PRODUCING COLLEGE-READY STUDENTS: THE PROMISE OF STANDARDS-BASED ACCOUNTABILITY, EVIDENCE FROM THE EDUCATION LONGITUDINAL STUDY OF 2002

A Thesis
submitted to the
Graduate School of Arts and Sciences
at Georgetown University
in partial fulfillment of the requirements for the
degree of
Master of Public Policy
in Public Policy

By

Anne F. Hyslop, B.A.

Washington, D.C.
April 15, 2011

PRODUCING COLLEGE-READY STUDENTS: THE PROMISE OF STANDARDS-BASED ACCOUNTABILITY, EVIDENCE FROM THE EDUCATION LONGITUDINAL STUDY OF 2002

Anne F. Hyslop, B.A.

Thesis Advisor: Thomas E. Wei, Ph.D.

ABSTRACT

Standards-based accountability reforms in education have been touted for several decades as necessary to improve student outcomes. Most existing research evaluating the impacts of standards-based reforms have focused on student test scores, with mixed findings. I instead focus on non-test score outcomes such as high school completion, college attendance, and college remediation in math, reading, and writing. In particular, I use the Education Longitudinal Study of 2002, a national study which follows students from their 10th grade year into college and their careers, to explore whether exposure to standards-based accountability systems is associated with improved student outcomes, and whether this association depends on the extent to which reforms have been implemented. I find that exposure to complete standards-based accountability systems is unrelated to student outcomes, and that contrary to theory, individual components of accountability systems can be related to student outcomes. Nevertheless, there does appear to be a limited, positive association between complete exposure to standards-based accountability and postsecondary reading and writing skills for low-income students.

ACKNOWLEDGEMENTS

I would like to dedicate this thesis to the many people who helped me throughout the research and writing process: first and foremost, my advisor, Tom Wei, whose assistance and guidance were invaluable; Carolyn Hill and Nora Gordon, for their expertise, understanding, and common sense when I needed it most; Amanda Jarrard and Sophie Kim, for sharing this experience with me and providing indispensable support and humor; and of course, my family and Stephen, whose contributions were the most important of all.

Anne F. Hyslop

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

I. Introduction

Since the 2002-2003 school year, the No Child Left Behind Act of 2001 (NCLB) has required states to enact standards-based accountability reforms in education. These reforms are built on the theory that student outcomes will improve if schools 1) adopt learning standards in core subjects, 2) measure progress through assessments that test students' mastery of the standards, and 3) hold schools accountable for all students and key subgroups achieving proficiency on the standards.

While standards-based accountability was institutionalized by NCLB, it is not a new idea. Published in 1983, *A Nation at Risk* embodied a decade of rising concerns that "the educational foundations of our society are…being eroded by a rising tide of mediocrity that threatens our very future" (National Commission on Excellence in Education 1983). The report noted that achievement on standardized tests had been declining and the performance of American students was falling behind international competitors like Japan and the Soviet Union (West and Peterson 2003). *A Nation at Risk* scrutinized the academic content schools were teaching, finding that "secondary school curricula have been homogenized, diluted, and diffused to the point that they no longer have a central purpose." Few schools were adequately preparing students for success in postsecondary education and on the job. For example, the report cited that the proportion of students enrolled in a general program of study, as opposed to specific vocational or college preparatory programs, had increased from 12 percent in 1964 to 42 percent in 1979.

The recommendations of *A Nation at Risk* called for states to adopt tougher standards, rigorous curriculum, and higher graduation requirements. This encouraged a number of reform-oriented governors to enact the first state-level standards-based accountability systems, focusing on student outcomes. These reforms were incorporated into President Clinton's agenda in the

1994 Improving America's Schools Act and Goals 2000 initiative, which called for schools to demonstrate annual student progress towards state standards. However, there was little enforcement of this provision and no sanctions for states that failed to comply, leading to uneven implementation until NCLB (Goertz, Duffy, and Le Floch 2001, McDermott and Jensen 2005, Hess and Petrilli 2006, Pew 2006, and Hamilton, Stecher, and Yuan 2008). Even after NCLB, wide variation remained between states; as the National Center for Education Statistics (NCES) reported in 2007, states differed along several dimensions including 1) when they adopted their standards, 2) the rigor of standards compared to national benchmarks, and 3) how well standards-based reforms were implemented in terms of providing support to staff in using the standards and the student-level data it generated. This variation between states continues today (Phillips 2010).

The Obama administration's *Blueprint for Reform* builds upon the foundation of standards-based accountability; specifically, President Obama has framed NCLB reauthorization around leading the world in college completion and raising expectations to "ensure that every student graduates from high school well prepared for college and a career" (Blueprint for Reform 2010). Additionally, 41 states have endorsed the Common Core State Standards Initiative (Common Core State Standards Initiative 2010). Common Core replaces existing state standards with a national set of academic standards in reading and math. Participation in Common Core was a prerequisite for states' applications to Race to the Top, a competitive grant program for states that are implementing ambitious and comprehensive reforms in education (Strauss 2010). Further signaling the Obama administration's support of Common Core, the President has proposed that states would need to adopt Common Core, or work with an institution of higher education to certify that their own standards are "college- and career-ready," to receive Title I funds, which are the main source of federal funding for education (McNeil 2010).

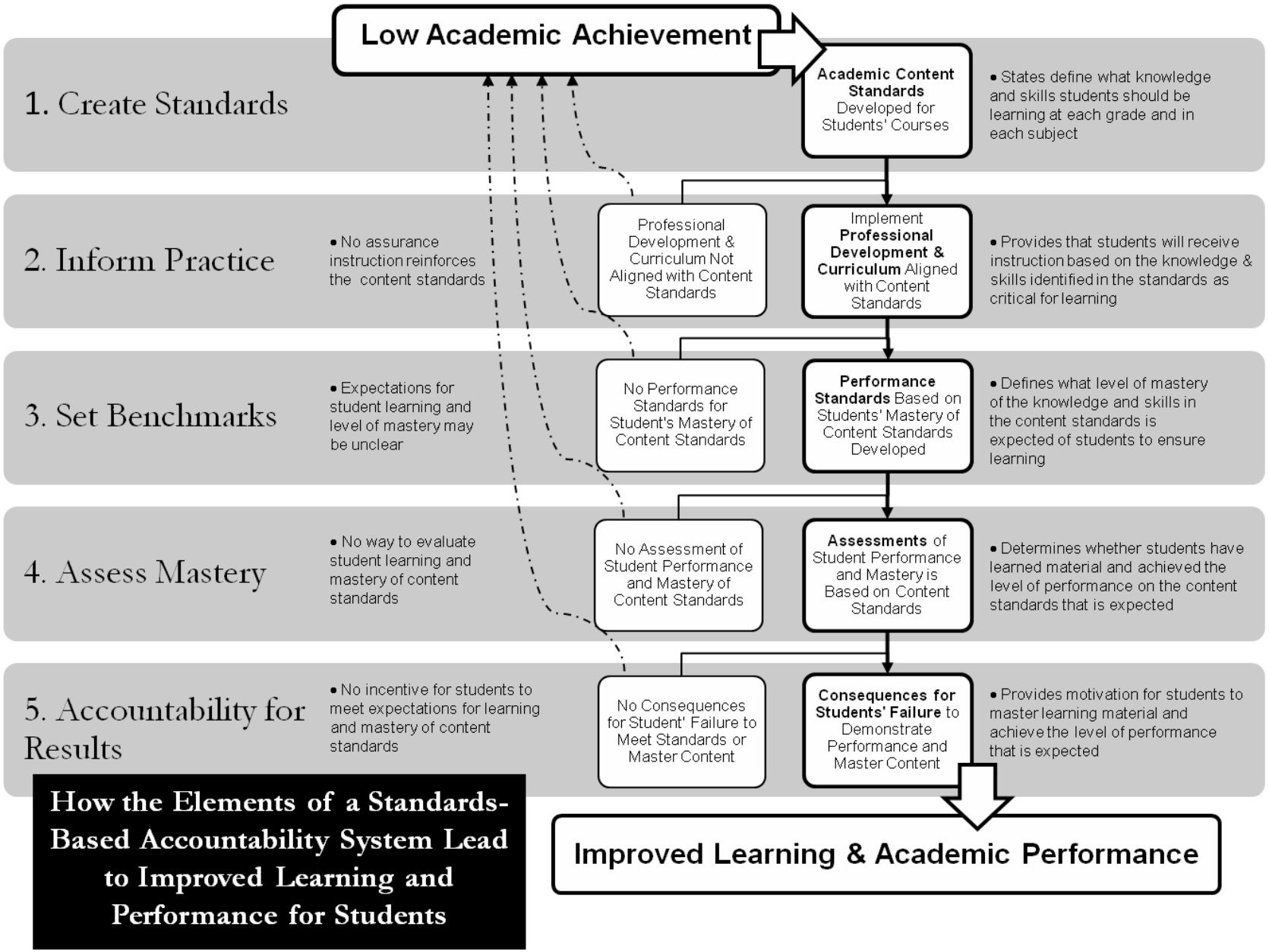
Despite the faith of education leaders in standards-based reforms and the billions of dollars allocated to these reforms, research has been mixed in whether the adoption of standards-based accountability has accomplished its purpose of improving student outcomes. As Grover Whitehurst (2010) has noted, "the lack of evidence that better content standards enhance student achievement is remarkable given the level of investment in this policy and high hopes attached to it." One reason for the lack of evidence is that until recently, data were not available except in a handful of states or districts. Additionally, since each state was allowed discretion in designing standards and assessments, there is wide variation between states, making direct comparisons of systems difficult. Finally, certain aspects of standards-based reforms are difficult to measure, particularly how they translate from classroom instruction into student outcomes. Whitehurst emphasizes this final shortcoming, describing standards as "such a leaky bucket with respect to classroom instruction that any potential relationship dissipates before it can be manifest."

