CCRI College and Career Readiness Initiative

High School Predictors of College Readiness:

Determinants of Developmental Course Enrollment and Second-Year Postsecondary Persistence in Virginia

December 2011

Texas Schools Project • Virginia Department of Education • National Center for Educational Achievement/ACT, Inc

High School Predictors of College Readiness:

Determinants of Developmental Course Enrollment and Second-Year Postsecondary Persistence in Virginia

Texas Schools Project
Marshall Garland
Jason LaTurner
Angelica Ware Herrera
Anne Ware

Virginia Department of Education
Deborah Jonas

National Center for Educational Achievement/ACT, Inc
Chrys Dougherty

December 2011

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

Introduction

In 2007, the Virginia Board of Education directed the Virginia Department of Education (VDOE) to study academic indicators that are associated with high school students' successful preparation for college and careers. VDOE then embarked on a multi-year effort to identify available data sources and conduct research on high school indicators that are associated with enrollment and success in credit-bearing courses in college. VDOE's initial studies linked high school indicators with students' enrollment in four-year colleges and universities across the country that was acquired from the National Student Clearinghouse.

Through this research, VDOE identified indicators of preparation for college that are independently associated with a high probability of enrollment and persistence in four-year postsecondary institutions. The Virginia-specific indicators identified in these initial studies are:

- Participating in a college preparatory course curriculum that includes Algebra II and chemistry.
- Earning advanced proficient scores on Virginia's statewide end-of-course assessments (Standards of Learning (SOL) assessments) in mathematics, reading, and writing.
- Earning an Advanced Studies diploma.
- Participating in the Virginia Early College Scholars program.[1]

Other indicators of students' preparedness for credit-bearing courses in postsecondary education include:

- Participation in Advanced Placement, International Baccalaureate, and dual-enrollment courses.
- Earning college ready scores on tests such as the SAT and ACT.[2]

The initial work primarily focused on enrollment in four-year colleges in large part because these institutions are less likely to enroll students in developmental (remedial) educational courses. A summary of these initial research findings is available from VDOE's College and Career Readiness Web site (see Appendix of the CCRI briefing for research summary, Virginia Department of Education, 2010).

The initial studies were limited, however, because the data sets did not include information about students' course enrollment and outcomes once they were in college; nor did the analysis address student success in two-year colleges. The latter limitation was significant because it was not clear whether the same factors associated with success in four-year colleges applied to two-year colleges. By linking high school data with information about enrollment and outcomes in college courses, including those in two-year colleges, VDOE sought to further understand how high school indicators are associated with success in postsecondary education. In particular, VDOE was interested in understanding how indicators that are readily available for nearly all of Virginia's high school students are associated with enrollment in credit-bearing courses in college.

This report is the first in a two-part series connecting high school program and outcomes data to course enrollment, grades, and persistence in Virginia's two- and four-year institutions of higher education (IHE). This first report addresses the following questions:

- How did students who enroll in Virginia two- and four-year colleges differ from those with no record of enrollment in a Virginia college?

[1] *The Early College Scholars program allows eligible high school students to earn at least 15 hours of transferable college credit while completing the requirements for an Advanced Studies Diploma or an Advanced Technical Diploma. More information is available at: http://www.doe.virginia.gov/instruction/graduation/early_college_scholars.*

[2] *The College Board and ACT have developed scores on their respective tests that represent the minimum required to have high probability of success in the first year of college.*

- To what extent were better prepared students less likely to enroll in developmental courses?
- To what extent did better high school academic preparation close income and ethnic gaps in student enrollment in developmental courses?
- To what extent were better prepared students more likely to persist into their second year in the Virginia higher education institution in which they enrolled in the first year?
- To what extent did better high school preparation close income and ethnic gaps in second-year persistence?

Following a brief description of the high school and higher education data files used in the analysis, the remaining sections of this report address each of these five questions. Throughout the report, we present both descriptive statistics that compare higher education outcomes for different groups of students, and results from statistical analyses that estimate the incremental association between a given high school variable and higher education outcomes after controlling for other variables. For example, students who earned a proficient score of 450 on the Algebra I SOL are likely to differ demographically from students who earn an advanced proficient score of 550. However, we are interested in the association between higher SOL scores and college outcomes for students with a given fixed set of demographic characteristics. Multivariate statistical models allow us to estimate the associations among variables of substantive interest—such as Algebra I SOL performance and college outcomes—while controlling for other differences among students.

Data Sources and Study Limitations

VDOE provided researchers at the Texas Schools Project (TSP) with four files. Each file contained deidentified records at the student level that could be linked to assemble a master data file. These files consisted of:

- A data file from VDOE with de-identified records on 86,312 high school graduates and non-graduate completers (students who earned GEDs and certificates of program completion) from the 2007-2008 school year.[3] This file contained information on the type of diploma the student earned (e.g., Advanced or Standard)[4], the student's high school SOL test scores, gender, ethnicity, economic disadvantaged status, month and year of birth, Limited English Proficient (LEP) status, dual enrollment status, and whether one or more Advanced Placement (AP) courses were taken. In addition, this file contained a binary (yes-no) indicator of whether a student was identified as enrolled in a Virginia postsecondary institution as a first-time entering freshman in the fall of 2008.
- A data file of de-identified records from the State Council of Higher Education for Virginia (SCHEV) that included data from 32,614 first-time entering freshmen at Virginia's IHE for the 2008-2009 school year. This file contained information on which students enrolled in which institutions in that year. A persistence flag was also included in the file, representing whether a student returned to the institution in the subsequent school year (fall 2009). The file included

[3] *This file in turn was excerpted from a file of multiple graduating cohorts, including the 2006, 2007, 2008, and 2009 graduating year, with a total of 334,174 observations.*

[4] *To earn a Standard Diploma, a student must earn at least 22 standard units of credit. These include four credits in English; three credits each in mathematics, laboratory science, and history and social studies; two credits in health and physical education; one credit in fine arts or career and technical education; and six credits in electives. For an Advanced Studies Diploma, a student must earn at least 24 standard units of credit. Of these credit hours, students must earn at least four credits each in English, mathematics, laboratory science, and history and social studies; three credits in foreign language; two credits in health and physical education; one credit in fine arts or career and technical education; and two credits in electives. (http://www.doe.virginia.gov/instruction/graduation/)*

students who attended private high schools or were home-schooled during high school, and therefore not all of these records could be matched to the VDOE high school graduating student records.

- A file with information of all students enrolled in Virginia's IHE (not just first-time freshmen) in 2008-2009, with information on the institution in which they were enrolled, the institution type (i.e., two-year, four-year, or private non-profit (PNP)), the student's enrollment status (i.e., full-time or part-time), demographic characteristics, financial aid recipient status, first and second academic major, and type of higher education degree earned, if any.

- A file on the English and mathematics courses taken by students in each semester of the 2008-2009 school year and the grades students earned in the courses. These files included both developmental and credit-bearing courses in those two subject areas.

Merging these four files produced a cohort of 31,434 unduplicated, matched students with information on their demographics, high school SOL scores, diploma types, AP and dual enrollment participation, higher education enrollment and second-year persistence, and English and mathematics developmental and credit-bearing courses taken in their freshman year. Appendix A includes a detailed description of the data sources and the matching process. Appendix B provides a more detailed description of the statistical models and techniques used to estimate the relationship between high school and higher education academic outcomes for students with a given set of demographic characteristics.

Findings from this study should be interpreted with caution due to several limitations in the data. First, the SCHEV enrollment file did not include students who enrolled in out-of-state higher education institutions. With the data available for this report, out-of-state enrollees could not be counted separately from students not enrolling in any higher education institution. Thus, this report uses "not enrolled" to mean "not enrolled in a Virginia higher education institution."

Second, the second year persistence flag provided by SCHEV only identified students who returned to the same institution in the subsequent fall semester (fall 2009). Consequently, students who left college could not be distinguished from those who transferred between institutions. This is an important limitation for two-year institutions with high percentages of students transferring to four-year or other two-year institutions after their first year.

Third, postsecondary enrollment and course records were provided only for the 2008-2009 academic year. Because of this, students who were required to enroll in a developmental course but postponed enrollment in the course to the following academic year were classified as having not enrolled in a developmental course. The prevalence of this situation could not be determined with the available data.

Study Findings

1. How did students who enroll in Virginia two- and four-year colleges differ from those with no record of enrollment in a Virginia college?

The cohort of students who enrolled in Virginia's two- and four-year colleges were better prepared and less likely to belong to minority and/or disadvantaged groups than were their counterparts who did not enroll in Virginia IHE in the 2008-2009 school year. Since the "non-enrolled" group also includes well-prepared students who enrolled out of state, the gaps in academic preparation between Virginia enrollees and students who did not enroll in any higher education institution are likely to be even larger than the ones shown in this report.

Demographics of enrollees and non-enrollees

As shown in Table 1, 71 percent of graduates who enrolled in a Virginia postsecondary institution in 2008 were White, with 19 percent African American, 5 percent Asian, and 3 percent Hispanic. Compared with the 2008 high school graduates who did not enroll in Virginia IHE, white students, particularly, were overrepresented (13.5 percentage points) in the enrolled student population, while African American students had the largest rate of underrepresentation (-9.3 percentage points). Hispanic students were underrepresented in the enrolled student cohort as well (3.5 percentage points).

