalism and Mass Communication and affiliated with the Center for the Study of the American South at the University of North Carolina at Chapel Hill.

The Program on Public Life wishes to recognize and thank its funders that have provided support for research and white papers:

Z. SMITH REYNOLDS FOUNDATION
TRIANGLE COMMUNITY FOUNDATION

DIRECTOR'S NOTE

As the white paper that follows explains, North Carolina serves as a nearly ideal state in which to take a look at how efforts to propel higher performance by public schools and their teachers, and higher achievement by their students, have played out.

For one thing, North Carolina has an advanced, extensive system of data collection that affords scholars and policymakers the informative ability to track schools, teachers and students. For another thing, since 2003, North Carolina schools have operated under two different systems for measuring progress or the lack of it.

Under the ABCs program, the state has its own accountability system that measures annual gains in student achievement. Under the federal No Child Left Behind law, the state's schools also fall under a system that measures student performance in relation to a defined goal of proficiency.

This white paper is a somewhat condensed version of a detailed academic study forthcoming in the June issue of the *Journal of Policy Analysis and Management*. It is important because our public schools are important—to the future prospects of our young people as earners and as citizens, to our communities' quality of life, and

to the vibrancy of North Carolina's economy and democracy.

The authors are Helen F. Ladd, the Edgar Thompson Professor of Public Policy and Professor of Economics at Duke University, and Douglas L. Lauen, assistant professor of public policy at the University of North Carolina at Chapel Hill. They received research assistance from Steven M. Gaddis. The Spencer Foundation funded their research.

They drew on the North Carolina Education Research Data Center, housed at Duke University. For further information on the calculations, sources and background for this paper, you may go to this website: www.unc.edu/~dlauen.

We are grateful to Professors Ladd and Lauen for giving the Program on Public Life an opportunity to collaborate with them in disseminating their findings to legislators and other state policy makers, as well as more widely to the academic community. We hope, along with the authors, that their work will contribute to a deeper understanding of the effects of school reform policies and strategies.

— FERREL GUILLORY

Director, Program on Public Life



Status vs. Growth: Comparing Strategies for School Improvement

HELEN F. LADD AND DOUGLAS L. LAUEN

High on the U.S. educational policy agenda is how best to hold schools accountable for the performance of their students. The theory behind educational accountability is that by setting standards and measuring performance relative to standards, teachers will work harder and students will learn more. With the passage of the federal No Child Left Behind Act (NCLB) in early 2002, student test scores in math and reading have come to represent the outcome of interest, regardless of their relationship to any specific curriculum standard, and schools are judged primarily on the percent of students at grade level or above, or academic *status*. Increasingly, however, observers have argued for shifting the metric for school accountability away from the achievement status of a school's students, as is the case under NCLB, in favor of a metric based on students' *growth* in achievement during the year — how much they have improved.

The argument for using achievement growth rather than achievement status as the basis of school accountability is two-fold. First, because children come to school with different degrees of readiness to learn and prior achievement levels, many people believe it is unfair, and potentially counterproductive, to expect schools alone to offset the effects of the background characteristics of their students. Instead, the argument goes, schools should be held accountable for outcomes over which they have more control, such as how much the children learn during the year, typically measured by their gains in test scores.

Second, the focus on achievement status, as defined by a proficiency threshold, provides a strong incentive for schools to focus attention on students

near the threshold to the potential disadvantage of students far below the threshold and of those above the threshold. At the same time, some growth models have been criticized for lack of transparency and their failure to require students to meet specific standards.

The effects on students at all three levels (above the threshold, at the threshold, and below the threshold) within schools that failed to meet the required standards of these two types of accountability systems are the focus of this policy brief. In contrast to recent research, which has focused almost exclusively on the distributional effects of status programs such as NCLB, we compare the within-school distributional effects of a system based on achievement status to one based on achievement growth. Our study is based on longitudinal data from North Carolina.