As reauthorization of NCLB moves forward and the federal government considers providing more resources to states to refine and utilize information from their standards-based accountability systems, it is crucial to know what impact, if any, standards-based reforms have had on student outcomes. Therefore, this paper seeks to shed some light on the question: do standards-based reforms improve student outcomes? Specifically, does exposure to standards-based accountability lead to improved secondary completion and postsecondary outcomes? And does the effect of these reforms depend on the extent to which they have been implemented?

II. Theoretical Framework

Standards-based reform can be summarized with a theory of action (Figure 1). The theory links academic standards, assessments, and accountability to positive student outcomes (Hess

Figure 1. Theory of Action



and Petrilli 2006, Ladd 2007, Hamilton, Stecher, and Yuan 2008, Figlio and Ladd 2008, and Cohen and Moffitt 2009). Given *A Nation at Risk*'s emphasis on raising expectations, the logical first step for states was to define what these expectations should be – specifically, the knowledge and skills students are expected to master as they progress from grade to grade. These **content standards** then "become the focal point for changing other elements of the educational system" (Hamilton, Stecher, and Yuan 2008) through the alignment of subsequent layers of standards-based accountability with the content standards.

Once content standards are established, the next step is ensuring that teachers and staff are given **professional development, curriculum, and other instructional resources** that reflect the content standards, so that learning and activities inside the classroom emphasize the skills and knowledge students are expected to learn.

After states have created content standards and used the standards to inform the practice of their teachers and school leaders, the next element in a standards-based accountability system is for states to set **performance standards** – the specific benchmarks that indicate what level of achievement students are expected to attain on the content standards. It is unreasonable to expect student outcomes to improve if states do not first establish a) what students should be learning – the content standards – and b) how well they should be learning it – the performance standards.

Once aligned content and performance standards are established, the theory of action calls for **assessments aligned with the content and performance standards**. Appropriate assessments must be developed so states can determine whether students are proficient and demonstrate mastery of the established standards.

Even with well-established content standards aligned with instructional and professional development, performance standards, and student assessments, proponents of standards-based

reforms argue that **accountability measures** are needed to ensure that student outcomes will actually improve. Unless there are tangible consequences for students who fail to demonstrate mastery of the content standards, it is unlikely that the standards and assessments will be rigorously followed because there are fewer incentives to meet the higher standards. Accountability measures can involve direct consequences for students or indirect consequences for their teachers, principals, or schools.

The alignment of instructional activities with content standards is central to the theory of action for standards-based accountability. Proponents of standards-based accountability believe America has a fragmented and disjointed education system (Hess and Petrilli 2006, Ladd 2007) that sets unacceptably low expectations of students. Further, they argue that lack of strong incentives and accountability can inhibit schools from properly applying and focusing on their academic content and performance standards (Dee 2009). Strong accountability measures are needed to ensure that school districts focus on the content standards and what is being assessed. Because weak performance will trigger negative consequences for schools, stronger incentives exist for schools to examine their assessment data and consider altering how they deliver education to students (Figlio and Ladd 2008).

With a focused, aligned system that incorporates standards, assessments, and accountability, the theory of action explains how incentives could be utilized to improve student achievement. Likewise, if any element is out of alignment, the theory of action posits that student performance would likely not improve. When all of the elements of the accountability system have been implemented and aligned with content standards, proponents of the theory of action argue that students are the ultimate beneficiaries; they hypothesize that students would achieve at higher levels and graduate the K-12 system with a more complete knowledge base and set of

skills. The increase in achievement would then allow students to excel in college and in their careers.

III. Existing Empirical Research

Early studies of the effect of standards-based accountability on student outcomes examined one or several states that were early adopters of reforms (Dee 2003, Hess 2003, and Hanushek and Raymond 2003). These early adopters are unlikely to be nationally representative because standards-based accountability was not mandated until NCLB. Existing studies of standards-based accountability also typically have focused on the relationship between standards-based reforms and changes in short-term academic outcomes on standardized tests (Carnoy and Loeb 2002, Kane and Staiger 2003, Jacob 2005, Figlio and Rouse 2006, Springer 2008, and Ballou and Springer 2008). However, these studies do not track students' long-term academic outcomes or examine whether standards-based accountability improved student college readiness.

Figlio and Ladd (2008) provide a useful summary and review of the most relevant studies exploring the impact of standards-based accountability on student outcomes. They conclude that the existing evidence is mixed, but four general findings emerge:

1) Positive achievement effects "emerge far more clearly and frequently" in student math performance than in reading performance (Carnoy and Loeb 2002, Jacob 2005, Cronin 2005, Figlio and Rouse 2006, and Dee and Jacob 2009);

2) The effect size on student achievement is "modest at best" after introducing standards-based accountability (Hanushek and Raymond 2005);

3) There is mixed evidence about the impact of standards-based accountability on the academic performance of specific racial subgroups, such as blacks, Hispanics, and white students (Carnoy and Loeb 2002, Kane and Staiger 2003, Dee 2003, and Hanushek and Raymond 2005); and

4) Some studies find that the lowest-achieving schools or students demonstrate larger achievement gains than those whose performance was at the higher end of the distribution (Jacob 2005, Cronin 2005, Figlio and Rouse 2006, Springer 2008, Ballou and Springer 2008, and Dee and Jacob 2009).

From Figlio and Ladd's review, I delve further into the three studies (Carnoy and Loeb 2002, Jacob 2005, and Hanushek and Raymond 2005) they describe as the most robust and empirically-sound.

Rather than studying state-based systems, Jacob (2005) examined Chicago Public Schools' district-level standards-based accountability system, which was implemented in 1996. He found that reading scores and math scores in particular increased on a high-stakes exam following the adoption of standards-based accountability relative to the gains expected in the district and relative to gains seen in similar, large urban Midwestern districts. However, he also found that for elementary school students the new accountability policy did not improve achievement on a low-stakes, state-administered math exam where students were not held accountable for their results. Further, the observed gains on high-stakes tests were driven by an increased focus on test-specific skills, with teachers responding to incentives the accountability system created. This included increasing special education placements, preemptively holding back students, and substituting away from non-tested subjects.

