



11-25-2013

# How (and Why) NCLB Failed to Close the Achievement Gap: Evidence from North Carolina, 1998-2004

Roslyn Mickelson

*UNC Charlotte, rmicklson@uncc.edu*

Jason Giersch

*University of North Carolina at Charlotte, jgiersch@uncc.edu*

Elizabeth Stearns

*University of North Carolina at Charlotte, Elizabeth.Stearns@uncc.edu*

Stephanie Moller

*University of North Carolina at Charlotte, Stephanie.Moller@uncc.edu*

---

## Recommended Citation

Mickelson, Roslyn; Giersch, Jason; Stearns, Elizabeth; and Moller, Stephanie (2013) "How (and Why) NCLB Failed to Close the Achievement Gap: Evidence from North Carolina, 1998-2004," *ECI Interdisciplinary Journal for Legal and Social Policy*: Vol. 3: Iss. 1, Article 1.

Available at: <http://ecipublications.org/ijlsp/vol3/iss1/1>

## **How (and Why) NCLB Failed to Close the Achievement Gap: Evidence from North Carolina, 1998-2004**

The twin goals of educational excellence and equity have inspired federal educational reforms since the middle of the last century thanks to Sputnik and *Brown* (1954). While there has been notable movement toward both goals, attaining educational equity and excellence has proven to be elusive. International comparisons indicate overall US student performance in science and mathematics is mediocre, lagging behind that of students in many advanced industrialized nations. Racial and socioeconomic gaps in educational performance and other outcomes persist despite decades of reforms designed to eliminate them.

The political compromise that became No Child Left Behind (NCLB 2002) was designed to address both goals of excellence and equity but, as this article will show, there are reasons to believe it has achieved neither goal. The linchpin of this reform dynamic is high-stakes tests used to hold educators accountable. This article focuses on one aspect of this theory of reform in relation to the equity goal of NCLB—high-stakes tests and the gaps in achievement along the lines of race and socioeconomic status (SES). If NCLB is successful, the long-standing gaps in test scores between middle class and low-income youth, and among White, Black, Latino/a, Native American, and Asian students should shrink or disappear.

In this article we present findings from our study of student achievement in North Carolina using standardized test score data from the state's population of approximately 76,000 high school seniors who graduated in 2004. Using test score data for each individual from 1998 through 2004, along with indicators of family background, individual student factors, and school characteristics we show that testing and accountability did not eliminate achievement gaps. Moreover, we show that North Carolina's standardized test scores actually contribute to the perpetuation of gaps in achievement because of (a) testing's likely effects distorting curriculum and instruction and (b) the use of test scores to assign students to racially stratified tracks in core classes where students are exposed to very different curricula and have starkly different opportunities to learn. Black, Latino, Native American and low-income students are more likely to be assigned to lower tracks than their middle-class, White, and Asian counterparts. We argue that given the stratified nature of tracking, standardized tests used in conjunction with tracking and curricular distortions triggered by the pressure on educators to improve low-scoring students' test performance (so-called 'drill and kill') contribute to the perpetuation of achievement gaps. It is through the quantifiable measurement of the interplay between high-stakes testing and tracking that this paper makes an important contribution to the literature on education policy.

The article proceeds as follows. We begin with a discussion of the theory of reform that underlies NCLB, contrasting its market-inspired principles with the equity-inspired reforms of the 1950s through 1980s. Next, we describe our North Carolina study that assesses whether

NCLB has narrowed achievement gaps in the state. We conclude the article with a discussion of the findings in light of the title's claim that NCLB has failed to close achievement gaps, and we examine the implications of our findings for policy including the likely reauthorization of NCLB.

### NCLB and the Choice Theory of School Improvement

The United States, unlike many European countries with centralized national control of educational policy, has no history of federal control of local schools. State and local control of public schools are a hallmark of the U.S. educational system, but the federal government became increasingly involved in public education during the second half of the 20<sup>th</sup> century. Since then the federal courts, beginning with landmark decisions such as *Brown* (1954), pushed both the federal and state governments to remove race, disability, poverty, and gender barriers to educational opportunities. A number of presidents enforced judicial rulings and issued executive orders broadening educational access, most notably President Lyndon B. Johnson's Executive Order 11246 (1965) that established the policy groundwork for affirmative action in higher education. In the 20<sup>th</sup> century Congress passed sweeping education legislation, such as the GI Bill in 1945, the Civil Rights Act of 1964, the Elementary and Secondary Education and Higher Education Acts of 1965, the 1972 amendment to the higher education act known as Title IX, and the Education for Children with Disabilities Act of 1975 (Public Law 94-142), reenacted in 1997 as the Individuals with Disabilities Education Act (Cross 2004, Manna 2006). Federal involvement in U.S. public education accelerated with the No Child Left Behind Act (NCLB 2002). Passed as part of the 2001 reauthorization of the Elementary and Secondary Act and signed by President Bush in 2002, NCLB is one of the most comprehensive federal reforms governing state and local educational policies in U.S. history (NCLB, Section 1116 subsection 115 STAT 1484).

The current national educational reform movement, epitomized by NCLB, is a response to one of the core conservative critiques of public education, namely that it lacked standards and accountability. From the 1950s to the end of the 1980s, federal governmental efforts to reform education focused on equity-inspired legislation, policies, practices, and judicial decisions including school desegregation, Title IX (Heckman 1992), bilingual education (Troike 1978), compensatory education programs like Head Start (Garces, E., Thomas, D., and Currie, J. 2002), and Title I (Borman, Stringfield, and Slavin 2001). Beginning in the late 1980s during the George H.W. Bush administration, federal legislative and executive efforts turned to market principles such as choice, competition, standards, and accountability to achieve both goals (Petrovich and Wells 2005). The No Child Left Behind Act is emblematic of such market-inspired federal school reform efforts. Its formal goals are to improve academic achievement while simultaneously advancing equality of educational opportunity and closing racial and

socioeconomic achievement gaps in outcomes. The law seeks to achieve these ends through market principles of choice, competition, standards, and accountability. Advocates of market-inspired reforms argue that market dynamics will motivate principals and teachers to change their behaviors and improve educational efficiencies or go out of business because if they do not perform well, families will exercise their choice and move to schools that offer better options. Motivated to keep their jobs, teachers will improve and streamline their performance. Administrators will lead, inspire, and supervise by rewarding or sanctioning their staff accordingly. At the school level, administrators will hold teachers accountable for student outcomes and parents will hold them both accountable for their child's and the school's overall performance based largely on standardized test scores that ostensibly measure how well children have learned the formal curriculum standards set by the state.

