AMERICAN INSTITUTES FOR RESEARCH®

Using the National Assessment of Educational Progress as an Indicator for College and Career Preparedness

Mark Schneider, Sami Kitmitto, Helen Muhisani, and Bo Zhu

OCTOBER 2015

Using the National Assessment of Educational Progress as an Indicator for College and Career Preparedness

October 2015

Mark Schneider, Sami Kitmitto, Helen Muhisani, and Bo Zhu

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000 | TTY 877.334.3499

www.air.org

3741_10/15

Contents

Introduction

For more than 10 years, the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES) have been exploring whether the Grade 12 National Assessment of Educational Progress (NAEP) can be used to assess the college and career preparedness of high school students (Fields, 2014).

The more widely used term of "college and career readiness" is broadly defined as the level of academic skills and knowledge that a student needs to succeed, without remediation, in regular courses in a degree-granting postsecondary institution or program of study (e.g., a certificate program) in a career that offers the potential for advancement (Conley, 2010). In contrast, NAGB defines "preparedness" more narrowly, as "reading and mathematics knowledge and skills needed for placement into entry-level, credit-bearing, non-remedial courses in broad access 4-year institutions and, for 2-year institutions, the general policies for entry-level placement, without remediation, into degree bearing programs designed to transfer to 4-year institutions." (Fields, 2014, p. 7) As the only nationally representative measure of what 12th-grade students know and can do in reading and mathematics, NAEP provides a unique opportunity for measuring the college and career preparedness of the nation's 12th graders.[1]

A straightforward way of using an assessment, such as NAEP, to provide information about the level of preparedness of America's 12th graders is to establish a point on the NAEP scale, a cut score, as a threshold dividing those who are likely to succeed and those who are not. Such a benchmark allows NAEP to report how well prepared America's students are for college and careers across NAEP's geographic reporting levels (the nation, states, and select urban districts) as well as variation in preparedness across subpopulations (students with disabilities, English language learners, students who are National School Lunch Program eligible, and students in different racial or ethnic categories).[2]

This study contributes to the research on NAEP as an indicator of college and career preparedness by using longitudinal data on individual student achievement to relate NAEP scores to subsequent observed levels of postsecondary success. Using postsecondary institution records and state unemployment system data from the state of Virginia, this study examines the relationship between NAEP scores and a set of the most important markers of postsecondary success, ranging from progression to retention to completion, plus a critical measure of student success in the labor market after college completion. We use these data to examine student success in four measures of college and career performance and how the levels of student success are related to different NAEP cut scores:

- Progression: Earning enough credits in the first year of college to be "on track" to complete a bachelor's degree in four years;
- Retention: Enrolling in college for a second year
- Completion: Receiving a bachelor's degree within six years

[1] NAEP results and reports can be found at: https://nces.ed.gov/nationsreportcard/
[2] NAEP does not report scores for individual students or schools.

- Earnings: Earning an income of more than 200 percent of the poverty line in the year after completing a bachelor's degree ("labor market success")[3]

Using data on individual students spanning seven years for the cohort of recent high school graduates who entered four-year institutions in Virginia in the 2006–07 school year, this study addresses two questions:

1. What is the empirically determined optimal (by criteria discussed below) NAEP mathematics cut score for distinguishing success?
2. What are observed levels of success in college and in the labor market at various potential cut scores?

Background

Since 2004, when it released its report on *12th-Grade Student Achievement in America: A New Vision for NAEP*,[4] NAGB has investigated how NAEP could be used to report on the college and career preparedness of 12th graders. These efforts have included measuring the statistical relationships between NAEP and other assessments, content alignment studies, standard-setting studies, higher education surveys, and benchmarking studies.[5]

These studies culminated in NAGB setting a preparedness cut score of 302 for 12th-grade reading, which is the same as the NAEP *Proficient* cut score in reading, and a cut score of 163 for 12th-grade mathematics, which is 13 points lower than the NAEP *Proficient* cut score in mathematics (176).[6] These choices were made based on a statistical linking study the established a relationship between NAEP and SAT (Moran, Oranje, & Freund, n.d.) and found these NAEP scores to be equivalent to the SAT benchmarks for career and college readiness (Fields, 2014).[7]

As evident across the breadth of NAEP studies commissioned by NAGB as well as studies by researchers and developers using other assessments, there are a number of different approaches that can be used to set benchmarks that indicate readiness for college (Kobrin, 2007). Some approaches have sought to set a benchmark based on professional judgment of what is required for success after high school, typically with reference to criteria that colleges use for admissions (Berkner & Chavez, 1997; Greene & Winters, 2005). The companies that have developed the SAT (Kobrin, 2007)[8] and ACT (Allen & Sconing, 2005),[9] have based their benchmarks on

[3] These are wages earned by students in the calendar year after graduation, giving the typical student graduating in May or June approximately six months immediately after graduation seeking employment.

[4] Available at http://www.nagb.org/publications/12_gr_commission_rpt.pdf

[5] Studies are available at http://www.nagb.org/what-we-do/preparedness-research.html

[6] The NAGB resolution from August 3, 2013, can be found here https://www.nagb.org/what-we-do/preparedness-research/nagb-motion.html.

[7] College entrance exams such as the SAT and ACT have had lines of research establishing career and college readiness cut points similar to NAEP. See Allen and Sconing (2005).

[8] "The SAT Benchmark score of 1550 is associated with a 65 percent probability of obtaining a first year GPA (FYGPA) of a B− or higher" (College Board, 2012, p. 21).

[9] The ACT College Readiness Benchmarks "are scores on the ACT subject area tests that represent the level of achievement required for students to have a 50% chance of obtaining a B or higher or about a 75% chance of

observed levels of success in college. This study also uses observed outcomes to research benchmarks, but it employs different methods.

This study is closely related to a statistical relationship study commissioned by NAGB that used observed outcomes to provide information on setting a benchmark for NAEP Grade 12. That study, by Moran, Freund, and Oranje (2012), analyzed postsecondary outcomes for the sample of students in Florida who took the 12th-grade NAEP in 2009 to assess the validity of potential cut scores. Moran et al. (2012) plotted the range[10] of NAEP scores for students who had met a given benchmark or level of success outside of NAEP against the range of scores for students who did not; it found that the range of potential preparedness cut scores under consideration at the time (164–175 in mathematics) was reasonable. Among the success measures Moran et al. studied were college readiness benchmarks on the SAT and ACT, college enrollment, no remedial coursework in the first year, and a first-year grade point average (GPA) of a B− or greater. However, they found that their outcome indicators had "relatively weak relationships with NAEP results" (Moran et al., 2012, p. 12).

This present study is similar to the work of Moran et al. (2012) but differs in the data available for analysis: our data contain information on the population of college-bound students in Virginia in the 2005–06 school year rather than a sample. Additionally, our data contain different indicators than Moran et al.'s work, and although our study examines proximal and intermediate outcomes (earning enough credits in the first year, retention into the second year), we also examine more long-term outcomes (college graduation, post-college earnings). Finally, although similar to Moran et al. in that the data are used to evaluate success at various cut scores, we use observed outcomes to define optimal cut scores.

obtaining a C or higher in corresponding credit-bearing first-year college courses. These college courses include English Composition, College Algebra, Biology, and an introductory social science course" (ACT, 2011, p. 24).

[10] Specifically, the interquartile range.

Data and Methods

Data

The data for this study were obtained from the State Council of Higher Education for Virginia (SCHEV), which maintains a student-level administrative records database that tracks students enrolled in Virginia's colleges and universities. SCHEV data also contain information on student wages after graduating, based on data from the state's unemployment insurance (UI) record system. The data used in this study included all first-time enrollees in a higher education institution in Virginia for academic year 2006–07, including public and private four-year colleges. (See Appendix A for more details about the data.)