The Carnoy and Loeb (2002) study used a cross-state analysis of achievement data from the National Assessment of Educational Progress (NAEP) to rank the "strength" of each state's standards-based accountability system on a 0 to 5 scale by counting their school and student accountability components. They found that between 1996 and 2000, large within-state growth on the 8th grade mathematics exam was correlated with "stronger" levels of accountability. They also found that high-accountability states were no more likely to retain students in 9th grade or to see lower high school completion rates compared to states without accountability in place.

Because they examine multiple states, their findings "are likely to be less idiosyncratic and more generalizable than those that emerge from the analysis of a specific program" (Figlio and Ladd 2008). Nevertheless, this study is unlikely to be fully nationally-representative since it is based on data from fewer than half of the states. Only 22 states had uniform standards for Title I and non-Title I schools in 2000 (Goertz, Duffy, and Le Floch 2001), and only 31 states even had content standards in 1997 (Pew 2006).

Hanushek and Raymond (2005) also examined the effect of standards-based accountability on state NAEP math and reading scores. They created four cohorts by comparing the difference between the test scores of a state's 8^{th} graders relative to the test scores of 4^{th} graders in the same state four years earlier; math growth was measured with a 1992-1996 cohort and a 1996-2000 cohort, while reading growth compared a 1994-1998 cohort and a 1998-2002 cohort. They also categorize each state's accountability system, if they had one, as "report-card accountability" or "consequential accountability." The difference between these two classifications is that while both publicly report school-level achievement data, there are accountability measures tied to school-level performance in "consequential accountability" states. According to their classifications, 25 states had consequential accountability systems prior to NCLB. They find that there is an association between the adoption of consequential accountability and increases in test scores for student cohorts on the NAEP, although the effects were only statistically significant for Hispanic and white students.

More recently, Dee and Jacob (2009) estimated the impact of NCLB on student achievement using state-level panel data with a comparative interrupted time series analysis. This enables a comparison of the NAEP test score changes from 1992-2007 in states that already had standards-based accountability prior to NCLB with the change in test scores for states that

had not already implemented standards-based reforms. They find that NCLB led to statistically significant increases in average math performance for 4th graders, including improvements in the highest and lowest percentiles. They also found weaker gains in 8th grade math, particularly among low-achieving students. However, they found no evidence that NCLB improved reading achievement in either grade.

IV. Research Questions and Hypotheses

This study builds on the existing empirical literature by providing new evidence from the Education Longitudinal Study of 2002 (ELS: 2002), a nationally representative survey that has not yet been explored in relation to the effect of standard-based accountability on student outcomes.

To complement existing studies that have focused on student test scores on state assessments and the NAEP, I specifically examine student outcomes that are directly related to college-readiness – high school completion, postsecondary access, and postsecondary academic outcomes. While improving student test scores is a valid and worthwhile objective, the ultimate objective is to improve the likelihood that students graduate high school prepared for college-level work. Given that NCLB placed increasing "pressures on… schools to reach a new, unprecedented goal: to prepare all students to master a demanding set of academic standards that will prepare them for postsecondary education and training" (Murnane 2009), my focus on high school completion and postsecondary access and outcomes is particularly appropriate.

I address the following research questions:

1) Does exposure to complete standards-based accountability systems lead to improved secondary completion and postsecondary academic outcomes for students?

2) Does the effect of standards-based reform on student outcomes depend on the extent to which reforms have been implemented?

The theory of action for standards-based accountability hypothesizes that students who experience fully-developed standards-based accountability systems will have better high school completion and postsecondary outcomes compared to similar students without standards-based accountability or with an incomplete system in place. The theory also predicts that there should be no statistically significant difference between the effects of individual elements or combinations of elements of standards-based accountability on student outcomes; the presence of a complete system should be the only significant indicator associated with more positive student outcomes.

V. Estimation Strategy

To address my first hypothesis that students who experience fully-developed standards-based accountability systems will have better secondary completion and postsecondary outcomes than students without standards-based accountability or with an incomplete system, I estimate the following Ordinary Least Squares (OLS) linear probability model:

Model 1: $Y_i = \beta_0 + \beta_1 SBA + \beta_2 YD + \beta_3 FD + \beta_4 SD + \varepsilon$

where Y_i is one of eight binary variables measuring high school completion, college readiness, and postsecondary outcomes for student i. Two of the dependent variables are related to "High School Completion" – if the student is a high school *dropout* (1) and graduated high school *on time* (2) A student was classified as graduating on time if they had completed high school within two years of the baseline survey. The next three dependent variables are related to "Postsecondary Attendance" – if the student *applied* (3) for postsecondary education in high school, ever *attended* (4) postsecondary education, and attended postsecondary education *full-*

time and on-time (5). Both two- and four-year institutions were included in these measures. Students were measured as attending college full-time and on-time if there was no gap between their high school completion and college attendance *and* if they were a full-time student. The last three dependent variables are related to "Postsecondary Performance" – if the student ever needed college remedial coursework in *reading* (6), *writing* (7), and *math* (8).

The *SBA* variable is a dummy variable equal to 1 if the student attended a high school in 2002 with a complete state-level standards-based accountability system and equal to 0 if one or more of these elements were not implemented in their school. The ELS:2002 base year administrator survey specifically asked and linked questions about whether students experienced content standards, performance standards, proficiency tests, and accountability measures *and* whether these elements are aligned. Therefore, students are identified as having a complete *SBA* system if their school administrator indicated the following on the ELS:2002 survey:

- there were **content standards** for academic subjects defined by the state; and
- content standards were linked with **performance standards**; and
- state **competency tests** linked with content standards were given; and
- **accountability measures** were in place including any of the following: taking a remedial class, completing a test preparation class, attending tutoring or another individualized academic program, attending summer school, or principal evaluations are based on student standardized test scores.

Because many students are missing information on at least one of the components of my main independent variable – standards-based accountability (see section VI, part b for further details) – a missing data indicator for *SBA* is also included in the model, which may help improve estimation precision.

All but one element of standards-based accountability is measured in the ELS:2002: using content standards to inform the practice of teachers and school leaders through aligned curriculum and professional development. According to the theory of action, if this component is

not well-functioning, an entire standards-based accountability system will be less likely to improve student outcomes. Proponents argue that this element is crucial because it enables teachers to translate the new content standards into instruction and student learning within their classrooms. Without it, standards-based reforms may be, using Whitehurst's terminology, "a leaky bucket." The fact that this element is missing from the *SBA* variable raises an important caveat for my regression results.

Three sets of control variables (listed in Tables 1a-1c) are included in the model and described further in Section IV. The *YD*, or youth demographic, set of covariates controls for differences between students that could explain part of the variation in high school completion and postsecondary educational outcomes including race/ethnicity, socioeconomic status quartile, special education status, ELL status, and number of academic risk factors (e.g., repeating a grade or repeatedly changing schools). Similarly, the *FD*, or family demographic, set of covariates controls for variation at the family level that could affect students' secondary or postsecondary outcomes. These factors include parents' educational attainment, family composition (e.g., two-parent households as opposed to other kinds of households), and parental English fluency.

The last set of control variables, *SD*, or school demographic, accounts for differences at the school-level other than the presence of standards-based accountability systems that could explain variation in high school completion, college readiness, and postsecondary outcomes, since certain kinds of schools may be less likely to have the resources or expertise to implement reform. Variables include school urbanicity, percent of students receiving free lunch, total school enrollment, and percent of students in the school with limited English proficiency. Finally, β_0 represents the intercept term, while ε is a random error term.

