ACT Research Report Series 2016 (10)

Are Value-Added Measures of High School Effectiveness Related to Students' Enrollment and Success in College?

Dina Bassiri, PhD

Dina Bassiri is a senior research scientist in the Statistical and Applied Research Department, specializing in educational outcomes research and student growth models.

Acknowledgments

The author is grateful to Jeff Allen, Dick Buddin, Richard Sawyer, Jim Sconing, Teri Fisher, and Kurt Burkum for their review and helpful comments and suggestions on an earlier draft of this report.

MS488

Contents

Abstract

One outcome of the implementation of the No Child Left Behind Act of 2001 and its call for better accountability in public schools across the nation has been the use of student assessment data in measuring schools' effectiveness. In general, inferences about schools' effectiveness depend on the type of statistical model used to link student assessment results to schools. For example, characteristics beyond a school's control (e.g., entering achievement level and socio-economic status of the students served by the school) can strongly influence simple proficiency rates. In contrast, measures derived from growth and value-added models potentially estimate school effects more accurately.

This study investigated the predictive strength of value-added measures of high schools' performance on their students' enrollment and success in college. It is based on the data of 263,000 students who graduated in 2004 through 2009 from 1,119 high schools across the United States. The students had test scores from two time points (ACT Explore® in 8th grade and the ACT® college readiness assessment in 11th/12th grades).

The findings indicate that value-added school effect estimates predict college enrollment and retention, as well as grades in first-year college courses in English/Language Arts, Mathematics, Natural Sciences, and Social Sciences, even after adjusting for student-level and other school-level characteristics. This study provides evidence that some high schools are much more successful than others at moving their students towards success in college. It does not, however, look at the potential determinants that make these schools more successful than others.

Introduction

Educational accountability in the nation's public schools has gained considerable attention over the last decade, primarily because it became the basis for rewarding or sanctioning teachers or schools. This is in large part due to the 2001 reauthorization of the federal Elementary and Secondary Education Act (ESEA) known as the No Child Left Behind Act (National Association of State Boards of Education, 2002) which provided a specific framework within which states must develop their educational accountability system. Much has been written about what works and what does not work to improve the delivery of education. Different types of statistical models for attributing student assessment results to schools can lead to quite different conclusions about school effectiveness. Status measures, such as simple proficiency rates derived from status and improvement models, can be heavily influenced by characteristics outside of the school's control—specifically, the entering achievement level and socioeconomic status of the students served by the school. Other measures, such as those obtained from value-added modeling, can potentially estimate school effects more accurately.

This study investigated whether school value-added measures are potentially useful predictors of success in college. If so, they should be related statistically to college enrollment, college retention, and first-year course grades in college. Moreover, the statistical relationships should persist even after adjusting for student and school level characteristics.

Report Organization

This report begins with a brief overview of the school accountability models (i.e., status, improvement, growth, and value-added) and the need to validate accountability measures, followed by a description of the sample of high school cohorts, students, and college courses used in this study. Next, it describes how value-added measures of schools' performance are generated and the ways the predictive strength of these measures on students' success in college can be investigated. Finally, examples demonstrating the predictive strength of value-added school effect estimates related to college enrollment, retention percentages, and average course grades are provided for six variants of school types. These include high- and low-performing schools that either serve students from high-poverty and high-minority schools, and/or from lower-poverty and lower-minority schools. The report concludes with a summary of findings and limitations of the study.

No Child Left Behind and Every Student Succeeds Act

The No Child Left Behind Act (NCLB) held schools and school districts accountable for 100% proficiency of all their students by 2014, regardless of subgroup status, and adequate yearly progress (AYP) in attaining this goal (Lockwood, 2005). From the very beginning, however, many critics questioned the way AYP was determined and whether 100% proficiency could be achieved by 2014 (Doran & Izumi, 2004). In accordance with the new Every Student Succeeds Act (ESSA), the U.S. Department of Education, on December 10, 2015, made significant modifications to NCLB accountability provisions. ESSA, which will take full effect in the 2017–18 school year, aims to enhance the authority of states and school districts that had been restricted by NCLB (Klein, 2016).[1]

Before growth models were allowed, one significant criticism was that AYP was determined jointly by a status model (i.e., AYP) and an improvement model (i.e., Safe Harbor), both of which use very simplistic measures. Status models only measure students' current achievement status, which reflects their family's socio-economic status (SES) more than school effectiveness (Teddlie, Reynolds, & Sammons, 2000). Measurement without taking the SES of the student body into consideration is seriously biased; it puts schools in poor districts at a disadvantage, because their students are more likely to fail on exams even if they receive the same quality of instruction as students who come from affluent families (Doran & Izumi, 2004).

Improvement models, also known as "status-change" models, are not much superior to status models (Hanushek & Raymond, 2002). Improvement models compare the same grades in two

[1] The new ESSA affects many aspects of public schooling, including accountability and testing, teacher quality, research, regulation, funding, early-childhood education, and issues involving underperforming student groups. There is also provision allowing local school districts to offer a state-approved alternative high school assessment, like the ACT and SAT tests, in place of the established statewide assessment. ESSA retains the requirement that states test all students in reading and math in grades three through eight and once in high school, as well as the requirement that states ensure those tests align with states' college- and career-ready standards. However, the law makes significant changes to the role of tests in state education systems. For example, ESSA requires states to include a broader set of factors in school accountability systems rather than just test scores, provides funding for states and districts to audit and streamline their testing regimes, and allows states to cap the amount of instructional time devoted to testing. It also eliminates the requirement under the Obama administration's NCLB waiver program that states evaluate teacher performance based on, in part, student test score growth. Under ESSA, states are required to adopt "challenging" academic standards. However, states are not forced or even encouraged to pick a particular set of standards (including the Common Core State Standards).

different years (e.g., the performance of this year's fourth graders compared with last year's fourth graders). However, two cohorts of fourth graders can be quite different from one another in important characteristics, such as proportion of students with limited English proficiency or living in poverty. This is particularly true for schools and districts with high rates of student mobility (Goldschmidt et al., 2005).

Growth and Value-Added Models

As dissatisfaction with simplistic accountability models under the NCLB Act increased, growth models were proposed as a useful addition to the AYP and Safe Harbor models. The main characteristic of growth models is their use of longitudinally linked data (i.e., panel data). Growth models can also use vertically scaled measurements (i.e., a continuous measurement scale that allows comparison of scores from one grade to the next), as well as more advanced statistical modeling techniques (e.g., hierarchical linear modeling [HLM]). These techniques help eliminate many extraneous variables that lead to biased conclusions regarding a school's effectiveness, and potentially make growth models more desirable for accountability (Lissitz et al., 2006).

Growth models are praised for their fairness, but they are also criticized for setting lower standards in evaluating disadvantaged students. This is the case for a special class of growth models known as "value-added models". In value-added models (VAMs), expectations for growth are based on student demographics and other variables over which a school has no control. The model therefore allows these characteristics to be taken into account when looking at the performance of student "subgroups"—students in poverty, students with disabilities, students with limited English proficiency, and students in racial and ethnic minorities. Indeed, students who score below proficiency may still be making substantial growth from year to year, according to value-added models.

Value-added models decompose the variance of test scores into components that are explained by student inputs (i.e., adjusting for student backgrounds), and those that are presumed to be directly related to school performance. Somewhat similar to regression models that were used in school effects research in the past, value-added models hold schools accountable only for the portions of variance over which they have control (Lissitz et al., 2006); the principal difference is that value-added models normally use panel data and hierarchical modeling techniques.

In response to growing criticism of NCLB's accountability requirements, the U.S. Department of Education in 2005 announced inclusion of a growth model pilot project as part of NCLB's accountability objectives. Many states responded by proposing growth models (excluding value-added models) that include many NCLB-required goals, the most important being closing achievement gaps among groups and attaining one hundred percent proficiency by the 2013–14 school year. While holding all students to the same standards, these newly proposed growth models demonstrate flexibility by allowing disadvantaged students to catch up by the 2013–14 school year. In support of these objectives, schools and districts that are on track to achieving the long-range goals of accelerated growth and complete proficiency are exempted from negative classification even if they fall short of the intermediate goals.

Value-added accountability measures (i.e., growth measures that take into account student background characteristics), however were not accepted under NCLB, but were used for other purposes, such as: evaluating teacher performance (Ballou, 2002), improving school practice (Hershberg, Simon, & Lea-Kruger, 2004), and focusing on areas for improvement (Schatz, Von Secker, & Alban, 2005). For example, the principal intended use of the Tennessee Value-Added Assessment System was to help identify teachers who could benefit from targeted professional development, to measure teacher effectiveness, and to provide information to teachers, parents, and the public on how well schools are helping students learn (Ballou, Sanders, & Wright, 2004). For the purposes of this report, we use value-added models to estimate the effectiveness of high schools.

School districts and state education agencies across the country have relied on VAMs to measure school and teacher performance, sometimes with high stakes attached. Using value-added measures to inform high-stakes decisions is controversial, and there is not currently a consensus in the research community on the use of value-added measures for evaluation and decision making. Some of the disagreement is rooted in technical aspects and statistical properties of VAMs and their use in accountability (Harris, 2009; Braun, Chudowsky, & Koenig, 2010). In addition, the American Statistical Association (2014) has issued an official statement on the use of VAMs. It urges states and school districts to exercise caution in the use of VAM scores for high stakes purposes and offered reasonable guidelines for practice. Some have expressed skepticism (Amrein-Beardsley, 2014; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; National Research Council, 2010), but others, including prominent foundations and some think tanks, have been more positive (Bill & Melinda Gates Foundation, 2012; Glazerman et al., 2010; Gordon, Kane, & Staiger, 2006; Hanushek & Rivkin, 2004). It appears that teachers and principals trust classroom observations more than VAMs. This conclusion emerges from Goldring et al. (2015) analysis across eight districts and an in-depth case study of Chicago Public Schools (Jiang, Sporte, & Luppescu, 2015).

The general consensus is that a set of VAM scores does contain some useful information that meaningfully differentiates among teachers, especially in the tails of the distribution. However, individual VAM scores do suffer from high variance and low year-to-year stability as well as an undetermined amount of bias (Goldhaber & Hansen, 2013; Kane, McCaffrey, Miller, & Staiger, 2013; McCaffrey, Sass, Lockwood, & Mihaly, 2009), leading some to suggest that it is not reliable enough to be used for high-stakes purposes (Darling-Hammond et al., 2012). Consequently, if VAM scores are to be used for evaluation, they should not be given inordinate weight and certainly not treated as the "gold standard" to which all other indicators must be compared (Braun, 2015).

In view of the importance of VAMs, a special issue of the Journal of Educational and Behavioral Statistics (Wainer, 2004) was devoted to the careful examination of this methodology; the issue focused almost entirely on the statistical properties of the measures. More recently, a special issue of Educational Researcher (Harris & Herrington, 2015) was devoted to the careful examination of value-added–based teacher accountability. It focused on the effects of teaching and learning that come from embedding VAMs into policies like teacher evaluation, tenure, and compensation. Although teacher accountability and school accountability systems based on VAMs both use student test scores, there is much wider acceptance of school-level measures generally, and they have better statistical properties.

Results of the studies indicate that teacher value-added scores are unreliable, are sensitive to modeling choices and unstable across statistical models, years, and classes that teachers teach (Wei et al., 2012; Briggs & Domingue, 2011).

The Validity of Accountability Systems

The NCLB regulations gave states flexibility in decision making, but stipulated that they must make reliable and valid decisions regarding their accountability systems. Using "minimum cell size" and confidence intervals, most states have attended to the reliability requirement (Marion et al., 2002).

In contrast to states' attention to reliability issues, they have not equally attended to validity requirements as stipulated by the USED mandate. Validity is the most significant technical criterion to defend the quality, credibility, and fairness of an accountability system (Marion & Gong, 2003). In general, accountability system validity focuses on the accuracy and consistency of school classifications (i.e., are the "right" schools/subgroups being labeled as passing or failing?), the consequences—both positive and unintended negative—of the accountability system, and the subsequent interventions to help students, schools, and districts succeed (Marion & Gong, 2003).