The study population included the 35,293 students who had graduated high school within the previous 12 months, were under the age of 21, and enrolled in a 4-year institution in the 2006–07 academic year. A key part of the analysis is linking higher education success to NAEP scores. NAEP does not report individual student scores. However, SCHEV data include SAT mathematics scores. We therefore assigned a NAEP score based on observed relationships been SAT and NAEP as described in Appendix B. Because two-year colleges do not require or record SAT scores, those students (14,655 students) in the SCHEV data were dropped from the analysis. Therefore, the study population likely contains a more successful, more motivated population of students than if the study looked at all of Virginia's recent high school graduates or at all of Virginia's Grade 12 students from the previous school year (2005–06 school year). Of the students in the study population who enrolled in a 4-year school, 13.1 percent were missing SAT scores leading to an analysis data set containing 30,652 students.[11]

The success indicator "earning more than 200 percent of the poverty line in the year after graduation" was, by definition, limited to students who graduated within the time frame of the study, 22,301 students. The analysis was further limited to observations for which the data included an income measure from the state's Unemployment Insurance (UI) wage record system (i.e., worked in Virginia in the calendar year after graduation for an employer that participated in the state's UI system). Hence, this part of the analysis was limited to 13,331 observations.

Exhibit 1 provides the key student demographic characteristics of the data used for the study. Consistent with national averages, the analytic data set is majority female (55 percent to 45 percent for males). The analytic data set is composed of mostly White, Caucasian Americans (67 percent), with the next largest racial/ethnic group comprising African Americans (13 percent). Fewer than one out of five students in the analytic data set are Pell Grant recipients (16 percent). The mean age at college entry for the analytic data set is 18 years. Approximately seven out of 10 students in the analytic data set are in-state students (71 percent).

[11] Thirty students with SAT mathematics scores 250 or less also were dropped because they could not be linked to a NAEP score.

Exhibit 1. Demographic Characteristics of the Analysis Data

Student Demographic Characteristic	Percentage
Gender	
Male	45%
Female	55%
Race/Ethnicity	
African American or Black	13%
American Indian/Native American[12]	<1%
Asian and Pacific Islander	7%
Hispanic	3%
White, Caucasian American	67%
Unknown	8%
Pell Grant recipient	16%
Mean student age at college entry	18
In-state student	71%

Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Using SCHEV data, we created four measures of success:

- **Progression—Earning enough credits in the first year of college to be "on track" to complete a bachelor's degree in four years.** In Virginia, a typical bachelor's degree requires 120 credits. For this study, earning 30 or more credits in the first year, combining credits taken in the fall, spring, and summer terms for the 2006–07 school year, was counted as a success.

- **Retention—Returning to college for a second year.** Success was measured in the data as returning to a four-year Virginia public or private institution for the 2007–08 fall term.

- **Completion—Receiving a bachelor's degree within six years of entry.[13]** Success for this measure was earning a bachelor's degree by the end of the 2012–13 school year at a Virginia public or private institution.

- **Earnings—Earning more than 200 percent of the poverty threshold the year after attaining a bachelor's degree.[14]** Success here was measured as earning at least twice the poverty threshold for a single person household in Virginia during the calendar year after

[12] Given study confidentiality guidelines, the size of this subgroup is considered too small for reporting purposes. Hence, results of the analysis for this subgroup will not be provided in this report.
[13] Bachelor degree completion in six years is a benchmark set by the federal Student Right to Know Act of 1990. This represents 150 percent of the expected four years it would take a full-time student to earn a bachelor's degree.
[14] We used 200 percent of the poverty line to align with SCHEV's definition of "sustainable wages" (e.g., see Blake, Kang, & Massa, 2015).

obtaining a degree. Earnings are available only for graduates who worked in Virginia after graduation for employers covered by the UI system.

Exhibit 2 presents basic descriptive data on all four measures of college and career success. Forty-one percent of the analytic data set earned at least 30 course credits during their first year of college. Nine out of 10 students in the analytic data set returned for a second year of study at a Virginia higher education institution. Almost three out of four students in the analytic data set obtained a bachelor's degree within six years. Almost half of graduates earned more than 200 percent of the poverty threshold the year following college graduation.

Exhibit 2. Proportion of Students Meeting Success Indicators

Measure	Proportion	Number of Observations
Earned 30 credits in the first year	0.41	30,625
Retained into the second year	0.90	30,652
Received a bachelor's degree within six years	0.73	30,652
Earned more than 200 percent of the poverty threshold for a household size of one the year after graduation	0.47	13,331

Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Caveats

Although the SCHEV data present a unique opportunity to analyze postsecondary success, there are limitations to the data that should be considered when interpreting results. First, as is common in this type of study the data include only students who enroll in postsecondary institutions (Allen & Sconing 2005; Kobrin 2007). Although NAEP is representative of all 12th-grade students, the results here are only representative of those who go on to a four-year college within one year of high school graduation. Hence, the population being studied is likely to be more successful than the entire 12th-grade population represented in NAEP. Second, because we have data only from Virginia institutions, some of the students whom we observe as failing to persist or complete college may have transferred to an institution out of state. Third, the earnings data that we have are similarly limited to the state of Virginia. Students who might have been working out of state are not included in our analysis.

Methods

Recall our two research questions:

- What is the empirically defined optimal NAEP cut score for distinguishing success?
- What are observed levels of success in college and in the labor market at different cut scores?

To answer these questions, we use two different approaches to relate test scores to measures of student success. As noted, we first converted the SAT scores in the Virginia data into equivalent

NAEP scores using a concordance table provided to the study team (see Appendix B). We focus on mathematics, because the relationship between SAT mathematics and NAEP mathematics is strong enough to support the translation of SAT scores to NAEP scores (Moran et al., n.d.), which is not the case for SAT critical reading and NAEP reading scores. (See Appendix B for details of the methods used.)

Question 1: What is the optimal NAEP mathematics cut score for distinguishing success?

To answer this question, we need to empirically identify the score on the NAEP scale above (or below) which students are optimally predicted to be successful (or not successful) for a given outcome, where we define the term "optimal" below. The method used for this analysis, receiver operating characteristic (ROC) analysis (Gönen, 2007), is a common approach for letting observed outcomes identify the best cut score for predicting success (see, e.g., Vivo & Franco, 2008). Note that a perfect cut score, if it exists, would be a single score where:

1. Every student predicted to be successful (i.e., above the cut score) actually was successful
2. Every student predicted to be unsuccessful (i.e., below the cut score) actually was unsuccessful

Because no point will perfectly classify all students, two types of errors result: first, some students predicted to be successful because they scored higher than the cut score will actually not be successful; and, second, some students predicted to be unsuccessful because they scored lower than the cut score will actually be successful. Higher cut scores will minimize the first type of error, and lower cut scores will reduce the second type. The "optimal" cut score in ROC analysis is the score that balances the two types of error.

The ability to calculate an optimal cut score this way does not necessarily mean that it is a good predictor of success, however. To assess the ability of the test score in predicting each of the four outcomes, two measures of predictive performance and corresponding minimum criteria were set in advance.[15]

Question 2: What are observed levels of success in college and in the labor market at various potential cut scores?

The analysis for this question examines the different levels of success that students at different cut scores actually achieve. To calculate the levels of success at different levels of achievement, a statistical relationship was estimated between NAEP Grade 12 mathematics scores and each of the four outcome measures. The estimated relationship needed to demonstrate a minimal correlation between NAEP score and outcomes to be useful for this analysis and criteria were specified in advance.[16] The estimated relationship was then used to calculate predicated levels of

[15] The two criteria were (1) the percent of all students classified correctly (also known as the "accuracy") must be greater than 50 percent and (2) the area under the ROC curve (AUC) must be greater than .6. See Appendix C for details.