Tables 1a-1c. Summary Statistics for Control Variables Included in Models 1-4

Table 1a. Youth Demographics

Variable Name	Mean
male	0.472
gender missing	0.051
White	0.609
black	0.116
hispanic	0.130
asian	0.033
other race	0.052
race missing	0.059
English Language Learner	0.015
ELL missing	0.016
lowest income quartile	0.235
2nd lowest income quartile	0.244
2nd highest income quartile	0.243
highest income quartile	0.219
income missing	0.059
no academic risk factors	0.318
one risk factor	0.238
two risk factors	0.123
three risk factors	0.056
four or more risk factors	0.029
risk factors missing	0.237
special education	0.074
special education missing	0.110

Table 1b. Family Demographics

Variable Name	Mean
parent speaks English as 2nd language	0.110
parent ESL missing	0.166
non-2 parent household	0.410
household composition missing	0.054
parent lacks 4 year degree	0.606
parent education missing	0.054
mother lacks high school degree	0.125
mother has high school degree	0.281
mother attended some college	0.325
mother has BA degree	0.145
mother has more than a BA degree	0.070
mother education missing	0.054

Table 1c. School Demographics

Variable Name	Mean
school enrollment < 400	0.089
school enrollment 400-800	0.169
school enrollment 800-1200	0.182
school enrollment 1200-1600	0.180
school enrollment 1600-2000	0.122
school enrollment > 2000	0.222
school enrollment missing	0.036
urban school	0.237
suburban school	0.529
rural school	0.234
free lunch students < 10%	0.286
free lunch students 10-30%	0.360
free lunch students 30-50%	0.156
free lunch students 50-75%	0.092
free lunch students > 75%	0.050
free lunch missing	0.056
% students Limited English Proficient	4.56

**Notes: Sample size = 8,001 students. Sample does not include non-public school students, students who were not sophomores in the base year, and students missing any dependent variables or continuous independent variables. Student sampling weights included in all statistics. The omitted baseline category for dummy variables included in regressions is highlighted in bold text. Academic risk factors include students: 1) in a single-parent home; 2) with 2 parents without a high school diploma; 3) with a sibling who dropped out of school; 4) who have changed schools 2 or more times; 5) who have repeated a grade; and 6) whose family is below the federal poverty threshold.

Because I am exploring the relationship of *SBA* and these eight secondary and postsecondary outcomes, β_1 is the variable of interest in Model 1. My hypothesis suggests that the coefficient on *SBA* should be positive and significant, meaning that students experiencing full *SBA* are likely to have more positive college readiness outcomes. In the case of undesirable student outcomes, like *dropout* and *remedial college reading, writing,* and *math*, this means that

I expect the coefficient on *SBA* to be negative. Thus, holding the *YD, FD,* and *SD* sets of control variables constant, Model 1 estimates the relationship between students experiencing fully-developed *SBA* – compared to students having no standards-based accountability or an under-developed system – and the probability of attaining key secondary and postsecondary outcomes.

My second hypothesis implies that there should be no statistically significant difference between effects of individual elements or combinations of elements of standards-based accountability on student outcomes. For policymakers and educators implementing standards-based reform, it is critical to determine if any one element is most significantly related to student outcomes, or whether the standards-based system as a whole is necessary. To explore this question, I modify my OLS regression in Model 1 to measure standards-based accountability "intensity" as the primary independent variable, with four indicator variables measuring the number of components of *SBA* a student experienced – from one to all four elements (zero elements is the omitted baseline category):

Model 2: $Y_i = \beta_0 + \beta_1 \text{SBAindex1} + \beta_2 \text{SBAindex2} + \beta_3 \text{SBAindex3} + \beta_4 \text{SBAindex4} + \beta_5 \text{YD} + \beta_6 \text{FD} + \beta_7 \text{SD} + \varepsilon$

The *YD*, *FD*, and *SD* sets of covariates are maintained from Model 1. As in Model 1, β_0 represents an intercept term, while ε accounts for random error in the model.

Here, the *SBAindex* coefficients are of interest. By comparing the magnitude of these coefficients, it is possible to determine whether different combinations of exposure to standards-based accountability have similar associations with student outcomes. Based on my hypotheses, I expect the coefficient on *SBAindex4* to be positive and significant both on its own and in comparison to the other *SBAindex* coefficients. Again, for undesirable student outcomes, theory suggests that the coefficient on *SBAindex4* should be negative. Secondly, I expect the other *SBAindex* coefficients should be statistically insignificant and also not significantly different

from one another. In this way, Model 2 parses out the relationship between high school completion and postsecondary student outcomes and different types of partially-developed standards-based accountability systems.

Finally, it is also important to determine whether the relationship between having a full *SBA* system and these particular student outcomes differ for key at-risk subgroups. Therefore, I augment Model 1 to include interaction terms, allowing for the relationship between standards-based accountability and student secondary and postsecondary outcomes to vary between minority and white students and between students in the lowest socio-economic quartile and higher quartiles.

Model 3: $Y_i = \beta_0 + \beta_1 SBA + \beta_2 YD + \beta_3 FD + \beta_4 SD + \beta_5 (SBA*black) + \beta_6 (SBA*hispanic) + \varepsilon$

Model 4: $Y_i = \beta_0 + \beta_1 SBA + \beta_2 YD + \beta_3 FD + \beta_4 SD + \beta_5 (SBA*lowincome) + \varepsilon$

In Model 3, the omitted baseline category is white students who did not experience complete *SBA*, as *black* and *hispanic* are dummy variables indicating whether a student is black or Hispanic. Therefore, β_5 and β_6 are both coefficients of interest in Model 3; if either of these is significant, it implies that there is heterogeneity based on race in the association between *SBA* and student outcomes. If β_6 is positive and significant, it suggests that there is a stronger association with desired student outcomes from experiencing full standards-based accountability for Hispanic students.

In Model 4, *lowincome* is an indicator for students who are in the lowest family income quartile, so the omitted baseline category is students who are in the 3 highest income quartiles and did not experience fully-developed standards-based accountability. Therefore, β_5 is the main coefficient of interest in Model 4 and indicates whether the relationship between *SBA* and student outcomes depends on family income. For example, if β_5 is positive and significant, it implies that the association between fully-implemented *SBA* and positive student outcomes is stronger for

low-income students. In both Models 3 and 4, β_0 is an intercept term, while ε captures random error in the model. In this way, it is possible to determine whether there is significant heterogeneity in the effect of *SBA* on these particular groups of interest.

VI. Data

a. Structure of the ELS:2002

The Education Longitudinal Study of 2002 (ELS:2002), conducted by the National Center for Education Statistics at the U.S. Department of Education, is a nationally representative, longitudinal sample of a cohort of over 15,000 high school sophomores following their educational outcomes from high school to postsecondary education and the workforce. The ELS:2002 is available in both a public-use and restricted-use format; for purposes of this study, I was only able to access the public-use version, which excludes state identifiers. I revisit this limitation in Section VIII. Although the dataset follows the same cohort of students over time, the ELS:2002 is not a panel in the sense that many questions are only asked in one year, including those related to standards-based accountability. It is impossible to know whether students were exposed to standards-based accountability before 2002, and whether students' exposure to standards-based accountability systems changed between 2002 and 2004, their senior year in high school.

While it is impossible to address the first problem of prior exposure to standards-based accountability with the public-use data, the second problem is more tractable. This issue is particularly apparent for the students who transferred schools between the base year and follow up years; I cannot be sure that their exposure to standards-based accountability was constant throughout the survey. While some students may have transferred within district and experienced

similar accountability systems, some students may have moved between states. As a sensitivity analysis, I re-estimate each model after dropping these observations from the sample. Unfortunately, this still cannot account for whether a state changed its accountability system between 2002 and 2004. In that case, even if a student did not move, it is possible he or she may have experienced a change in exposure to standards-based accountability.

The base year (BY) data in ELS:2002 were collected from students, school administrators, and parents in spring of 2002, providing a "comprehensive picture of the home, school, and community environments and their influences on the student" (NCES). BY data include student surveys, cognitive tests, and questionnaires completed by parents, teachers, principals, and librarians at 750 schools. ELS:2002 used a stratified sampling technique, so that schools were selected at random first, followed by a random selection of sophomores at those schools. Because observations could be related within schools, clustered standard errors are utilized to address the group component of the error. In addition, Asian students were over-sampled. Thus, probability weights are applied in all analyses to keep the sample representative.