There have been a number of studies evaluating the consistency and/or predictive strength of value-added models. A number of studies used a value-added approach to evaluate teachers based on their effects on their students' test scores (Hanushek, 1971; Murnane, 1975; Rockoff, 2004; Rivkin, Hanushek, & Kain, 2005; Aaronson, Barrow, & Sander, 2007; Kane & Staiger, 2008). Kane and Staiger (2008) used a random-assignment experiment in Los Angeles Unified School District to predict student achievement following random assignment of teachers to classrooms. They found that: (1) teacher effect estimates were significant predictors of student achievement, (2) those estimates that were adjusted for prior student test scores yielded unbiased predictions, and (3) those estimates that were further adjusted for mean classroom characteristics yielded the best prediction accuracy. Chetty, Friedman and Rockoff (2011), by analyzing school district data from grades 3–8 for 2.5 million children linked to tax records on parent characteristics and adult outcomes, presented evidence that value-added measures are informative about teachers' long-term impacts. For example, students assigned to high value-added teachers are more likely to attend college, attend higher-ranked colleges, earn higher salaries, live in higher SES neighborhoods, save more for retirement, and are also less likely to have children as teenagers. Allen, Bassiri and Noble (2009) evaluated the reliability of value-added measures by observing the autocorrelations of adjacent cohorts' value-added measures (one-year apart), as well as those of cohorts that are two or three years apart. They found that, in general, the correlations were larger for adjacent cohorts, and were smaller for cohorts that were two or three years apart, suggesting consistency of the measures over time.

Probably the most sophisticated value-added model is the Education Value-Added Assessment System (EVAAS) (see Ballou, Sanders, & Wright, 2004). However, a major criticism of EVAAS is that too few studies have examined the model's validity and the inferences made in its value-added reports (Braun, 2005; Amrein-Beardsley, 2008). Critics emphasize the necessity of validating estimates of teacher effects against external measures of teacher effectiveness, particularly if the inferences made are used for high-stakes decisions (Braun, 2005).

Research Question

In this study, we examine the predictive strength of value-added measures of schools' performance that are based on the ACT college and career readiness system on three measures of students' success in college, all external to the value-added measures of school effectiveness. These measures of students' success in college include:

- college enrollment in the fall after high school graduation,
- grades in first-year college courses from four core content areas (English/Language Arts, Mathematics, Natural Sciences, and Social Sciences), and
- college retention to year two.

The overarching research question is: To what extent do value-added measures of high school effectiveness predict college success? On a similar topic, ACT's cofounder E.F. Lindquist, over a half century ago when hierarchical modeling as we know it had not been developed, used various pooling strategies to determine whether scaled high school grades resulting from internal scaling (by taking into account high school attended) would improve the prediction of college grades (Lindquist, 1963). Additional prior research statistically controlled for high school attended as a joint predictor (with test scores) to predict ACT scores (Schiel, Pommerich, & Noble, 1996; Noble et al., 1999; Noble, Roberts, & Sawyer, 2006; Noble & Schnelker, 2007). The 1996 research report categorized high school attended into five categories. The categories were determined by comparing the predicted outcome (ACT score) for the students in a particular high school to the predicted outcome based on the sample pooled over high schools. The 1999 research report used effect-coded dummy variables (fixed effects) to represent each high school in traditional multiple linear regression models. The 2006 report used multilevel structural equation modeling and the 2007 report used hierarchical linear regression modeling.

Results from the ACT college and career readiness assessment system are reported on a single score scale designed to inform students, parents, teachers, counselors, administrators, and policymakers about students' strengths and weaknesses. The ACT college and career readiness system consists of ACT Explore (for eighth and ninth graders), ACT Plan® (for tenth graders), and the ACT (for eleventh and twelfth graders).[2] All three components of the ACT college and career readiness system measure academic achievement (in English, mathematics, reading, and science), and each is firmly based on the curriculum of the grade level for which it is intended and are related to the skills needed for academic success in college (ACT, 2007a, 2007b). The ACT college and career readiness system represents a consensus among educators and curriculum experts about what is important for students to know to be ready for college-level work (ACT, 2004, 2006).

In this report, high schools' effects on ACT scores (conceptualized as school's contribution to students' academic growth/performance in high school) are estimated using a value-added model. This model explicitly controls for student characteristics (incoming academic achievement level as measured by the same students' ACT Explore scores in eighth grade), the number of months between ACT Explore and ACT testing, gender, race/ethnicity, and school contextual characteristics (school size, proportion of students tested, poverty level,

[2] The ACT Explore and ACT Plan assessments are only available to existing customers through 2016 and will be replaced by the ACT Aspire™ system afterward.

proportion of racial/ethnic minority students, and mean ACT Explore scores). The analyses are based on a large sample of high school cohorts with students who took the ACT Explore and ACT tests in grades eight and eleven/twelve, respectively.

The new ESSA requires that states incorporate at least four indicators into their accountability systems, one of which could be growth on state tests (Klein, 2016). This accountability measure is closely aligned with the principal variable in the present study: high schools' effects on ACT scores (conceptualized as their contribution to students' academic growth/performance in high school) using a value-added model.

Data

School-level Data

There were 1,119 high schools for which there were up to six cohorts of available data for graduating classes of 2004 to 2009. In all, there were 2,707 cohort-by-high school combinations; on average, there were 2.4 cohorts per high school. To estimate the model, all cohorts from each individual high school were pooled together, without distinguishing which cohort they belong to. To be included in our study sample, the proportion of students who took ACT Explore and the ACT must have been at least 0.50 for a given high school cohort and the cohort size must have been at least ten. Here, *proportion tested* was defined as $N \div (Enroll_{11} + Enroll_{12})/2$, where N is the number of students who took both assessments (ACT Explore and the ACT), $Enroll_{11}$ is the high school enrollment count as of 11th grade, and $Enroll_{12}$ is the high school cohort's enrollment count of the same group of students as of 12th grade.[3] With this inclusion criterion, the sample was restricted to high school cohorts where the majority of students were represented. It is particularly noteworthy that maximizing student representation is a crucial element of any accountability system.[4]

Of the 1,119 high schools, 406 had one cohort that met the inclusion criterion, 270 had two, 169 had three, 131 had four, 128 had five, and 15 had six. Among the 2,707 high school cohorts, the median proportion tested was 0.59; the 25th percentile was 0.53; and the 75th percentile was 0.66. The mean sample size was 97; the median sample size was 51; and the 25th and 75th percentiles were 26 and 135, respectively.

In Appendix A, a US map displays the frequency of the 2,707 high school cohorts, by state. Much of the sample comes from the Midwestern and south-central US, with little representation from the eastern and western states. This is due to the fact that most schools that use both ACT Explore and the ACT are from Midwestern and south-central states. The states with the most high school cohorts represented include Illinois (479), Louisiana (381), Arkansas (380), and Oklahoma (374). To assess how well the sample represents the population of public high schools with respect to school locale, we compared the sample to all high schools in the NCES Common Core of Data for 2005–06 (Sable, Gaviola, & Garofano, 2007). Relative to the population, the sample has more high school cohorts from rural (56% versus 39%) and small

[3] High school enrollment counts of 11th and 12th grade are derived from NCES Common Core of Data for 2005–06 (Sable, Gaviola, & Garofano, 2007).

[4] If data are not available for a significant portion of students in a school, there could be concern that the resulting accountability measures are not an accurate reflection of the school's effects. Perhaps the most obvious requirement of a value-added accountability model is longitudinal test score data for students. Moreover, the standard errors of accountability measures will be larger when many students are missing from the analysis—the consequence of this is greater uncertainty about the school's effects.

town locales (15% versus 10%); relative to the population, the sample has fewer high school cohorts from the urban fringe of a city (21% versus 27%), mid-size cities (6% vs. 12%), and large cities (1% versus 11%).

In Table 1, high schools are described in terms of school size (mean of grades 11 and 12 enrollment[5]), poverty level (school's proportion of students eligible for free or reduced lunch), and proportion of minority students (school's proportion of students who are African American, American Indian, or Hispanic). Again, the sample can be compared to the general population of public high schools in the US. In the sample, the mean school size is 161.3 (standard deviation = 169.4, median = 87). The high school cohorts in the sample are somewhat larger than the typical school in the US, where the average school size is 154.5, with median 83. In the sample, the average poverty level is 0.35, with median of 0.34. These are similar to the US population average of 0.38 and median of 0.35. The sample's average proportion minority is 0.19, with median of 0.10. The sample of high school cohorts has relatively fewer high-minority schools than the US, where the mean proportion minority is 0.33, with median of 0.20.

Table 1. Summary Statistics for High Schools in Study Sample and the US

Variable	Group	Mean	SD	Min	P_{25}	Med	P_{75}	Max
School size	Sample	161.3	169.4	9	42	87	226	1,031
	US	154.5	174.2	1	27	83	232	1,617
Poverty level	Sample	0.35	0.20	0.00	0.21	0.34	0.48	1.00
	US	0.38	0.26	0.00	0.18	0.35	0.54	1.00
Proportion minority	Sample	0.19	0.22	0.00	0.03	0.10	0.27	1.00
	US	0.33	0.32	0.00	0.05	0.20	0.55	1.00

Note. n = 1,119 high schools; SD = standard deviation; min = minimum; P_{25} = 25th percentile; med = median; P_{75} = 75th percentile; max = maximum. Sample and population total adapted from NCES Common Core of Data for 2005–06 (Sable et al., 2007).

In summary, the sample of high schools is similar to the population of public high schools in the US with respect to poverty level, but has relatively fewer small and high-minority schools.

Student-level Data

Nested within the 2,707 high school cohorts there were 1,119 high schools and 263,737 students. As was noted earlier, all cohorts from each individual high school were pooled together, without distinguishing which cohort they belong to. Table 2 compares the gender and racial/ethnic group breakdowns for the sample and population of 11th grade public high school students in the US. White students are over-represented in the sample (72% versus 61%), while Hispanic (5% versus 17%), African American (8% versus 15%), and Asian American students (3% versus 5%) are under-represented. A portion of the sample (7%) has unknown or missing race/ethnicity. Females are slightly overrepresented (53% versus 49%) and males conversely are slightly underrepresented (46% versus 50%).

[5] From population of all public high schools with grades 11 and 12 enrollment that could be located in NCES Common Core of Data for 2005–06 (Sable, Gaviola, & Garofano, 2007).

Table 2. Race/Ethnicity and Gender in the Sample and the US

	Gender			Total	
Race/Ethnicity	**Female**	**Male**	**Missing**	**Sample**	**US**
African American	12,852 9%	8,955 7%	52 1%	8%	15%
Asian American	3,758 3%	3,796 3%	43 1%	3%	5%
Hispanic	6,764 5%	5,699 5%	53 1%	5%	17%
White	101,780 73%	88,471 73%	446 11%	72%	61%
Other	6,403 5%	5,468 5%	36 1%	3%	2%
Missing	7,427 5%	8,063 7%	3,671 85%	7%	0%
Sample total	53%	46%	1%	100%	
US total	49%	50%	1%		100%

Note. n = 263,737; Population total adapted from 11th grade totals in NCES Common Core of Data for 2005–06 (Sable et al., 2007).

Our research is based on data from students who took ACT Explore and the ACT. We only included students who took ACT Explore and the ACT[6] (pre and near-end high school assessment) within 28 to 72 months apart, inclusively. For students who had duplicate/multiple records, we only kept the record closest to 48 months apart. With this criterion, relatively few students were excluded for taking ACT Explore and the ACT too far apart or too close in time. The number of months between ACT Explore and ACT testing ranged from 28 to 66; the median was 45 months, the 25th percentile 41 months, and the 75th percentile was 49 months.

In Table 3, the student sample is described with respect to ACT Explore and ACT test scores, as well as composite scores. The average ACT Explore scores in the sample range from 15.9 in reading to 17.5 in science. Nationally, for 2009 ACT Explore-tested eighth grade students, the mean scores ranged from 14.1 in reading to 16.2 in science (ACT, 2009a) and for grade nine ACT Explore-tested students, the mean scores ranged from 15.1 in reading to 17.0 in science (ACT, 2009b). The average ACT scores in the sample range from 21.2 in mathematics to 21.6 in reading. Nationally, for 2009 ACT-tested high school graduates, the mean scores ranged from 20.6 in English to 21.4 in reading (ACT, 2009c); the student sample appears to be quite typical of ACT Explore-tested or ACT-tested populations in terms of academic achievement.