[16] Specifically, a logit regression model was used to estimate the relationship between NAEP Grade 12 mathematics scores and each binary success measure. Both the Pearson chi-square and Hosmer-Lemeshow chi-square goodness-of-fit statistics were used to test whether the model meets minimal criteria for fitting the data. See Appendix C for details.

success for different scores, specifically, for each of the three NAEP achievement level cut scores (141 for *Basic*, 176 for *Proficient*, and 216 for *Advanced*) as well as for the NAGB college academic preparedness cut score (163).

What is the optimal NAEP mathematics cut score for distinguishing success?

Results for the first research question are presented in Exhibit 3, where the vertical axis measures NAEP scores. The NAEP achievement levels are included in the exhibit to give the results context. The dashed horizontal lines at 141, 176, and 216 represent the well-known *Basic*, *Proficient*, and *Advanced* performance levels, respectively. The NAEP college academic preparedness cut score as specified by NAGB (163) is represented by a solid horizontal line. Against this backdrop, the results of the analysis, the empirically determined optimal cut scores for each success measures are represented as data points on the exhibit:

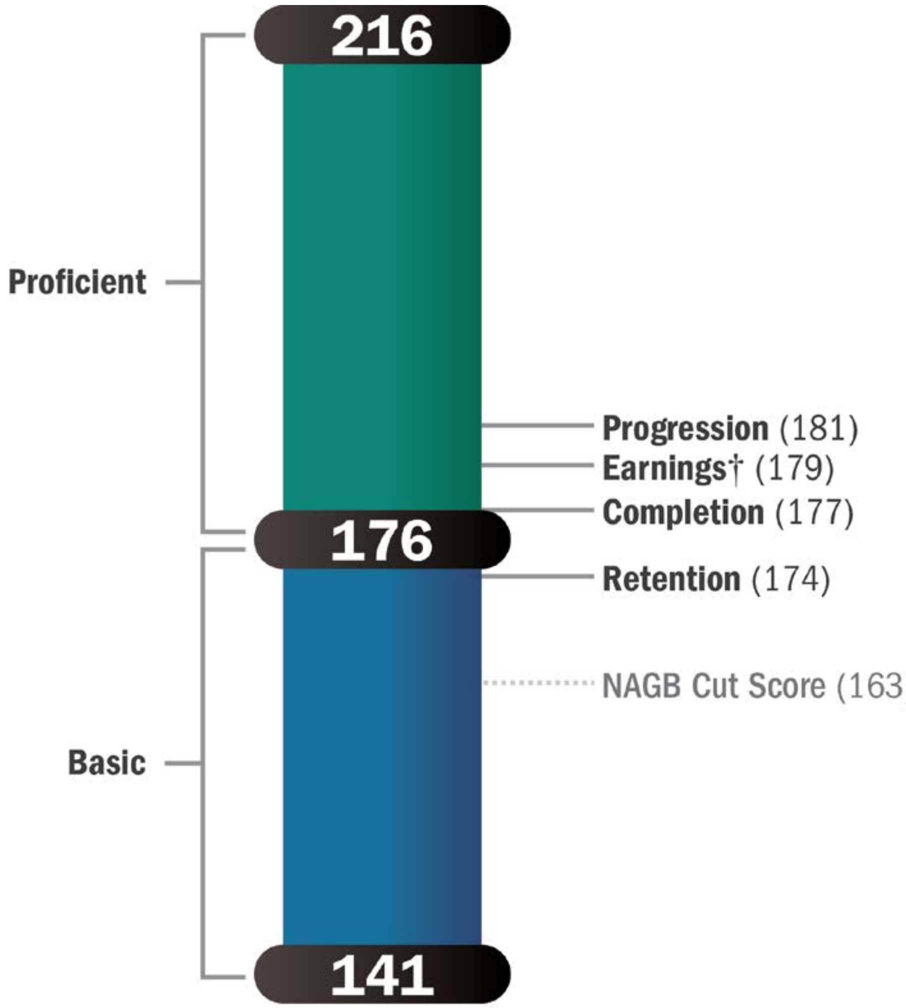
- For Progression (earning more than 30 credits in the first year), the optimal cut score for predicting success was 181, or 5 points above the cut score for NAEP *Proficient*.
- For Retention (returning for a second year), the optimal cut score was 174, or 2 points below the NAEP *Proficient* cut score.
- For Completion (earning a bachelor's degree within six years), the optimal cut score was 177, or 1 point above the NAEP *Proficient* cut score.
- For Earnings (earning more than twice the poverty line in the year after graduation), the optimal cut score was 179, or 3 points above the NAEP *Proficient* cut score of 176; however, the estimated relationship between earnings and NAEP scores did not meet one of the study's two pre-established criteria for accuracy and results should be interpreted with caution (see Appendix C for details).

The results from this approach to selecting an optimal cut score using observed outcomes results in suggested cut scores that are higher than the NAGB college academic preparedness cut score and cluster around the NAEP *Proficient* cut score.

To put these two different cut scores in perspective, NAGB preparedness (163) and NAEP *Proficient* (176), can be compared to results for students by the highest level mathematics course they had taken.[17] The NAGB preparedness cut score is close to the average of students whose highest course was pre-calculus. In contrast, the NAEP *Proficient* cut scores is about half way between the averages for students whose highest course was pre-calculus and students whose highest course was calculus.

[17] For the 2013 NAEP grade 12 mathematics assessment, students whose highest course was Algebra II scored 143 on average, students whose highest course was pre-calculus scored 165 on average, and students whose highest course was calculus scored 187 on average (http://www.nationsreportcard.gov/reading_math_g12_2013/#/learning-context).

Exhibit 3. Optimal Cut Score, by Success Indicator



Note: The NAEP mathematics Grade 12 achievement level cut scores are 141 for *Basic*, 176 for *Proficient*, and 216 for *Advanced*; the NAGB college academic preparedness cut score is 163; the empirically optimal cut scores from the analysis were 181 for "earned at least 30 credits in the first year," 174 for "retained into the second year," 177 for "earned a bachelor's degree within six years," and 179 for "earned more than 200 percent of poverty line in the first year after graduation."

†The model for "Earned >200 percent of poverty line in the year after graduation" did not meet the criteria for good model fit and should be interpreted with caution.

Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

What are observed levels of success in college and in the labor market at various potential cut scores?

In the second part of this analysis, we examine the levels of student success at different points on the NAEP scale. In particular, given that results in the first part of the analysis clustered around the NAEP *Proficient* cut score of 176, it is important to compare success at that score with success at the NAGB college academic preparedness cut score of 163.[18] The levels of success for each of the four outcomes at these two points, as well as at the NAEP *Basic* and *Advanced* cut scores, are displayed in Exhibit 4, where the rates of success are measured on the vertical axis.