North Carolina is a good state to examine these two types of accountability because its schools have been subject to the state's growth-based accountability system since the academic year 1996/97 and the federal No Child Left Behind (NCLB) status-based accountability system since 2002/03. Because the two systems use different methods for judging the effectiveness of schools, some schools that appear to be performing well under one system may do poorly under the other. In contrast to most other states, North Carolina has long used tests that are aligned with the state's standard course of study, with test scores reported on a developmental scale. As a result, the tests measure what teachers are expected to teach and students to learn, and students in any grade are less likely to reach a ceiling test score than would be the case with a maximum score in each grade.



THE NORTH CAROLINA ABCS PROGRAM — THE GROWTH APPROACH

The North Carolina accountability program — referred to as the ABCs program—was part of a broader state effort to improve the academic performance of the state’s children throughout the 1990s. First implemented in 1996–97, the ABCs program was intended to hold teachers in individual schools accountable for the overall performance of their students. Though the program applies to high schools as well, the present study focuses on schools serving students in grades three through eight. Under the ABCs program, schools are judged primarily on the annual achievement gains of their students from one year to the next. This growth approach to accountability was feasible because the state had been testing all students in grades three through eight annually in math and reading since the early 1990s — long before it was required to do so under the federal No Child Left Behind legislation of 2001.

From 1996/97 to 2005, an expected average gain in test scores was predicted for each student, and the school was deemed effective or not depending on how the actual gains of its students compare to their predicted gains. If a school raised student achievement by more than was predicted for that school, all the school’s teachers received financial bonuses — \$1,500 for

federal Title 1 grants to pay for supplemental services, including those from private providers. A school with five years of failure is subject to takeover by the state, an outcome that, to date, has been rare across the country. Under NCLB, North Carolina policy makers must set annual proficiency targets—defined in terms of the percentages of students who are at grade level—that will assure that each school is on target toward the 2013/14 goal of 100 percent proficiency. The result is that under the federal law each school faces an annual target defined in terms of achievement *status* rather than in terms of achievement *growth* as under the state accountability system. Not surprisingly, a school that performs well under the state’s accountability system may do poorly under the federal system, and vice versa.

ACCOUNTABILITY PRESSURE AND SCHOOL FAILURE

Under both approaches, all primary and middle schools feel pressure from accountability requirements. Central to the analysis of the within-school effects we describe below is the assumption that the accountability pressure is far more explicit and immediate for some schools in some years than for those same schools in other years or in other schools. The key determinant of accountabil-

The theory behind educational accountability is that

by **SETTING STANDARDS** and **MEASURING PERFORMANCE** relative to standards,
teachers will work harder and students will learn more.

achieving high growth and \$750 for meeting expected achievement growth. Schools that did not achieve their expected growth were publicly identified and in some cases subject to intervention from the state. The intent of the program was to induce each school to provide its students with a year’s worth of learning for a year’s worth of education. In 2005, the formula for calculating growth was changed, but the focus on holding schools accountable for achievement growth, rather than levels, remained.

In addition to their growth rankings, schools also receive various designations, such as schools of excellence, schools of distinction, and priority schools, based on the percentages of students meeting grade-level standards, which carry with them no financial bonuses. In addition, some schools are labeled “low performing” based on their high failure rates as well as their poor growth performance. Thus the ABCs program does not completely ignore achievement status. At the same time, the teachers’ bonuses are based solely on the growth in student achievement.

NO CHILD LEFT BEHIND (NCLB) — THE STATUS APPROACH

The federal government started holding schools accountable for student achievement with the 2001 reauthorization of the federal Elementary and Secondary Education Act, called No Child Left Behind. Started in the 2002/03 academic year, NCLB requires states to test students annually in reading and mathematics in grades 3–8, and assesses schools on the basis of whether their students are making adequately yearly progress (AYP) toward the ultimate goal of 100 percent proficiency by 2014. Moreover, each school must meet annual proficiency targets not only for the student body as a whole, but also for various subgroups defined by race, socio-economic status, and disability within the school.

Failure to meet AYP brings with it consequences, such as the right of children to move to another school and the requirement that districts use their

ity pressure is the clarity and salience of the adverse consequences associated with failing the standard. These consequences are very clear and immediate for school failure under the growth standard in that teachers in failing schools are denied a bonus in that year.