Although business leaders, governors, and other state actors began educational reforms three decades ago (Mickelson 1999), the contemporary alliance that dominates federal and state educational reform initiatives also includes several loosely related conservative advocacy groups, think tanks, philanthropies, and corporate reforms (Broad, Gates, and Walton—see Ravitch 2010) that, while differing from each other in a number of respects, share a belief that education ought to be operating more as a market (including privatization, educational management organizations, vouchers, charters, and choice); that curricula, teaching, and achievement standards ought to be established; that systematic assessments ought to be used to determine if standards have been met; and that responsible parties (students, teachers, educators, schools, and ultimately the states) ought to be held accountable and sanctioned if standards are not met (Apple 2001, Ehrenreich 1989, Smith and Miller-Kahn 2004). Another faction supporting standards-based reforms includes Congressional Democrats, like Representative George Miller and the late Senator Edward Kennedy who both sponsored NCLB. These Congressional liberals were motivated by a different set of priorities from those of the conservative alliance. Their concerns lie more with the disparate outcomes of schooling by race, class, disability, and language, whereas conservatives tend to be motivated by concerns about enhancing efficiency and market principles, and loosening state control of public education.

The formal goals of NCLB are to improve the overall quality of education, to provide equality of educational opportunity to all students, and to eliminate group differences in educational outcomes. Advocates of NCLB believe its requirements for rigorous curricula, teaching, and learning standards and accountability strike the right balance of incentives and sanctions (Ali 2004). NCLB's supporters believe accountability will force states to align course content, instruction, and assessment, thereby improving instruction and enhancing learning. The law's advocates maintain that the legislation's requirements for annual testing and disaggregation of scores by subpopulations will reduce the achievement gaps because schools will not be able to hide the low achievement of disadvantaged students behind the higher achievement of advantaged children. They assert public access to reports will help parents make

informed decisions about their children's education and stimulate greater parental involvement. Advocates believe that giving parents information *and* the right to transfer their child will force educators in low performing schools to improve their capacity to educate all children by creating a competitive atmosphere where schools must improve to retain students (Chubb and Moe 1990, Powers and Cookson 1999).

Some educators and civil rights advocates supported NCLB because they view it as a federal policy that expresses the moral and legal imperative to provide equality of educational opportunity to all youth (Ali 2004, Taylor 2003). Senator Edward Kennedy co-sponsored NCLB because he believed the transparency of disaggregated test scores would advance the public's demand for equity (Kennedy 2005). The late civil rights attorney William Taylor argued that, if properly and fully implemented, NCLB had the potential to empower parents through the requirement that school district authorities permit student transfers from low-performing schools to better ones irrespective of the latter's seat capacity or other potential constraints (Taylor 2003).

Yet, several weaknesses are endemic to the operation of NCLB. School choice theory posits that any of NCLB's negative consequences are necessary in the short-term to trigger the kinds of changes in behavior that result in long-term improvements in low performing schools (see Chubb & Moe 1990, Fiske and Ladd 2000, Godwin and Kemerer 2002, Henig 1995, McEwan and Carnoy 2000).

**Requiring Schools to Solve Societal Inequality.** A general critique of NCLB notes that the legislation fails to take into account how the social, political, and economic sources of unequal educational achievement that lie beyond school walls affect what goes on within them. This point is relevant to schools in rural and central city communities where the material realities of daily life make teaching and learning extremely difficult (Anyon 1997, Rebell and Wolff 2009).

**Accountability Undermines Teacher Quality for the Neediest Students.** Schools with concentrated poverty tend to have high student mobility rates because poor people are more likely to be residentially mobile than other subpopulations. Student mobility and concentrated poverty in schools destabilize teacher quality because the best educators tend to transfer to work environments where students will meet adequate yearly progress (AYP) (Jackson 2009, Lankford, Loeb, and Wyckoff 2002).

**Pedagogy of Drill and Kill.** Because assessments of student progress are based on annually established (and therefore variable) benchmarks of achievement rather than value-added systems that measure improvement, schools whose students show significant progress yet are still below proficiency are labeled as failing to meet AYP. To avoid this sanction, low-

performing schools are pushed to concentrate on test preparation to ensure that students pass state standardized tests. This means that in underperforming schools substantial amounts of classroom time must be devoted to preparation for math and language tests, rather than to learning art, music, science, and social studies in creative ways that engage students and stimulate their higher order thinking (Nichols and Berliner 2007, Watanabe 2008).

**Test Preparation Driving the Curriculum.** For some observers, test preparation is a positive use of instructional time, especially for low-performing students, but skeptics raise the issue of what does *not* get taught while students are preparing for tests (Nichols and Berliner 2007). Critics point to two aspects of this problem: first, when test preparation drives teaching, the curriculum gets narrowed and simplified to only those topics found on the standardized test; and second, science, social studies, art, music and other elements of the curriculum are often given short shrift during periods when schools focus single-mindedly on test scores (Ravitch 2010).

**Perpetuating Achievement Gaps by Exposure to Differential Curricula.** Critics note that in schools that easily make AYP, students are taught science, social studies, art, and music because there is less time devoted to test preparation. Moreover, the nature of the pedagogy and breadth and depth of curricular coverage to which such students are exposed stimulates higher order thinking (Lee 2002). These dynamics contribute to perpetuating the racial, ethnic, and social class gaps in outcomes because poor and disadvantaged minority youth are over-represented in low performing schools where they receive the narrowed curriculum and are drilled in test taking skills in lieu of a rich, broad, and engaging curriculum taught through active learning activities.

**Opt-out Provisions Exacerbating Concentrations of Low Performing Students.** NCLB's opt-out provision is also a source of concern to many observers. According to Taylor (2003), a number of factors may prevent parents from transferring their children from low-performing to high-performing schools and thus limit NCLB's ultimate utility as a vehicle for enhancing equality of educational opportunity and for school improvement. Districts reluctant to fully implement transfer policies can limit transfers to certain schools in the district, reduce the number of spaces available, delay notification to parents, or provide tepid announcements without affirmative outreach to parents of eligible students.

Given the self-selection mechanisms at work with any choice system, those who opt-out may be from more academically-oriented families with higher-achieving students. When the better students leave one school for another, the sending school loses funding and parental support as well as the better students' contributions to the school's positive academic climate. Teachers at the sending school may become demoralized and move to another school, thereby contributing to staff instability that often plagues low-performing schools. Students who remain

in low-performing schools are likely to be less proficient and from lower-income families, thereby exacerbating their school's problems in achieving AYP.

### Findings about NCLB and Achievement Gaps

High-stakes tests are nothing if not controversial. McDermott's (2007) historical case study of high-stakes tests suggests that advocates of accountability policies expected high-stakes tests to be a fair and efficient way to inspire improvements in both educational equity and achievement. However, in all four of the states in McDermott's study, the results fell short of the expectations. As this article will show, North Carolina could serve as the fifth example of the failure of accountability to improve outcomes.

On the one hand, accountability policies are flawed in that they demand standard outcomes without regard to differences in inputs. On the other hand, the tests themselves are problematic in terms of their distortion of educational goals and teaching strategies. Once high stakes consequences are attached to standardized tests, maintaining a valid accountability program becomes even more challenging. The principle known popularly as Campbell's Law (Campbell 1979) states that whenever important decisions are based upon the outcomes of some quantitative measure, the measure itself becomes corrupted. Indeed, Jones, Jones, and Hargrove (2003) found that once states implemented high-stakes testing policies, schools' curricula shrank, teaching practices worsened, motivation suffered, retention increased, teachers became more difficult to recruit, and school reputations deteriorated.