[6] Note that these students are more likely to attend college than are students generally.

Table 3. Summary Statistics of Students' ACT Explore and ACT Test Scores

Test	Mean	SD
ACT Explore		
English	16.2	3.9
Mathematics	16.5	3.4
Reading	15.9	3.7
Science	17.5	2.8
Composite	16.6	3.0
ACT		
English	21.4	5.9
Mathematics	21.2	5.1
Reading	21.6	5.9
Science	21.3	4.7
Composite	21.5	4.9

Note. n = 263,737

College Enrollment and Retention Data

Data from the National Student Clearinghouse (NSC) were used to identify students who enrolled in college the fall after high school graduation (first year enrollment) and who re-enrolled at the same or a different postsecondary institution the second fall after high school graduation (retention). Enrollment information from the NSC is limited to participating postsecondary institutions.[7] To account for these limitations, as discussed below, indicator variables were created to represent whether a student's first or second choice college was excluded from NSC data. Retention data is available for 2,701 high school cohorts in the sample. Overall, 71% of the students in the sample enrolled at an NSC institution their first year after graduation from 2004 to 2009. Twenty-five percent of these enrollees (47,491 students) enrolled in 2009 for which year-two enrollment data were not available at the time the analyses were done. Thereby, we excluded 2009 enrollees from our retention analyses and only considered enrollees of 2004 to 2008. Likewise, 85% of the students returned to an NSC institution their second year following their high school graduation (from 2005 to 2009); the majority of these students (87%) reenrolled at the same college and only 13% reenrolled at a different college.

When students register for the ACT, they specify their first and second-choice college. Students whose first and second-choice colleges were not among those included in the NSC data were identified and indicator variables were created to represent whether a student's first or second choice college was excluded. By doing so, the analysis was adjusted to accommodate for the fact that not all enrollments were included in the NSC data set. Of those whose first- and second-choice college was not included in the NSC data, 61% and 65% enrolled at a NSC institution, respectively. For those whose first- and second-choice colleges

[7] As of 2013, more than 3,400 colleges and universities, enrolling over 96% of all students in public and private U.S. institutions, participate in the NSC.

were not included in the NSC data, the retention percentage was 84%. Because, college choice variables are included in the model to account for limitations of the NSC data and to calibrate the effects of the predictors of interest, the relationship of the choice variables to college enrollment should be strong and positive by design.

College Course Grade Data

First-year college course grade data were collected across multiple years from postsecondary institutions participating in ACT's Course Placement Service (CPS) or ACT's Prediction Service (hence, college course grade data were available for a subset of college enrollees). As part of ACT's CPS, a list of course content areas and placement level codes are given to participating postsecondary institutions. Given the variety of names for similar courses, ACT requests that each institution assign a course content code for each course and specify whether the course is a developmental, standard, or honors course. Assignment of course content and placement level codes to data received through Prediction Service is done at ACT based on the course title provided by the participating institutions.

First-year college courses are classified into eight content areas: Fine Arts (five courses), Business (four courses), English/Language Arts (eight courses), Foreign Languages (eight courses), Mathematics (13 courses), Social Sciences (13 courses), Natural Sciences (15 courses), and Miscellaneous (eight courses). Placement codes are classified as Developmental/Remedial (DR), Standard (ST), and Honors (HO).

For various reasons listed below, we estimated models from a subset of the course grade data set:

- We only included records of courses that had valid course grades (A^+ to F)[8] and valid scores for all predictor variables in the model.
- We only included institutions with known two-year or four-year type (institution type was unknown for some institutions).
- Some students take a course more than once, sometimes under different placement codes (DR, ST, and HO). We eliminated duplicate records according to the following rubric (while ordering placement codes from lowest to highest as DR, ST, and HO for duplicate records having multiple placement codes): (a) for duplicate records without enrollment dates, the record with the lowest course grade from the lowest placement code was kept; and (b) for duplicate records with enrollment dates, we kept the record from the earliest enrollment date and from the lowest placement code.
- Finally, we eliminated records of courses with one or more intervening developmental courses that either preceded their enrollment date or enrollment dates were missing. For example, reading is considered an intervening developmental course for all social science courses. If a student took a course in reading that was classified as Developmental/Remedial prior to taking an American History course or took them concurrently, the American History course for that student was flagged and eliminated from the course data set. In all, we eliminated about 4% of the course data (5,255 records).

[8] Breakdown of the letter grades A^+ to F, numerically correspond to: A^+ = 4.33, A = 4.00, A^- = 3.67, B^+ = 3.33, B = 3.00, B^- = 2.67, C^+ = 2.33, C = 2.00, C^- = 1.67, D^+ = 1.33, D = 1.00, D^- = 0.67, F = 0.00.

In all, there were 26,863 students with first-year college course data; the median number of courses taken was four, the 25th percentile was two, and the 75th percentile was seven. Of 231 participating postsecondary institutions, 139 were four-year colleges (versus 92 two-year colleges). The majority of courses (71%) were taken from four-year colleges (versus 29% from two-year colleges). The four-year colleges had mostly traditional (65%) and selective (24%) admissions policies, about 11% had open or liberal admissions policies, and less than 1% were highly selective. In contrast, the majority of two-year colleges had open admissions policies (97%) and only 3% had traditional or liberal admissions policies.[9] The states with the most college course data represented include Arkansas (51%) and Oklahoma (38%). Illinois and Louisiana, the other two states with the most high school cohorts, only represented 4% and 2% of college course data, respectively.

In our study sample, we only included courses from four core content areas (English/Language Arts, Mathematics, Natural Sciences, and Social Sciences); courses from the non-core content areas were not considered. As expected, percentages of course grade data from core content areas were higher than that from non-core content areas (83% versus 17%). The 83% core courses taken include 26% from English/Language Arts, 17% from Mathematics, 14% from Natural Sciences, and about 26% from Social Sciences. The 17% non-core courses taken include 6% from Fine Arts, 1% from Business, about 1% from Foreign Languages, and 9% from the Miscellaneous category. With this inclusion criterion, the number of postsecondary institutions included in the course grades analyses was reduced to 226, only affecting four-year institutions (reduced to 134 from 139); the number of two-year community colleges remained at 92.

Table 4 lists first-year college courses in the four core content areas, total students' enrollment in each course, percentages of students' enrollment at two-year versus four-year colleges, and their success percentages in each course (defined as earning a course grade of 3.0 [B] or better). The highest enrollment was in Composition I (15%), followed by Composition II, College Algebra, and American History (7% each); the lowest enrollments were in Archaeology and Geometry.

[9] According to the Institutional Data Questionnaire (IDQ), the admission policy categories are defined as: Highly selective (the majority of students rank in the top 10% of their high school class), selective (the majority of students rank in the top 25% of their high school class), traditional (the majority of students rank in the top 50% of their high school class), liberal (the majority of students rank in the top 75% of their high school class), and open admissions (that virtually anyone is eligible for general admission, regardless of previous academic record or grades).

Table 4. List of First-Year College Courses in Core Content Areas

	Enrollment (N)		Institution Type (%)		Grade Status (%)	
Course Content	**Student**	**Institution**	**2-Year**	**4-Year**	**Failure**	**Success**
English/Language Arts						
Grammar	683	38	76	24	55	45
Reading	2,451	74	64	36	44	56
Composition I (1st writing course)	15,153	162	36	64	37	63
Composition II (2nd writing course)	7,551	66	31	69	32	68
Literature	1,225	42	27	73	36	64
Speech/ Rhetoric	4,572	76	30	70	27	73
Film Criticism/History	505	13	9	91	40	60
Other	336	28	40	60	52	48
Mathematics						
Arithmetic Skills	445	34	81	19	56	44
Elementary Algebra	2,394	85	67	33	63	37
Intermediate Algebra	2,248	74	41	59	63	37
College Algebra	7,569	108	26	74	50	50
Geometry	4	3	75	25	25	75
Analytic Geometry	40	2	0	100	38	63
Trigonometry	727	39	14	86	51	49
Pre-Calculus/Finite Math	817	21	2	98	54	46
Calculus	1,652	42	2	98	47	53
Computer Science	3,810	62	49	51	37	63
Statistics/Probability	157	14	48	52	44	56
Logic	93	7	1	99	46	54
Other	1,159	55	25	75	55	45
Natural Sciences						
General Science	65	6	60	40	26	74
Biology/Life Sciences	5,707	92	24	76	51	49
General Chemistry	3,357	53	11	89	46	54
Physics (without Calculus)	378	23	62	38	36	64
Physics (with Calculus)	208	13	4	96	27	73
Botany	114	15	21	79	56	44
Conservation/Ecology	56	11	43	57	48	52
Engineering	708	14	3	97	23	77
Zoology	717	23	15	85	54	46
Anatomy/Physiology	689	44	61	39	63	37
Health Sciences	3,524	58	24	76	30	70
Astronomy	200	11	17	83	52	48
Geology	679	23	7	93	42	58
Meteorology	22	4	5	95	41	59
Other	877	37	29	71	38	62

(continued)

Table 4. (continued)

	Enrollment (N)		Institution Type (%)		Grade Status (%)	
Course Content	**Student**	**Institution**	**2-Year**	**4-Year**	**Failure**	**Success**
Social Sciences						
American History	7,165	73	31	69	45	55
Other History (World, Western, etc.)	3,527	54	22	78	48	52
Psychology	6,127	89	41	59	38	62
Sociology	4,632	66	23	77	35	65
Geography	907	33	15	85	51	49
Anthropology	638	17	5	95	30	70
Archaeology	1	1	0	100	0	100
Political Science	4,940	56	34	66	41	59
Economics	896	33	18	82	51	49
Law	635	36	20	80	50	50
Philosophy	1,811	44	13	87	35	65
Religion	339	17	11	89	40	60
Other	152	18	6	94	29	71
Total	102,662					

Notes. N = Number; % = Percent.

With the exceptions of Grammar, Reading, Arithmetic Skills, Elementary Algebra, Geometry, General Science, Physics (without calculus), and Anatomy/Physiology, most of the sample came from four-year colleges. The highest percentages of enrollment at two-year colleges were in Grammar (76%), Geometry (75%), and in Arithmetic Skills (81%). All enrollments in Analytic Geometry (n = 40) and Archaeology (n = 1) were at four-year colleges (100%); and at least 95% of enrollments in Pre-Calculus/Finite Math, Calculus, Logic, Physics (with Calculus), Engineering, Meteorology, and Anthropology were at four-year colleges.

The overall percentage of success in first-year college courses was 60%. As shown in Table 4, success percentages for the first-year courses in the four content areas (excluding courses with low n-counts) ranged from 45% to 73% in English/Language Arts and 49% to 71% in Social Sciences, compared to 37% to 63% and 37% to 77% in Mathematics and Natural Sciences, respectively. The lowest success percentages were in Elementary and Intermediate Algebra, and in Anatomy/Physiology (each 37%). In College Algebra and Law, success and failure percentages were equally divided. Overall success percentages in the four core content areas were highest in English/Language (65%) and lowest in Mathematics (50%); in Social Sciences and Natural Sciences they were 58% and 56%, respectively.

Generally, with respect to success percentages, the four core content areas may be ordered (from highest to lowest), as English/Language Arts, Social Sciences, Natural Sciences, and Mathematics. It should be noted that the differential success percentages by content areas could well be affected by differential grading practices across departments in postsecondary institutions. Bassiri and Schulz (2010) showed that scaling approaches to college grade data might be useful in identifying differential grading practices within and across departments in postsecondary institutions.