Although the bonuses are not huge, they are still large enough to be meaningful. A \$1,500 bonus on a \$30,000 salary, for example, represents a five percent boost in salary. Although the consequences of school failure under the NCLB legislation may be less salient in that the ultimate sanction of school takeover may still be several years in the future, the negative publicity — sometimes referred to as “naming and shaming” — associated with being in a failing school is likely to be felt immediately and to provide a spur to action.

The chart on page 5 illustrates the variation over time in the percentages of schools failing to meet either or both of accountability standards by year. The bottom line is that for the past 11 years, many schools in North Carolina have not met one or both of the standards for student achievement. How a school’s failure to meet a specific standard has affected students at different points in the prior year achievement distribution is the subject of the following sections.

WHAT RESPONSES WOULD WE EXPECT?

The status approach, as epitomized by NCLB, sets a target rate of proficiency. The school has little or no incentive to invest additional effort in students who are expected to exceed the target level of achievement. The school has an incentive to invest in students below the target but close enough that investing in their achievement results in reduced penalty to the school. There could well be some students at the bottom of the expected performance distribution for whom the additional effort on the part of the school would simply be too costly relative to the benefits for the school to make the additional effort worthwhile.



Thus, the status model generates one clear distributional prediction. Students whose expected achievement is *below* the proficiency level will receive more attention — and hence should achieve at higher levels than they otherwise would have — than those above the proficiency level. Less clear is whether there will be a group of students at the very bottom who are left behind because of the high costs of raising them to the standard.

For students whose expected achievement is well above the proficiency standard, in contrast, the question becomes whether they will receive less attention — and hence will achieve at lower levels than they otherwise would have. If additional effort for the students at the bottom is redistributed from students at the top, achievement of the higher performing students would fall.

The growth approach sets a target for the average rate growth of student achievement during the year. It is difficult to predict which students will benefit most because differential benefits depend on the relationship between additional effort and student achievement. One possibility is that the additional effort needed to raise student achievement by a given amount is uniform across students defined by their prior-year achievement. In that case, a school under pressure to raise its average achievement growth has no incentive to invest any more in one group of students more than in another. Alternatively if additional effort generates greater gains for low-performing students than for high-performing students, a growth-based accountability system would give schools an incentive to invest more in the students at the bottom of the distribution than at the top. A third possibility is that, consistent with the observation that students at the high end of the achievement distribution have made greater gains in the past than those at the bottom end, it may be easier to generate larger additional gains at the top of the distribution than at the bottom. In that case, schools under pressure would have an incentive to invest in the higher performing students, with larger gains for that group than for other groups.

WHAT THE DATA SHOW: BASIC RESULTS

We compare results of schools reacting to three types of pressure — failing only the state’s growth standard, failing the federal status standard, and failing both the growth and status standards — to the baseline case of schools meeting both the growth and status standards. We examine the results in relation to students performing above the performance threshold, those just below the threshold, and those far below the threshold for each approach to accountability.

In schools failing only the ABCs growth standard, for example, students with low prior achievement tend to gain more in math than students just

above grade level. Students with the lowest levels of prior achievement exhibit the largest gains. Among high achievers in math, students in the top category appear to gain somewhat relative to students just above grade level. This pattern suggests that in response to the pressure arising from failure to meet the growth standard, schools apparently find it easier to raise math test scores at the bottom and at the very top relative to students close to grade level. The responses to failing the growth standard differ somewhat for reading, but mainly for lower-achieving students. Again we find no evidence of adverse effects on students at the very bottom of the distribution. For this subject, however, no significant effects, either positive or negative, emerge for students below grade level. As was true for math, positive effects emerge for the top performing students.

With respect to NCLB status pressure, our results indicate that schools failing only AYP generate positive gains for low-performing students in both subjects, with far larger effects in math than in reading. Of interest is that negative effects emerge for high-performing students, with the negative effects larger in reading than in math. The pattern for lower-achieving students is fully consistent with the prediction that under a status approach to accountability, schools facing pressure to raise student achievement to the proficiency standard would focus attention on students below it. The negative distributional effects for high performing students indicates that the gains at the bottom in both subjects have come at the expense of higher achieving students in the affected schools.