Approximately 12 percent of all enrolled students were economically disadvantaged, compared to 25 percent of all non-enrolled students. **Figure 1** depicts the percentage of enrollees and non-enrollees who were classified as economically disadvantaged in high school, by ethnic group membership. In all ethnic groups, high school graduates and completers who were economically disadvantaged were less likely to enroll in a Virginia IHE than students who were not economically disadvantaged. For example, 45 percent of Hispanic students who were not enrolled were economically disadvantaged, compared to 30 percent of enrolled Hispanic students. Similar differences between non-enrollees and enrollees existed for African American students (42 percent economically disadvantaged not enrolled compared to 31 percent enrolled) and White students (15 percent not enrolled compared to 6 percent enrolled).

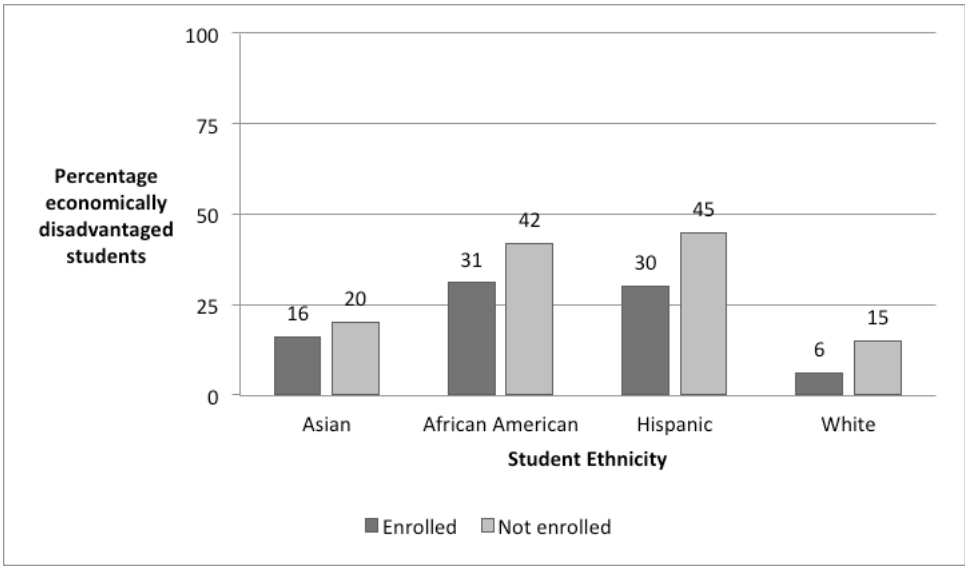
Another variable used in the analysis was whether a student was ever classified as LEP in a Virginia high school. (Students classified as LEP in elementary or middle school but not in high school were not recorded as LEP in this dataset.) Students who were identified as LEP in high school were less likely to enroll in a Virginia IHE. Among students enrolling in a Virginia higher education institution, 23 percent (**Figure 2**) of Asian students and 21 percent of Hispanic students were classified LEP at some point in a Virginia high school. However, nearly 44 percent of non-enrolled Hispanic students and 37 percent of non-enrolled Asian students had been classified as LEP in high school.

Table 1. Ethnic composition of fall 2008 graduates and non-graduate completers

	Percentage of enrolled student population	Percentage of not-enrolled student population	Disproportionate representation between not-enrolled and enrolled
Asian	5.2	5.8	-0.6
African American	18.7	28.0	-9.3
Hispanic	3.4	6.9	-3.5
White	71.2	57.7	13.5

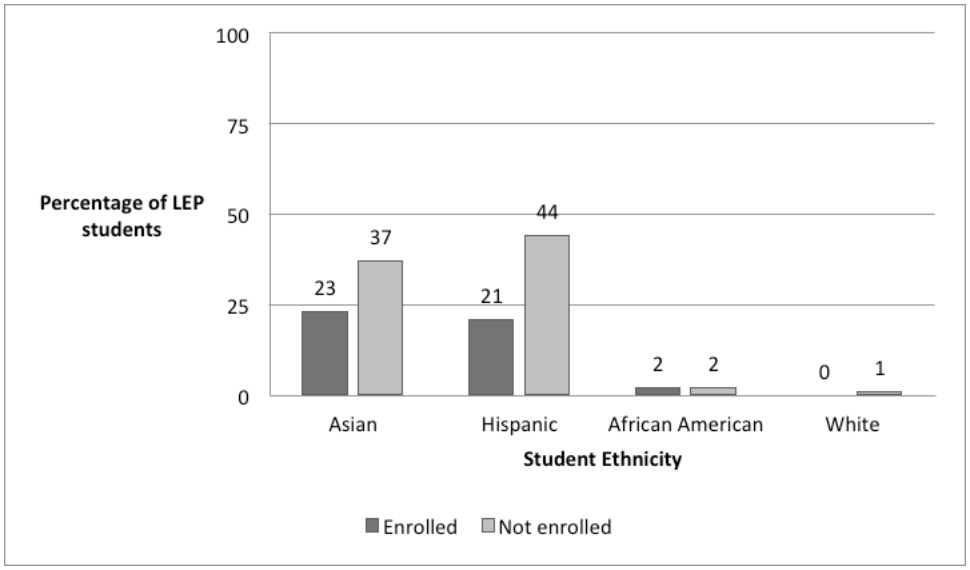
Source. Data taken from VDOE graduate file linked to SCHEV enrollment file.

Figure 1. Percentage of fall 2008 graduates and non-graduate completers classified as economically disadvantaged, by ethnicity



Source. Data taken from VDOE graduate file and SCHEV enrollment file.

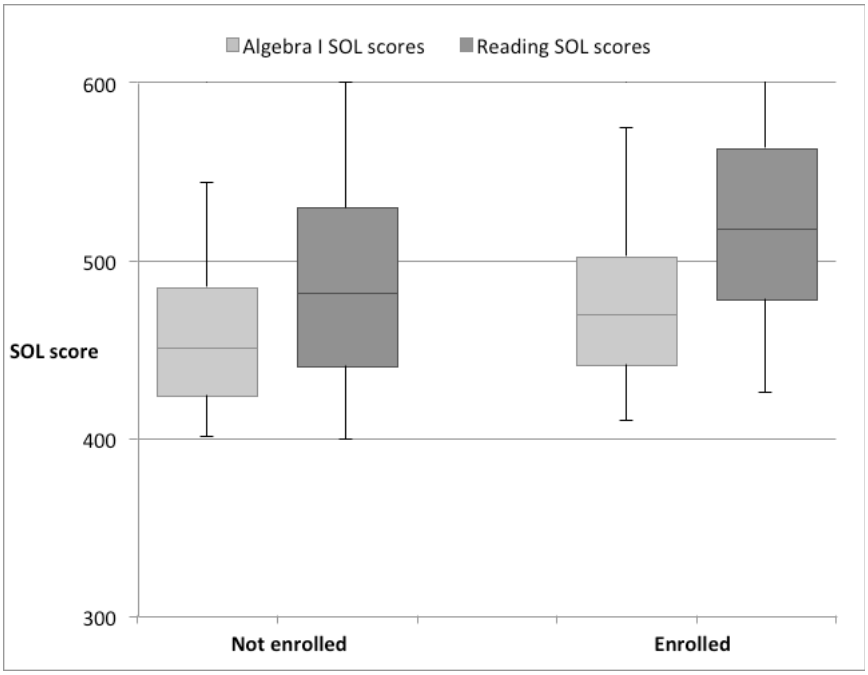
Figure 2. Percentage of fall 2008 graduates and non-graduate completers classified as Limited English Proficient at any time during secondary school, by student ethnicity



Source. Data taken from VDOE graduate file and SCHEV enrollment files.

Figure 3 shows how the distribution of English and Algebra I SOL scores differed between students who enrolled and did not enroll in a Virginia IHE. First, the median scores (denoted by the horizontal line in each box) for the Algebra I and reading SOL were higher among enrolled students than among non-enrolled students. The difference in median performance between enrolled and non-enrolled students was markedly greater on the reading SOL exam (36 points) than the Algebra I SOL (19 points). In general, reading SOL scores were closer to the top of the scale for both groups; for example, approximately 25 percent of all scores

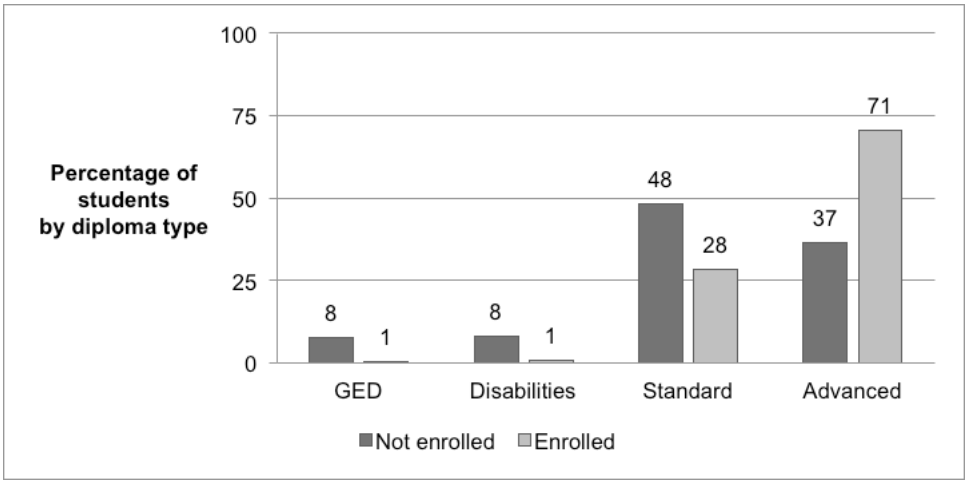
Figure 3. Distribution of Algebra I and reading Virginia Standards of Learning test scores for fall 2008 graduates and non-graduate completers, by Virginia IHE enrollment status



Source. Data taken from VDOE graduate file.