Value-Added Model

High schools' effects on ACT scores (conceptualized as the school's contribution to students' academic growth in high school) are estimated using a value-added model (Equation 1). The schools' effects are estimated by explicitly controlling for student-level covariates X_1, X_2, X_3, X_4 (incoming academic achievement level as measured by the same students' ACT Explore subject area test scores in grade eight), the number of months between ACT Explore and ACT testing X_5, and school-level covariates S_1, S_2, S_3, S_4, S_5 (school size, proportion of students tested, poverty level, proportion of racial/ethnic minority students, and mean of ACT Explore scores).[10] Additionally, the school effect is denoted as τ, and ε is the residual error for the regression model. The school effects and residual errors are assumed to be normally distributed and independent with mean 0.0 and unknown variances.

In this study, we considered students who had tested in eighth and eleventh or twelfth grade. In order to measure the effect that high schools have on student learning, an entry and an exit score are necessary. Because students typically take ACT Explore in eighth grade, those scores are the natural choice for an entry score; likewise, ACT scores are the natural choice for exit scores. Ideally, ACT Explore would be taken at the end of eighth grade; otherwise, the measured high school effect would include the portion of learning that took place in grade eight after ACT Explore was taken. Similarly, the ACT would ideally be taken at the end of 12th grade; otherwise, the measured high school effect would not include the portion of learning that took place in grade 11 or 12 after the ACT was taken. Because of the requirements of college applications, very few students choose to take the ACT at the end of grade 12. Therefore, it is likely that measures of high school effects would only include the portion of learning that took place through the time of ACT testing. In order for accountability measures to be truly comparable across schools, it is necessary for the assessments to be spaced in a similar fashion. For example, it might be misleading to compare academic growth from the beginning of grade eight to the beginning of grade 12 at school "A" to academic growth from the end of grade 8 to the middle of grade 11 at school "B". In this case, students at school "A" might be expected to show greater growth due to the larger time span. When implementing a value-added model, care should be taken to account for different time spacing of assessments across schools. In this study, this problem is addressed by introducing a covariate in the model (Equation 1) that accounts for varying time spans. This model is a special case of a HLM (Raudenbush & Bryk, 2002) and can be fit using statistical software packages such as HLM® or SAS®. Because under the assumed value-added model, the "average" school effect is always zero, the school effect can also be interpreted here as the number of ACT score points attributable to a school, above and beyond what can be attributed for the average school. Note that this method only requires ACT Explore and ACT scores and does not utilize the vertical scaling of the ACT college and career readiness system test scores.

$$ACT_{ij} = \beta_0 + \sum_{p=1}^{5} \beta_p X_{ijp} + \sum_{r=1}^{5} \gamma_r S_{jr} + \tau_j + \varepsilon_{ij} \tag{1}$$

In Equation 1, ACT_{ij} is the ACT score for the i^{th} student from the j^{th} high school, β_0 is the overall intercept term, X_{ijp} (p = 1,2,3,4,5) are the prior test scores in the four subject areas and the

[10] Clearly, fully accounting for contextual characteristics requires additional student-level data, such as absenteeism, dropouts/transfer, family income, and parent's education level, that may not be readily available or reliably measured.

number of months between ACT Explore and ACT testing for the i^{th} student from the j^{th} high school, β_p (p = 1,2,3,4,5) are their corresponding regression coefficients. Similarly, S_{jr} (r = 1, 2, 3, 4, 5) are the five school-level covariates for the j^{th} high school and γ_r are their corresponding regression coefficients. Finally, τ_j is the school effect for the j^{th} high school and ε_{ij} is the random error term for the i^{th} student from the j^{th} high school.[11]

Because the model adjusts for student and school characteristics that are outside of the school's control, the resulting school effect estimate is usually interpreted as the high school's contribution to students' academic performance. This model can be fit for each of the four ACT subject tests, resulting in estimated school effects on students' academic performance in English, mathematics, reading, and science.

Table 5 summarizes the distributions of the value-added measures generated by Equation 1 for each of the four ACT subject tests, as well as ACT Composite. It is apparent that there is considerable variation across high school cohorts in value-added measures of school effects on the ACT (ranging from SD = 0.47 for Science to SD = 0.73 for English). This suggests that there is greater consistency across high schools in their influence on science performance relative to English performance. Under the assumed value-added model, using the standard deviation of each ACT test score,[12] the difference between attending a high school at the 25th percentile relative to the 75th percentile is about 0.17 standard deviations in English, 0.18 in Mathematics, 0.13 in Reading and Science, and 0.15 in ACT Composite test scores.

Table 5. Distributions of Estimated School Effects on ACT Scores

	Estimate of School Effect on ACT Score					
Subject	**Min**	**P_{25}**	**Med**	**P_{75}**	**Max**	**SD**
English	-2.27	-0.49	-0.04	0.46	2.95	0.73
Mathematics	-2.02	-0.46	-0.03	0.43	2.38	0.67
Reading	-1.79	-0.39	-0.01	0.37	2.21	0.57
Science	-2.35	-0.31	-0.01	0.29	1.93	0.47
Composite	-1.96	-0.36	-0.03	0.35	2.11	0.55

Note. n = 2,707 high school cohorts: Min = minimum; P_{25} = 25th percentile; Med = median; P_{75} = 75th percentile; Max = maximum; SD = standard deviation.

Maximizing student representation is a crucial element of any accountability system. Moreover, the standard errors of accountability measures will be larger when many students are missing from the analysis—the consequence of this is greater uncertainty about the school's effects. The standard errors of accountability measures can be quite large, even when all students are counted in the calculations. Naturally, this problem is more pervasive at smaller schools. Because of this problem, standard errors of accountability measures should be reported, especially when the measures are used for high-stakes decisions.

[11] There were 1,119 high schools for which there were up to six cohorts of available data for graduating classes of 2004 to 2009. In all, there were 2,707 cohort-by-high school combinations. To estimate the model, all cohorts from a given high school were pooled together, without distinguishing which cohort they belong to.

[12] Standard deviation of ACT test scores adopted from the ACT technical manual are 5.70 in English, 5.00 in Mathematics, 5.82 in Reading, 4.51 in Science, and 4.68 in ACT Composite (ACT, 2007c).

Under the assumed value-added model, the “average” school effect is always 0. The 25th and 75th percentiles of the estimated school effects can give us a rule of thumb of what constitutes a “good” score for a high school and what constitutes a “poor” score. For example, only 25% of the high schools have an English effect larger than 0.46; 0.46 could be considered a good score for the number of ACT English score points that could be attributed to a high school, over and above what could be expected of an “average” high school. Similarly, -0.31 could be considered a poor score for the number of ACT Science score points that could be attributed to a high school.

Table 6 contains the inter-correlations of the estimated school effects on ACT scores. These correlations suggest that value-added measures are correlated across subject areas, and that high schools that score well in one area will likely score well in other areas.

Table 6. Intercorrelations of School Effects on ACT Scores

Estimated School Effect on . . .	1.	2.	3.	4.	5.
1. English	1.00				
2. Math	0.58	1.00			
3. Reading	0.75	0.55	1.00		
4. Science	0.67	0.67	0.75	1.00	
5. Composite	0.87	0.81	0.88	0.88	1.00

Note: n = 1,119 high school.

Relationships with Prior Mean Academic Achievement and School Contextual Characteristics. Because the model adjusts for school characteristics (Equation 1), the school effect measures generated by the model are unrelated to school size, proportion of students tested, poverty level, and proportion of racial/ethnic minority students, and prior mean academic achievement. In fact, because context-adjusted value-added measures have no association with school characteristics that are included in the adjustment, they are more likely to be accepted as fair measures of school effects. Allen, Bassiri and Noble (2009) showed that schools’ value-added scores (estimated effects on ACT scores) are not likely to be influenced much by whether contextual characteristics are statistically controlled. In other words, schools that are considered above average using the context-adjusted model will most likely be considered above average using the non-context-adjusted model.

Assessing the Predictive Strength of Value-added Scores on College Success

Statistical evidence can be used to support the argument that school value-added scores can be used as markers of schools’ effects on college readiness. An example of such evidence would be regression model results that show that value-added scores are significant predictors of college outcomes within a multiple regression model that accounts for prior student achievement and other characteristics. In our investigation of the predictive strength of value-added measures of schools’ performance on students’ success in college, we examined three college outcomes: (a) college enrollment in the fall after high school graduation; (b) grades in first-year college courses from four core content areas (English/Language Arts, Mathematics, Natural Sciences, and Social Sciences); and (c) college retention to year two. As described earlier, college course grade data were only available for a sample of the original group, thus the course grade analyses have a much smaller sample size (n = 26,863) than the enrollment and retention analyses (n = 263,737). In this study the nesting structures for the college outcomes are conceptualized as:

a) College enrollment—Students are nested within high schools

b) College retention—Students are nested within colleges

c) First-year college course grades—Students are nested within colleges

College Enrollment and College Retention

This section begins with a brief overview of the two-level hierarchical logistic models that are used to model college enrollment and college retention, followed by results of the analyses investigating the predictive strength of value-added measures on college enrollment and college retention.

Two-Level Hierarchical Logistic Models. Two-level hierarchical logistic regression with random intercepts was used to model college enrollment and college retention (binary outcomes). The hierarchical logistic model differs from the hierarchical linear model (HLM) only in the specification of the response distribution and the link function. The logistic hierarchical model uses a Bernoulli response model and a logit link function instead of a normal response model with a linear link function. (For detailed information about these models, see Raudenbush & Bryk, 2002; Snijders & Bosker, 1999.)

College Enrollment. A two-level hierarchical logistic regression model with random intercepts was used to model college enrollment (Equation (2)). Note that I am re-using symbols and Greek letters across models for convenience (though they are not the same parameters because the outcome variables in the models are different). Gender, race/ethnicity, and students' first- and second-choice college are introduced as student-level covariates ($B_1, B_2, \ldots, B_7$). Note that B_1 is the coefficient for male gender, thus making female students the reference group; racial/ethnic minority students is a five-category nominal variable (African American, Hispanic, Asian American/Pacific Islander, Other including American Indian, and White), fully captured with four dummy-coded covariates and making white students the reference group. Finally, B_6 and B_7 are the coefficients associated with indicator variables for whether first and second college choice was or was not included in the NSC data.

$$\eta_{enrolled ij} = \beta_0 + \sum_{p=1}^{5} \beta_p X_{ijp} + \sum_{q=1}^{7} \lambda_q B_{ijq} + \sum_{r=1}^{5} \gamma_r S_{jr} + \theta_j + \alpha \hat{\tau}_j \qquad (2)$$

In Equation 2, $\eta_{enrolled ij}$ is the log of the odds of enrollment for i^{th} student in the j^{th} high school and θ_j is the high school random intercept that can vary among high schools, $\hat{\tau}_j$ are estimated high school value-added measure (the predictors of interest in this model), and α is the regression coefficient (associated with the value-added measures) which is the parameter of interest. The probability P of college enrollment can be obtained by algebraic manipulation of $\eta_{enrolled ij}$, that is $P = \dfrac{1}{1 + \exp(-\eta_{enrolled ij})}$.

College Retention. Two-level hierarchical logistic regression with random intercepts was used to model college retention.

$$\eta_{re\text{-}enrolled_{ik}} = \beta_0 + \sum_{p=1}^{5} \beta_p X_{ip} + \sum_{q=1}^{5} \lambda_q B_{iq} + \sum_{r=1}^{5} \gamma_r S_r + \theta_k + \alpha \hat{\tau}_k \quad (3)$$

Note that Equation 3 for college retention ($\eta_{re\text{-}enrolled_{ik}}$) is very similar to Equation 2 except it doesn't include B_6 and B_7 covariates for students' first- and second-choice college. In Equation 3, $\eta_{re\text{-}enrolled_{ik}}$ is the log of the odds of re-enrollment for i^{th} student in the k^{th} college, θ_k is the college random intercept that can vary among colleges, $\hat{\tau}_k$ are estimated college value-added measure (the predictors of interest in this model), and α is the regression coefficient (associated with the value-added measures) which is the parameter of interest. Note that the term θ_k in Equation (3) (a college random effect) is not the same as θ_j in Equation (2) (a high school random effect), as the outcome variables in the two models are different. The probability P of college re-enrollment can be obtained by algebraic manipulation of $\eta_{re\text{-}enrolled_{ik}}$ that is $P = \frac{1}{1 + \exp(-\eta_{re\text{-}enrolled_{ik}})}$.