Finally, schools simultaneously failing to meet both the state’s growth standard and the federal AYP standard positive show gains for low achieving students in reading but not in math. Consistent with the findings for status pressure, negative distributional effects emerge for high achieving students in both math and reading, but with the magnitudes far larger for reading.

In summary, two main conclusions emerge from this analysis. The first is that, with a few exceptions, students below grade level typically benefit relative to students close to grade level in schools responding to some form of achievement pressure. The main exceptions are low-achieving students in reading in schools failing to meet the growth standard alone and low-achieving students in math in schools facing pressure from both programs. The second is that status pressure exerted by NCLB appears to generate more within-school shifting of resources away from higher achievers to low achievers than is the case for ABCs-type accountability. In response to pressure from the growth approach, schools appear to focus additional attention both on students who are below grade level and on those at the top of the achievement distribution.

SUMMARY OF RESULTS

This table summarizes whether the test scores of students above and below grade level increase, decrease, or remain unchanged as a result of attending a school which faced various types of accountability pressure (failed the state’s growth standard, failed the federal NCLB standard, or both).

	Math		Reading	
	Below Grade Level	Above Grade Level	Below Grade Level	Above Grade Level
Failing the Growth Standard (ABCs)	Increase	Small increase	No change	No change
Failing the Status Standard (NCLB)	Increase	No change	Increase	Small decrease
Failing both Standards (ABCs & NCLB)	Increase	No change	Increase	Small decrease

DATA AND METHODS

Our student-level longitudinal regression model predicts test score gain as a function of accountability pressure, student prior achievement, and the interaction of accountability pressure and student prior achievement. We use a statistical technique called fixed effects analysis to control for student and school factors that might confound our estimates. Our models seek to uncover evidence of any within-school educational triage in schools facing three types of accountability pressure: 1) pressure from the status model only (NCLB), 2) pressure from the growth model only (ABCs), and 3) pressure from both status and growth models (both NCLB and ABCs).

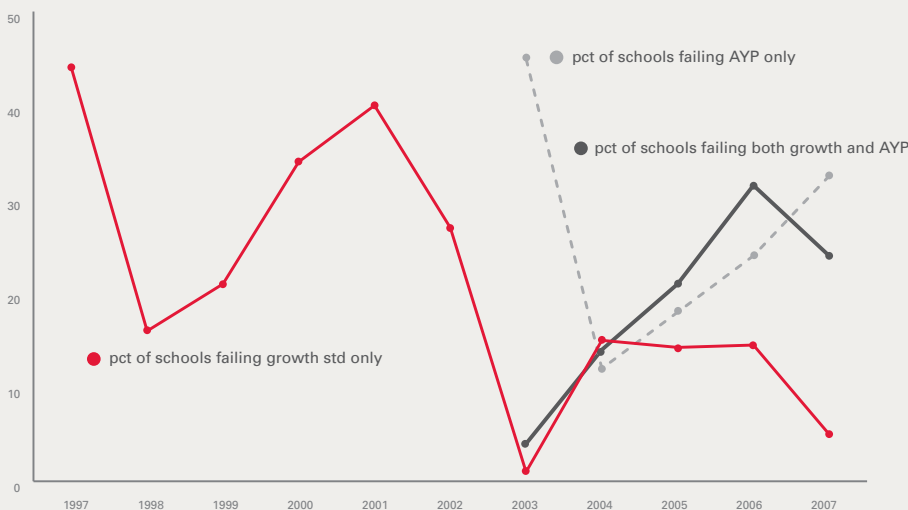
We start with data on all students in North Carolina public schools in grades 3–8 from 1996/97 to 2006/07 for whom test scores are available in either math or reading. These data are available through the North Carolina Education Research Data Center, housed at Duke University. The total panel data set includes more than 6.8 million student-year observations, with more than 1.9 million unique students and 2,129 unique elementary and middle schools. Test scores are standardized to ensure comparability across time.