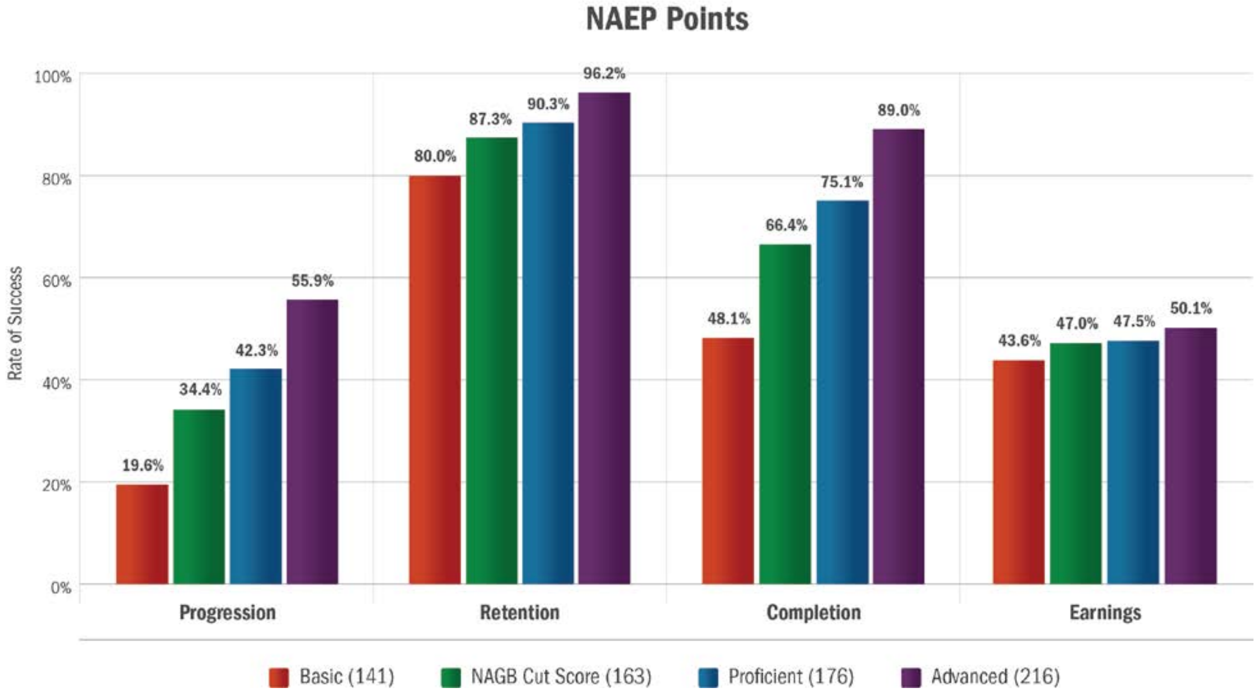
As one would expect, the levels of success across cut points for various outcomes differ. What is key in this figure, however, is for a given measure, how different the levels of success are at the NAEP *Proficient* cut score (176) and the NAGB college academic preparedness cut score (163). The results indicate that between these two scores, the rates of success differ by the following:

- 7.9 percentage points for Progression
- 3.0 percentage points for Retention
- 8.6 percentage points for Completion
- 0.5 percentage points for Earnings

The largest difference occurred in the levels of Completion: at the NAGB college academic preparedness cut score, about two thirds of the students were counted as successful; at the NAEP *Proficient* cut score, about three fourths were counted as successful.

[18] In our sample, about 18 percent of the students scored 163 or higher but less than 176.

Exhibit 4. Levels of Success at Various NAEP Mathematics Grade 12 Cut Scores, by Indicator



Note: Students scoring 176 on the NAEP mathematics Grade 12 assessment, the cut score for NAEP *Proficient*, were estimated to have a 42 percent probability of progressing in the first year (earning at least 30 credits), 90 percent probability of retention (enrolling in the second year), 75 percent probability of completion (earning a bachelor's degree within six years), and 47 percent probability of labor market success after graduation (earning more than 200 percent of poverty line in the first year after graduation).
Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

In summary, although the relationship between NAEP scores and success met the study's minimum criteria for reporting, two success measures—retention and labor market success—indicate that NAEP has little relation to success. Large differences do not appear in the levels of success at an NAEP mathematics score of 163 compared with a score of 176. For progression and completion, the levels of success differ by more than 7 or 8 percentage points. These differences raise a normative question that this report does not answer: are the observed levels of success at 163 or 176 more desirable for a college academic preparedness cut score?[19]

---

[19] The analysis for Research Question 1 provided an empirical approach to choosing a cut score (and suggested a cut score of around 176). In contrast, when studying a different measure of success, course GPA, Allen and Sconing (2005) set, as a normative threshold of success for readiness, the point where students had a 50 percent probability of earning a GPA of B or higher; Kobrin (2007) considered normative thresholds both where students had a 65 percent probability of earning a C or higher and where students had a 65 percent probability of earning a B− or higher.

Conclusion

To add to the literature about using NAEP as an indicator of college and career preparedness, we examined NAEP Grade 12 mathematics scores in relation to four observed measures of postsecondary success (progression, retention, completion, and earnings post college). We analyzed data for a cohort of almost 31,000 recent high school graduates under the age of 21 who enrolled in four-year higher education institutions in Virginia. We found that the optimal preparedness cut scores ranged from 174 to 181, closer to NAEP's long-established Grade 12 mathematics cut point for *Proficient* than NAGB's new college academic preparedness cut score of 163.

In further analysis, two of the measures analyzed, retention and labor market success, showed little relationship to test scores. For these measures, the levels of success at the NAEP Grade 12 mathematics scores of 163 (college academic preparedness cut score set by NAGB) and 176 (NAEP *Proficient* cut score) differed by 3 percentage points or less. For the two other measures, progression and completion, the levels of success differed by more than 7.9 and 8.6 percentage points, respectively. These differences raise the normative question: What level of success in these measures is most desirable for a preparedness indicator?

Although our study population was similar to that of other studies of college preparedness, it is likely not reflective of the entire national grade 12 population (i.e., the population that NAEP is designed to represent) for two reasons. First, the population is limited to students who graduated high school and enrolled in a four-year institution, leaving out who either did not finish high school, did not enroll in college in Virginia, or enrolled in a Virginia community college. Second, the data are from Virginia, a state with a relatively affluent and well-educated population. This matters because the methodology used is sensitive to the population included in the study: including a lower achieving population will likely result in lower empirically optimal cut scores (using ROC analysis) and lower rates of observed success.

Though limitations of these data prevent generalization to the entire NAEP Grade 12 population, the analyses demonstrate how longitudinal data on outcomes can be used to determine preparedness indicators, and the results suggest that using NAEP *Proficient* as a cut score for preparedness may be more than appropriate for NAEP. For more generalizable results, these analyses should be replicated with a full population of 12th graders at the state level or with a nationally representative sample of 12th graders.

References

ACT. (2011). *The condition of college & career readiness*. Iowa City, IA: Author. Retrieved from http://www.act.org/research/policymakers/cccr11/pdf/ConditionofCollegeandCareerReadiness2011.pdf

Allen, J., & Sconing, J. (2005). *Using ACT assessment scores to set benchmarks for college readiness (ACT Research Series 2005-3)*. Iowa City, IA: ACT. Retrieved from http://www.act.org/research/researchers/reports/pdf/ACT_RR2005-3.pdf

Berkner, L., & Chavez, L. (1997). *Access to Postsecondary Education for the 1992 High School Graduates. (NCES 98-105)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Blake, P., Kang, W., & Massa, T. (2015, September 14). *Demography, planning and measuring success: Why they matter*. Presentation to the Joint Subcommittee on the Future Competitiveness of Virginia Higher Education. Retrieved from http://www.schev.edu/council/presentations/JointSubHigherEd91415.pdf

Conley, D. (2010). *College and career ready: Helping all students succeed beyond high school*. San Francisco, CA: Jossey-Bass.

Fields, R. (2014). *Towards the National Assessment of Educational Progress (NAEP) as an indicator of academic preparedness for college and job training*. Washington, DC: National Assessment Governing Board. Retrieved from http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/NAGB-indicator-of-preparedness-report.pdf

Greene, J. P., & Winters, M. A. (2005). *Public high school graduation and college-readiness rates: 1991–2002* (Education Working Paper No. 8). Retrieved from http://www.manhattan-institute.org/html/ewp_08.htm

Gönen, M. (2007). *Analyzing receiver operating characteristic curves with SAS.* Cary, NC: SAS Institute.

Hosmer, D. W., & Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics-Theory and Methods*, *9*(10), 1043–1069.