Predictive Strength of Value-Added Measures on College Enrollment and College Retention. Table 7 contains raw (unmodeled) enrollment and retention percentages by race/ethnicity and gender. In all, a greater percentage of female than male students enrolled and reenrolled in college the first and second year after high school graduation. Overall, enrollment and retention percentages were the highest for Asian American students followed by White students; Hispanic students enrolled at the lowest percentage, but their retention percentage was higher than that of African Americans, who had the lowest retention percentage across ethnic groups. Note that students re-enrolled at the same college at a lower percentage than they do at any college the second fall after high school graduation (by 8% to 12% across racial/ethnic group). This was expected as the former cohort is a subset of the latter cohort.

Table 7. College Enrollment and Retention Percentages by Race/Ethnicity and Gender

	College Enrollment (%)			College Retention (at any institution) (%)			College Retention (at same institution) (%)		
Race/Ethnicity	**Female**	**Male**	**Overall**	**Female**	**Male**	**Overall**	**Female**	**Male**	**Overall**
African American	70	64	67	79	76	78	68	64	66
Asian American	81	78	79	93	91	92	84	83	84
Hispanic	57	53	55	81	78	80	72	67	70
White	75	71	73	87	84	86	76	73	75
Other	64	60	62	82	80	81	70	68	0.69

Note. Enrollment data are based on high school graduates who enrolled in a postsecondary institution from 2004 to 2009 covered in National Student Clearinghouse data (n = 186,632). Retention data (n = 118,139) are based on students who reenrolled in a postsecondary institution from 2004 to 2008 following their second year graduation from high school.

Table B-1 in Appendix B contains point-biserial correlations between college enrollment and retention and characteristics of high schools and students (all statistically significant at $p < .0001$). As we can see, the high school effect measures are positively related to college enrollment and retention. The correlation coefficients also indicate that students' eighth grade academic achievement has the strongest relationships with enrollment and retention.

The correlations are larger for enrollment (0.19 to 0.21) than for re-enrollment (ranging from 0.14 to 0.16), suggesting that retention is less influenced by students' prior academic achievement.

As we will discuss later, the majority of overall variance in college enrollment and retention is due to students' characteristics; less of the variance is due to the characteristics of high schools or colleges. Enrollment and retention are also positively related to prior mean academic achievement, school size and proportion of students tested, and as expected, are negatively related to schools' poverty level and proportion of minority students. Not surprisingly, college enrollment is positively related to whether students' first and second choice colleges were included in the NSC enrollment data.

If the value-added measures are valid as markers of a school's effect on college readiness, they should have statistical relationships with enrollment and retention percentages. Further, if the value-added measures are truly measuring the high school's contribution to college readiness, the statistical relationships should persist after adjusting for the high school cohort's prior mean academic achievement, as well as contextual characteristics (school size, proportion tested, high school poverty level, and proportion of racial/ethnic minority students in the school).

We used logistic regression to evaluate the effect of value-added estimates of school performance on the probability of enrollment and retention, controlling for selected student and school-level covariates. The response variable for college enrollment is a dichotomy distinguishing between enrolled (1) and not enrolled (0); and college retention is a dichotomy distinguishing between re-enrolled (1) and not re-enrolled (0). Note that the nesting structure for enrollment data is students nested within high school (n = 1,119), with a small but statistically significant intra-class correlation coefficient (ICC[13]) estimate of 0.08. Similarly, for re-enrollment at any institution or at the same institution the nesting structure is students nested within college (n = 1,701), with statistically significant ICC estimates of 0.13 and 0.21, respectively. Thus, the majority of overall variance in the propensity to enroll and re-enroll in college is due to students' characteristics; less of the variance is due to high schools (for college enrollment) or colleges (for retention).

Table 8 summarizes logistic regression estimates and standardized logistic regression estimates (beta coefficients) for college enrollment and college retention at any (or at the same) postsecondary institution. All school-level and student-level predictors were grand-mean centered. Associated with each estimated fixed logistic regression effect is an estimated standard error (SE), which measures the precision of the estimated fixed effect. The results show that coefficients for student and school-level covariates are predominantly positive (statistically insignificant coefficients, $p > 0.05$ are marked as 'ns').

The third columns under the college enrollment and retention headings in Table 8 contain the beta weights (standardized regression coefficients) for the school and student level covariates. The beta weights not only tell us each characteristic's association with the outcomes, beyond that explained by other student and school characteristics, but also the relative importance of these characteristics/covariates (after transforming all covariates to have variance of 1).

[13] ICC for enrollment and retention are calculated by setting Level 1 variance to be $\sigma^2 = \pi^2/3$, the variance of the logistic distribution (Snijders & Bosker, 1999).

The school effect estimates (value-added measures) are predictive of college enrollment and retention percentages beyond what is already predicted by the other model variables.

Table 8. Estimated Coefficients for Predicting Enrollment and Retention

	College Enrollment			College Retention (at any institution)			College Retention (at same institution)		
Predictor	**Estimate (unstd.)**	**SE**	**Estimate (std.)**	**Estimate (unstd.)**	**SE**	**Estimate (std.)**	**Estimate (unstd.)**	**SE**	**Estimate (std.)**
Intercept ($\hat{\beta}_0$)	0.88	0.02	1.07	1.85	0.03	1.95	0.93	0.03	0.98
High School-level Covariate									
School effect on ACT Composite	0.23	0.02	0.14	0.12	0.02	0.07	0.05	0.01	0.03
Mean ACT Explore Composite	0.04	0.01	0.04	0.02^{ns}	0.01	0.02^{ns}	-0.02^{ns}	0.01	-0.02^{ns}
School size	0.00	0.00	0.16	0.00	0.00	0.06	0.00	0.00	0.04
Proportion tested	0.28	0.10	0.02	0.67	0.13	0.05	0.51	0.11	0.04
Poverty level	-0.58	0.10	-0.11	-0.86	0.09	-0.16	-0.55	0.07	-0.10
Proportion minority	0.17^{ns}	0.09	0.03^{ns}	0.34	0.07	0.07	0.18	0.06	0.03
Student-level Covariate									
ACT Explore									
English	0.04	0.00	0.16	0.01	0.00	0.05	0.00^{ns}	0.00	0.02^{ns}
Mathematics	0.07	0.00	0.23	0.06	0.00	0.18	0.03	0.00	0.09
Reading	0.03	0.00	0.12	0.02	0.00	0.07	0.01	0.00	0.02
Science	0.05	0.00	0.12	0.03	0.00	0.08	0.02	0.00	0.06
Time span	0.07	0.00	0.32	0.03	0.00	0.12	0.01	0.00	0.02
First college choice	0.61	0.01	0.29	NA	NA	NA	NA	NA	NA
Second college choice	0.13	0.01	0.07	NA	NA	NA	NA	NA	NA
Male	-0.18	0.01	-0.09	-0.25	0.02	-0.12	-0.20	0.01	-0.10
African American	0.07	0.02	0.02	-0.16	0.03	-0.05	-0.17	0.03	-0.04
Hispanic	-0.45	0.02	-0.10	-0.22	0.04	-0.04	0.01^{ns}	0.04	0.00^{ns}
Asian American	0.14	0.03	0.02	0.28	0.06	0.05	0.30	0.05	0.05
Other	-0.24	0.02	-0.05	-0.20	0.04	-0.04	-0.12	0.03	-0.02
Variance of intercept ($\hat{\tau}_{00}$)	0.27	0.02		0.51	0.04		0.86	0.06	
Intraclass correlation coefficient (ICC)	0.10	0.00		0.22	0.01		0.27	0.01	
N	243,946			127,783			127,783		

Note. unstd = unstandardized; std = standardized.; NA = not applicable. Statistically non-significant coefficients (p > 0.05) are marked as 'ns.'

The baseline predicted probability of college enrollment (for a White female student with average values for all student and school covariates and whose first and second choice college was included in the NCS data set) is 0.74 suggesting that on average 74% of such students enroll in college the fall after high school graduation.[14] The probability of enrollment increases with the high school effect estimate and for students with higher ACT Explore scores, longer time span between ACT Explore and ACT testing whose first and/or second college choice are included in the college enrollment data. The effect of proportion minority in high school is positive but not statistically significant.

[14] The expected log-odds of enrollment is 1.07, corresponding to a probability of 1/(1 + exp{-1.07}) = 0.74.

We hypothesized that higher enrollment and retention percentages will be associated with the higher *school effect estimate on ACT Composite*. The model suggests that *school effect* is associated with a higher log-odds of enrollment, 0.14 (holding constant the other variables in the model and the random effect, ε). For each one standard deviation increase in the *school effect*, the log-odds increase by 0.14, leading to a higher predicted log-odds of 1.07 + (0.14) = 1.21, associated with a higher predicted probability of 0.77. These are the results for a typical White female student with average student and school characteristics whose first and second choice college was included in the NCS data set.

The baseline predicted probabilities of re-enrolling at any or at the same institution are 0.88 and 0.73, respectively.[15] Again, as expected, students re-enroll at the same college at a lower percentage than they do at any college: On average, 88 out of 100 re-enroll at any college and 73 out of 100 students re-enroll at the same college the second fall after high school graduation. The retention probabilities increase with the high school effect estimate, school size, proportion tested, and proportion minority, and is greater for those with higher ACT Explore scores and longer time span between ACT Explore and ACT testing. The model suggests that *school effect* is associated with a higher log-odds of retention at any or at the same institution, by 0.07 and 0.03, respectively (holding constant the other variables in the model and the random effect, ε). The expected change in the retention at any or at the same institution, given two standard deviations increase in the *school effect*, while holding the other predictors constant lead to higher predicted log-odds of 1.95 + (2) * (0.07) = 2.09 at any institution and .98 + (2) * (.03) = 1.04 at the same institution, associated with higher predicted probabilities of 0.89 and 0.74, respectively.

Additionally, we hypothesized that higher enrollment and retention percentages will be associated with the higher ACT Explore scores. The model suggests that ACT Explore scores especially in Mathematics, are positively and significantly related to enrollment (log-odds = 0.23) and re-enrollment percentages at any (log-odds = 0.18) and the same college (log-odds = 0.09). One standard deviation increase in ACT Explore Mathematics scores, will lead to predicted probability of 0.79 for enrollment, 0.89 for re-enrollment at any college and 0.74 for re-enrollment at the same college. We also expect that schools with a higher poverty level will have lower enrollment and retention percentages, indicating that the probabilities are negatively related to schools' proportion eligible for free or reduced lunch, controlling for other covariates in the model. For example, increasing *poverty level* by one standard deviation will lower the predicted probability to 0.72 for enrollment; and to 0.86 and 0.71 for re-enrollment at any or the same college, respectively. Again, these are the results for a typical White female student with average student and school characteristics whose first and second choice college was included in the NCS data set.

As we can see in Table 8, male students enrolled and re-enrolled at any (or at the same) postsecondary institution at a lower percentage than female students. The model suggests that being male is associated with a lower log-odds of enrollment (-0.09), a lower log-odds of re-enrollment at any institution (-0.12), and lower log-odds of re-enrollment at the same institution (-0.10), holding constant the other covariates in the model. Additionally, the model suggests that compared to White students, Asian American students have higher enrollment

[15] The expected log-odds of re-enrolling at any institution is 1.95, corresponding to a probability of 1/(1 + exp{-1.95}) = 0.88. Similarly, the expected log-odds of re-enrolling at the same institution is 0.98, corresponding to a probability of 1/(1 + exp{-0.98}) = 0.73.

and retention percentages; Hispanic students and students from Other ethnic groups have lower percentages in both enrollment and retention; and African American students enroll at a higher percentage, but their retention percentages are lower. Furthermore, we see here that schools' grade eight mean academic achievement had a significant positive relationship with enrollment percentages, but the effects on re-enrollment were not strong enough to reach statistical significance. Interestingly, school size (average enrollment in grades 11 and 12) are positively and significantly related to enrollment percentages, but re-enrollment percentages are less affected by them. Note also that after controlling for other covariates in the model, time span has the strongest effects on college enrollment percentages.