Follow-up surveys and cognitive tests were completed in 2004 (F1) when most students in the sample were high school seniors. Dropouts and early completers were also surveyed. 12,400 students from the BY remained in the sample, in addition to 1,100 BY students who had transferred schools and 1,300 dropouts or early completers. To ensure the F1 survey was still nationally representative, the original cohort was "freshened" to include 2004 seniors at sampled schools who were not in 10^{th} grade at that school due to skipping or failing a grade or being out of the country. A second round of follow-up surveys was completed in 2006 (F2) to track college application rates, enrollment in postsecondary education, employment, and other postsecondary outcomes. All students who were included in the BY and/or F1 were surveyed as part of the F2.

Despite its limitations, the ELS:2002 is in some ways an ideal data source to investigate the relationship between standards-based accountability and college readiness. BY data were taken during the spring of the school year that immediately preceded NCLB implementation when all states were mandated to adopt standards-based systems, including strict accountability measures. Even though many states began the process of standards-based reforms in the 1990s, "by the early 2000s… fewer than one-half of these systems were in full compliance with the… standards and testing requirements at the time" (Hamilton, Stecher, and Yuan 2008). Therefore, the BY data from ELS:2002 taken in the 2001-2002 school year is likely the last school year in which significant variation still exists between the level of standards-based reforms students experienced, with some schools implementing only content standards without assessments or accountability mechanisms and other schools implementing more robust and complete systems.

b. Missing Data

While over 15,000 students were surveyed as part of the ELS:2002, the focus of this study is on students who experienced state-level standards-based accountability systems as a sophomore in high school, like those mandated by NCLB. Therefore, I drop students who either did not attend public schools or were not 10^{th} graders in the base year survey, reducing the sample to 12,614 students. Many students were also missing one or more of my dependent variables. 1,807 students were missing data on the High School Completion and Postsecondary Attendance outcomes; similarly, only students who had ever attended college provided information about needing college-level remediation on the follow up survey, so an additional 3,480 students were missing data on the Postsecondary Performance dependent variables.

As previously mentioned, a large proportion of students were missing data on at least one of the dependent variables *and* one of the main independent variables – the components of

standards-based accountability – which were obtained from a survey given to school administrators. To prevent 2,796 student observations from dropping out of the sample, I include a missing data dummy in all regressions to identify students who are missing information on *SBA* and retain these observations. I also utilize missing data indicators for all binary covariates that measure student, family, or school demographic characteristics of students in the sample. As indicated in Tables 1a-1c, for some covariates there is a significant amount of missing data. For example, about 11 percent of the sample is missing information on whether they ever were placed in special education. The largest percentage of missing data exists within the academic risk factors variable, where nearly a quarter of students in the sample are missing information.

After dropping all students who are missing any of my High School Completion and Postsecondary Attendance dependent variables and creating missing data dummies for the covariates and main independent standards-based accountability variables, my final sample includes 8,001 students, with 5,410 students included in the Postsecondary Performance analyses (with the difference being that the latter sample further excludes those students who did not attend college). As Table 2 indicates, nearly a quarter of the sample experienced no elements of standards-based accountability, while a third of the sample experienced fully-implemented

Table 2. Summary of ELS:2002 Students' Exposure to Standards-Based Accountability

contentstds	*perfstds*	*assessments*	*accountability*	**Percent**
				24.27
X				2.95
X	X			34.33
X	X	X		2.09
X		X	X	3.01
X	X	X	X	33.35

** Notes: Sample size = 8,001 students. Sample excludes non-public school students, non-sophomores in the base year survey, and students missing information on dependent variables or continuous independent variables. *Contentstds* indicates student experienced state academic content standards. *Perfstds* indicates student experienced performance standards aligned with state content standards. *Assessments* indicate student experienced state-required 10^{th} grade assessments linked to content standards. *Accountability* indicates student experienced accountability measures tied to students' results on state-required, content-aligned 10^{th} grade assessments. Student sampling weights included in all statistics.

standards-based reforms. Another third of the sample was exposed only to state-level academic content standards aligned with performance standards. Smaller proportions of the sample (about 3 percent) either experienced only content standards; aligned content standards, performance standards, and assessments; or content standards aligned with assessments and accountability measures without performance standards.

VII. Results

a. Main Results

Table 3 presents regression results from Models 1 and 2, examining the relationship between standards-based accountability and high school completion and postsecondary student outcomes. Overall, the results from Model 1 do not support my hypothesis. As Table 3 indicates, after accounting for differences in youth, family, and school demographic factors, students who experienced standards-based accountability systems did not have statistically significantly different outcomes than students who did not experience standards-based accountability or only experienced a handful of the components of standards-based reforms. These findings held for the high school completion and postsecondary attendance outcomes, as well as for the postsecondary performance outcomes which were based on the sample of college attendees only.

The results from Model 2 fail to support the hypothesis that the effect of standards-based accountability on student outcomes is invariant to the extent to which reforms have been implemented. Reinforcing the findings from Model 1, the coefficient on *sbaindex4* is not statistically significant in any of the eight regressions. Further, the sign of this coefficient is opposite what would be expected if fully-developed accountability systems were associated with improved student outcomes for all iterations of the model except remedial college reading. In

Table 3. OLS Coefficient Estimates of the Relationship between Standards-Based Accountability Measures and College Readiness Outcomes

	HIGH SCHOOL COMPLETION		POSTSECONDARY ATTENDANCE			POSTSECONDARY PERFORMANCE		
	dropout	*ontime grad*	*ever applied*	*ever attend*	*attend FTOT*	*remedial col read*	*remedial col write*	*remedial col math*
Baseline	0.1255	0.8688	0.7665	0.6911	0.5330	0.1972	0.2570	0.2873
Model 1								
Sba	0.00605	-0.0122	-0.0112	0.0149	0.00583	-0.0153	-0.0264	0.0125
	(0.0118)	(0.0122)	(0.0144)	(0.0154)	(0.0166)	(0.0146)	(0.0193)	(0.0186)
R-squared	0.098	0.102	0.123	0.175	0.195	0.037	0.032	0.028
Baseline	0.1094	0.8803	0.7885	0.7258	0.5636	0.1918	0.2793	0.3021
Model 2								
sbaindex1	0.151***	-0.187***	0.0999***	0.0600*	-0.0203	-0.0267	0.0481	0.00609
	(0.0379)	(0.0356)	(0.0282)	(0.0331)	(0.0313)	(0.0340)	(0.0353)	(0.0308)
sbaindex2	0.0184	-0.0105	-0.00893	-0.0413	-0.0364	0.180***	0.174***	-0.0220
	(0.0187)	(0.0200)	(0.0228)	(0.0266)	(0.0270)	(0.0419)	(0.0405)	(0.0398)
sbaindex3	0.0224	-0.0103	0.00468	-0.0151	-0.0373	-0.0255	0.00168	-0.0118
	(0.0232)	(0.0247)	(0.0407)	(0.0389)	(0.0356)	(0.0344)	(0.0360)	(0.0335)
sbaindex4	0.0144	-0.0175	-0.0108	-0.00159	-0.0106	-0.0340	0.00198	0.0102
	(0.0156)	(0.0166)	(0.0197)	(0.0198)	(0.0213)	(0.0323)	(0.0321)	(0.0288)
F-statistics								
sbaindex1 vs. 2	12.87***	27.38***	15.16***	8.72***	0.23	41.64***	22.97***	0.07
sbaindex2 vs. 3	0.03	0.00	0.11	0.42	0.00	0.55	0.00	0.12
sbaindex3 vs. 4	0.14	0.10	0.16	0.14	0.66	1.11	0.00	0.13
R-squared	0.099	0.103	0.123	0.175	0.195	0.038	0.033	0.028