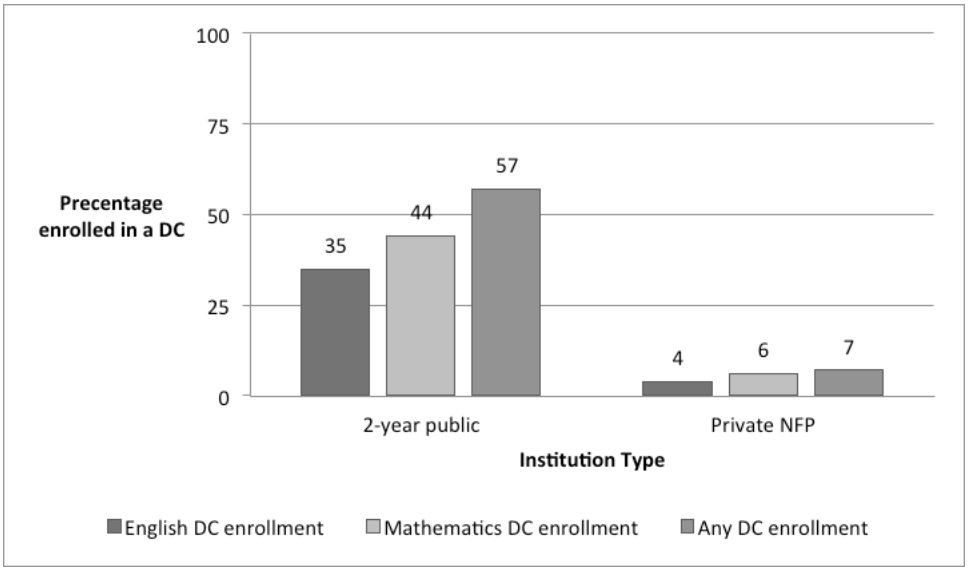
Note. Performance levels for the Virginia SOL assessments are: 0 to 400, fail; 400 to 500, pass proficient; 500 to 600, advanced proficient. The boxplot conveys several pieces of information about the distribution of a variable. In this graphic, the top and bottom of the boxes represent the 25th and 75th percentiles, meaning that one-quarter of the Algebra I or English SOL scores lie below or above those values. The whiskers represent the 5th and the 95th percentiles, meaning that five percent of the Algebra I or English SOL scores are above or below those values.

Figure 4. Diploma and Completer status of members of the high school class of 2007-2008



Source. Data taken from VDOE graduate file.

Figure 5. Developmental course enrollment rates in two-year and private NFP institutions in Virginia, by DC subject



Source. DC data taken from SCHEV enrollment records.

on the reading SOL (aggregating both groups) fell above 563, whereas 25 percent of Algebra I SOL scores were above 502. Among enrolled students, approximately 14 percent of students who took the reading SOL earned a perfect score.

Enrollees also differed from their non-enrolled counterparts by the types of diplomas they earned. The vast majority (71 percent) of enrollees earned an Advanced Studies diploma[5], while 37 percent of non-enrollees earned similar credentials (**Figure 4**). Among non-enrollees, 48 percent earned a Standard diploma, compared to 28 percent of graduates who enrolled in a Virginia IHE in the fall of 2008. In addition, GED and Disabilities diploma earners comprised a much larger proportion of the non-enrolled student group than the enrolled student population. For instance, although combined, students who earned a GED or Disabilities diploma constituted approximately 10 percent of the entire population of spring 2008 high school graduates in Virginia, only one percent of enrolled students earned a GED or Disabilities diploma.

[5] *Approximately 50 percent of Virginia's graduates and non-graduate completers earned Advanced Studies diplomas in 2008.*

2. To what extent were better prepared students less likely to enroll in developmental courses (DC)?

Overall DC enrollment by institution type

Four-year public IHE in Virginia are statutorily prohibited from offering developmental courses. However, 35 percent of students enrolling in two-year colleges in fall 2008 enrolled in at least one English developmental course in the fall or spring of the 2008-2009 school year, 44 percent in a mathematics developmental course, and 57 percent in at least one developmental course in either subject. In private, not-for-profit (NFP) institutions, the percentages enrolling in English and mathematics developmental courses were 4 and 6 percent, respectively (**Figure 5**).

SOL performance and DC enrollment[6]

A major purpose of this study was to provide information to policymakers, program leaders,

[6] *The data from students enrolling in four-year IHE are included in the statistics presented in this and the following sections of this report: one of the ways that better academic preparation reduces the remediation rate in Virginia is by increasing the odds that a student will attend a four-year institution.*

Table 2. Developmental course enrollment rates by Standards of Learning quartiles and developmental course subject area

SOL scaled score quartile	Mathematics developmental course	English developmental course	Any developmental course*
Top quartile	3.8%	1.2%	5.5%
Third quartile	10.0%	4.1%	14.4%
Second quartile	20.9%	12.3%	26.5%
Bottom quartile	38.3%	39.4%	47.5%
Total	**18.6%**	**14.6%**	**23.9%**

Source. Developmental course data taken from SCHEV enrollment records and VDOE graduate file.

**Note: Algebra I SOL quartile used for segmenting quartiles for this category.*

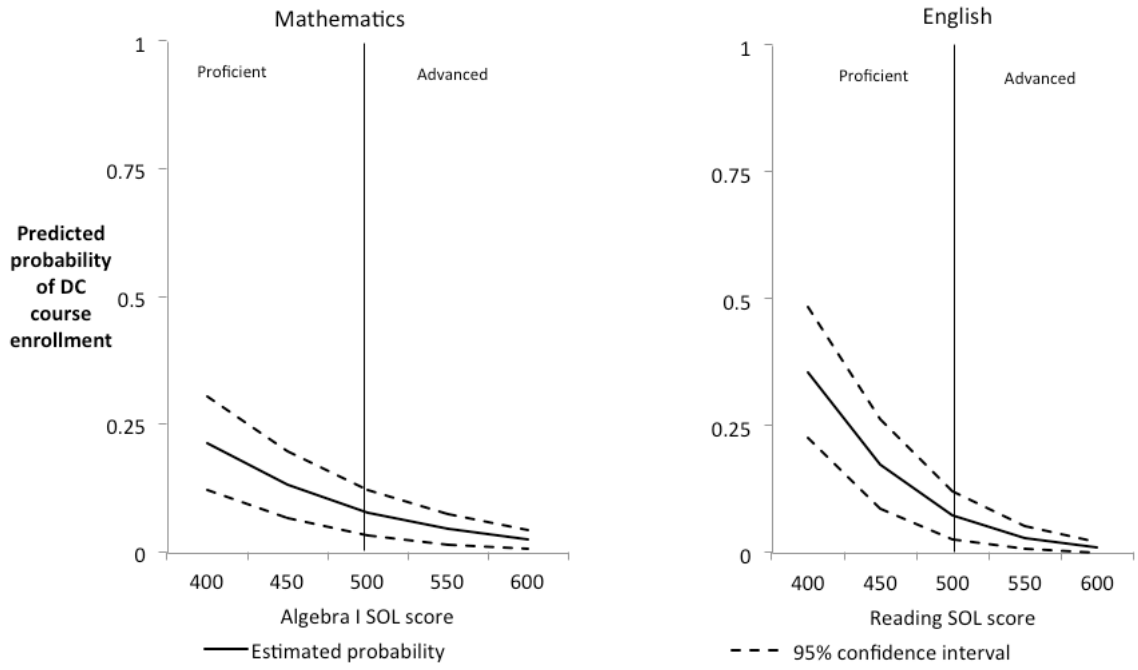
educators, and parents on the levels of academic preparation needed to be ready to enroll directly in credit-bearing courses in college. Among the measures of academic preparation used in this study were SOL test scores in reading and mathematics. **Table 2** shows mathematics and reading development course enrollment rates by students' SOL score quartiles. Quartiles in the Reading SOL test were used to predict English developmental course enrollment, while Algebra I quartiles were used for the mathematics and "any developmental course" categories. For example, over 38 percent of students who scored in the bottom quartile of the Algebra I SOL test in high school enrolled in mathematics developmental courses in college, versus less than four percent of students scoring in the top quartile. The odds that a top-quartile student in reading would be assigned to a reading developmental course were especially low.