Berliner (2011) argues that outcomes observed above are the unfortunate result of completely rational behavior by educators and politicians. High-stakes testing creates incentives that reward teachers and administrators for getting high scores and not for developing, inspiring, or guiding the student, or even for teaching a broad array of academic skills or knowledge. Whether a teacher "teaches to the test" (Perna and Thomas 2009), a principal excludes students from the test (Roderick and Engel 2001), or a school engages in outright cheating (Ravitch 2010), the reasons for doing so are always to game the system by getting student scores that are as high as possible (in contrast to scores that are as accurate as possible).

Twenty years of trend lines showing the performance of students on the National Assessment of Education Progress (NAEP), the nation's 'report card,' are instructive. Proponents of market-inspired reforms anticipated they would improve both educational equity and excellence. Although performance among all racial and SES groups shows a slight uptick, a review of the gaps in reading and mathematics from 1992 to 2011 shows no change in the *size* of the disparities between the performances of Blacks and Whites, something we could expect to see if NCLB were operating to eliminate inequities as advocates expected it would. Importantly,

the *rate* of improvement among all groups is the same before and after the passage of NCLB, suggesting that any improvements in performance cannot be attributed to the policy.

One reason for NCLB's apparent failure to narrow race gaps is that the law does not influence the core organizational features of schooling that contribute to the creation of race gaps in student performance. A well-known organizational source of gaps is curricular differentiation associated with the practice of ability grouping in elementary schools and academic tracking in secondary schools. The reason for this dynamic is that the students who receive the worst of the narrowed curricula, the weakest pedagogy, "drill and kill" lessons, and limited learning opportunities are those in lower-track classes. A growing body of research indicates that such students have long gotten the short end of the academic stick (Lucas 1999, Mickelson and Everitt 2003, Oakes 2005). New research indicates the introduction of high-stakes tests has exacerbated the problem. For example, in her case study of North Carolina schools, Watanabe (2008) observed that the very same teachers spent, on average, nearly 50% of their class time on test preparation in the lower track classes but only 15% of their class time on test preparation in their upper-level classes. What were the honors students doing while other students practiced multiple choice questions? They were doing speeches, debates, reports, creative writing, and other activities that include the higher-order thinking skills that are important to the quality of education. High-stakes tests, it would seem, have become one more mechanism by which schools provide unequal opportunities to learn.

### **The Present Study: Testing and the Race Gap in North Carolina, 1998-2004**

Several state governments instituted standards and accountability reforms for K-12 and higher education years before NCLB required all states receiving federal aid to do so. Under the leadership of former Governor Jim Hunt, North Carolina became one of the first of these early states (Heubert and Hauser 1999). The centerpiece of North Carolina's own public education restructuring efforts is its 1996 school-based management and accountability program known as the New ABCs of Public Education framework (NCDPI 2011). ABC stands for accountability and high standards, the basics, and local control. In fact, many of NCLB's standards-based reforms are foreshadowed in North Carolina's reforms (Mickelson, Smith, and Southworth 2007).

Between 1995 and 1998 the state implemented the ABCs program in every public school from kindergarten to 12<sup>th</sup> grade. The program continues to evolve, but the use of standardized test results to evaluate, reward, and punish students and schools has been a part of the ABCs from the mid-1990s. The ABC framework established growth and performance standards beginning in 1996. Students in grades 3 through 8 are assessed by annual End-of-

Grade (EOG) tests in reading, mathematics, and writing. High school students graduating in 2004 were tested in one English course, two social studies courses, three math courses, and four science courses.<sup>1</sup> High school assessments are called End-of-Course (EOC) tests. EOG and EOC tests are aligned with the state's curricular standards in the tested subjects.

Schools with students who meet growth standards are rewarded with accolades from the state and monetary bonuses for their staff. Schools failing to meet growth and performance standards receive help from assistance teams sent by the North Carolina's Department of Public Instruction (NCDPI). Ultimately, schools that do not improve will be taken over by the state. School results are sent home for parents, and school by school ABC results are reported to the public in *A Report Card for the ABCs of Public Education* made available on the NCDPI's website. After two consecutive years of low performances, the ABC Framework permits parents to transfer their children to a high-performing school from their low-performing school (Mintrop and Trujillo 2005).

Following the passage of NCLB in 2002, several components of the ABC Plan were modified or added to conform to the federal law (NCDPI 2011). The two assessments and standards programs remain complementary but not identical. For instance, the ABC rating can conflict with NCLB labels. Elementary and middle school EOG tests in reading and mathematics are used for both NCLB and ABCs rating, but results are calculated differently. NCLB looks at performance at the school level disaggregated by group; the ABC Plan assesses school wide performance averages in addition to growth scores. Therefore, it is possible for a school to meet its ABC goals but fail to make AYP according to NCLB's standards (Helms 2004).<sup>2</sup>

### Research Questions

Given the stated objective of NCLB (and the ABCs program) to close achievement gaps and the various reasons for skepticism about its ability to do so, we pose three research questions. First, do North Carolina students' scores on high-stakes tests indicate gaps in achievement associated with race or class? Ample evidence suggests a long history of racial/ethnic minority students and low-SES students, on average, earning lower scores on state tests than their White and high-SES counterparts, respectively. Second, if those gaps exist, do those gaps shrink as a cohort of North Carolina students progresses through school? If accountability policies are having the desired outcome, we should see that the gaps between

---

<sup>1</sup> In the 2012-2013 school year, North Carolina cut back its EOC tests to one science, one math, and one English course.

<sup>2</sup> Discrepancies between state and NCLB proficiency results occur elsewhere across the nation (Hornbeck 2013, Linn 2005, Pinzur 2003).

Black and White and between rich and poor shrink as the pressure on educators to meet standards inspires them to change their behavior to boost the achievement of chronically underperforming groups. Third, if gaps persist or increase as a North Carolina cohort moves through school, does a difference in opportunities to learn explain that difference? The third question raises the issue of “why”—that is, what are the likely mechanisms that contribute to the failure of the current approach, as reflected in the NCLB and ABCs programs, to improve learning outcomes for all children and to closing the achievement gap. Decades of research suggest a multiplicity of individual, family, and school factors are responsible for the persistence of the racial achievement gap. In this study, we examine the contribution of one aspect of the *organization* of teaching and learning: academic tracking. A well-established line of research has shown the experiences of students in higher-level and lower-level academic tracks are quite different, with honors classes engaging in higher-order thinking activities such as synthesis and evaluation and lower tracks focusing on individual seat work and memorization. A growing body of scholarship finds that NCLB may be exacerbating that difference, with lower track classes also emphasizing basic test preparation lessons at the expense of richer curricular and instructional activities involving higher-order thinking (Dorn 2010, Giersch 2012, Nichols and Berliner 2010, Watanabe 2008).