Kobrin, J. L. (2007). *Determining SAT benchmarks for college readiness* (Research Notes RN-30). New York, NY: College Board Office of Research and Analysis. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-2007-30-sat-benchmarks-college-readiness.pdf

Moran, R., Freund, D., & Oranje, A. (2012). *Analyses relating Florida students' performance on NAEP to preparedness indicators and postsecondary performance*. Washington, DC: Author. Retrieved from https://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/Florida_Statistical_Study.pdf

Moran, R., Oranje, A., & Freund, D. (n.d.). *NAEP 12th grade preparedness research: Establishing a statistical relationship between NAEP and SAT*. Retrieved from http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/SAT-NAEP_Linking_Study.pdf

National Commission on NAEP 12th Grade Assessment and Reporting. (2004). *12th grade student achievement in America: A new version for NAEP.* Washington, DC: Author. Retrieved from https://www.nagb.org/content/nagb/assets/documents/publications/12_gr_commission_rpt.pdf

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157–175.

Radunzel, J., & Nobel, J. (2013). *Differential effects on student demographic groups of using ACT college readiness assessment composite score, ACT benchmarks, and high school grade point average for predicting long-term college success through degree completion* (ACT Research Report Series). Iowa City, IA: ACT.

UCLA. (2007). *Introduction to SAS.* Retrieved from http://www.ats.ucla.edu/stat/sas/notes2/

Uekawa, K., Merola, S., Fernandez, F., & Porowski, A. (2010). *Creating an early warning system: Predictors of dropout in Delaware* (Technical Assistance Brief). Calverton, MD: Regional Educational Laboratory Mid-Atlantic. Retrieved from http://dedoe.schoolwires.net/cms/lib09/DE01922744/Centricity/Domain/91/MA1275TAFINAL508.pdf

Vivo, J-M., & Franco, M. (2008). How does one assess the accuracy of academic success predictors? ROC analysis applied to university entrance factors. *International Journal of Mathematical Education in Science & Technology*, *39*(3), 325–340.

Wyatt, J., Kobrin, J., Wiley, A., Camara, W. J., & Proestler, N. (2011). *Development of a College Readiness Benchmark and Its Relationship to Secondary and Postsecondary School Performance.* New York, NY: College Board. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-5-sat-college-readiness-benchmark-secondary-performance.pdf

Wyatt, J. N., Remigio, M., & Camara, W. J. (2012). *SAT subject area readiness indicators: Reading, writing, and STEM.* New York, NY: College Board Office of Research and Analysis. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2013/1/researchnote-2013-1-sat-subject-readiness-indicators_0.pdf

Appendix A. Data

The State Council of Higher Education for Virginia analysis data contains 30,652 records. The individuals in the analysis data set were first-time enrollees in a higher education institution in the state of Virginia for academic year 2006–07, had graduated high school within the previous 12 months of college entry, and were under the age of 21 at the time of college entry. The higher education institutions include four-year public and four-year not-for-profit private colleges. The students included in the data include both residents of Virginia and nonresidents. The analysis data includes only those records with an available SAT mathematics score of at least 260.[20]

Table A1 compares the gender and race distribution of the analytic data set to two external reference points:

- The 2005–06 Virginia 12th-grade population (according to fall enrollment counts)
- The 2008–09 Virginia college-bound graduates (high school students who entered high school in 2004, earned a diploma by the 2008–09 school year, and entered any higher education institution included in the National Student Clearinghouse within 16 months of graduation)[21]

Compared with these two reference points, the students included in the analysis contained:

- A percentage of females that was more than 4 percentage points higher than the 12th-grade population but less than 1 percentage point higher than the college-bound seniors
- A percentage of African-American students that was more 7 percentage points lower than either of the reference points

The results shown in the table imply that the findings of this study should be interpreted with caution and cannot be generalized to all 12th graders in Virginia or all college-bound graduates. .

Table A1. Comparison of the Analytic Data Set and 2005–06 Virginia 12th Graders (Population)

Student Characteristic	Analysis Data (Students entering VA 4-year intuitions in 2005–06, graduated in past year, under age 21, had SAT score 260 or higher)	Virginia 12th-Grade Students 2005–06	Virginia College-Bound High School Graduates
Gender			
Male	45.15%	49.86%	45.56%
Female	54.85%	50.14%	54.44%

[20] An SAT mathematics score less than 260 cannot be linked to an NAEP score.
[21] Information on student entering high school in 2004 was the earliest available.

Student Characteristic	Analysis Data (Students entering VA 4-year intuitions in 2005–06, graduated in past year, under age 21, had SAT score 260 or higher)	Virginia 12th-Grade Students 2005–06	Virginia College-Bound High School Graduates
Race/ethnicity			
African American or Black	13.26%	23.69%	21.10%
American Indian/Native American[a]	<1%	<1%	<1%
Asian and Pacific Islander	7.27%	5.44%	6.71%
Hispanic	3.45%	5.49%	4.30%
White, Caucasian American	67.25%	64.23%	66.32%
Unknown	8.39%	<1%	1.57%

Note. "Virginia College Bound High School Graduates" are students who entered high school in 2004, earned a diploma by 2008–09, and entered any higher education institution include in the National Student Clearinghouse database within 16 months of graduation.

Source. Analysis data characteristics were from authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year. Information on "Virginia 12th-Grade Students" was obtained from the Virginia Department of Education (VDOE) fall membership reports for 12th graders in school year 2005–06 (http://www.doe.virginia.gov/statistics_reports/enrollment/index.shtml). Information on "Virginia College-Bound High School Graduates" was obtained from the VDOE High School Graduates Postsecondary Enrollment Reports webpage (https://p1pe.doe.virginia.gov/postsec_public/postsec.do?dowhat=LOAD_REPORT_C11).

[a] Given study confidentiality guidelines, the size of this subgroup is considered too small to analyze as a subgroup.

Appendix B. Methodology

Statistical Linkage Between the National Assessment of Educational Progress and the SAT

For the purposes of examining the connection between NAEP and college or career preparedness, students in the data were assigned NAEP mathematics scores based on the SAT mathematics score recorded in their college admissions records. Although the data did not contain NAEP scores, they did include SAT scores, which we link to NAEP scores following Moran et al. (2012). Moran et al. (2012), as part of a study for the National Assessment Governing Board (NAGB), found that the correlation between NAEP mathematics and SAT mathematics scores was strong enough to support the creation of a concordance table, while the correlation between NAEP reading and SAT critical reading scores was not. Although the concordance tables established by Moran et al. (n.d.) were not available, Tirre (personal communication, January 8, 2015) used similar methods to create concordance tables for this study using the 2009 High School Transcript Study data. As with Moran et al. (2012), the correlations between NAEP reading and SAT critical reading were too low to support the use of a concordance table.

With this linkage between NAEP and SAT, we were able to conduct analyses to answer the two research questions, which are discussed in separate sections that follow. Analyses were conducted with SAT scores, and the results were then translated into NAEP scores for reporting.

Statistical Analysis of Outcomes and SAT Scores

Research Question 1: What is the empirically determined optimal NAEP mathematics cut score for distinguishing success?