Notes: *dropout* = ever dropped out of high school, *ontimegrad* = graduated high school on-time, *everapplied* = ever applied for postsecondary education, *everattend* = ever attended postsecondary education, *attendFTOT* = attended postsecondary education on-time, full-time, *remedialcolread* = needed remedial college reading, *remedialcolwrite* = needed remedial college writing, *remedialcolmath* = needed remedial college math. For all regressions, *sba* is a dummy indicating fully-implemented standards-based accountability, as described in Table 2. *sbaindex* represents a count of the elements of standards-based accountability experienced based on values for *contentstds, perfstds, assessments, and accountability*. For all models, right-hand-side control variables include the *YD*, *FD*, and *SD* control variables– see Tables 1a-1c. For high school completion and postsecondary outcomes specifications, sample size = 8,001; for postsecondary performance, sample size = 5,410. Student sampling weights are included in all regressions. Baseline in Model 1 is students who did not experience fully-developed *sba*, while the baseline in Model 2 reflects students who experienced 0 elements of *sba.* Only relevant standards-based accountability coefficients are reported above. Parentheses correspond to clustered standard errors by school. F-statistics included for comparisons of the difference in magnitude between standards-based accountability coefficient estimates in Model 2 (i.e. *sbaindex1*compared to *sbaindex2*). *** p<0.01, ** p<0.05, * p<0.1

addition to the contradictory findings on *sbaindex4*, the other *sbaindex* coefficients are at odds with my hypothesis in Model 2. Specifically, *sbaindex1* and *sbaindex2* are individually significant in several of the regressions. It also appears that the effect of *sbaindex1* on student outcomes is larger in magnitude than the effect of *sbaindex2* for high school completion and postsecondary attendance variables, while the reverse is true for postsecondary performance outcomes. Secondly, for all college readiness outcomes except attending college full-time and on-time and needing remedial college math, there is a statistically significant *difference* between

the effect of having experienced one element of standards-based accountability and the effect of having experienced two elements (F-test p-value < 0.01). However, there are no statistically significant differences between the magnitude of the coefficients for two and three elements or three and four elements.

While the OLS coefficient estimates and F-statistics comparing the *sbaindex* variables show that individual elements of standards-based accountability and partially-implemented systems have a significant effect on high school completion and postsecondary outcomes, the direction of this effect is inconsistent. In some cases, exposure to partially-implemented systems improves student outcomes, but in others the likelihood of a positive outcome decreases. For example, holding all other variables constant, the probability of dropping out of high school is about 15 percentage points *higher* for students who experienced content standards only compared to students who experienced no elements of standards-based accountability. However, the reverse is true for ever attending college; the probability of attending college is approximately 6 percentage points higher for students who were exposed to content standards compared to students who were not exposed to standards-based accountability. While the direction of the effects are sometimes inconsistent, the results from Model 2 suggest that student outcomes can be correlated with standards-based accountability systems, even if the complete system has not been implemented.

b. Distributional Differences

Because of the theory of action that standards-based accountability should improve student outcomes and the focus on at-risk student subgroups in prior research, Table 4 explores the relationship between standards-based accountability and key secondary and postsecondary outcomes for three at-risk subgroups: black students, Hispanic students, and low-income

Table 4. OLS Coefficient Estimates of Standards-Based Accountability on College Readiness Outcomes for At-Risk Student Subgroups

	HIGH SCHOOL COMPLETION		POSTSECONDARY ATTENDANCE			POSTSECONDARY PERFORMANCE		
	dropout	*ontime grad*	*ever applied*	*ever attend*	*attend FTOT*	*remedial col read*	*remedial col write*	*remedial col math*
Model 3								
sba	0.00447	-0.00454	-0.0142	0.0133	0.00605	-0.00339	-0.0233	0.00326
	(0.0121)	(0.0128)	(0.0153)	(0.0178)	(0.0201)	(0.0182)	(0.0231)	(0.0216)
black	0.0224	-0.0124	0.0686***	0.0204	-0.00729	0.0585**	0.00900	0.0232
	(0.0226)	(0.0224)	(0.0221)	(0.0239)	(0.0256)	(0.0295)	(0.0317)	(0.0349)
*sba*black*	-0.0123	-0.0178	-0.00962	0.0123	-0.000520	-0.0259	-0.0598	-0.0604
	(0.0364)	(0.0369)	(0.0354)	(0.0403)	(0.0425)	(0.0425)	(0.0433)	(0.0488)
hispanic	0.0365	-0.0542**	-0.0434	-0.0608**	-0.132***	0.0800**	0.0829**	0.0449
	(0.0237)	(0.0241)	(0.0299)	(0.0273)	(0.0256)	(0.0344)	(0.0389)	(0.0353)
*sba*hispanic*	-0.00882	0.0117	0.0108	-0.0185	-0.0215	-0.0448	-0.0191	0.0525
	(0.0311)	(0.0330)	(0.0383)	(0.0382)	(0.0369)	(0.0460)	(0.0527)	(0.0504)
R-squared	0.099	0.103	0.123	0.175	0.195	0.038	0.033	0.030
Model 4								
sba	0.00808	-0.0105	-0.0101	0.00832	0.00397	-0.00514	-0.0154	0.00889
	(0.0126)	(0.0127)	(0.0151)	(0.0159)	(0.0187)	(0.0161)	(0.0207)	(0.0206)
lowincome	0.0319*	-0.0304*	-0.0514**	-0.0853***	-0.0567***	0.0334	0.00765	-0.0164
	(0.0188)	(0.0184)	(0.0214)	(0.0222)	(0.0219)	(0.0237)	(0.0268)	(0.0304)
*sba*lowincome*	-0.00852	-0.00737	-0.00439	0.0278	0.00813	-0.0592*	-0.0635*	0.0211
	(0.0268)	(0.0272)	(0.0275)	(0.0291)	(0.0324)	(0.0323)	(0.0355)	(0.0391)
R-squared	0.098	0.102	0.119	0.171	0.188	0.038	0.032	0.027

Notes: *dropout* = ever dropped out of high school, *ontimegrad* = graduated high school on-time, *everapplied* = ever applied for postsecondary education, *everattend* = ever attended postsecondary education, *attendFTOT* = attended postsecondary education on-time, full-time, *remedialcolread* = needed remedial college reading, *remedialcolwrite* = needed remedial college writing, *remedialcolmath* = needed remedial college math. For all regressions, *sba* is a dummy variable indicating fully-implemented standards-based accountability. For all specifications, right-hand-side control variables include the *YD*, *FD*, and *SD* control variables – see Tables 1a-1c. For high school completion and postsecondary outcomes, sample size = 8,001; for postsecondary performance, sample size = 5,410. Student sampling weights included in all regressions. Only standards-based accountability, *black*, *hispanic*, or *lowincome* dummy, and *black-sba*, *hispanic-sba*, or *lowincome-sba* interaction coefficients reported above. Parentheses correspond to clustered standard errors by school. *** p<0.01, ** p<0.05, * p<0.1

students. In general, there appears to be no interactions between standards-based accountability exposure and these at-risk groups on college readiness outcomes. Only two of the 16 OLS regressions displayed in Table 4 contain a statistically significant interaction term between *SBA* and the at-risk student subgroup; both of these effects were for low-income students in postsecondary performance outcomes. In both cases, standards-based accountability exposure had a positive effect: low-income students exposed to complete standards-based accountability have a 5.9 percentage point lower probability of needing remedial college reading than low

income students who did not experience accountability systems. Similarly, low-income students that experienced complete standards-based accountability have a 6.4 percentage point lower probability of needing writing remediation compared to similar students who did not experience fully-developed systems. The results from Model 3 indicate no distinctive effect of standards-based accountability reforms for minority students.

c. Sensitivity Tests

To check the precision of my analyses, I performed a few sensitivity checks on each of the four models. First, because each of the eight dependent variables is a dummy variable, I re-estimated the OLS linear probability model used in Models 1 and 2 as a probit model. For Model 1, the signs and statistical significance of all regression coefficients using probit did not differ from the OLS results presented in my main findings. For Model 2, the statistical significance for each F-statistic comparing the coefficients obtained via probit did not differ from the F-statistics reported for Model 2 in Table 3, and individual coefficients on *sbaindex1* and *sbaindex2* remained statistically significant.