### Design of the Study

We answer these questions using a longitudinal dataset from North Carolina. North Carolina data are an excellent source for examining the questions that this study investigates. First, North Carolina employed curriculum standards and standardized tests for several years prior to the implementation of NCLB in 2002. This means the available data offer a six-year window from 1998 through 2004 during which the essential dynamic of standards and accountability through high stakes testing was operating. Second, North Carolina resembles the United States in terms of student achievement, student-teacher ratio, and education expenditure. In 2004, North Carolina also resembled the United States in terms of racial diversity and SES. In 2005, North Carolina was 68% White while the United States was 67% White.<sup>3</sup> In the same year, 13% of North Carolina residents lived in poverty, matching the rate for the United States overall (Hovey and Hovey 2007). These similarities make North Carolina’s population useful in drawing conclusions that may be suggestive of national trends.

---

<sup>3</sup> While the majority population in North Carolina resembles the United States, the minority population is disproportionately black. Although the Latino/a population of North Carolina is growing rapidly, in 1998 the baseline year for data used in this study the number of Latino/a students was relatively small.

## Sample

The sample used for this research includes North Carolina's entire 2004 high school graduating class (more than 76,000 students). This was one of the state's first cohorts of students to take high-stakes tests in both middle and high school. Our data include students' test scores, academic tracks, and family background information as well as data about the schools they attended from middle school through high school. These data are part of the North Carolina Roots of STEM dataset.<sup>4</sup> The Roots of STEM project seeks to understand what factors increase the likelihood that women and minorities will declare and complete majors in the science, technology, engineering, or math (STEM) disciplines. This study is part of that larger project.

The analysis in this study focuses on student-level variables, but the statistical models control for the effects of high schools to account for variance that may be associated with attendance at a particular high school (Laird and Ware 1982). This approach is important because if we do not control for the effects of different high schools on student test scores, we cannot accurately assess the influence of student characteristics on achievement. Table 1 provides an overview of the student-level variables. Most measures we employ are conventionally found in social science research about educational outcomes, but a few constructs deserve some explanation.

We wanted to capture two aspects of socioeconomic status (SES), both parent education and income. The link between parent education (a variable included in the dataset) and SES is fairly well established in the literature (Davis-Kean 2005, Lareau 2003, Magnuson 2007). The Roots of STEM data also included a variable for participation in the federal Free and Reduced Lunch (FRL) program while in middle school.<sup>5</sup> While FRL is far from a perfect indicator of SES, it remains widely used in education research (Harwell and LeBeau 2010). For this reason we used these two measures to create a low-SES variable that is coded 1 for any student who received free or reduced lunch in middle school or had parents without high school diplomas.

---

<sup>4</sup> This article utilizes the unique Roots of STEM dataset (<http://clas-pages.uncc.edu/rootsofstem/blog/tag/roots-of-stem/>) that includes indicators of individual, family, community characteristics, and secondary school and college experiences among the population of 2004 high school graduates from the state of North Carolina (NC) who entered one of the 16 campuses of the University of North Carolina system. This article is drawn from Jason Giersch's doctoral dissertation, "Standardized Distraction: Why the Emphasis on High-Stakes Testing Can't Resolve Educational Inequality" (2012).

<sup>5</sup> Our data do not include information about free or reduced lunch in high school.

**Table 1.** Core concepts, variable names, and the operationalization of variables.

<b>Core Terms</b>	<b>Variable</b>	<b>Operationalization</b>
High School Achievement	EOC Scores	Mean score of student's required EOC tests (algebra I, geometry, algebra II, ELP, US history, English I, biology)
Opportunities to Learn	Track Placement	Proportion of EOC courses taken as honors, AP, or IB
Middle School Achievement	EOG Scores	Mean scale scores for math and reading EOGs
Student Background	Race	Dummy variables for each
	Sex	Dummy variable in which male=1
	Low SES	Dummy variable in which 1 indicates either free/reduced lunch while in middle school or parents did not graduate high school
	Limited English Ability	Dummy variable in which 1= limited English skills

The dependent variables in this study are measures of mean EOC scores and mean EOG scores. Although North Carolina's requirements for EOC tests are always in flux, at the time of this cohort's high school education students in the Core curriculum were required to take high-stakes tests in Algebra I, Geometry, Algebra II, ELP (Economics, Law, and Politics), U.S. History, English I, and Biology. Middle school students took reading comprehension and mathematics EOG tests in both 7<sup>th</sup> and 8<sup>th</sup> grades. All of these tests follow a multiple choice format. North Carolina's Department of Public Instruction converts each student's raw score into a scale score that can be compared to the student's scale score on the same exam in previous (and later) years. Scale scores are reported to parents, along with a percentile rank score and an achievement level, the latter of which fits into one of four pre-determined performance benchmarks. Each item on the EOG and EOC tests is directly tied to the goals found in the state standards. For example, Goal 5 of North Carolina's 8<sup>th</sup> grade language arts curriculum is for students to "use interpretive and evaluative processes to analyze texts and their characteristics" (NCDPI 2012). On the EOG, students needed to read a passage from "The Final Memo," a short story by Paul Stewart, and answer five multiple choice questions about characters, their personalities, and word usage in the reading.

To create the mean EOG variable, we averaged students' reading and math EOG scores. To create the mean EOC variable, we averaged students' required EOC test scores, including Algebra 1, Geometry, Algebra 2, English 1, ELP, U.S. History, and Biology. Table 4 shows that the EOG math and reading scores are scaled to similar means, ranges, and standard deviations. Likewise, the EOC tests are scaled to similar means, ranges, and standard deviations.

The EOC data also offer ways to identify students’ academic tracks. All EOC courses have an indicator of course level. From these variables we created a proportion honors classes variable by taking the number of EOC-tested honors-, AP-, or IB-level EOC courses a student had taken among and dividing it by the total number of EOC courses they had taken. Although this method only incorporates a small portion of the high school career, it covers multiple disciplines and the most important courses, if only because accountability policies give them such importance. The role of high-stakes tests also makes this variable a conservative estimate of the effects of tracking generally. Because they are state-tested, these courses are especially tied to the core curriculum, so any differences we see as a result of tracking are probably tied to differences in how they are taught and not due to differences in content.