To answer this research question, we conducted receiver operating characteristic (ROC) curve analyses (Gönen, 2007) to calculate the optimal point on the NAEP scale, that is, the score above (or below) which students could be expected to be successful (or not successful). An optimal cut score in this analysis is a trade-off between two concepts: sensitivity and specificity. Sensitivity is a ratio of the number of correctly classified (or predicted) successes over the total number of actual successes. Specificity is a ratio of the number of correctly classified nonsuccesses over the number of actual nonsuccesses. The perfect sensitivity would be 1, which would indicate that all the actual successes were correctly classified (or predicted) as successes. The perfect specificity also would be 1, which would indicate that all the actual nonsuccesses are correctly classified (or predicted) as nonsuccesses. There is (in almost all real-life cases) a trade-off between sensitivity and specificity: as one increases the other decreases.[22] Hence, ROC analysis defines the optimal

[22] For example, if one chose a score above the maximum NAEP score as the cut score, all students would be predicted to be "nonsuccesses" and sensitivity would be 0 (none of the actual successes would be correctly classified) and specificity would be 1 (all of the actual nonsuccesses would be correctly classified). In contrast, if one chose a score of 0 as the cut score, all students would be predicted to be "successes" and sensitivity would be 1 (all of the actual successes would be correctly classified) and specificity would be 0 (none of the actual nonsuccesses would be correctly classified).

cut score on the NAEP scale (or on its equivalent SAT scale) as the cut score where the differences between the actual sensitivity and sensitivity from their optimal values is minimized. Specifically, the optimal cut score is defined by the following equation:

$$OptimalCutScore = \min_{s}\{(1 - sensitivity_s)^2 + (1 - specificity_s)^2\}$$

Where:

s is a test score;

$sensitivity_s$ is the sensitivity at scores s; and

$specificity_s$ is the specificity at scores s.

For each of our outcomes measures, we calculated an optimal SAT cut score and translated that into an optimal NAEP cut score using the concordance tables. We also calculated the sensitivity and specificity for each measure.

In addition, we assessed the quality of cut scores. Following Uekawa et al. (2010), we set two minimum criteria that the optimal cut scores were expected to meet based on the overall accuracy (number of true positives and true negatives divided by the total number of observations) of the optimal cut score as well as the area under the ROC curve (AUC), which is a general measure of predictive ability of the model:

1. The accuracy must both be greater than or equal to .5. The rationale for this is the cut score should be at least robust enough to accurately predict outcomes for half the students.
2. The AUC must be greater than .6. The rationale is that AUC ranges from 0 to 1, and random guessing would result in an AUC of .5. This should, however, be understood as a minimum cut off. Others (e.g., Vivo, 2008) have used .7 as a cut off for a reliable indicator.

Research Question 2: What are observed levels of success in college and in the labor market at various potential cut scores?

To answer this research question, we followed techniques used in similar studies that examined the association between Grade 12 achievement (as measured by the SAT and ACT) and college and career success (Kobrin, 2007; Radunzel & Nobel, 2013; Wyatt, Remigio, & Camara, 2012). In this study, we used four binary success indicators (i.e., measure of success on obtaining more than 30 credits in the first year, measure of success on retention into the second year, measure of success on earning a bachelor’s degree within six years, and measure of success on earning higher than twice the poverty threshold for a household size of one the year after degree attainment), one at a time, as the dependent variable in a logistic regression model with the SAT measure as the sole independent variable:

$$y_i = \frac{1}{1+e^{-(\beta_0+\beta_1 SAT_i+\beta_2 SAT_i^{2}+\beta_3 SAT_i^{3})}} \quad (1)$$

Where:

y_i = the binary outcome indicator of success for student *i*;

SAT_i = the SAT score for student *i;*

β_0 = an intercept term (the hypothetical estimated log-odds of success if the SAT score were equal to zero);

β_1 = the increase in estimated log-odds of success for a 1-point increase in the SAT score;

β_2 = the increase in estimated log-odds of success for a 1-point increase in the square term of SAT score; and

β_3 = the increase in estimated log-odds of success for a 1-point increase in the cubic term of SAT score.

The maximum likelihood estimation method was used to estimate these models.

In determining the optimal functional form of SAT in the model, we used the likelihood ratio test[23] (UCLA, 2007) to compare the linear, quadratic, and cubic specifications. As suggested by the results, the optimal specification for most outcome measures[24] was the model with cubic terms. For the purpose of consistency in our analysis, we used this functional form for analysis of all outcome measures.

We also used regression diagnostic statistics (e.g., goodness-of-fit measures) to assess how accurate the model was in predicting success, or, in other words, how much noise there was in our estimated relationship between 12th-grade achievement and each measure of college and career success. In particular, we used two goodness-of-fit tests—the Hosmer-Lemeshow test[25] (Hosmer & Lemeshow, 1980) and Pearson chi-square test (Pearson, 1900)—which tested the extent to which the observations for which we *predicted* success on the basis of 12th-grade achievement closely matched actual *observed* success.

After the regression models were estimated, they provided a means for calculating the probability of success for students at each NAEP/NAGB proficiency cut score. First, each cut score was translated into an SAT equivalent using the correspondence tables. Second, the estimated model (Equation 1) was evaluated at each of the four SAT equivalents for the four proficiency-level cut scores to obtain estimated probabilities of success at each.

[23] The likelihood ratio test statistic is calculated in the following way: LR = -2 ln(L(m1)/L(m2)) = 2(ll(m2)-ll(m1)), where L(m*) denotes the likelihood of the respective model, and ll(m*) the natural log of the models' likelihood.
[24] For the measure on retention into the second year, the optimal model was the linear model.
[25] The Hosmer-Lemeshow goodness-of-fit test is based on splitting the sample into 10 groups according to their predicted probabilities.

Appendix C. Regression Results

The results for regression and diagnostic tests are presented separately by research questions.

Research Question 1: What scores on the NAEP scale are related to measures of career and college success?

Tables C1 through C4 present results for the analysis of optimal cut scores on the NAEP scale related to the study's measures of success. Analysis was conducted both for the entire study population as well as separately for the populations of subgroups defined by Pell Grant receipt and race/ethnicity. One would expect ROC results to change when the data used for analysis change. Results by subgroup illustrate how sensitive results are to the population used for analysis.

The tables present results in terms of both SAT scores ("Optimal SAT Cut Score" row) and NAEP scores ("Optimal NAEP Cut Score" row) as determined by the concordance tables. For example, as shown in Table C1, the optimal cut score for success in progression (earning 30 or more credits in the first year) was calculated to be 580 on the SAT scale, or 181.3 on the NAEP scale. This optimal cut score on the NAEP scale is about 5 points higher than the NAEP *Proficient* cut score of 176.

Results from the subgroup analyses show that variation in the identified optimal cut scores. For example, as displayed in Table C1, the optimal "progression" NAEP cut score for Pell Grant recipients was 167.5, while the optimal cut score for students who did not receive a Pell Grant was 183.6—a difference of about 16 NAEP points. Differences across subgroups reflect the specific method used to obtain the optimal cut score: when limited to a lower (or higher) performing population, the analysis will produce a lower (or higher) optimal cut score where sensitivity and specificity are balanced. The tables also report the sensitivity and specificity of each optimal cut score where the perfect sensitivity or specificity would be 1.

Finally, the tables present our measures of the quality of the optimal cut scores: accuracy and AUC. Note that although accuracy is related to sensitivity and specificity, it is a different concept based on the correct classification all students in the analysis while sensitivity and specificity are based on the correct classification of successful and unsuccessful students, respectively. For analysis of all students the accuracy and AUC were above our predetermined thresholds for all measures except labor market success (Table C4), where the AUC was below .6. For subgroup analysis, the accuracy and AUC exceed the thresholds for almost all subgroups for progression (Table C1), retention (Table C2), and completion (Table C3). However, when looking at labor market success (Table C4), as with the entire population, the AUC threshold was not exceeded for any subgroup.