On average only 22 percent and 19 percent of students in reading

and math, respectively, were below grade level between 1997 and 2007, a finding consistent with the observation that North Carolina’s proficiency level had been set at a relatively low level. Comparing results from the state assessment to NAEP scores is one way to determine the relative rigor of North Carolina’s proficiency levels. On NAEP in 2007, 34 percent of 8th graders were at or above proficient in math, and 28 percent were at or above proficient in reading. On the North Carolina state assessments in 2007, 63 percent of 8th graders were above grade level in math, and 86 percent were above grade level in reading.

On average over the period 1997 to 2006, 22 percent of students attended a school that failed AYP, 32 percent attended a school that failed the growth standard and 10 percent attended a school that failed both standards. During the period 2003–06, 50 percent of students attended a school that failed AYP, 42 percent attended a school that failed the growth standard, and 29 percent attended a school that failed both standards. The sample includes 30 percent black students, 5 percent Hispanic, 5 percent Asian and American Indian) — and 46 percent received a free or subsidized lunch.

PERCENTAGE OF SCHOOLS FAILING GROWTH AND AYP STANDARDS



Across the post-NCLB years the percentages of elementary and middle schools not meeting AYP ranged from a low of 26 in 2004 to a high of 56 in 2007. The variation in the growth failure rate across years is even greater, in part because of an anomalous outcome in 2003. Due to changes in the state assessments in 2003, only five percent of the schools failed to make their expected growth in that year compared to 27 and 29 percent in the prior and the following years, respectively. The highest failure rate over the entire period was 43 percent in 1997; as of 2007, it was about 28 percent.



UNC
CENTER FOR THE STUDY
OF THE AMERICAN SOUTH

NON-PROFIT ORGANIZATION
U.S. POSTAGE

PAID

PERMIT NO. 177
CHAPEL HILL, NC

context

CAROLINA

PROGRAM ON PUBLIC LIFE

Campus Box 3365
Chapel Hill, NC, 27599-9127
southnow@unc.edu

DISCUSSION AND CONCLUSIONS

1. Using a ten-year panel data set, we find that neither type of school-based accountability system generates distributionally neutral effects on student achievement in the schools subject to accountability pressure. Moreover, consistent with our predictions, the effects on students at different performance levels differ depending on whether the system holds schools accountable for the growth or the status of their students' learning. This conclusion simply reflects the fact that educators do indeed respond to incentives. The policy challenge is to design a system consistent with the goals of the policy.

2. We find that under both approaches to accountability, students below the proficiency standard typically benefit relative to those just above the standard, although that pattern is clearer and more consistent for math than for reading. In the growth approach, the overall effect within the affected schools is to raise student achievement in the aggregate somewhat more for high-performing students than for low-performing students. This outcome occurs because of the far larger number of students above grade level than below in the relevant schools. This pattern could represent a shortcoming of the policy if the main goal of the program were to close achievement gaps within such schools. No such concern arises if the main goal is to raise achievement among all levels of students within the low-performing schools.

3. In reading, status-based accountability pressure appears to generate negative within-school effects for students above the proficiency threshold. No such negative effects emerge with respect to the growth approach. One possible

explanation is that schools may try to improve reading scores of low performers by using student-specific strategies that reduce resources available to other students, while in math they may use more general strategies, including better instruction for all students. Whether the gains to students below proficiency are worth the costs to students above proficiency is a question of values.

4. We find little or no evidence that students at the very bottom are "left behind" in connection with either approach to accountability. In particular, under both types of accountability in North Carolina, such students experience positive (or in some cases, zero) achievement gains on average relative to students slightly above grade level. This finding contrasts with much, but not all, of the prior literature that examines how schools respond to status based accountability systems. One possible explanation for the difference is North Carolina's relatively low proficiency standard. It may well be that in this state, raising students up to the proficiency standard is more feasible than in other states with higher standards.

We believe the within-school patterns highlighted in this study are relevant to policy debates about school accountability. One key question is whether the goal of school-based accountability is to make low performing schools better for everyone or to narrow achievement gaps by raising the performance of students at the bottom relative to those at the top in the low-performing schools. To the extent that it raises achievement for some students, but lowers it for others, as appears to be the case in status-based accountability system in reading, there are clear tradeoffs that require additional policy discussion and debate.