To examine the ability of SOL test scores to predict enrollment in credit-bearing (non-developmental) courses independent of student demographics, we used logistic regression statistical models to estimate the probability that a "typical" Virginia student would enroll in a developmental mathematics or English course, varying only the student's test scores (**Figure 6**). The figure depicts those estimated probabilities along with their 95 percent confidence bands. These results show a strong and statistically significant association between SOL test scores in each subject and the probability of enrolling in developmental coursework. For students who scored just at the advanced proficient threshold of 500, the estimated probability of enrolling in developmental coursework was less than 10 percent, and these probabilities declined further for students higher in the advanced proficient range. These patterns were similar across the two subjects.

For most students, the Algebra I SOL was taken earlier in a students' school tenure than was the reading SOL. This delay may have led to the weakening in the association between Algebra I SOL performance and postsecondary outcomes related to mathematics achievement because students may not have retained the information required to achieve proficiency on the Algebra I SOL. To explore this hypothesis, we added students' Algebra II SOL scores to the statistical prediction model.[7] The probability of developmental course enrollment decreased even more sharply for students with higher Algebra II scores, suggesting that Algebra II SOL performance was a more decisive predictor of mathematics developmental course enrollment than Algebra I. In addition, when Algebra I scores were the only mathematics indicator used in the statistical model, the error rate in predicting which students would need remediation was 19 percent, but this rate fell to 14 percent when

[7] *The correlation between Algebra I and Algebra II SOL scores was 0.58. Variance inflation factors (VIFs) did not exceed customary thresholds (10) of excessive collinearity.*

Figure 6. Predicted probability of mathematics and English developmental course enrollment, by respective SOL score



Source. Developmental course data taken from SCHEV enrollment records and VDOE graduate file.

Note. Predicted probabilities derived from logistic regression with standard errors adjusted for clustering at the IHE level. Variables included in the estimation were: ethnicity, LEP status, gender, economically disadvantaged status, high school attendance rate, dual enrollment participation, chemistry and Algebra II SOL participation, Algebra I and Reading SOL scaled scores, diploma type, and AP participation category.

Algebra II scores were also used.[8]

When interpreting the results in **Figure 6**, it is important to keep in mind that relatively few students who enrolled in a Virginia IHE scored substantially below the advanced proficient cut score of 500 on the English reading SOL (see Figure 3). This results in a less accurate estimation of the probability of enrollment in a developmental course for students who scored less than 500 on the reading test. This lower precision is reflected in the wider confidence bands in that score range in Figure 6.

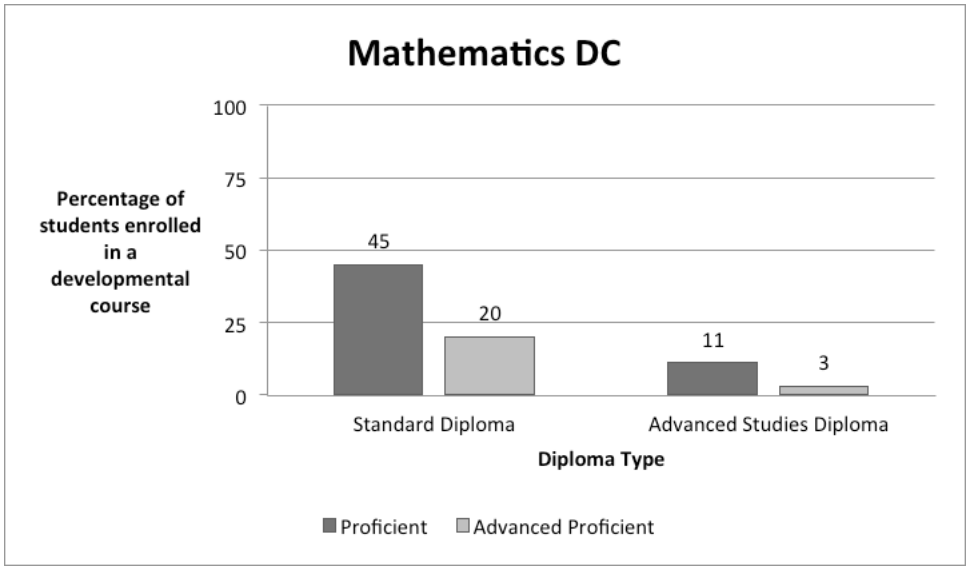
Diploma type and DC enrollment

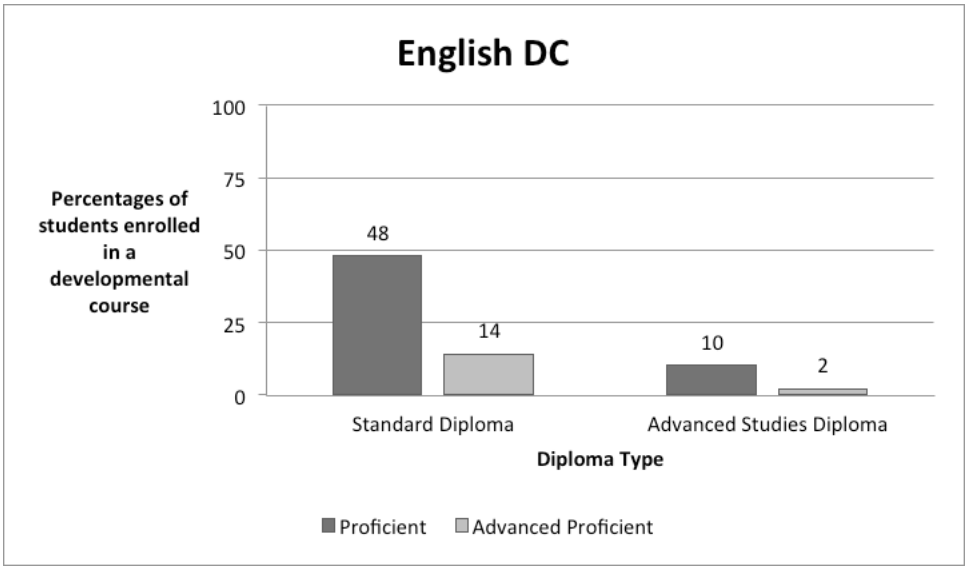
The minimum course or credit requirements necessary to obtain an Advanced Studies Diploma in Virginia are higher than for a Standard Diploma.[9] **Figure 7** disaggregates developmental course enrollment rates not only by whether students scored advanced proficient or proficient on the SOL, but also by whether they earned an Advanced Studies or a Standard Diploma. Thus, we can see that students who only scored proficient on the Algebra I SOL and earned a Standard diploma have a developmental course enrollment rate of 45 percent. This rate declined to 20 percent for students who earned Standard diplomas and scored advanced proficient on the Algebra I SOL, 11 percent for students who were only proficient in Algebra I but earned an Advanced Studies diploma, and 3 percent for students who were both advanced proficient and earned an Advanced Studies diploma. Thus, both diploma types and SOL scores were strong indicators of academic preparation for college.

[8] *Outcomes are classified as positive (here, enrollment in a DC course) if the predicted probability of a positive outcome is greater than 50 percent. The full classification results, including a brief interpretative section, are presented in Appendix E, Table E1.*

[9] *See the requirements for Advanced Studies and Standard diplomas described in Footnote 4 or see http://www.doe.virginia.gov/instruction/graduation/index.shtml*

Figure 7. DC enrollment rates for the fall 2008 semester, by Diploma status, SOL proficiency, and DC subject area





Source. Developmental course data taken from SCHEV enrollment records and Diploma or Completer status and SOL proficiency data taken from VDOE graduate file.

3. To what extent did better high school academic preparation close income and ethnic gaps in student enrollment in developmental courses?

When students are disaggregated by ethnicity, African American and Hispanic students participate in developmental courses at significantly higher rates than do their White or Asian peers. Likewise, economically disadvantaged students are more likely to take these courses than are students who are not economically disadvantaged. For example, in mathematics, the gap in developmental course participation rates is 12 percentage points (28 percent versus 16 percent) for African American compared to White students, and 13 percentage points (30 percent versus 17 percent) for economically disadvantaged compared to non-economically disadvantaged students (**Figure 8**). In English, the gap in DC enrollment rates is widest (18 percentage points) between Hispanic and White students (30 percent versus 12 percent; see **Figure 9**).