Table C1. Optimal Cut Scores for Progression (Earning 30+ Credits in the First Year of College), Grade 12 Mathematics

	All	Pell Grant Recipients	Pell Grant Nonrecipients	African American or Black	Asian and Pacific Islander	Hispanic	White, Caucasian American	Unknown Race/Ethnicity
Optimal SAT cut score (from ROC analysis)	580	520	590	480	630	560	590	610
Optimal NAEP cut score (from concordance table)	181.3	167.5	183.6	158.4	192.7	176.7	183.6	188.2
Sensitivity (from ROC analysis; goal = close to 1)	0.58	0.61	0.56	0.54	0.56	0.59	0.56	0.59
Specificity (from ROC analysis; goal = close to 0)	0.39	0.38	0.39	0.4	0.39	0.4	0.4	0.37
Accuracy	0.6	0.62	0.59	0.58	0.58	0.6	0.58	0.61
AUC	0.63	0.65	0.62	0.6	0.62	0.63	0.61	0.66
N	30,625	4,755	25,363	4,057	2,226	1,058	20,597	2,567

Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Table C2. Optimal Cut Scores for Retention (Enrolling in the Second Year of College), Grade 12 Mathematics

	All	Pell Grant Recipients	Pell Grant Nonrecipients	African American or Black	Asian and Pacific Islander	Hispanic	White, Caucasian American	Unknown Race/Ethnicity
Optimal SAT cut score(from ROC analysis)	550	490	570	460	610	540	570	580
Optimal NAEP cut score (from concordance table)	174.4	160.7	179	153.8	188.2	172.1	179	181.3
Sensitivity (from ROC analysis; goal = close to 1)	0.6	0.59	0.57	0.56	0.56	0.59	0.57	0.6
Specificity (from ROC analysis; goal = close to 0)	0.39	0.4	0.37	0.42	0.41	0.34	0.38	0.38
Accuracy	0.6	0.59	0.57	0.56	0.56	0.59	0.58	0.6
AUC	0.65	0.62	0.64	0.6	0.58	0.67	0.63	0.66
N	30,652	4,757	25,370	4,063	2,227	1,058	20,613	2,571

Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Table C3. Optimal Cut Scores for Completion (Earning a Bachelor's Degree in Six Years), Grade 12 Mathematics

	All	Pell Grant Recipients	Pell Grant Nonrecipients	African American or Black	Asian and Pacific Islander	Hispanic	White, Caucasian American	Unknown Race/Ethnicity
Optimal SAT cut score(from ROC analysis)	560	510	570	470	620	550	570	590
Optimal NAEP cut score (from concordance table)	176.7	165.2	179	156.1	190.4	174.4	179	183.6
Sensitivity (from ROC analysis; goal = close to 1)	0.62	0.59	0.60	0.56	0.56	0.58	0.61	0.60
Specificity (from ROC analysis; goal = close to 0)	0.36	0.34	0.37	0.39	0.37	0.33	0.37	0.35
Accuracy	0.62	0.62	0.61	0.58	0.57	0.61	0.61	0.61
AUC	0.68	0.66	0.66	0.62	0.61	0.68	0.66	0.67
N	30,652	4,757	25,370	4,063	2,227	1,058	20,613	2,571

Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Table C4. Optimal Cut Scores for Labor Market Success (Earning Higher Than 200 Percent of Poverty for a Household Size of One the Year After Graduation), Grade 12 Mathematics

	All	Pell Grant Recipients	Pell Grant Nonrecipients	African American or Black	Asian and Pacific Islander[a]	Hispanic	White, Caucasian American	Unknown Race/Ethnicity
Optimal SAT cut score(from ROC analysis)	570	530	580	480	620	560	580	600
Optimal NAEP cut score (from concordance table)	179	169.8	181.3	158.4	190.4	176.7	181.3	185.9
Sensitivity (from ROC analysis; goal = close to 1)	0.47	0.5	0.5	0.52	0.53	0.49	0.5	0.51
Specificity (from ROC analysis; goal = close to 0)	0.47	0.47	0.49	0.44	0.47	0.44	0.5	0.48
Accuracy	0.51	0.52	0.5	0.54	0.47	0.53	0.5	0.52
AUC	0.51	0.51	0.51	0.53	0.53	0.54	0.51	0.52
N	13,331	1,834	11,433	1,464	1,005	418	9,299	1,085

[a] For the Asian and Pacific Islander subgroup, the relationship between the outcome measure and SAT scores is negative.

Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Research Question 2: How successful in college and in careers are students at each NAEP proficiency level (*Basic, Proficient,* and *Advanced*)?

Tables C5–C8 present results for the analysis of the levels of success in college and in careers at multiple NAEP scores: the NAEP *Basic* cut score, the NAGB college academic preparedness cut score, the NAEP *Proficient* cut score, the NAEP *Advanced* cut score, and the optimal cut score for each given measure as determined by the analysis for research question 1 using all students. As with the first research question, the analysis was conducted for the entire study population as well as separately for each subgroup (i.e., Pell Grant status and race/ethnicity).

In Tables C5–C8, the column "SAT Equivalent of Cut Score" shows the SAT mathematics scores equivalent to each NAEP proficiency cut score per the concordance tables (W.C. Tirre, personal communication, January 8, 2015). The column "All" presents results for the entire analysis population. Results indicate the probability of success in a given measure for a student if the student has a specific NAEP score (or SAT equivalent). For example, as shown in Table C5, a NAEP Grade 12 mathematics score of 176, that is, the NAEP *Proficient* cut score, is equivalent to a score of 557 on the SAT mathematics test. A student obtaining this score was estimated to have a 42.3 percent chance of earning more than 30 credits in the first year of college. In contrast, a student who scored at 141 on NAEP, at the NAEP *Basic* cut score and equivalent to an SAT score of about 404, was estimated to have about a 19.6 percent chance of earning more than 30 credits in the first year of college—about 20 percentage points less than a student who was scoring at the NAEP *Proficient* cut score.

Similarly, when broken down by Pell Grant recipient status or race/ethnicity, the results show the probability of success for a given outcome for each subgroup. For example, as shown in Table C5, at the NAEP *Proficient* cut score, students who did not have Pell Grants have a higher chance of earning more than 30 credits in the first year of college (43.75 percent) than the Pell Grant recipients (35.88 percent), and White students (44.54 percent) have a slightly higher probability than Asian and Pacific Islanders students (43.99 percent). Differences in success across subgroups are likely due to factors other than Grade 12 achievement as measured by the test scores used in the analysis, and they reflect the extent to which other factors, such as socioeconomic status, influence outcomes.

The tables also provide results of regression diagnostic statistics (i.e., goodness-of-fit test) that indicate whether the regression model is a satisfactory fit to the data. Two goodness-of-fit measures were used to assess the models, both using a null hypothesis that the fitted model is correct. Test statistics are reported with *p* values in parentheses below them. *P* values equal to or greater than .05 indicate that the regression model is consistent with the data, whereas *p* values lower than .05 imply that the model does not fit the data well. A *p* values of zero means that the model has no predictive power. As shown in Tables C5–C8, the results of both Pearson chi-square test and Hosmer-Lemeshow[26] test indicate that the regression model is a satisfactory fit to the data across outcome measures and subgroups.

[26] The Hosmer-Lemeshow goodness-of-fit test is based on splitting the sample into 10 groups according to their predicted probabilities.