The estimated proportion of variance[16] (based on the log likelihood ratio R^2) in enrollment percentages between enrollees' high schools explained by the model is 0.07. The proportion of variance in re-enrollment percentages between colleges explained by the model are 0.02 and 0.01, for re-enrollment at any college and at the same college, respectively.

College Course Grades

Two-level HLMs with random intercepts were used to model college course grades. In this model, the grades in first-year college courses were regressed on the school effect estimates and student and school level covariates (Equation 4). Separate models were fit for each of the four core college content areas (English/Language Arts, Mathematics, Natural Sciences, and Social Sciences). In this model, the terms X_1, X_2, X_3, X_4, X_5, as well as the terms S_1, S_2, S_3, S_4, S_5, are described in Equation 1, and the terms B_1, B_2, B_3, B_4, B_5 are described in Equation 2 (here S_5 is the subject area-specific mean ACT Explore score of the high school). In Equation 4, $Grade_{ik}$ is the grade in a core college content area for the i^{th} student in the k^{th} college, θ_k is the college random intercept that can vary among colleges, $\hat{\tau}_k$ are estimated college value-added measures (the predictors of interest in this model), α is the regression coefficient (associated with the value-added measures) which is the parameter of interest, and ε_{ik} is the random error term for the i^{th} student from the k^{th} college assumed to be normally distributed with mean 0 and unknown variances. Note that the term θ_k (college random effect) used in Equations (3) and (4) are not the same, as the outcome variables in the two models are different.

$$Grade_{ik} = \beta_0 + \sum_{p=1}^{5} \beta_p X_{ip} + \sum_{q=1}^{5} \lambda_q B_{iq} + \sum_{r=1}^{5} \gamma_r S_r + \theta_k + \alpha \hat{\tau}_k + \varepsilon_{ik} \quad (4)$$

In Table 9, the student sample is described with respect to aggregated first-year college course grades in the four core content areas. As we can see, more data are available for courses in English/Language Arts and Social Sciences than in Mathematics or Natural Sciences, and lower grades tend to be assigned to courses taken from the latter two content areas. The smaller samples for mathematics and natural science courses may be a direct result of differential grading practices (Larkey & Caulkins, 1992), or perhaps colleges have lower core requirements in these areas, or students are less interested in these courses. Furthermore, some students may have satisfied mathematics and science requirements with their high school course work.

[16] Proportion of variance (R^2 measure) for enrollment and retention are based on the log likelihood ratio of outcome variable (Menard, 2000).

Table 9. Distribution of Estimated First-Year College Course Grades

Content Area	N	Mean	SD	Min	P_{25}	Med	P_{75}	Max
English/language arts	32,387	2.65	1.35	0.00	2.00	3.00	4.00	4.00
Mathematics	21,041	2.18	1.50	0.00	1.00	3.00	3.00	4.00
Natural sciences	17,020	2.45	1.38	0.00	2.00	3.00	4.00	4.00
Social Sciences	31,517	2.51	1.37	0.00	2.00	3.00	4.00	4.00

Note. N = number of students; SD = standard deviation; Min = minimum; P_{25} = 25th percentile; Med = median; P_{75} = 75th percentile; Max = maximum.

Table B-2 in Appendix B contains correlations between aggregated first-year college course grades in each core content area and contextual characteristics at the school and student levels. Surprisingly, the high school effect estimates are not statistically significantly correlated with college course grades in English/Language Arts, Mathematics, and Natural Sciences, and are only weakly correlated with grades in Social Sciences. The value-added measures of school effects have only a small correlation with college course grades. On the other hand, ACT Explore scores (at the student level) have the strongest correlations with grades in each content area, ranging from 0.19 to 0.25. As we will discuss later in this report, the majority of the overall variance in first-year course grades is due to students' characteristics; less variance is due to the characteristics of high school or college.

At the school level, the mean ACT Explore scores are correlated with college course grades in each respective subject area, with correlations of 0.09 or 0.10. High school poverty level, proportion minority, and time between ACT Explore and ACT testing are negatively related to college course grades across all content areas.

We now assess the statistical relationships between the subject-specific school effect estimates and aggregated grades in each of the four core content areas (English/Language Arts, Mathematics, Natural Sciences, and Social Sciences), adjusted for student-level and school-level covariates using a multilevel linear regression model. For each content area, Table 10 contains estimated linear regression coefficients, estimated standard error (SE), and standardized linear regression coefficients (beta coefficients). Again, all school-level and student-level predictors are grand-mean centered. A two-level hierarchical model was specified with students nested within combination of course and college. (So, for example, a college with data from College Algebra and Geometry would have two clusters represented). We obtained statistically significant ICC estimates of 0.10, 0.13, 0.18, and 0.09, for random-intercept models for English/Language Arts, Mathematics, Natural Sciences, and Social Sciences, respectively. Thus, the majority of overall variance in first-year course grades in college for the four core content areas is due to students' characteristics, and less variance is attributed to college/course combinations.

With this design, we would expect the high school effect estimates to be positively related to college grades if they are truly indicators of school's effects on students' college readiness. The third columns under each content area of Table 10 contain the beta weights for the school and student level covariates. All regression coefficient estimates of the high school effects are positive and statistically significant; ranging from 0.02 in English/Language Arts (significant at $p < .05$); 0.04 in Natural Sciences and Social Sciences; to 0.07 in Mathematics (all significant at $p < .01$). These results suggest that the value-added measures representing school effects on ACT scores are predictive of first-year college course grades in selected core content areas and that the statistical relationships persist even after adjusting for student and school level characteristics. Hence, the results support the proposition that the value-added measures are

Table 10. Estimated Coefficients for Predicting First-Year Course Grades in Core Content Areas

	English/Language Arts			Mathematics			Natural Sciences			Social Sciences		
Estimate	**Estimate (unstd.)**	**SE**	**Estimate (std.)**	**Estimate (unstd.)**	**SE**	**Estimate (std.)**	**Estimate (unstd.)**	**SE**	**Estimate (std.)**	**Estimate (unstd.)**	**SE**	**Estimate (std.)**
Intercept ($\hat{\beta}_0$)	2.64	0.02	2.60	2.28	0.03	2.24	2.34	0.04	2.43	2.45	0.02	2.53
School-level Covariate												
School effect on ACT . . .	0.03	0.01	0.02	0.10	0.02	0.07	0.09	0.02	0.04	0.09	0.02	0.04
Mean ACT Explore . . .	0.00^{ns}	0.01	0.00^{ns}	-0.01^{ns}	0.01	-0.01^{ns}	0.00^{ns}	0.02	0.00^{ns}	-0.01^{ns}	0.01	-0.01^{ns}
School size	0.00	0.00	-0.03	0.00^{ns}	0.00	0.01^{ns}	0.00^{ns}	0.00	0.01^{ns}	0.00	0.00	0.02
Proportion tested	-0.26^{ns}	0.13	-0.02^{ns}	0.17^{ns}	0.17	0.01^{ns}	0.36	0.18	0.02	0.01^{ns}	0.14	0.00^{ns}
Poverty level	-0.17	0.07	-0.03	-0.24	0.09	-0.04	-0.25	0.09	-0.04	-0.18	0.07	-0.03
Proportion minority	-0.03^{ns}	0.06	-0.01^{ns}	-0.03^{ns}	0.08	-0.01^{ns}	0.00^{ns}	0.08	0.00^{ns}	-0.14	0.06	-0.02
Student-level Covariate												
ACT Explore												
English	0.02	0.00	0.06	0.01^{ns}	0.00	0.02^{ns}	0.03	0.00	0.11	0.02	0.00	0.09
Mathematics	0.03	0.00	0.09	0.07	0.00	0.22	0.05	0.00	0.15	0.05	0.00	0.13
Reading	0.02	0.00	0.07	0.01	0.00	0.04	0.02	0.00	0.06	0.03	0.00	0.11
Science	0.04	0.00	0.10	0.06	0.01	0.14	0.05	0.01	0.14	0.06	0.00	0.14
Time span	0.00	0.00	0.02	0.01	0.00	0.04	0.00^{ns}	0.00	0.01^{ns}	0.01	0.00	0.03
Male	-0.36	0.01	-0.18	-0.37	0.02	-0.19	-0.14	0.02	-0.07	-0.20	0.01	-0.10
African American	-0.12	0.03	-0.04	-0.13	0.04	-0.04	-0.28	0.04	-0.07	-0.28	0.03	-0.07
Hispanic	-0.06^{ns}	0.05	-0.01^{ns}	0.11^{ns}	0.06	0.02^{ns}	-0.17	0.06	-0.03	0.01^{ns}	0.05	0.00^{ns}
Asian American	0.21	0.06	0.03	0.18	0.08	0.02	0.19	0.06	0.03	0.33	0.05	0.05
Other	-0.08	0.03	-0.02	-0.15	0.04	-0.04	-0.02^{ns}	0.04	-0.00^{ns}	-0.07	0.03	-0.02
Variance of intercept ($\hat{\tau}_{00}$)	0.13			0.23			0.35			0.16		
Intraclass correlation coefficient (ICC)	0.10			0.13			0.18			0.09		
N	32,387			21,041			17,020			31,517		

Note. unstd = unstandardized; std = standardized.; NA = not applicable. Statistically insignificant coefficients (p > 0.05) are marked as 'ns.'

markers of schools' effects on college readiness. Earlier, we reported zero correlations between high school effect estimates and college course grades in three content areas (see Table A-2), yet we see that regression coefficients relating high school effect estimates to college grades are positive and statistically significant (Table 10). The results suggest that, while the simple correlations between various covariates are insignificant (possibly due to confounding variables), it is still plausible to obtain significant high school effects after controlling for other covariates.

Now, consider estimated intercepts for the four content areas. For example, we see an intercept of 2.60 (which corresponds approximately to letter grade B–) for English/Language Arts. This is the predicted grade average in English/Language Arts for a white female first-year college student of average achievement as an eighth grader and with average values for the other covariates in the model. The model suggests that attending a high-poverty school is associated with a reduction of 0.03 in overall English/Language Arts grade, being male is associated with a further reduction of 0.18, and being African American is associated with an additional reduction of 0.04. English/Language Arts grades increase with the school effect on ACT English performance and with higher ACT Explore scores. Also, being Asian American is associated with an increase of 0.03, on average, in English/Language Arts grades.

These findings generally held for the other content areas: Mathematics, Natural Sciences, and Social Sciences. School poverty level was negatively associated with first-year college grades, male and African American students tended to under-perform compared to female and White students, and Asian American students tended to over-perform relative to White students.

Furthermore, higher ACT Explore Mathematics and Science scores were associated with higher grades in first-year college courses in all four content areas; grades in all four core content areas were less influenced by ACT Explore English and Reading scores. Results show that coefficients for student-level covariates are generally positive and statistically significant. The largest beta weights were for ACT Explore Mathematics, ranging from 0.09 to 0.22; ACT Explore Science, ranging from 0.10 to 0.14; and for male gender, ranging from -0.07 to -0.19. The high school-level characteristics generally were not predictive of college course grades, except for the school effect estimates and schools' poverty level coefficients, indicating that estimated grades in all four core content areas are negatively related to schools' proportion eligible for free or reduced lunch, controlling for other covariates in the model. The estimated proportions of variance in first-year college grades between colleges explained by the model were 0.05 for English/Language Arts, 0.06 for Mathematics, 0.09 for Social Sciences, and only 0.08 for Natural Sciences (corresponding to the estimated multiple correlation coefficient (R) of 0.23, 0.25, 0.30, and 0.28, respectively). As was mentioned earlier, college course grade data were available only for a subset of college enrollees (postsecondary institutions participating in ACT's Course Placement Service or ACT's Prediction Service). Admittedly, our sample was not representative of all postsecondary institutions in the United States (about 90% of college course grade data were from southwestern states).