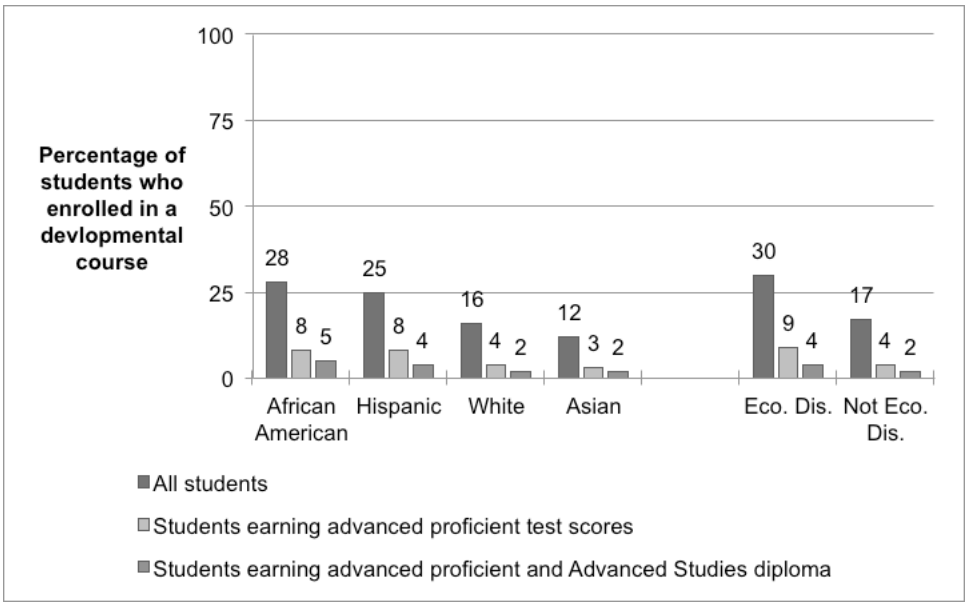
Because the need for remediation in higher education is an indicator of the need for better academic preparation before college, the question is how well the high school indicators in this study (e.g., SOL scores and diploma type) capture the differences in academic preparation that lead to different enrollment rates in developmental courses. One approach is to examine how much income and ethnic gaps in developmental course enrollment decline after SOL scores and diploma type are statistically taken into account. Figure 8, for example, shows that the gap in mathematics DC enrollment between African American and White students declines from 12 percent to 4 percentage points in English for students scoring at the advanced proficient level on the Algebra I SOL, and further to 3 percentage points for students who earned advanced proficient and who earned Advanced Studies diplomas. In addition, the mathematics DC enrollment rate disparity between economically and non-economically disadvantaged students narrowed from 13 percentage points overall to 5 percentage points for those who scored advanced proficient on the Algebra I SOL, to 2 percentage points for students who scored advanced proficient and earned an Advanced Studies Diploma.

Although Hispanic students were 18 percentage points more likely to enroll in an English DC than were White students, this enrollment gap declined to 7 percentage points for students who scored advanced proficient on the English reading SOL and 3 percentage points for students who both scored advanced proficient on the English reading SOL and earned an Advanced Studies diploma. In sum, for Hispanic students who scored advanced proficient on the English SOL and earned an Advanced Studies diploma, the likelihood of enrolling in an English DC was only about one-sixth as high (5 percent versus 30 percent) compared to all Hispanic students. Collectively, Figure 8 and Figure 9 illustrate how ethnic and income gaps in developmental enrollment rates narrow for students with higher SOL scores and increased course preparation as measured by the Advanced Studies diploma.

As a further look at gap closing, we identified the minority ethnic group that had the highest participation in developmental courses (African American students in mathematics courses and Hispanic students in English courses) and disaggregated those students into groups representing those who were economically disadvantaged and those who were not economically disadvantaged. We further disaggregated economically disadvantaged Hispanic students into those who were LEP at some time in high school and those who were not. Then we compared the developmental course enrollment rates of those subgroups with those of non-economically disadvantaged White students (**Figure 10 and Figure 11**).

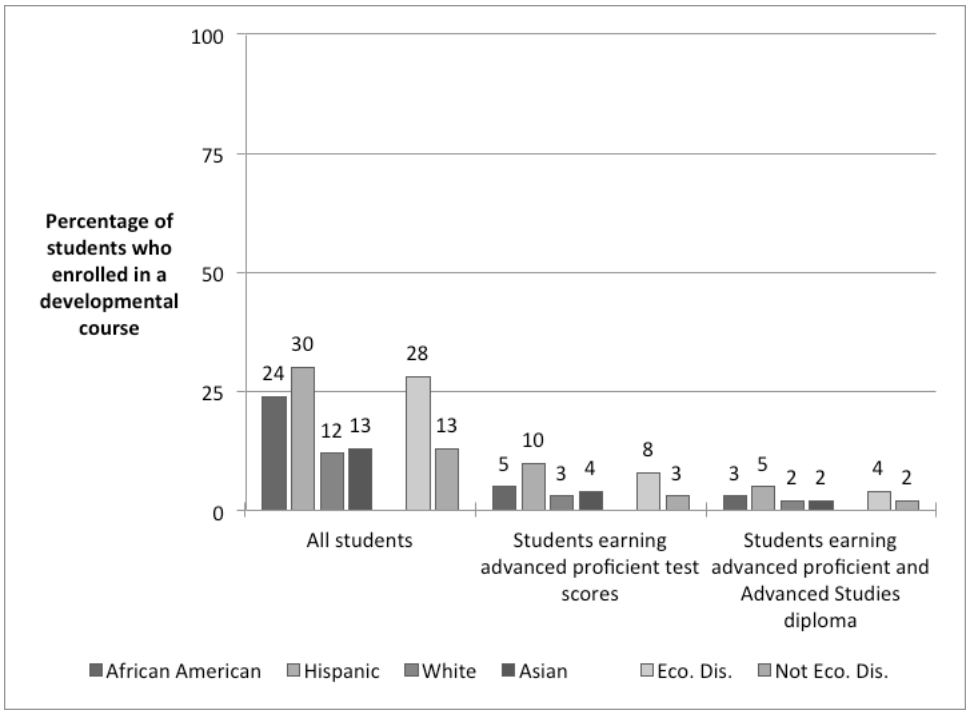
If measures of academic preparation are ignored, economically disadvantaged African Americans are 17 percentage points more likely (32 percent versus 15 percent) to enroll in a mathematics developmental course than are non-economically disadvantaged White students ("all students in group" in Figure 10). Gaps decrease among African American students with higher levels of high school academic achievement. For example, the gap for

Figure 8. Mathematics DC enrollment rates for the fall 2008 cohort, by student ethnicity, economic disadvantage status, Algebra I SOL performance, and diploma type



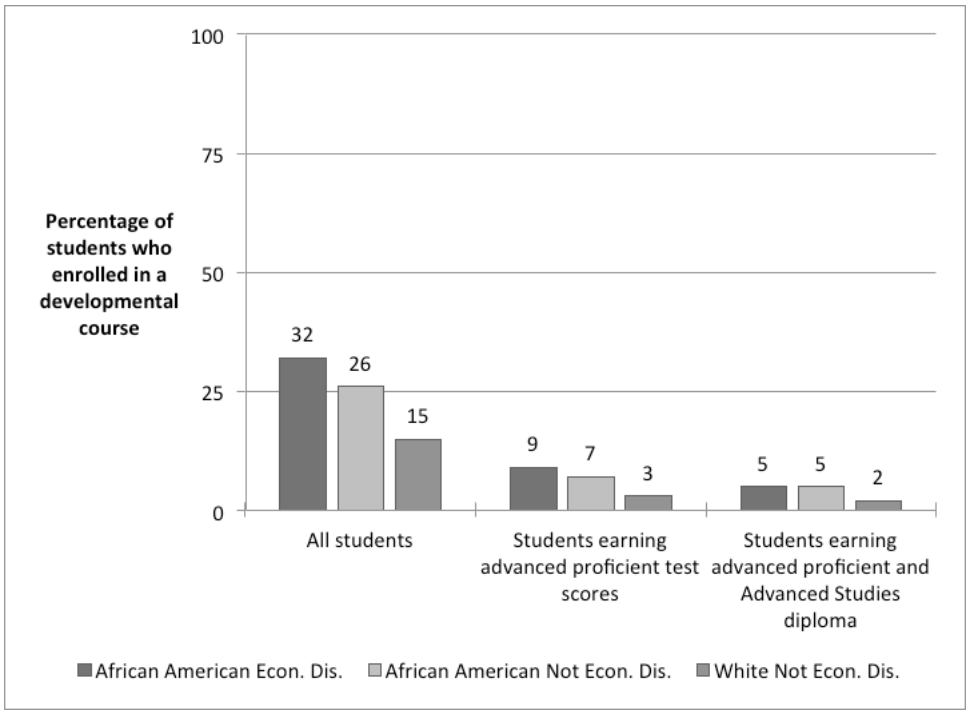
Source. Developmental course data taken from SCHEV enrollment records and linked VDOE graduate file. Algebra I SOL performance, economic disadvantage status, and diploma type from the VDOE graduate file.