**Table 2.** Descriptive summaries for middle and high school analyses.

<b>Variable</b>	<b>Observations</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
<i>Dependent Variables</i>					
Mean EOG score	70,710	166.609	8.316	133.5	191
Mean EOC score	70,710	57.948	8.310	21	87.0
<i>Independent Variables</i>					
Male	70,710	0.489	0.500	0	1
White	70,710	0.666	0.472	0	1
Black	70,710	0.270	0.444	0	1
Hispanic	70,710	0.020	0.141	0	1
Asian	70,710	0.018	0.132	0	1
American Indian	70,710	0.016	0.124	0	1
Other race	70,710	0.000	0.018	0	1
Low SES	70,710	0.323	0.467	0	1
Proportion honors classes	70,710	0.220	0.310	0	1

#### Analytic Steps

The analyses follow tens of thousands of individual North Carolina students through middle school and high school. We use several multilevel models that test the effects of student background and prior achievement on academic performance while controlling for the effects of track placement during high schools. These models allow the effects of individual characteristics to be separated from the effects of schools. We employ the software package Stata (Rabe-Hesketh and Skrondal 2012) and its xtmixed procedure to run multilevel regression models. To account for possible collinearity problems in the analysis, we centered the variables by their grand means by subtracting the mean value of the observations in the analytic sample from each observation.

## Results

The first question of this study asks if students from minority and low-SES families score lower than those from White and more prosperous backgrounds. Table 3 addresses this question in terms of high-stakes test scores through the results of multilevel regression models predicting mean EOG and EOC scale scores only by student background factors. Students are clustered by high school to account for school effects. Both models show that low-SES students score, on average, about 3.5 scale-score points lower than other students on both the seventh grade EOG tests and high school EOC tests. Males score lower than females by about a half-point (-.522\*\*\*) in middle school and a third of a point (-.332\*\*\*) in high school. Whites (the reference category for race) tend to score higher than Blacks in middle school (-5.165\*\*\*) and high school (-4.462\*\*\*), Hispanics in middle school (-.616\*\*\*), and American Indian students in middle school (-3.395\*\*\*) and high school (-3.279\*\*\*), but lower than Asian students in middle school (1.486\*\*\*) and high school (2.901\*\*\*). None of the results is unexpected given prior research on the race gaps in performance. The gap is widest between Blacks and Asians. The performance of Asians and Hispanics is higher, relative to other groups, on the high school tests than on the middle school tests. One explanation for their improved performance could be the substantial number of immigrants within those groups who learn English as they progress through the grades. Although the limited English variable accounts for those students who qualified for language services, becoming acclimated to America's culture and public schools likely accounts for some of the improvement among Asians and Hispanics. Overall, the relationships between student background and middle school tests (model 1) are similar to those between student background and high school tests (model 2). These results suggest that achievement gaps were fairly static for this cohort between middle and high school.

Table 4 uses middle school test scores as a predictor variable for high school scores. Logically, a student who does well in middle school will also do well in high school. The first model in Table 4 confirms that this is true. For each point improvement a student made above average on the middle school EOG tests, he or she could expect an additional eight tenths of a point (.806\*\*\*) on the high school EOC tests. Model 2 adds the student background variables to that model. The coefficient for the EOG tests becomes slightly smaller as a result (.787\*\*\*), confirming that background indeed plays a role in a student's continued success from middle school into high school. Relative to females, males' scores, on average, improve nearly a tenth of a point as they transition from middle to high school. Relative to Whites, Blacks (-.411\*\*\*) and Native Americans (-.618\*\*\*) do worse while Asians (1.806\*\*\*) and Hispanics (.397\*\*\*) do better. Students who are low-SES show a lower average high school test score compared to their counterparts and controlling for middle school achievement.

**Table 3.** Student Characteristics and the Prediction of EOG and EOC Test Scores.

	Model 1: Predicting EOG scores		Model 2: Predicting EOC scores	
	Estimate	S.E.	Estimate	S.E.
Low SES	-3.543***	0.066	-3.439***	0.065
Male	-0.522***	0.054	-0.332***	0.052
Black	-5.165***	0.072	-4.462***	0.070
American Indian	-3.395***	0.240	-3.279***	0.234
Hispanic	-0.616**	0.210	-0.087	0.204
Asian	1.486***	0.213	2.901***	0.207
Other race	2.002	1.452	3.052*	1.413
Limited English	-6.381***	0.290	-4.644***	0.282
Constant	170.265***	0.184	62.264***	0.200
Log-likelihood	-239283.0		-237368.1	
N	70710		70710	

\* significant at .05

\*\* significant at .01

\*\*\* significant at .001

**Table 4.** Student Characteristics, Prior Achievement, and Tracking EOG and the Prediction of EOC Test Scores.

	Model 1		Model 2		Model 3	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Mean EOG score	.806***	.002	.787***	.002	.716***	.002
Male			.087**	.031	.194***	.030
Black			-.411***	.043	-.483***	.042
American Indian			-.618***	.122	-.699***	.135
Asian			1.744***	.122	1.473***	.120
Hispanic			.397***	.043	.402***	.118
Other race			1.470	.834	1.453	.815
Limited English			.369*	.167	.270	.163
Low SES			-.666***	.039	-.509***	.038
Proportion of classes taken as honors classes					3.632***	.062
Constant	58.23***	.089	58.267***	.085	58.033***	.080
Log-likelihood	-200462		-200049		-198355	
N	70710		70710		70710	

\* significant at .05

\*\* significant at .01

\*\*\* significant at .001

Model 3 adds one more variable to the mix: academic track. The result is a small decline in the effect of previous achievement (from .787\*\*\* to .716\*\*\*). Although the academic track variable has a mean, variance, and range that is quite similar to the other background variables, its coefficient is much larger than the others, attesting to the academic benefits of enrolling in upper-level class. In this model, an honors-track student will outscore a non-honors track student by more than three points (3.632\*\*\*), even when middle school test scores, race, and SES are all held to their means. This suggests that students in honors tracks learn more than their otherwise comparable peers in lower tracks.

Accounting for academic track, however, does not explain away differences associated with race and SES. In fact, when academic track is held to the mean, the negative effects of being Black (-.483\*\*\*) or Hispanic (-.699\*\*\*) become slightly stronger. The gain that boys (.194\*\*\*) make on girls between middle and high school tests also becomes stronger when track is held constant. A VIF test confirmed that these results were not corrupted by multicollinearity.

Three important findings come from these results. First, we have confirmed that student background continues to play a statistically significant role in the performance of students on high-stakes tests, both in middle school and in high school. Second, we find that while on the path from middle school test scores to high school test scores, students will follow a slightly different trajectory depending on their gender, race, and SES. Third, while academic track does not account for the differences associated with student background, it does play a significant role in standardized test scores, and thus deserves more attention than it often receives.

Enrolling in honors classes clearly has benefits for student outcomes, and while such higher track classes are in theory open to all students, not all students enroll in them. We now investigate what roles student background and middle school test scores play in high school track placements. The answer to this question is critical to narrowing the race gap in achievement because prior research suggests that test scores are used for placement decisions.