Case Examples: Predictive Strength of Value-Added Measures of School Performance for Selected School Types

In this section, examples are provided that illustrate the predictive strength of the school effect estimates. Predicted college outcomes are given for four types of high schools, where school type is determined by the school effect estimate (high vs. low performing) and demographics of the student population (high-poverty/high-minority population vs. lower-poverty/lower-minority population). High performing schools are defined as having an overall school effect estimate one standard deviation above the mean; low performing schools are defined as having an overall school effect estimate one standard deviation below the mean. High-poverty/high-minority schools are defined as having 75 percent of students eligible for free or reduced lunch with 75 percent concentration of racial/ethnic minority students; lower-poverty/lower-minority schools are defined as having 25 percent of students eligible for free or reduced lunch with 25 percent concentration of racial/ethnic minority students. For each of the four school types, estimates are given for three types of students: those whose eighth grade achievement level measured by their ACT Explore score suggested that they were on track (met the ACT Explore Benchmark), below track (two points below the ACT Explore Benchmark), or above track (two points above the ACT Explore Benchmark) for college readiness. This analysis allows us to contrast the estimated effect sizes of the school effect estimates to those of eighth grade readiness status and high school type.

In Table 11, we present predicted first-year enrollment and retention percentages and predicted average grade in English/Language Arts, Mathematics, Natural Sciences, and Social Sciences for students from high-performing schools. The predicted values are based on the regression model estimates from Table 8 (for enrollment and retention) and Table 10 (for the course grades).

Table 11. College Success Outcomes for High-performing Schools, by Type of High School and Student Achievement Level on ACT Explore

	High School Characteristics					
	High-poverty, High-minority			Low-poverty, Low-minority		
	8th Graders Achievement Level					
College Success Outcomes	**Below Track**	**On Track**	**Above Track**	**Below Track**	**On Track**	**Above Track**
Adjusted first-year enrollment percentage	65%	73%	80%	74%	80%	85%
Adjusted retention percentage at same institution	67%	70%	73%	71%	74%	76%
Adjusted retention percentage at any institution	81%	84%	87%	85%	88%	90%
Adjusted average grade in English/Language Arts	2.40	2.62	2.83	2.51	2.72	2.93
Adjusted average grade in Mathematics	2.08	2.37	2.66	2.22	2.51	2.80
Adjusted average grade in Natural Sciences	2.05	2.35	2.65	2.17	2.47	2.77
Adjusted average grade in Social Sciences	2.07	2.39	2.70	2.23	2.55	2.86

From Table 11, one can see that among students from high-performing schools whose eighth grade achievement level suggests that they were on track for college readiness, the adjusted college enrollment percentage was 73% for students from high-poverty/high-minority schools and 80% for students from lower-poverty/lower-minority schools; the adjusted college retention percentage at the same college was 70% for students from high-poverty and high-minority schools and 74% for students from low-poverty and low-minority schools. Among on-track students from high-performing schools in mathematics, the adjusted average grade (on the four-point scale) in first-year college math courses was 2.37 for students from high-poverty and high-minority schools and 2.51 for students from lower-poverty and lower-minority schools.

Similarly, Table 12 provides predicted probabilities of college enrollment and retention and predicted grades in four core content areas for low-performing schools (defined as having an overall school effect estimate one standard deviations below the mean) serving similar groups of students. Similarly, school effect estimates for low-performing schools were also predictive of college enrollment and retention percentages and average grade in four core content areas.

Table 12. College Success Outcomes for Low-performing Schools, by Type of High School and Student Achievement Level on ACT Explore

	High School Characteristics					
	High-poverty, High-minority			Low-poverty, Low-minority		
	8th Graders Achievement Level					
College Success Outcomes	Below Track	On Track	Above Track	Below Track	On Track	Above Track
Adjusted first-year enrollment percentage	59%	68%	76%	70%	77%	83%
Adjusted retention percentage at same institution	66%	69%	71%	70%	73%	75%
Adjusted retention percentage at any institution college	79%	83%	86%	83%	86%	89%
Adjusted average grade in English/Language Arts	2.36	2.57	2.79	2.46	2.68	2.89
Adjusted average grade in Mathematics	1.95	2.24	2.53	2.09	2.38	2.66
Adjusted average grade in Natural Sciences	1.96	2.26	2.56	2.09	2.39	2.69
Adjusted average grade in Social Sciences	1.97	2.29	2.60	2.13	2.45	2.76

Contrasting the low-performing schools and the high-performing schools (Table 11 versus Table 12), we see that students from high-performing schools are more likely to have college success, regardless of high school poverty and minority level. Specifically, across different school types, enrollment and retention percentages for students from low-performing schools were lower than those for students from high-performing schools by two to six and by one to two percentage points, respectively, and had lower average grades in English/Language Arts by 0.04–0.05 points, in Mathematics by 0.13–0.14 points, in Natural Sciences by 0.08–0.09 points, and in Social Sciences by 0.10.

Discussion

The results of this study show that high school effect estimates, also known as value-added measures, are incrementally predictive of college enrollment, college retention, and grades in first-year college courses. The study provides evidence supporting the use of value-added measures as markers of schools' effects on college readiness. The analysis was based on 1,119 high schools (across 2,707 cohorts) of over 263,000 students with test scores from two time points, pre- and near-end high school assessments (eighth and eleventh or twelfth grades). For inclusion in the analyses, we required that at least 50% of each high school cohort completed ACT Explore and the ACT.

Because accountability measures are often used for high-stakes decisions (as the basis for rewarding or sanctioning schools), it is necessary to validate estimates of school effects against external measures of school effectiveness (Braun, 2005). Clearly, it is not "a given" that the value-added measures are measuring schools' contribution to learning; perhaps other, unmeasured, student characteristics account for variation in test scores among schools. It is also possible that high performing students are clustered in certain high schools and low performing students in certain other high schools. Cognizant of the confounding between gain attributable to schools and the gain attributable to students, we included the mean ACT Explore scores as a step towards trying to mitigate this possible confounding. Nevertheless, there is some inevitable uncertainty to a cause and effect argument.

One would expect that a high school's contribution to learning would extend beyond graduation. In other words, if the value-added measures are truly measuring the high school's contribution to college readiness, they should have statistical relationships with college enrollment, college retention, and first-year college course grades. Moreover, the statistical relationships should persist even after adjusting for student and school level characteristics.[17]

We investigated the predictive strength of school effect estimates against external measures of school effectiveness, namely students' success in college. In particular, we used two-level hierarchical logistic regression with random intercepts to evaluate the effect of value-added estimates of school performance on the probability of enrollment/retention, controlling for selected student and school-level covariates. We also used two-level HLM's with random intercepts to evaluate the effect of the school effect estimates on first-year college course grades in each of the four core college content areas (English/Language Arts, Mathematics, Natural Sciences, and Social Sciences), controlling for selected student- and school-level covariates.

The school effect estimates were predictive of enrollment and retention beyond what is already predicted by a host of high school characteristics and student academic achievement as of eighth grade. School effect estimates were also positive and statistically significant predictors of grades in each of four core content areas in the first year of college, beyond what is already predicted by the set of high school characteristics and student academic achievement as of eighth grade.

The predictive validity of the value-added measures presented in this study suggests that the measures have some merit regarding college success. This study provides evidence that some schools are more successful than others at moving students towards college success. However, this study does not begin the unpacking process of how unsuccessful (or

[17] Predictive relationships between the high school VAM and college outcomes *enhances* the plausibility of interpreting the VAM as an indicator of schools' contribution to college readiness, but it does not address the problem that the VAM score is partly the result of unmeasured variables. In principle, the same unmeasured variables that drive variation among high schools in their students' test scores could also drive variation in their students' success in college.

less successful) schools can be improved with respect to high school instruction, curriculum, student cohesion, environment, and a host of other characteristics that make some schools more successful than others.

Limitations

This study was based on value-added measures of schools' performance that were adjusted for school characteristics that were available to us (size, proportion of students tested, poverty level, proportion of racial/ethnic minority students, and mean eighth grade achievement scores). By design, such context-adjusted value-added measures are less likely to yield low scores for disadvantaged schools. Using a model similar to the one used in this study, Allen et al. (2009) found that the context-adjusted measures were highly correlated with unadjusted measures. Ballou et al. (2004) discuss how adjusting for contextual characteristics could distort the measurement of teacher or school effects. Further, fully accounting for contextual characteristics requires additional student-level data, such as absenteeism, dropout[18]/transfer, family income, and parent's education level that may not be readily available or reliably measured. Further research is needed to explore the virtues of context-adjusted versus unadjusted value-added measures of schools' performance.

Because college outcomes are also influenced by nonacademic student characteristics (e.g., parental support, motivation, study habits, interpersonal dynamics), additional research is needed to explore the predictive strength of school value-added measures by introducing covariates in the models that account not only for academic characteristics but also for nonacademic characteristics. Research has shown that psychosocial characteristics measured via student survey, such as motivation and social engagement, are predictive of college outcomes (Robbins et al., 2006).

Our sample was not representative of all public high schools in the United States. In particular, most of the high school cohorts were located in the Midwest and south-central states with little representation from the eastern and western states. Furthermore, high-racial/ethnic minority and small-enrollment high schools were under-represented. Finally, enrollment data, and by extension retention data, were not available for all students in the study sample. It is unlikely, however, that this under-representation has affected the primary findings of this study. Moreover, the predictive strength of school effect estimates on first-year college course grades might have been compromised due to the fact that college course grade data were available only for a subset of college enrollees. Additionally, this study was based on grade eight ACT Explore and the ACT data, thus the findings may not be generalized to all high school growth data using other assessment systems. Additionally, ACT-tested students are more likely to be college-bound than students generally. Therefore, the results for predicting enrollment in college might not generalize to all students. Furthermore, we only studied one type of high school effectiveness measure (that was based on ACT Explore and the ACT data). Alternative measures also need to be studied—for example, measures based on high school completion percentages (adjusted for entering student characteristics), measures based on other assessment systems, or measures based on student engagement in high school.

[18] The determinants of dropout are beyond the scope of this study, but differential dropout rates across high schools could distort (or bias) the value added estimates used in this study. For example, a school might have high value added estimates because it encourages low-performing students to dropout or transfer to another school. If so, the value added estimates may overstate the effectiveness of this school relative to other schools that encourage low-performing students to persist in high school.

References

ACT (2004). *Crisis at the core: Preparing all students for college and work.* Iowa City, IA: Author.

ACT (2006). *EPAS: A system that works*. Iowa City, IA: Author.

ACT. (2007a). *The ACT national high school profile report*. Iowa City, IA: Author.

ACT. (2007b). *Explore technical manual*. Iowa City, IA: Author.

ACT. (2007c). *The ACT technical manual*. Iowa City, IA: Author.

ACT. (2009a). *The Explore national profile report grade 8*. Iowa City, IA: Author.

ACT. (2009b). *The Explore national profile report grade 9*. Iowa City, IA: Author.

ACT. (2009c). *The ACT national high school profile report*. Iowa City, IA: Author.

Allen, J., Bassiri, D., & Noble, J. (2009). *Statistical properties of accountability measures based on ACT's Educational Planning and Assessment.* Iowa City, IA: ACT.

American Statistical Association. (2014). *ASA statement on using value added models for educational assessment*. Alexandria, VA: Author. Retrieved from http://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 37*, 65–75.

Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. New York, NY: Routledge.

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in Chicago public high schools. *Journal of Labor Economics, 24*(1): 95–135.

Ballou, D. (2002). *Sizing up test scores*. Retrieved from http://educationnext.org/files/ednext20022_10.pdf

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, *29*(1), 37–65.

Bassiri, D., & Schulz, E. M. (2010). Using adjusted GPA and adjusted course difficulty measure to evaluate differential grading practices in college. In M. L. Garner, G. Engelhard, W. P. Fisher, & M. Wilson (Eds.), *Advances in Rasch Measurement*, *1*, 205–220. Maple Grove, MN: JAM Press.

Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching*. Retrieved from http:// www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Braun, H. (2005). Value-added modeling: What does due diligence require? In R. W. Lissitz (Eds.), V*alue Added Models in Education: Theory and Applications,* 19–38. Maple Grove, MN: JAM Press.