Figure 9. English DC enrollment rates for the fall 2008 cohort, by student ethnicity, economic disadvantage status, and diploma type



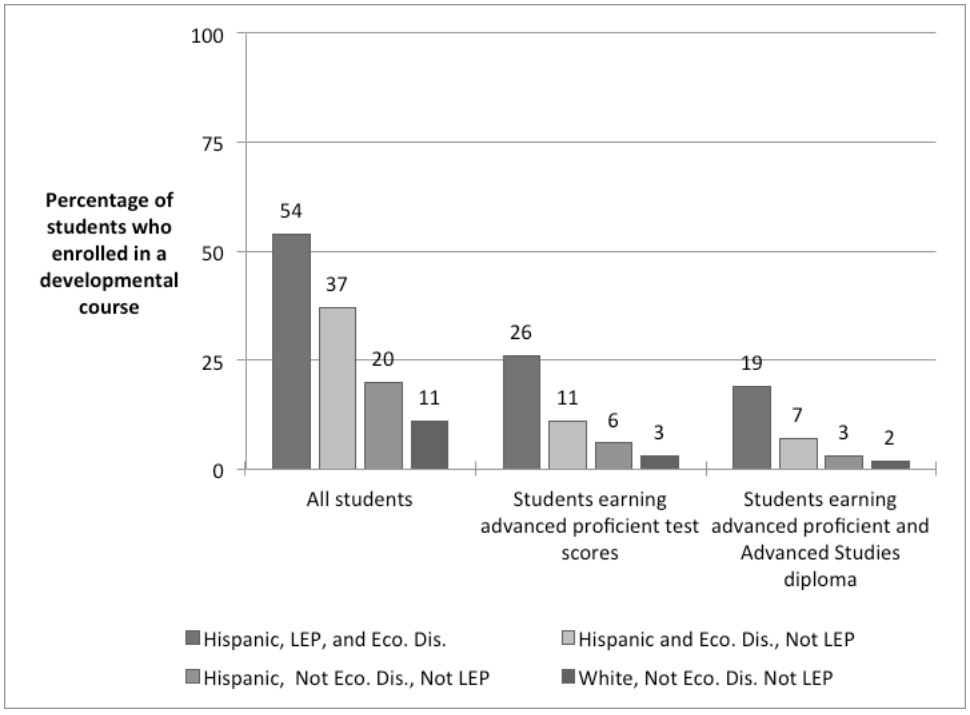
Source. Developmental course data taken from SCHEV enrollment records and linked VDOE graduate file. Reading SOL performance, economic disadvantage status, and diploma type from the VDOE graduate file.

Figure 10. Mathematics DC enrollment rates for African American students compared with non-economically disadvantaged White students, by Algebra I SOL performance and diploma type



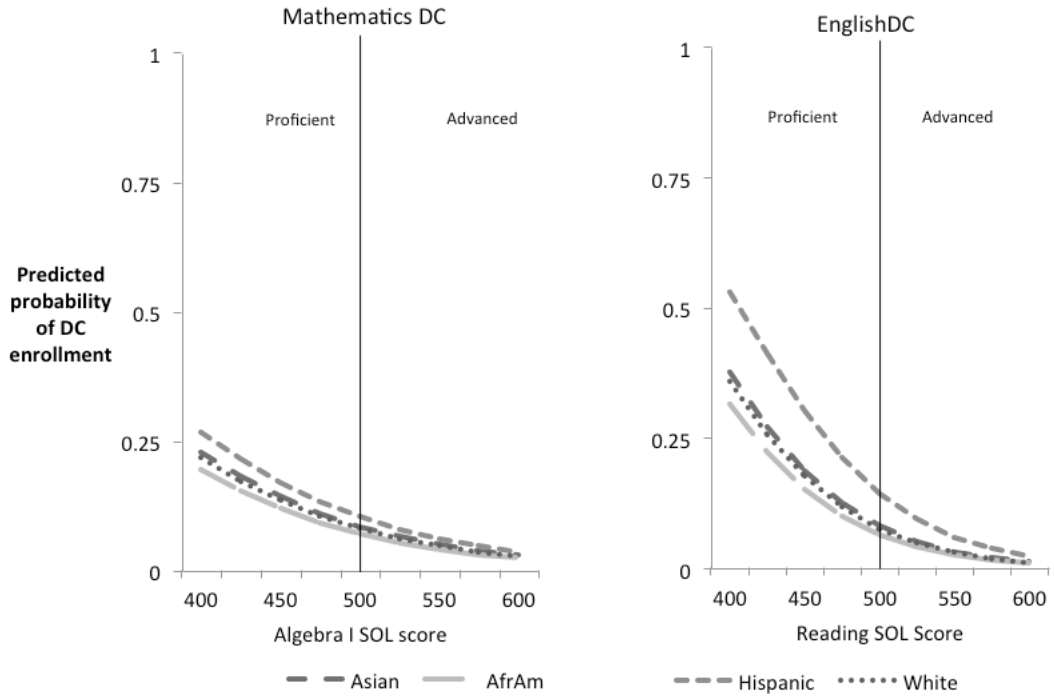
Source. DC data taken from SCHEV enrollment records and linked VDOE graduate file. Algebra I SOL performance, economic disadvantage status, and diploma type from the VDOE graduate file.

Figure 11. English DC enrollment rates for subgroups of Hispanic students compared to non-economically disadvantaged White students, by English SOL performance and diploma type



Source. DC data taken from SCHEV enrollment records and linked VDOE graduate file. Algebra I SOL performance, economic disadvantage status, and diploma type pulled from the VDOE graduate file.

Figure 12. Predicted probability of mathematics and English DC enrollment, by respective SOL score and student ethnicity



Source. SCHEV enrollment records and VDOE graduate file.

Note. Predicted probabilities derived from logistic regression with standard errors adjusted for clustering at the IHE level while holding all covariates at their means. Variables included in the estimation were: ethnicity, LEP status, gender, economically disadvantaged status, high school attendance rate, dual enrollment participation, chemistry and Algebra II SOL participation, Algebra I and English SOL scale scores, diploma type, and AP participation category.

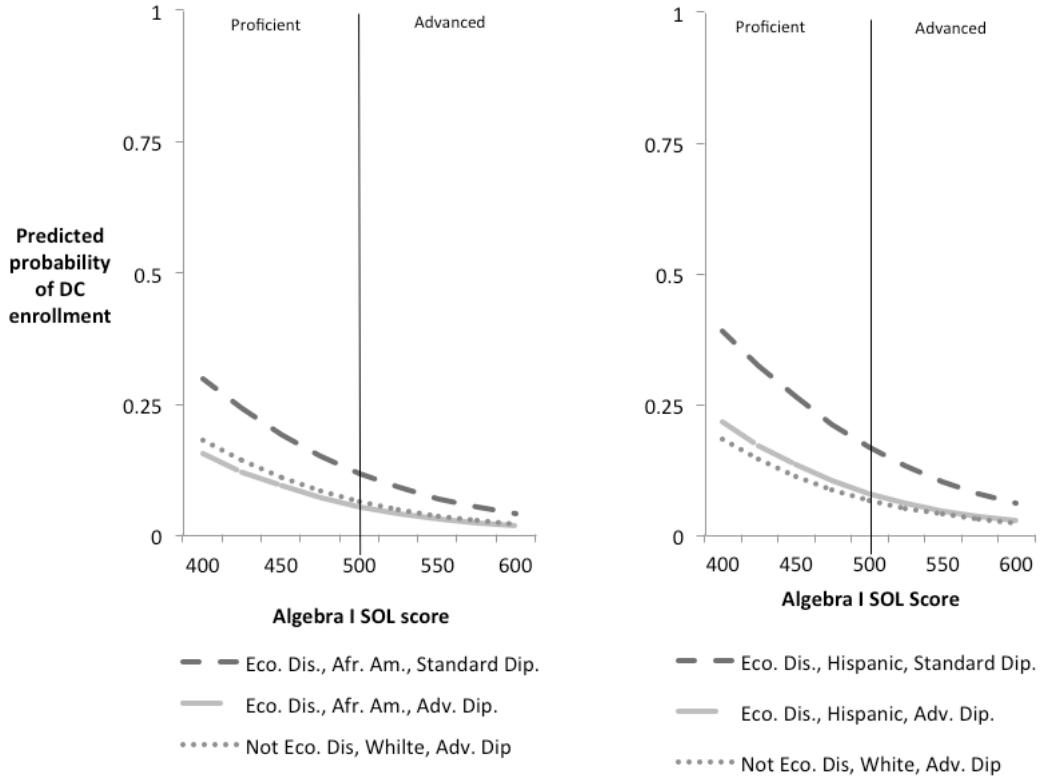
economically disadvantaged African Americans scoring advanced proficient on the Algebra I SOL is 6 percentage points (9 percent versus 3 percent), and 3 percentage points (5 percent versus 2 percent) for those who scored advanced proficient and earned an Advanced Studies diploma.

Likewise, Hispanic students who scored advanced proficient on the reading SOL and earned an Advanced Studies diploma had lower enrollment percentages in English developmental education courses than did Hispanic students in general (Figure 11). However, even among this group, significant gaps remained between economically disadvantaged, LEP Hispanic students (19 percent of whom enrolled in English developmental courses) and their non-economically disadvantaged English proficient White counterparts (2 percent).

It should be noted that LEP designation in high school may be an imperfect measure of students' language proficiency. For example, some students may be identified as LEP in elementary and/or middle school but not high school; yet many of these students may still have limitations in their academic English vocabularies or critical reading and writing skills that are needed for successful entry into credit-bearing courses in college. As well, at the time that this group of students was in high school, determining students' LEP status was based on a combination of English language proficiency assessments and a local body of evidence[10], which had the potential to increase the variability of English language skills for students exiting services in different regions of the state.