Our findings in this study offer further evidence that high-stakes test results in middle school lead to the sorting of students by race and SES into different academic tracks. Table 5 addresses this issue first with a model that predicts proportion of honors classes using student background data. Model 1 shows that males (-.040\*\*\*) take fewer honors classes than females, Blacks (-.080\*\*\*) and American Indians (-.044\*\*\*) take fewer honors classes than Whites who take fewer than Asians (.103\*\*\*). Students of low socio-economic status (-.111\*\*\*) take fewer honors classes than their counterparts. Model 2 adds middle school EOG test performance to the equation. Not surprisingly, doing well on the high-stakes test (.019\*\*\*) is associated with taking more honors classes in high school. And while the background variables that were statistically significant in the first model retain their signs and significance in the second model, their

coefficients each lose magnitude. By simply controlling for middle school test scores, more than 75% of the effects of being Black on taking honors classes are accounted for, half of the effect of being American Indian is accounted for, a quarter of the effect of being Asian is accounted for, and nearly two-thirds of the effect of being low-SES is accounted for. In other words, the segregation of students by race and SES into different high school tracks is in no small part associated with their scores on middle school high-stakes tests. Given the significant effect of track placement on high school test scores, and the role of EOG tests in high school track placements, the analyses reveal at least one way that NCLB’s mandate for standardized testing contributes to the persistence of the race gaps in achievement.

**Table 5.** Student Characteristics, Middle School Prior Achievement, and the Prediction of High School Track Placement.

	<b>Model 1</b>		<b>Model 2</b>	
	<b>Estimate</b>	<b>S.E.</b>	<b>Estimate</b>	<b>S.E.</b>
Male	-.040***	.002	-.030**	.002
Black	-.080***	.003	-.019***	.003
American Indian	-.044***	.009	-.022**	.008
Asian	.103***	.008	.074***	.007
Hispanic	-.015	.008	.003	.007
Other race	.038	.057	-.000	.049
Limited English	-.094***	.011	.029**	.010
Low SES	-.111***	.003	-.043***	.002
Mean EOG score			.019***	.000
Constant	.363***	.011	.267***	.009
Log-likelihood	-10,207		-454	
N	70,710		70,710	

\* *significant at .05*

\*\* *significant at .01*

\*\*\* *significant at .001*

### Discussion

No Child Left Behind mandates standards, testing, and accountability to improve student outcomes and narrow race and SES gaps in achievement. Based on market theories, the explicit theory of change advanced by the designers of high stakes tests is that sanctions and incentives will alter educators’ behaviors in ways that boost performance among previously low-achieving students. This expectation hinges upon the presumption that without high-stakes tests the low-achieving students and their teachers simply did not have enough motivation to learn or teach.

The findings of this study offer no evidence that NCLB's theory of reform has been successful in North Carolina.<sup>6</sup> In fact, we offer ample evidence to make the opposite case. Our findings show that not only did testing not close the race gaps, but also to the extent that middle school standardized test scores were part of the practice of assigning students to high school tracks, they contributed to maintaining them. Even controlling for other factors, students' EOG standardized test scores influenced their high school track placements. Those who did well on their middle school EOG test were more likely to be placed in higher tracks for the rest of their high school careers. As a result of widespread tracking in North Carolina, males, Blacks, and low-income youth are much less likely to learn in college prep tracks than otherwise comparable females, Whites, and Asian youth. Not surprisingly, the findings show that students in higher tracks score better on their EOC tests, therein reflecting one of the organizational mechanisms that contribute to the perpetuation of the notorious race gaps.

Our findings show the high-stakes accountability policy in place in North Carolina has not closed the achievement gaps. Using the policy's own measures of achievement, we find that in making the transition from middle school to high school, Asians extend their academic lead over White students, and Hispanics get slightly closer to the average score of Whites. Blacks and Native Americans, however, see their gap with White and Asian peers grow larger. The gap associated with SES also increases. Boys, while scoring lower than girls in both middle and high school, do close the gap slightly in high school, although it is also likely that the weakest students have dropped out (Stearns and Glennie 2006).

The fact that lower-track, Black, Native American, and low-SES students all did worse on their high school tests than their middle school tests is cause for alarm. Of even greater concern is the fact that these gaps in achievement are apparent on the very instruments that their teachers and schools have been focusing on for weeks as they prepared for exams. If disadvantaged students are falling behind on these measures, students whose daily classwork heavily emphasized test preparation, one can only imagine on what other areas of learning these students are being left even further behind.

Finally, the effects of high-stakes tests on academic tracking cannot be ignored. If test scores are being used to sort students into academic tracks and different tracks offer different opportunities to learn, then high-stakes tests are assisting in the stratification and segregation of opportunities to learn within schools, not fostering the equity advocates of NCLB promised. This

---

<sup>6</sup> Lauen and Gaddis (2013) examined test scores for multiple cohorts of North Carolina 3<sup>rd</sup> through 8<sup>th</sup> graders from 2002 to 2009 to assess NCLB's effects on closing SES gaps in outcomes. They report NCLB lead to a reduction in the SES achievement gap. They find that poor students benefit from NCLB subgroup-specific accountability pressure, while non-poor students suffer no significant adverse effects. The disparities in findings between this article and their findings may be due to the different age of their sample (elementary through middle school while we examined middle school through high school outcomes), the different outcome variable (EOG compared to our EOC score), the specific gap they examined (SES compared to our examination of race gap), and the later years of their study (2002 to 2009 compared to our 1998 through 2006 window).

pattern is particularly bothersome in light of the fact that standardized tests are poor instruments for measuring higher-order thinking skills, which honors classes tend to emphasize.

### Limitations

As with all research, there are several limitations to this study. One is its use of high-stakes tests to assess the academic progress of students. For reasons outlined in the beginning of this article, standardized test scores are limited in their usefulness, and attaching *any* high stakes to them as we have done in using them to assess NCLB's role in the race gap's persistence is potentially problematic. We employ them in this study for other reasons. One is that the data are readily available in this form and for better or worse, standardized tests remain the 'coin of the realm.' The other reason is their importance to the prevailing education policies under US and NC law. Whether they actually measure student learning in a rigorous and complete manner, EOGs and EOCs are being used as though they do, and as such the scores are being used to calculate grades, assess progress, compare performance, and evaluate teachers and students. Because of the importance attached to these test scores, we feel it is important to understand the factors that influence them.

The second limitation to this study is its reliance on statistical calculations from administrative data rather than qualitative investigations that capture and describe the mechanisms that influence student achievement on tests or their selection into academic tracks. Past research tells us that student performance influences decisions about tracking and that tracking in turn influences student performance, but we have not conducted the extensive interviews or observations that would be required to strengthen or explicate the argument. Nonetheless, we are confident that EOGs contribute heavily to high school track placements and that track plays a role in EOC scores.

### Conclusion

Market-inspired school reforms including NCLB have been implemented in North Carolina for almost two decades in order to attain greater equity and excellence in public school outcomes. If North Carolina is any indication of national trends, NCLB will not be the policy solution that raises the performance of chronically low-performing groups. NCLB does not introduce any new pedagogical techniques to educators; rather than adding resources, it threatened to take them away in low performing schools, along with sanctions on administrators and district authorities. Its fundamental premise is that low achievers have been capable of performing at the same level as their peers, but they lacked sufficient incentive to do so, or at least their teachers lacked sufficient motivation to teach them well. Differences in resources, neighborhoods, communities, and social contexts are irrelevant to the NCLB strategy; the emphasis is on getting every single student to "proficiency" on standardized tests. Whether

proficiency on standardized tests is a worthwhile goal is a long discussion best left for another time.