Table C5. Probabilities of Success for Earning 30+ Credits in the First Year of College, Grade 12 Mathematics

			Probability of Success at Cut Score (From Estimated Logistic Regression)							
NAEP Proficiency Level	**NAEP Proficiency Level Cut Score(Bottom of the Proficiency Level)**	**SAT Equivalent of Cut Score (From Concordance Table)**	**All**	**Pell Grant Recipients**	**Pell Grant Nonrecipients**	**African American or Black**	**Asian and Pacific Islander**	**Hispanic**	**White, Caucasian American**	**Unknown Race/Ethnicity**
Basic	141	404.3	19.60%	16.49%	21.32%	22.69%	14.77%	11.76%	17.14%	6.48%
NAGB cut score	163	500.4	34.40%	28.04%	36.16%	31.00%	35.41%	29.94%	36.15%	26.82%
Proficient	176	557	42.30%	35.88%	43.75%	35.58%	43.99%	37.69%	44.54%	38.50%
Optimal cut score from ROC analysis	181.3	580	45.07%	38.94%	46.39%	37.49%	46.53%	39.93%	47.06%	42.16%
Advanced	216	731.7	55.85%	50.22%	57.50%	54.44%	59.41%	50.30%	54.08%	57.17%
Goodness-of-fit test (Pearson chi-square)[a]			61.58 (0.15)	69.73 (0.04)	59.12 (0.20)	40.02 (0.82)	44.46 (0.41)	19.20 (1.00)	58.94 (0.21)	43.63 (0.65)
Goodness-of-fit test (Hosmer-Lemeshow chi-square)[b]			11.23 (0.19)	9.99 (0.27)	7.43 (0.49)	12.83 (0.11)	11.77 (0.16)	3.83 (0.87)	13.76 (0.09)	6.93 (0.54)
N			30,625	4,755	25,363	4,057	2,226	1,058	20,597	2,567

[a] The probability > chi-square is presented in parentheses.
Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Table C6. Probabilities of Success for Retention Into the Second Year of College, Grade 12 Mathematics

NAEP Proficiency Level	NAEP Proficiency Level Cut Score(Bottom of the Proficiency Level)	SAT Equivalent of Cut Score (From Concordance Table)	Probability of Success at Cut Score (From Estimated Logistic Regression): All	Pell Grant Recipients	Pell Grant Nonrecipients	African American or Black	Asian and Pacific Islander	Hispanic	White, Caucasian American	Unknown Race/Ethnicity
Basic	141	404.3	80.00%	77.04%	81.81%	79.99%	89.12%	75.40%	80.00%	79.77%
NAGB cut score	163	500.4	87.31%	84.26%	88.66%	85.62%	92.79%	84.16%	87.62%	86.05%
Optimal cut score from ROC analysis	174.4	550	89.97%	87.19%	90.97%	87.69%	93.82%	88.49%	90.11%	89.10%
Proficient	176	557	90.30%	87.56%	91.25%	87.97%	93.94%	89.04%	90.42%	89.51%
Advanced	216	731.7	96.21%	92.94%	96.46%	94.81%	96.00%	96.96%	96.08%	97.38%
Goodness-of-fit test (Pearson chi-square)[a]			48.03 (0.59)	54.85 (0.33)	36.16 (0.94)	55.12 (0.25)	57.45 (0.07)	33.81 (0.87)	52.22 (0.43)	46.98 (0.51)
Goodness-of-fit test (Hosmer-Lemeshow chi-square)[b]			2.87 (0.94)	6.66 (0.57)	5.28 (0.73)	2.25 (0.97)	4.34 (0.82)	5.72 (0.68)	7.62 (0.47)	12.30 (0.14)
N			30,652	4,757	25,370	4,063	2,227	1,058	20,613	2,571

[a] The probability > chi-square is presented in parentheses.

Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Table C7. Probabilities of Success for Earning a Bachelor's Degree in Six Years, Grade 12 Mathematics

NAEP Proficiency Level	NAEP Proficiency Level Cut Score(Bottom of the Proficiency Level)	SAT Equivalent of Cut Score (From Concordance Table)	Probability of Success at Cut Score (From Estimated Logistic Regression)							
			All	Pell Grant Recipients	Pell Grant Nonrecipients	African American or Black	Asian and Pacific Islander	Hispanic	White, Caucasian American	Unknown Race/Ethnicity
Basic	141	404.3	48.08%	40.70%	52.45%	45.43%	57.52%	40.19%	50.74%	44.32%
NAGB cut score	163	500.4	66.44%	56.94%	69.41%	59.25%	71.26%	65.88%	68.21%	66.63%
Proficient	176	557	75.05%	66.15%	76.87%	66.17%	76.75%	74.86%	75.87%	76.64%
Optimal cut score from ROC analysis	176.7	560	75.45%	66.60%	77.21%	66.49%	76.99%	75.22%	76.23%	77.08%
Advanced	216	731.7	88.97%	80.87%	89.51%	76.56%	85.73%	86.32%	90.14%	87.97%
Goodness-of-fit test (Pearson chi-square)[a]			45.30 (0.70)	51.98 (0.44)	50.90 (0.48)	64.88 (0.06)	61.60 (0.03)	40.41 (0.63)	49.31 (0.54)	38.53 (0.83)
Goodness-of-fit test (Hosmer-Lemeshow chi-square)[b]			5.68 (0.68)	7.55 (0.48)	5.55 (0.70)	9.34 (0.31)	13.32 (0.10)	12.18 (0.14)	10.09 (0.26)	8.83 (0.36)
N			30,652	4,757	25,370	4,063	2,227	1,058	20,613	2,571

[a] The probability > chi-square is presented in parentheses.
Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Table C8. Probabilities of Success for Earning Higher Than 200 Percent of Poverty for a Household Size of One the Year After Graduation, Grade 12 Mathematics

			Probability of Success at Cut Score (From Estimated Logistic Regression)							
NAEP Proficiency Level	**NAEP Proficiency Level Cut Score (Bottom of the Proficiency Level)**	**SAT Equivalent of Cut Score (From Concordance Table)**	**All**	**Pell Grant Recipients**	**Pell Grant Nonrecipients**	**African American or Black**	**Asian and Pacific Islander**	**Hispanic**	**White, Caucasian American**	**Unknown Race/Ethnicity**
Basic	141	404.3	43.62%	43.44%	43.44%	38.21%	47.89%	33.10%	45.99%	39.62%
NAGB cut score	163	500.4	47.00%	44.90%	47.52%	43.85%	55.37%	41.48%	48.43%	44.16%
Proficient	176	557	47.48%	44.98%	47.93%	47.69%	55.84%	40.30%	47.70%	45.35%
Optimal cut score from ROC analysis	179	570	47.52%	44.98%	47.93%	48.23%	55.64%	40.02%	47.48%	45.53%
Advanced	216	731.7	50.08%	46.66%	50.43%	27.81%	47.55%	64.72%	51.55%	47.50%
Goodness-of-fit test (Pearson chi-square)[a]			62.42 (0.11)	55.93 (0.26)	50.72 (0.41)	38.43 (-0.74)	36.45 (-0.49)	43.40 (0.33)	51.23 (0.39)	46.67 (0.32)
Goodness-of-fit test (Hosmer-Lemeshow chi-square)b			7.68 (-0.47)	5.11 (0.75)	4.67 (0.79)	9.19 (-0.33)	14.06 (-0.08)	2.71 (-0.95)	4.89 (0.77)	10.73 (0.22)
N			13,331	1,834	11,433	1,464	1,005	418	9,299	1,085

[a] The probability > chi-square is presented in parentheses.

Source. Authors' calculations using SCHEV data on students entering Virginia higher education institutions in the 2006–07 school year.

Appendix D. Poverty Thresholds

Table D1 shows the numbers for 200 percent of the federal poverty threshold for a household size of one in years 2007–2013.

Table D1. 200 Percent of Poverty Threshold in Years 2007–2013

Year	Threshold
2007	$20,400
2008	$20,800
2009	$21,660
2010	$21,660
2011	$21,780
2012	$22,340
2013	$22,980

Source. U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.

ABOUT AMERICAN INSTITUTES FOR RESEARCH

Established in 1946, with headquarters in Washington, D.C., American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally. As one of the largest behavioral and social science research organizations in the world, AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce, and international development.

LOCATIONS

Domestic

Washington, D.C.

Atlanta, GA

Baltimore, MD

Chapel Hill, NC

Chicago, IL

Columbus, OH

Frederick, MD

Honolulu, HI

Indianapolis, IN

Naperville, IL

New York, NY

Rockville, MD

Sacramento, CA

San Mateo, CA

Waltham, MA

International

Egypt

Honduras

Ivory Coast

Kyrgyzstan

Liberia

Tajikistan

Zambia

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000 | TTY 877.334.3499

www.air.org

Making Research Relevant

3741_09/15