[10] *Consistent with the Final Title III Interpretations dated October 2008, Virginia's LEP determinations were made with a consistent measure beginning in the 2008-2009 school year.*

Figure 13. Predicted probability of mathematics developmental course enrollment, by Algebra I SOL score, student ethnicity, diploma type, and economic disadvantage status



Source. DC data taken from SCHEV enrollment records and VDOE graduate file.

Note. Predicted probabilities derived from logistic regression with standard errors adjusted for clustering at the IHE level while holding all covariates at their means, with the exception of LEP status, which was held at not LEP (0). Variables included in the estimation were: ethnicity, LEP status, gender, economically disadvantaged status, high school attendance rate, dual enrollment participation, chemistry and Algebra II SOL participation, Algebra I and English SOL scale scores, diploma type, and AP participation category.

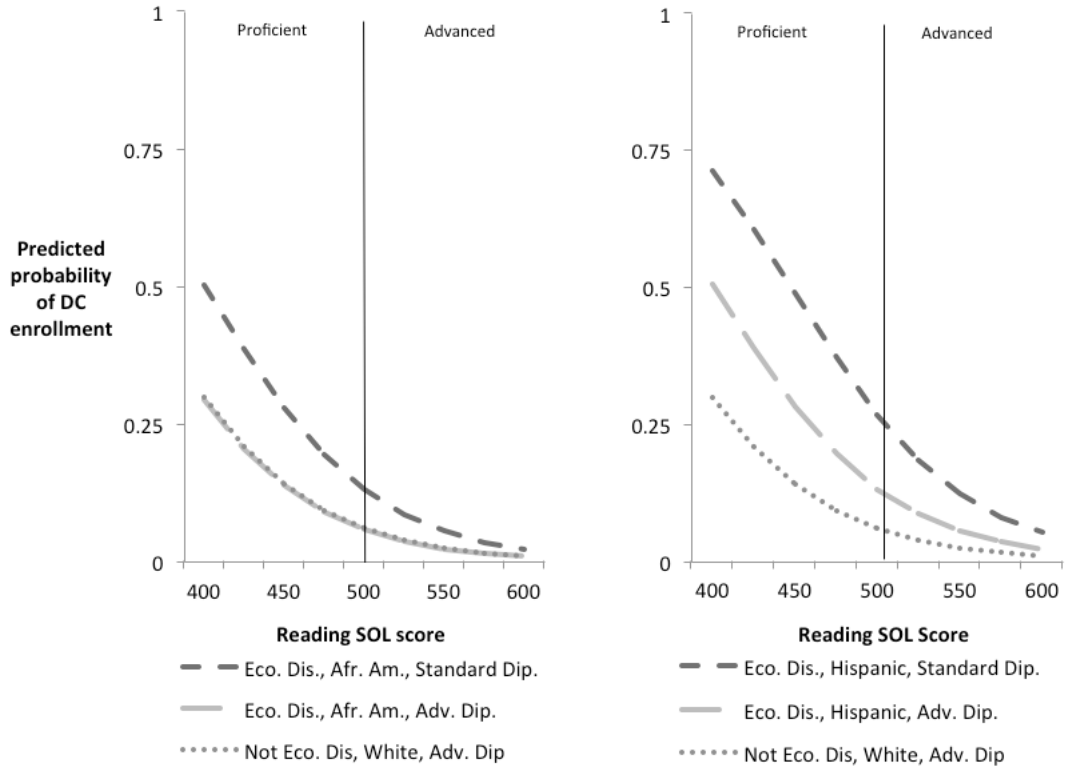
These factors could help account for the 7 percent of non-LEP (in our data) Hispanics who earned high Reading SOL scores and Advanced Studies diplomas but still participated in English developmental courses in college (Figure 11).[11]

To further investigate the closing of income and ethnic DC enrollment gaps, we used logistic regression statistical models to estimate the incremental association between each student characteristic and the probability that the student would enroll in developmental education in a given subject. (Two models were used: one to predict enrollment in English and one in mathematics developmental courses—see Appendix C.) The note to **Figure 12** contains a list of the student demographic and academic preparation variables in these models.

Figure 12 shows the association between SOL scores and DC enrollment rates for students from each ethnic group if all of the ethnic groups were assigned the state average value of the other student characteristics included in the model. The downward slope of the curves shows the association between higher SOL scores and lower enrollment in developmental courses when other student characteristics are statistically held constant. The narrowing of the gaps among the curves shows the association of higher SOL scores to closing gaps in developmental course taking, setting aside the effects of other student characteristics.

[11] *Perry, Bahr, Rosin, & Woodward (2010), for instance, found that Hispanic students were heavily overrepresented in reading developmental courses in California community colleges, but they were unable to determine whether this overrepresentation was a function of LEP status.*

Figure 14. Predicted probability of English DC enrollment, by reading SOL score, student ethnicity, diploma type, and economic disadvantage status



Source. DC data taken from SCHEV enrollment records and VDOE graduate file.

Note. Predicted probabilities derived from logistic regression with standard errors adjusted for clustering at the IHE level while holding all covariates at their means, with the exception of LEP status, which was held at Not LEP (0). Variables included in the estimation were: ethnicity, LEP status, gender, economically disadvantaged status, high school attendance rate, dual enrollment participation, chemistry and Algebra II SOL participation, Algebra I and English SOL scaled scores, diploma type, and AP participation category.

This suggests that racial/ethnic differences in DC enrollment can be almost entirely accounted for by differences in academic preparation and other student characteristics listed in Figure 12.

One pattern in Figure 12 is that Hispanic students had a slightly higher estimated probability of enrolling in an English DC even after controlling for their SOL scores and other measured characteristics. Hispanic students with state-average values of other student characteristics were 17 percentage points more likely than White students to enroll in an English DC when both had an English SOL score of 400, 6 percentage points more likely at 500, and 3 percentage points more likely at 550.

Figure 13 and **Figure 14** illustrate how both higher SOL scores and receiving an Advanced Studies diploma narrowed the gap in DC enrollment between economically disadvantaged minority students and non-economically disadvantaged White students who also received an Advanced Studies diploma. As in Figure 12, other student characteristics listed at the bottom of the table were held constant to isolate the effect of changes in SOL scores and diploma type. For Hispanics, the group simulated was economically disadvantaged but non-LEP students. With the exception of Hispanics enrolling in English developmental courses, nearly all of the combined income and ethnic gap is erased when students earn comparable academic credentials that equal their non-minority and non-economically disadvantaged

peers.[12] Economically disadvantaged but non-LEP Hispanic students who earned an Advanced Studies Diploma and scored at the cut point for advanced proficient (500) on the English SOL were more likely to enroll in a non-credit bearing English course than were non-economically disadvantaged White students with similar academic credentials (13 percent compared to 6 percent, respectively) (**Figure 14**). We recognize that Virginia's public educators have significant work to do to ensure that economically disadvantaged and minority students participate in the same rigorous courses and reach the same high achievement levels as their non-economically disadvantaged and white peers (see for example Jonas, 2011). However, these data show quite clearly that when the achievement gaps in high school are eliminated, all students have near equal likelihood of successful enrollment in credit-bearing courses in college.

4. To what extent were better prepared students more likely to stay into their second year in the Virginia higher education institution in which they enrolled in the first year?

Second year persistence rates across Virginia postsecondary institutions

Research in Virginia shows that persistence into the second year of college is a strong predictor of postsecondary graduation (SCHEV, 2009). VDOE was interested in learning whether academic preparation indicators in high school were associated not only with avoiding the need for developmental courses in the first year, but also with persisting into the second year of postsecondary education. Previous research in Virginia showed no association between high school achievement indicators and persistence in four-year schools nationwide (VDOE, 2010). This prior research used similar high school data linked to students' postsecondary enrollment in four-year IHE nationwide, acquired from the National Student Clearinghouse. The research showed that 87 percent of Virginia's high school students who enrolled in a four-year institution immediately after high school persisted into their second year of college. In the study, persistence was calculated based on students returning to any four-year college in the country in the second year, although nearly 94 percent of those who persisted were at the same institution in their second year.