As discussed earlier, the ELS:2002 does not capture how each variable changed from the base year through the follow up surveys. It is thus impossible to know whether students' exposure to standards-based accountability systems changed between 2002 and 2004, their senior year in high school. As a final sensitivity analysis, the coefficients in Models 1 and 2 were re-estimated after dropping any student from the sample who had changed schools between 2002 and 2004. As shown in Table 5, the overall conclusions from Model 2 were essentially the same as the original specification, indicating that one element or two elements of standards-based reforms are individually significant and associated with particular student outcomes. In this specification, *sbaindex1* is statistically significant for 6 of the 8 outcomes, compared to just 4 of

Table 5. OLS Coefficient Estimates of Standards-Based Accountability on College Readiness Outcomes of Non-Moving Students Only

	HIGH SCHOOL COMPLETION		POSTSECONDARY ATTENDANCE			POSTSECONDARY PERFORMANCE		
	dropout	*ontime grad*	*ever applied*	*ever attend*	*attend FTOT*	*remedial col read*	*remedial col write*	*remedial col math*
Baseline	0.0349	0.9569	0.8205	0.7562	0.6066	0.1968	0.2598	0.2825
Model 1								
sba	-0.00169	-0.0102	-0.00231	0.0156	0.00168	-0.0265*	-0.0423**	0.00248
	(0.00663)	(0.00910)	(0.0137)	(0.0135)	(0.0162)	(0.0148)	(0.0200)	(0.0183)
R-squared	0.048	0.072	0.137	0.201	0.237	0.062	0.046	0.050
Baseline	0.0270	0.9656	0.8323	0.7933	0.6363	0.1960	0.2900	0.3033
Model 2								
sbaindex1	0.0969	-0.176***	0.0230	0.106***	0.0639*	0.224***	0.152***	0.0880*
	(0.0594)	(0.0558)	(0.0276)	(0.0272)	(0.0336)	(0.0375)	(0.0431)	(0.0451)
sbaindex2	0.00902	-0.00715	-0.00361	-0.0487**	-0.0285	-0.0136	-0.0537*	-0.0247
	(0.0115)	(0.0140)	(0.0237)	(0.0240)	(0.0278)	(0.0235)	(0.0321)	(0.0309)
sbaindex3	-0.00473	0.00544	0.00447	-0.0306	-0.0413	0.0342	-0.0345	-0.0347
	(0.0162)	(0.0178)	(0.0408)	(0.0399)	(0.0437)	(0.0393)	(0.0403)	(0.0318)
sbaindex4	0.000761	-0.0141	-0.00180	-0.00738	-0.0154	-0.0189	-0.0577	-0.00599
	(0.00826)	(0.0115)	(0.0205)	(0.0182)	(0.0220)	(0.0188)	(0.0267)	(0.0242)
F-statistics								
sbaindex1 vs. 2	2.19	9.39***	0.95	27.82***	6.66***	38.89***	23.61***	5.66**
sbaindex2 vs. 3	0.60	0.46	0.04	0.19	0.08	1.41	0.21	0.08
sbaindex3 vs. 4	0.12	1.31	0.03	0.38	0.39	1.98	0.40	1.02
R-squared	0.083	0.955	0.843	0.805	0.692	0.245	0.287	0.328

Notes: *dropout* = ever dropped out of high school, *ontimegrad* = graduated high school on-time, *everapplied* = ever applied for postsecondary education, *everattend* = ever attended postsecondary education, *attendFTOT* = attended postsecondary education on-time, full-time, *remedialcolread* = needed remedial college reading, *remedialcolwrite* = needed remedial college writing, *remedialcolmath* = needed remedial college math. For all regressions, *sba* is a dummy indicating fully-implemented standards-based accountability. For all models, right-hand-side control variables include the *YD*, *FD*, and *SD* control variables – see Tables 1a-1c. For high school completion and postsecondary outcomes, sample size = 6,539; for postsecondary performance, sample size = 4,787. Student sampling weights are included in all regressions. Baseline in Model 1 is non-moving students who did not experience fully-developed *sba*, while the baseline in Model 2 reflects non-moving students who experienced 0 elements of *sba*. Only relevant standards-based accountability coefficients are reported above. Parentheses correspond to clustered standard errors by school, unless noted. F-statistics included for comparisons of the difference in magnitude between standards-based accountability coefficient estimates in Model 2 (i.e. *sbaindex1*compared to *sbaindex2*). *** p<0.01, ** p<0.05, *p<0.1

the outcomes when the sample was not limited to students that stayed in their high schools. Further, the association between experiencing one component and student outcomes appears stronger than the association between experiencing two components and student outcomes. The significance of the F-statistics comparing each of the *sbaindex* coefficients indicates that partially-developed systems are also associated with student outcomes. Specifically, the F-statistic comparing the coefficients on one element and two elements of standards-based accountability is statistically significant for six of the eight outcomes measured. The only

difference from the original model is that the F-statistic is now significant on attending college full-time and on-time and needing remedial college math, while the F-statistics on dropping out of high school and applying to college are no longer significant at conventional levels.

However, the results from Model 1 changed in more substantive ways once the sample was limited to students who did not change schools between the base year survey and the first follow-up. Among students who remained at their 2002 schools, exposure to fully-implemented standards-based accountability systems had a statistically significant effect on two of the three measures of postsecondary performance. As Table 5 shows, holding other variables constant, students exposed to complete standards-based accountability had a 2.7 percentage point lower probability of needing remedial college reading ($p<0.1$) and a 4.2 percentage point lower probability of needing remedial college writing ($p<0.05$) than students who were not exposed to complete standards-based accountability.

VIII. Discussion

This paper aims to clarify the mixed findings from prior studies on whether exposure to standards-based accountability improves student outcomes and to provide new evidence on how standards-based reforms are related to high school completion, postsecondary outcomes, and college readiness. The results presented in Table 3 suggest that there is little evidence to support either the theory that fully-implemented standards-based accountability systems are associated with better student outcomes or the theory that partially-developed systems or individual elements of standards-based accountability have no association with student outcomes. Additionally, the findings in Table 4 indicate little association between standards-based

accountability and at-risk subgroups' outcomes, with the exception of low-income students and college remediation in reading and writing.

One potential explanation for these findings is that exposure to standards-based accountability was not randomly assigned throughout the sample. In fact, minorities, low-income, and at-risk students were more likely to experience standards-based reforms than other students. These kinds of students may be more likely to experience standards-based accountability because they were the students these systems hoped to target and influence. The students who experienced standards-based accountability also tended to be lower-achieving than those who were not exposed to these kinds of systems. Thus, it was especially important to control for these differences in each of my models. While this was the intention of the three sets of control variables used in my regressions, it is unclear whether this effort was fully successful, since all of the relevant control variables may not have been included in the ELS:2002.

Further possible explanations for these findings is that the ELS:2002 does not measure a) whether the student's school actually implemented the four elements of standards-based accountability faithfully or just said they did for the purposes of the survey; b) whether the quality of this implementation was consistent across schools in the survey; or c) whether the school implemented the fifth element of standards-based reform: informing the practice of principals and teachers so they have the resources, tools, curriculum, and professional development to implement the system. According to the theoretical research, this component is critical for the success of the entire standards-based accountability system because it enables teachers to translate standards into instruction and learning in the classroom. The results presented here could exemplify Whitehurst's "leaky bucket" and explain sluggish academic improvement and persistent achievement gaps since NCLB implementation. Since this element

cannot be measured with the ELS:2002, further research should be conducted to analyze how curriculum, professional development systems, and classroom instruction within standards-based accountability systems are related to student outcomes.

While the findings from Table 3 do not support the theory of action behind standards-based accountability, these results changed once students who transferred schools between 2002 and 2004 were dropped from the sample. As Table 5 shows, exposure to complete standards-based accountability had a statistically significant effect on two of the three measures of postsecondary performance, lowering the probability students would need remedial college reading and writing courses. This finding is interesting considering most prior research on standards-based accountability had found positive effects on math achievement, with less consistent or significant effects on reading outcomes; (Carnoy and Loeb 2002, Jacob 2005, Cronin 2005, Figlio and Rouse 2006, and Dee and Jacob 2009); it also is consistent with the results presented in Table 4, where the probability low-income students exposed to standards-based reforms needed remedial college reading and writing decreased relative to other students. One possible explanation could be that while standards-based accountability has a limited effect on reading test scores, it has positive effects on reading and writing skills that are not captured in elementary and secondary reading assessments but are nevertheless important for postsecondary success. This relationship with postsecondary performance in reading and writing warrants further research, as standards-based reforms may have positive effects on both math *and* reading outcomes and help explain why standards-based reforms have typically shown little effect on reading skills as measured by standardized tests.