Braun, H., Chudowsky, N., & Koenig, J. (Eds.). (2010). *Getting value out of value-added: Report of a workshop*. Washington, DC: National Research Council, National Academies Press.

Braun, H. (2015). The value in value added depends on the ecology. *Educational Researcher*, *44*(2), 105–16. doi: 10.3102/0013189X15576341

Briggs, D., & Domingue, B. (2011). *Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles Unified School District Teachers by the "Los Angeles Times".* Retrieved from http://nepc.colorado.edu/files/NEPC-RB-LAT-VAM_0.pdf

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood.* National Bureau of Economic Research Working Paper No.17699. Cambridge, MA. http://obs.rc.fas.harvard.edu/chetty/value_added.pdf

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, R. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8–15.

Doran, H. C., & Izumi, L. T. (2004). *Putting education to the test: A value-added model for California.* Retrieved from http://www.pacificresearch.org/docLib/200702021_Value_Added.pdf

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D., & Whitehurst, G. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brookings Institution.

Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica*, *80*(319), 589–612.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, *44*, 96–104.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., & Williams, A. (2005). *Policymakers' guide to growth models for school accountability: How do accountability models differ?* Washington, DC: Council of Chief State School Officers.

Gordon, R., Kane, T., & Staiger, D. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: Brookings Institution.

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review Papers and Proceedings*, *61*(2): 280–88.

Hanushek, E. A., & Raymond, M. (2002). *Improving educational quality: How best to evaluate our schools?* Retrieved from http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%2BRaymond%202003%20Educ21stCent.pdf

Hanushek, R., & Rivkin, S. (2004). How to improve the supply of high-quality teachers. In D. Ravitch (Ed.), *Brookings papers on education policy* (pp. 7–25). Washington, DC: Brookings Institution.

Harris, D. N. (2009). Would accountability based on teacher value added be smart policy? An examination of the statistical properties and policy alternatives. *Education Finance and Policy, 4*(4), 319–350.

Harris, D. N., & Herrington, C. D. (2015). The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. *Educational Researcher*, *44*(2), 271–276. doi:10.3102/0013189X15576142

Hershberg, T., Simon, V. A., & Lea-Kruger, B. (2004). *Measuring what matters: How value-added assessment can be used to drive learning gains*. Retrieved from http://www.cgp.upenn.edu/pdf/measuring_what_matters.pdf

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, *44*(2), 105–116. doi:10.3102/0013189X15575517

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation*. National Bureau of Economic Research, Working Paper No. 14607.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment.* Seattle, WA: Bill & Melinda Gates Foundation.

Klein, A. (2016). The Every Student Succeeds Act: An ESSA overview. *Education Wee*k (April 10). Retrieved from http://www.edweek.org/ew/issues/every-student-succeeds-act/index.html?cmp=eml-enl-eu-news

Larkey, P. D., & Caulkins, J. P. (1992). *Incentives to fail.* Pittsburgh, PA: The Heinz School of Public Policy and Management, Carnegie Mellon University**.**

Lindquist, E. F. (1963). An evaluation of a technique for scaling high school grades to improve prediction of college success. *Educational and Psychological Measurement*, Vol. XXIII (4), 623–646.

Lissitz, R. W., Doran, H., Schafer, W. D., & Willhoft, J. (2006). Growth modeling, value-added modeling and linking: An introduction. In R. W. Lissitz (Ed.) *Longitudinal and Value Added Models of Student Performance*, (pp. 1–46). Maple Grove, MN: JAM Press.

Lockwood, A. T. (2005). Educators grapple with NCLB's demands for accountability. *Northwest Education*, *10*(4), 41–42.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The Inter-temporal variability of teacher effect estimates. *Education Finance and Policy*, *4*(4), 572–606.

Marion, S. F., & Gong, B. (2003). *Evaluating the validity of states' accountability systems.* Paper presented at the Reidy Interactive Lecture Series. Nashua, NH. Retrieved from http://www.nciea.org/publications/RILS2003_BGSM03.pdf

Marion, S. F., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decisions in the determination of adequate yearly progress*. A paper in the series: Implementing the state accountability system requirements under the No Child Left Behind Act of 2001. Washington, DC: Council of Chief State School Officers.

Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statisticians, 54*, 17–24.

Murnane, R. J. (1975). *The impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger.

National Association of State Boards of Education. (2002). *The No Child Left Behind Act: What states need to know?* Arlington, VA: National Association of State Boards of Education.

National Research Council. (2010). *Getting value out of value-added*. Washington, DC: National Academies Press.

Noble, J., Davenport, M., Schiel, J., & Pommerich, M. (1999). *Relationships between the noncognitive characteristics, high school course work and grades, and test scores of ACT-tested students.* Iowa City, IA: ACT.

Noble, J., Roberts, B., & Sawyer, R. L. (2006). *Student achievement, behavior, perceptions and other factors affecting ACT scores.* Iowa City, IA: ACT.

Noble, J., & Schnelker, D. (2007*). Using hierarchical modeling to examine course work and ACT score relationships across high schools.* Iowa City, IA: ACT.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models applications and data analysis methods.* Thousand Oaks, CA: Sage Publications.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools and academic achievement. *Econometrica*, *73,* 417.458.

Robbins, S., Allen, J., Casillas, A., Peterson, C., & Le, H. (2006). Unraveling the differential effects of motivational and skills, social, and self-management measures from traditional predictors of college outcomes. *Journal of Educational Psychology, 98*, 598–616.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94,* 247–252.

Sable, J., Gaviola, N., & Garofano, A. (2007). *Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2005–06* (NCES 2007–365). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Schatz, C., Von Secker, C., & Alban, T. (2005). Balancing accountability and improvement: Introducing value-added models to a large school system. In R. W. Lissitz (Ed.) *Value Added Models in Education: Theory and Applications* (pp. 1–18). Maple Grove, MN: JAM Press.

Schiel, J., Pommerich, M., & Noble, J. (1996). *Factors associated with longitudinal educational achievement, as measured by PLAN and ACT Assessment scores.* Iowa City, IA: ACT.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Modeling: An introduction to basic and advanced multilevel modeling.* London: Sage Publications.

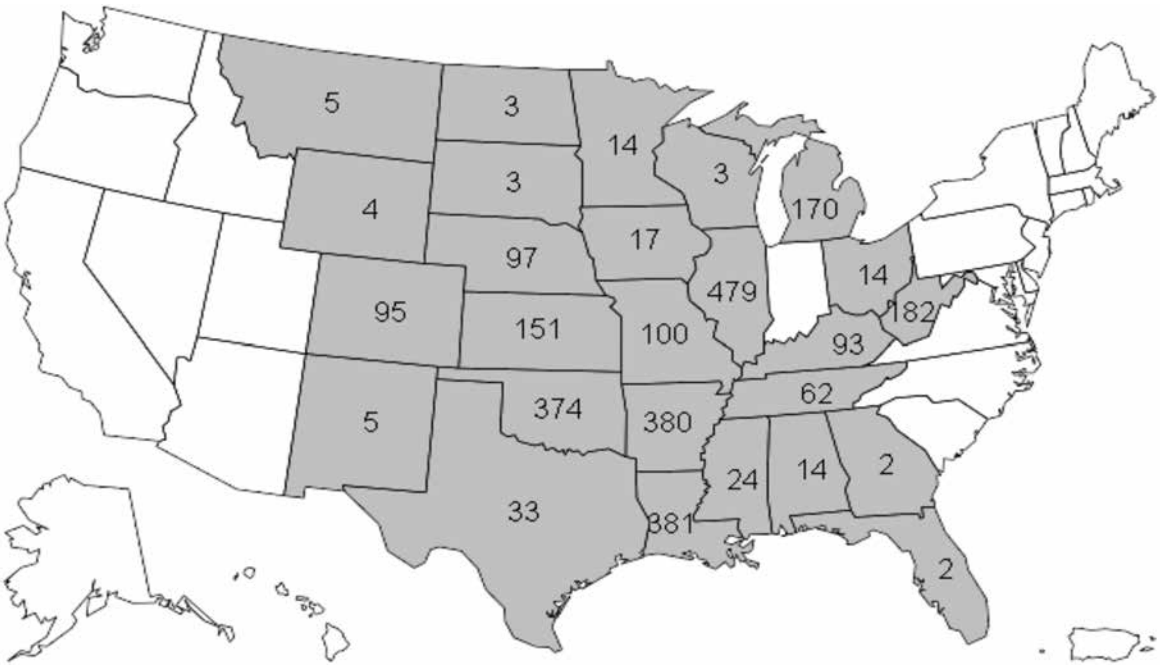
Teddlie, C., Reynolds, D., & Sammons, P. (2000). The methodology and scientific properties of school effectiveness research. In C. Teddlie and D. Reynolds (Eds.), *The International Handbook of School Effectiveness Research* (pp. 55–133). New York, NY: Falmer Press.

Wainer, H. (Ed.) (2004). Special issue of the Journal of Educational and Behavioral Statistics on value-added assessment. *Journal of Educational and Behavioral Statistics*, *29*(1).

Wei, H., Hembry, T., Murphy, D. L., & McBride, Y. (2012). *Value-added models in the evaluation of teacher effectiveness: A comparison of models and outcomes.* Retrieved from http://images.pearsonassessments.com/images/tmrs/ComparisonofValue-AddedModelsandOutcomes.pdf

Appendix A

Number of High School Cohorts Sampled Per State



Appendix B

Table B-1. Correlations of School-level and Student-level Characteristics with College Enrollment and Retention

Characteristics	College Enrollment	College Retention (at any institution)	College Retention (at same institution)
School-level			
School effect on ACT Composite	0.09	0.05	0.04
Mean ACT Explore Composite	0.14	0.12	0.10
School size	0.02	0.07	0.06
Proportion tested	0.06	0.08	0.07
Poverty level	-0.10	-0.12	-0.09
Proportion minority	-0.08	-0.06	-0.05
Student-level			
ACT Explore Scores			
English	0.20	0.15	0.15
Mathematics	0.21	0.16	0.16
Reading	0.19	0.14	0.14
Science	0.20	0.15	0.15
Time span	0.06	-0.02	-0.03
First college choice	0.14	NA	NA
Second college choice	0.10	NA	NA

Note. NA = not applicable. All point-biserial correlations reported in this table are statistically significant at $p < .0001$).

Table B-2. Correlations of First-Year College Course Grades and Student and School Characteristics

Characteristics	English/ Language Arts	Mathematics	Natural Sciences	Social Sciences
School-level				
School effect on ACT* . . .	-0.00 ns	-0.01 ns	0.00 ns	0.01
Mean ACT Explore** . . .	0.09	0.10	0.09	0.10
School size	0.02	0.01 ns	0.05	0.04
Proportion tested	-0.01	0.03	0.01 ns	-0.01 ns
Poverty level	-0.06	-0.06	-0.07	-0.07
Proportion minority	-0.06	-0.05	-0.03	-0.06
Student-level				
ACT Explore Scores				
English	0.21	0.20	0.22	0.25
Mathematics	0.19	0.23	0.24	0.24
Reading	0.20	0.19	0.21	0.24
Science	0.20	0.22	0.23	0.25
Time span	-0.02	-0.03	-0.04	-0.02
Count of students	32,387	21,041	17,020	31,517

* School effect on ACT English scores for English/Language Arts, on ACT Mathematics scores for Mathematics, on ACT Science scores for Natural Sciences, and on ACT Reading scores for Social Sciences.

** Mean ACT Explore English score for English/Language Arts, mean ACT Explore Mathematics scores for Mathematics, mean ACT Explore Science scores for Natural Sciences, and mean ACT Explore Reading scores for Social Sciences.

Note. Statistically insignificant coefficients (p > 0.05) are marked as 'ns.'

ACT is an independent, nonprofit organization that provides assessment, research, information, and program management services in the broad areas of education and workforce development. Each year, we serve millions of people in high schools, colleges, professional associations, businesses, and government agencies, nationally and internationally. Though designed to meet a wide array of needs, all ACT programs and services have one guiding purpose—helping people achieve education and workplace success.

For more information, visit www.act.org.

* 0 7 0 7 0 4 1 7 0 * Rev 1