After NCLB's implementation, all states and the District of Columbia created standards-based, high-stakes tests to be administered to students annually. Each year, states roll out more tests on more subjects. Results are publicized and schools, districts, and in some cases even teachers are judged, both formally and informally, on the outcomes. As predicted, educators scramble to find new, and in many cases illegal, ways to boost scores. Witness the high profile school districts faced with cheating scandals: Houston, Atlanta, and Washington, DC, for example. The scores on high-stakes tests increased in many states, but increased dropout rates and stagnant NAEP scores often accompanied those increases. The frequent mismatch between proficiency levels on a state's NAEP scores and proficiency levels on state standards is emblematic of the perversity of using standardized test scores as measures of learning.

This study corroborates the findings of other studies of inequalities associated with accountability policies in North Carolina (Watanabe 2008), Texas (Valenzuela 2005), and elsewhere (Nichols and Berliner 2007, Perna and Thomas 2009, Ravitch 2010) and provides new information about the drawbacks to NCLB. The findings show that race and SES gaps in student test scores remain, that the gaps do not substantially shrink as students progress through secondary school, and that the organization of teaching and learning in secondary schools—academic tracking—contributes to the failure of standards-based reforms to foster either educational equity or excellence in North Carolina.

Rather than contribute to achieving the twin goals of educational excellence and equity, our study suggests NCLB's focus on standardized testing and accountability harms the learning and teaching process. It narrows curricula and distorts teaching especially for low-performing students. Perhaps the most serious consequence of NCLB's theory of reform is that it ignores and obscures other school-based sources of inequality that include race and SES-linked access to teacher quality, continued segregation of schools along lines of race and class, ability grouping and curricular tracking, and resource inequality among the schools. Furthermore, NCLB's premise that claims competition increases performance is one that has no scientific warrant in the hundreds of empirical studies on the topic.

In contrast to theory, this study's empirical findings are relevant to reauthorization of No Child Left Behind. They reveal some of the problems inherent in an education system driven by competition. The first is that competition doesn't change anything. Bad teachers are still bad teachers. Poor schools are still poor schools. And students who are labeled as less capable are placed in classes with lower expectations, fewer learning opportunities, and diminished hopes for future success. While a standardized test may sound meritocratic and objective, the field on which the competition is held is far from level. Second, competition shifts the emphasis of

education away from authentic assessments of student growth, learning, and development and shifts it to a score on a multiple-choice test. For some students and schools, it is easy to score well on a test while also engaging in all the academic pursuits that will prepare someone for further education and a career. For other students, merely passing the exam requires that they engage in repetitive, lower-order mental tasks in a seemingly endless preparation and testing cycle.

The final, and perhaps most important, concern is how we treat the winners and losers of the competition. NCLB rewards winners and punishes losers at each stage of education. If students are being told beginning in elementary school that their abilities, their teachers, and their schools are below average, how should we expect them to respond? Under the guiding assumptions of NCLB, those students will be driven by the desire to have a better score next time, and if they fail in middle school they will try even harder in high school. Meanwhile, their best teachers have moved on to greener pastures, where students are likely to make “proficiency” and salaries are more competitive. Their parents have lowered their expectations. Their schools have seen their reputations drop along with nearby property values as families who have the means to do so flee to districts and schools with higher scores. And based on the results of this study, it is a safe bet that the child who does poorly in the early years will continue to do so in the later years.

## References

- Ali, R. (2004, August). Keynote remarks to sociology of education section, No Child Left Behind Conference. Meetings of the American Sociological Association. San Francisco, CA.
- Anyon, Jean. 1997. *Ghetto Schooling: A Political Economy of Urban Educational Reform*. Teachers College Press.
- Apple, Michael W. 2001. "Markets, Standards, Teaching, and Teacher Education." *Journal of Teacher Education* 52 (3) (May 1): 182–196. doi:10.1177/0022487101052003002.
- Berliner, David. 2011. "Rational Responses to High Stakes Testing: The Case of Curriculum Narrowing and the Harm That Follows." *Cambridge Journal of Education* 41 (3): 287–302. doi:10.1080/0305764X.2011.607151.
- Borman, Geoffrey D., Sam Stringfield, and Robert E. Slavin. 2001. *Title I: Compensatory Education at the Crossroads*. Taylor & Francis.
- Brown v. Board of Education*, 347 U.S. 483 (1954).
- Bush, George W. "No child left behind." (2001): 33-40.
- Campbell, Donald T. 1979. "Assessing the Impact of Planned Social Change." *Evaluation and Program Planning* 2 (1): 67–90. doi:10.1016/0149-7189(79)90048-X.
- Chubb, John E., and Terry M. Moe. 1990. *Politics, Markets, and America's Schools*. Brookings Institution Press.
- Cross, Christopher T. 2004. *Political Education: National Policy Comes of Age*. Teachers College Press.
- Darling-Hammond, Linda, Jacqueline Ancess, and Beverly Falk. 1995. *Authentic Assessment in Action: Studies of Schools and Students at Work*. Teachers College Press.
- Davis-Kean, Pamela E. 2005. "The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment." *Journal of Family Psychology* 19 (2): 294–304.
- Ehrenreich, Bárbara. 1989. *Fear of Falling: The Inner Life of the Middle Class*. Pantheon Books.
- Fiske, Edward B., and Helen F. Ladd. 2000. *When Schools Compete: A Cautionary Tale*. Brookings Institution Press.
- Garces E., Thomas D., and Currie J. 2002. "Longer-Term Effects of Head Start." *The American Economic Review* 92 (4): 999–1012.

- Gardner, David P. "A nation at risk." *Washington, D. C.: The National Commission on Excellence in Education, US Department of Education* (1983).
- Giersch, Jason. "Standardized distraction: Why the emphasis on high-stakes testing can't resolve educational inequality." Doctoral dissertation, Public Policy Program, UNC Charlotte, 2012.
- Godwin, R. Kenneth, and Frank R. Kemerer. 2002. *School Choice Tradeoffs: Liberty, Equity, and Diversity*. University of Texas Press.
- Harwell, Michael, and Brandon LeBeau. 2010. "Student Eligibility for a Free Lunch as an SES Measure in Education Research." *Educational Researcher* 39 (2) (March 1): 120–131. doi:10.3102/0013189X10362578.
- Heckman, Diane. 1992. "Women & (and) Athletics: A Twenty Year Retrospective on Title IX." *University of Miami Entertainment And Sports Law Review* 9: 1.
- Henig, Jeffrey R. 1995. *Rethinking School Choice: Limits of the Market Metaphor*. Princeton University Press.
- Heubert, Jay Philip, and Robert Mason Hauser. 1999. *High Stakes: Testing for Tracking, Promotion, and Graduation*. National Academies Press.
- Hornbeck, Joe. 2013. State Level Cheating on High Stakes Assessments. A Policy Study of NCLB. Unpublished doctoral dissertation. Department of Educational Leadership. University of Kansas.
- Hovey, Kendra A., and Harold A. Hovey. 2007. *CQ's State Fact Finder: Rankings Across America*. CQ Press.
- Jackson, C. Kirabo. 2009. "Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation." *Journal of Labor Economics* 27 (2) (April): 213–256. doi:10.1086/599334.
- Jones, M. Gail, Brett D. Jones, and Tracy Y. Hargrove. 2003. *The Unintended Consequences of High-Stakes Testing*. Rowman & Littlefield.
- Kennedy, Edward M. 2005. "No Child Left Behind Act - Fulfilling the Promise, The." *Human Rights* 32: 16.
- Laird, N. M, and J. H. Ware. 1982. "Random-effects Models for Longitudinal Data." *Biometrics* 38 (4) (December): 963–974.

- Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2002. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." *Educational Evaluation and Policy Analysis* 24 (1) (March 20): 37–62. doi:10.3102/01623737024001037.
- Lareau, Annette. 2003. *Unequal Childhoods: Class, Race, and Family Life*. University of California Press.
- Lauen, Douglas, and S. Michael Gaddis. 2013. "Can Schools Close SES Achievement Gaps? The Role of Accountability, Resources, and Teacher Quality." Paper presented at the meetings of the American Educational Research Association, San Francisco, April.
- Lee, Jaekyung. 2002. "Racial and Ethnic Achievement Gap Trends: Reversing the Progress Toward Equity?" *Educational Researcher* 31 (1) (January 1): 3–12. doi:10.3102/0013189X031001003.
- Levin, Henry. 2001. "Waiting for Godot: Cost-Effectiveness Analysis in Education." *New Directions for Evaluation* 2001 (90): 55–68. doi:10.1002/ev.12.
- Linn, Robert L. 2005. "Conflicting Demands of No Child Left Behind and State Systems: Mixed Messages About School Performance." *Education Policy Analysis Archives* 13 (33) (June 28). <http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ846738>.
- Magnuson, Katherine. 2007. "Maternal Education and Children's Academic Achievement During Middle Childhood." *Developmental Psychology* 43 (6) (November): 1497–1512.
- Manna, Paul. 2006. *School's In: Federalism and the National Education Agenda*. Georgetown University Press.
- McDermott, Kathryn A. 2007. "'Expanding the Moral Community' or 'Blaming the Victim'? The Politics of State Education Accountability Policy." *American Educational Research Journal* 44 (1) (March 1): 77–111. doi:10.3102/0002831206299010.
- McEwan, Patrick J., and Martin Carnoy. 2000. "The Effectiveness and Efficiency of Private Schools in Chile's Voucher System." *Educational Evaluation and Policy Analysis* 22 (3): 213–239. doi:10.3102/01623737022003213.
- Medina, Noe, and D. Monty Neill. 1988. "Fallout from the Testing Explosion: How 100 Million Standardized Exams Undermine Equity and Excellence in America's Public Schools." FairTest, Cambridge, MA.
- Mickelson, Roslyn Arlin. 1999. "International Business Machinations: A Case Study of Corporate Involvement in Local Education Reform." *Teachers College Record* 100 (3): 476–512.

- Mickelson, Roslyn Arlin, Stephen S. Smith, and Stephanie Southworth. 2007. "The Chimera of Choice. Post-unitary Charlotte's Rapid Resegregation." In *Court-Ended Desegregation: The Reconstitution of Schools and Communities*.
- Mintrop, Heinrich, and Tina Trujillo. 2005. "Corrective Action in Low Performing Schools: Lessons for NCLB Implementation from First-Generation Accountability Systems." *Education Policy Analysis Archives* 13 (48) (December 1): 1–30.
- Nichols, Sharon Lynn, and David C. Berliner. 2007. *Collateral Damage: How High-stakes Testing Corrupts America's Schools*. Harvard Education Press.
- North Carolina Department of Public Instruction. Evolution of the ABCs. Available at <http://www.ncpublicschools.org/docs/accountability/reporting/abc/2009-10/abcevolution.pdf>. Accessed 4/30/2013.
- North Carolina Department of Public Instruction. K-12 Curriculum and Instruction/NC Standard Course of Study. Available at <http://www.ncpublicschools.org/curriculum/>. Accessed 4/30/2013.
- Oakes, Jeannie. 1986. *Keeping Track: How Schools Structure Inequality*. Yale University Press.
- Perna, Laura W, and Scott L Thomas. 2009. "Barriers to College Opportunity The Unintended Consequences of State-Mandated Testing." *Educational Policy* 23 (3) (May 1): 451–479. doi:10.1177/0895904807312470.
- Petrovich, Janice, and Amy Stuart Wells. 2005. *Bringing Equity Back: Research for a New Era in American Educational Policy*. Teachers College Press.
- Pinzur, M. (2003, August 8). State schools fail to meet new federal test standards: Federal, state results differ. *Miami Herald*.
- Powers, Jeanne M., and Peter W. Cookson. 1999. "The Politics of School Choice Research: Fact, Fiction, and Statistics." *Educational Policy* 13 (1) (January 1): 104–122. doi:10.1177/0895904899131009.
- Rabe-Hesketh, Sophia, and Anders Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata*, 3rd Edition. Stata Press.
- Ravitch, Diane. 2010. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. Basic Books.
- Rebell, Michael A., and Jessica R. Wolff. 2009. *NCLB at the Crossroads: Reexamining the Federal Effort to Close the Achievement Gap*. Teachers College, Columbia University.

- Roderick, Melissa, and Mimi Engel. 2001. "The Grasshopper and the Ant: Motivational Responses of Low-Achieving Students to High-Stakes Testing." *Educational Evaluation and Policy Analysis* 23 (3) (September 21): 197–227. doi:10.3102/01623737023003197.
- Smith, Mary Lee, and Linda Miller-Kahn. 2004. *Political Spectacle and the Fate of American Schools*. Psychology Press.
- Stearns, Elizabeth and Elizabeth J. Glennie. "When and Why Dropouts Leave School." *Youth and Society*. 38: 29-57.
- Taylor, William. 2003. "Title I as an Instrument for Achieving Desegregation and Equal Educational Opportunity." *North Carolina Law Review*. 81 (4): 1751-1769.
- Troike, Rudolph C. 1978. *Research Evidence for the Effectiveness of Bilingual Education*. National Clearinghouse for Bilingual Education,
- Valenzuela, Angela 2005. *Leaving Children Behind: How "Texas-Style" Accountability Fails Latino Youth*. SUNY Press.
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).
- Watanabe, Maika. 2008. "Tracking in the Era of High Stakes State Accountability Reform: Case Studies of Classroom Instruction in North Carolina." *Teachers College Record* 110 (3): 489–534.