Another consideration for future research in this area would be to explore the restricted version of the ELS:2002. Notably, the public-use version used in this study lacks state

identifiers. There are many observable and unobservable differences between states, including when states adopted standards-based accountability reforms, the level of rigor in state academic standards, the value states place on education, states' investment in education, and dedication to education finance equity. These state-level factors are likely correlated with both secondary and postsecondary outcomes and the adoption of standards-based accountability, which introduces the possibility of omitted variable bias. The results in this study thus should be considered tentative and not be causally interpreted. However, access to the restricted-use ELS:2002 data would enable a researcher to better control for these state-level factors in future research.

As federal, state, and local resources continue to be channeled towards standards-based accountability systems and as new education policies are built around these systems, it is critical for researchers to continue to assess the effectiveness of standards-based reforms. As newer data, like the ELS:2002, are collected and distributed, they should be used to explore and further clarify the effects of accountability systems on student outcomes. More importantly, given the mixed evidence of existing research, future studies should focus not only on which elements of standards-based accountability are most influential, but also on the particular implementation mechanisms that are associated with positive outcomes for students. As the push for national academic content standards in math and reading continues, this line of research has the potential to influence how states build new systems or refine their existing standards-based accountability systems around the Common Core and whether these efforts are ultimately successful.

Bibliography

Ballou, Dale and Matthew G. Springer. 2008. Achievement Trade-offs and No Child Left Behind. Working paper.

Carnoy, Martin and Susanna Loeb. 2002. Does External Accountability Affect Student Outcomes? A Cross-State Analysis. Educational Evaluation and Policy Analysis. 24 (4), 305-331.

Cohen, David K. and Susan L. Moffitt. 2009. *The Ordeal of Equality: Did Federal Regulation Fix the Schools?* Cambridge, MA: Harvard University Press.

Common Core State Standards Initiative. "In the States." http://www.corestandards.org/in-the-states, Accessed 17 October 2010.

Cronin, John, G. Gage Kingsbury, Martha S. McCall, and Branin Bowe. 2005. The Impact of the NCLB Act on Student Achievement and Growth: 2005 edition. Technical Report. Northwest Evaluation Association.

Dee, Thomas S. 2003. The "First Wave" of Accountability. In *No Child Left Behind: The Politics and Practice of School Accountability*, ed. Paul E. Peterson and Martin R. West.

Dee, Thomas and Brian Jacob. 2009. The Impact of No Child Left Behind on Student Achievement. National Bureau of Economic Research Working Paper #15531.

Figlio, David N. and Cecilia Elena Rouse. 2006. Do Accountability and Voucher Threats Improve Low-Performing Schools? Journal of Public Economics 90(1), 239-255.

Figlio, David N. and Helen Ladd F. 2008. School Accountability and Student Achievement. In *Handbook of Research in Education Finance and Policy,* ed. Helen F. Ladd and Edward B. Fiske.

Goertz, Margaret E., Mark C. Duffy, and Kerstin Carlson Le Floch. 2001. Assessment and Accountability Systems in the Fifty States: 1999-2000. Consortium for Policy Research in Education, Number RR-046. http://www.cpre.org/images/stories/cpre_pdfs/rr46.pdf, Accessed 16 March 2011.

Hamilton, Laura S., Brian M. Stecher, and Kun Yuan. 2008. Standards-Based Reform in the United States: History, Research, and Future Directions. RAND Corporation. http://www.rand.org/pubs/reprints/2009/RAND_RP1384.pdf, Accessed 30 October 2010.

Hanushek, Eric A. and Margaret E. Raymond. 2005. Does School Accountability Lead to Improved Student Performance? Journal of Policy Analysis and Management 24(2), 297-327.

Hanushek, Eric A. and Margaret E. Raymond. 2003. Lessons about the Design of State Accountability Systems. In *No Child Left Behind: The Politics and Practice of School Accountability*, ed. Paul E. Peterson and Martin R. West.

Hess, Frederick M. 2003. Refining or Retreating? High-Stakes Accountability in the States. In *No Child Left Behind: The Politics and Practice of School Accountability*, ed. Paul E. Peterson and Martin R. West.

Hess, Frederick M. and Michael J. Petrilli. 2006. *No Child Left Behind: Primer*. New York: Peter Lang Publishing, Inc.

Jacob, Brian A. 2005. Accountability, incentives, and behavior: the impact of high-stakes testing in the Chicago Public Schools. Journal of Public Economics 89, 761-796.

Kane, Thomas J. and Douglas O. Staiger. 2003. Unintended Consequences of Racial Subgroup Rules. In *No Child Left Behind: The Politics and Practice of School Accountability*, ed. Paul E. Peterson and Martin R. West.

Ladd, Helen F. 2007. Holding Schools Accountable Revisited. 2007 Spencer Education Foundation Lecture on Education Policy and Management, Association for Public Policy Analysis and Management. https://www.appam.org/awards/pdf/2007Spencer-Ladd.pdf, Accessed 14 October 2010.

McDermott, Kathryn A. and Laura S. Jensen. 2005. Dubious Sovereignty: Federal Conditions of Aid and the No Child Left Behind Act. Peabody Journal of Education, 80(2), 39-56.

McNeil, Michele. 2010. "Obama Ups Ante for Common Core." Education Week. 22 February 2010. http://blogs.edweek.org/edweek/campaign-k-12/2010/02/some_states_are_stars_massachu.html, Accessed 15 March 2011.

Murnane, Richard J. 2009. Educating Urban Children. In *Making Cities Work, Prospects and Policies for Urban America*, ed. Robert P. Inman.

National Center for Education Statistics. 2007. *Mapping 2005 State Proficiency Standards Onto the NAEP Scales* (NCES 2007—482). US Department of Education. Washington, DC: US Government Printing Office.

National Center for Education Statistics. "Education Longitudinal Study of 2002 (ELS:2002) Overview: Purpose." http://nces.ed.gov/surveys/els2002/. Accessed 18 October 2010.

National Commission on Excellence in Education. 1983. *A Nation at Risk: The Imperative for Educational Reform.* U.S. Department of Education. http://www2.ed.gov/pubs/NatAtRisk/index.html, Accessed 1 November 2010.

Pew Charitable Trusts. 2006. "State Policies on Standards-Based Education Over the Past Decade Found to Have a Positive Relationship With Gains in Student Achievement." 4

January 2006. http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/News/Press_Releases /K-12_education/qcpress06-1_final.pdf, Accessed 16 March 2011.

Phillips, Gary W. 2010. International Benchmarking: State Education Performance Standards. American Institutes for Research. http://www.air.org/files/AIR_Int_Benchmarking_State_Ed__Perf_Standards.pdf. Accessed 21 November 2010.

Springer, Matthew G. 2008. The influence of an NCLB accountability plan on the distribution of student test score gains. Economics of Education Review 27 (2008), 556-563.

Strauss, Valerie. 2010. "What Common Core State Standards Are – and aren't." The Washington Post. 3 June 2010. http://voices.washingtonpost.com/answer-sheet/national-standards/what-the-common-core-standards.html, Accessed 16 March 2011.

US Department of Education. 2010. ESEA Reauthorization: A Blueprint for Reform. http://www2.ed.gov/policy/elsec/leg/blueprint/index.html. Accessed 18 October 2010.

West, Martin R. and Paul E. Peterson. 2003. The Politics and Practice of Accountability. In *No Child Left Behind: The Politics and Practice of School Accountability,* ed. Paul E. Peterson and Martin R. West.

Whitehurst, Grover. 2009. Don't Forget Curriculum. Brown Center Letters on Education #3, Brookings Institution. http://www.brookings.edu/papers/2009/1014_curriculum_ whitehurst.aspx, Accessed 18 October 2010.