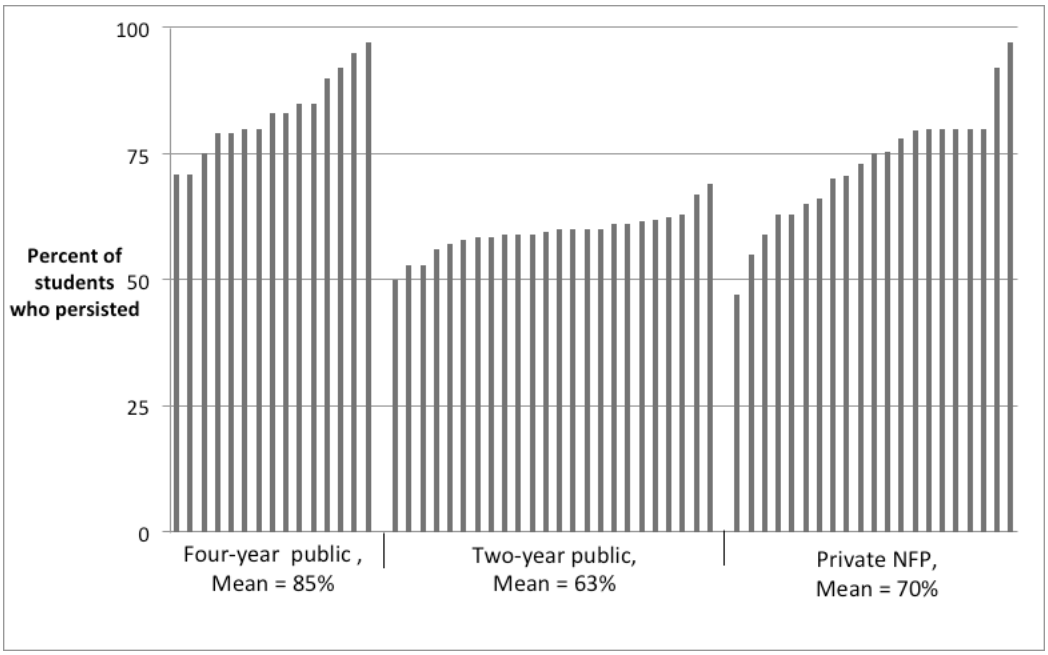
The data set available for the current study represents the first statewide dataset in Virginia that permitted an investigation of the associations between high school achievement indicators and persistence in both two- and four-year institutions. In this study, the overall percentage of students who persisted into their second year of college was 75 percent for students who graduated or completed high school in school year 2007-2008 and enrolled in a Virginia IHE in the fall of 2008. However, results of this study showed that student persistence into the second year varied by type of institution: four-year public universities averaged 85 percent persistence, two-year public institutions 63 percent and private NFP 71 percent.[13] In addition, persistence rates of students varied considerably by individual institution with each type (**Figure 15**). The analysis that follows examines the association between SOL scores, diploma types, and other high school variables on second-year persistence, but does not attempt to explain differences in persistence rates across individual institutions. It is important also to note that students who transferred between institutions after their first year were recorded in the data set as not persisting, as were students who left after completing one-year certificate programs in Virginia's Community Colleges. Thus, cross-institution transfers and students completing certificate programs could not be distinguished from college dropouts in the data used in this study. This could affect the comparison between two- and four-year institutions in **Figure 15**;

[12] *In this context, comparable academic credentials are represented by participation in Algebra II and chemistry and earning Virginia's Advanced Studies Diploma, which has increased course-taking requirements relative to Virginia's Standard and other diplomas.*

[13] *In the remaining figures, private NFP schools are combined with 4-year IHE for ease of presentation. The statistical models used an indicator variable for whether a four-year college was private or public in order to account for possible differences across these types of institutions.*

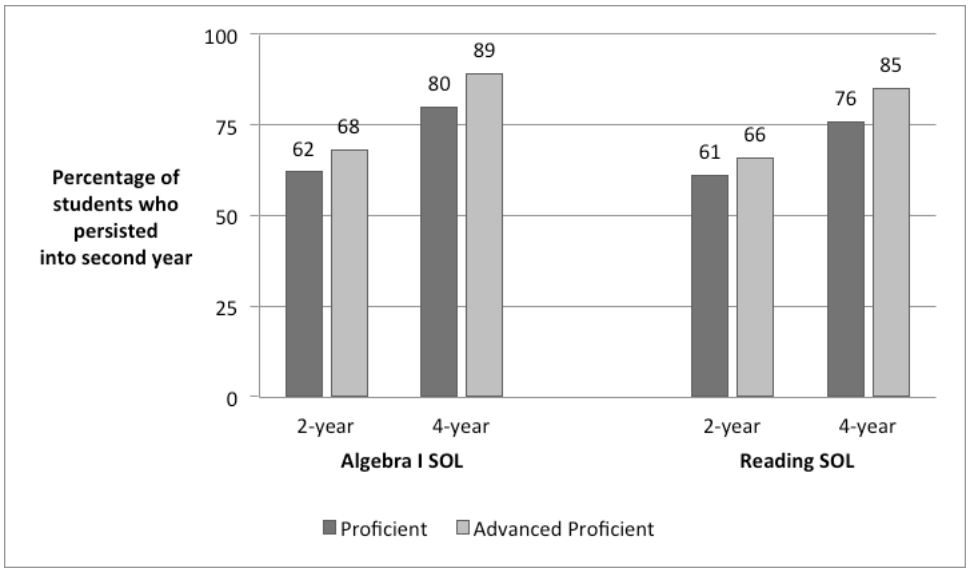
students in two-year schools are more likely to transfer from two-year institutions into four-year institutions or complete one-year certificate programs and therefore would not be expected to return to the same school for a second year. The information provided in this section of the report, therefore, must be validated with a more complete data set before it is conclusive.

Figure 15. Second year persistence rates in Virginia IHE for the fall 2008 cohort, by institution type



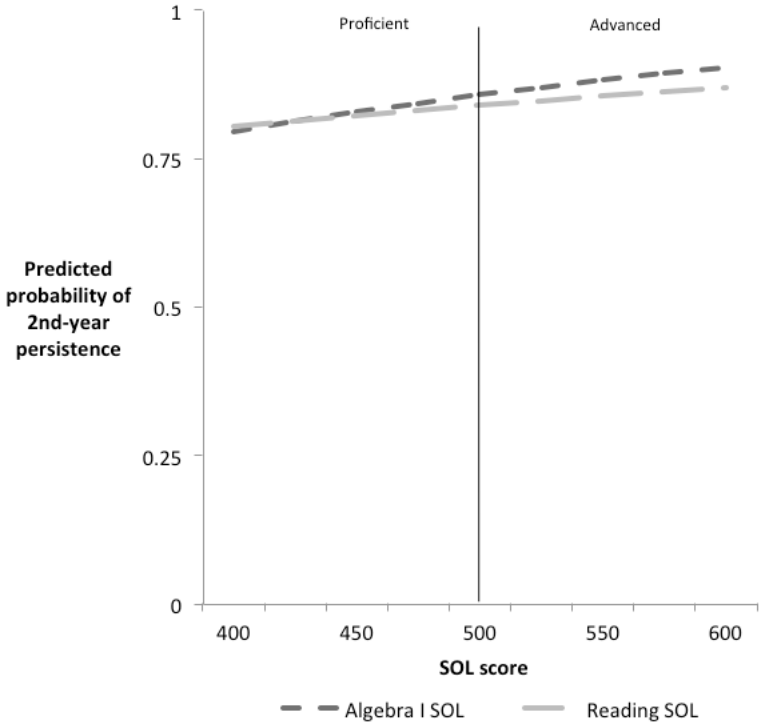
Source. SCHEV enrollment file and VDOE graduate records.

Figure 16. Second year persistence rates by SOL proficiency and IHE type



Source. Second year persistence data taken from the linked SCHEV enrollment file and the VDOE graduate records. Algebra I and Reading SOL performance were taken from the VDOE graduate records.

Figure 17. Predicted probabilities of second-year persistence in a four-year IHE, by Reading and Algebra SOL scores



Source. SCHEV enrollment records and VDOE graduate file.

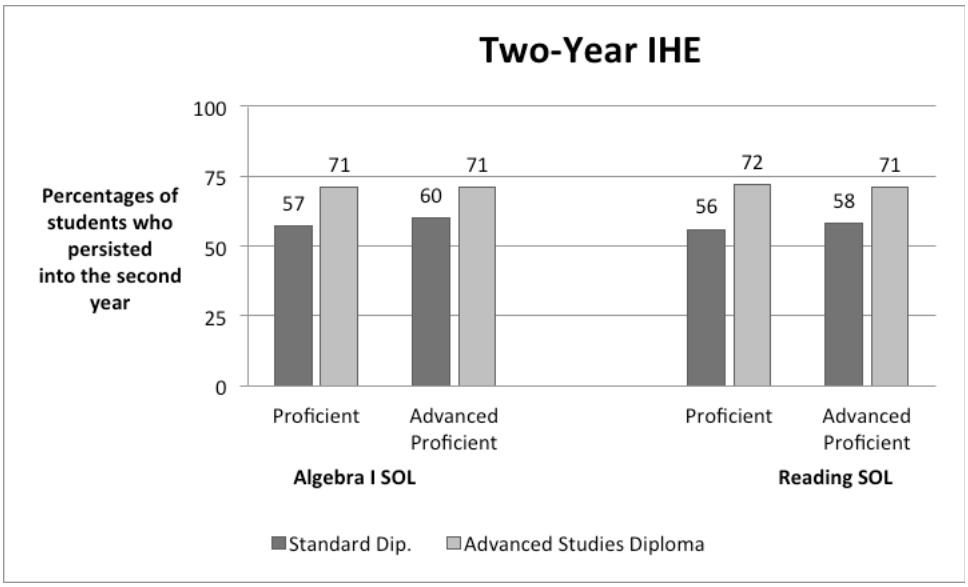
Note. Predicted probabilities derived from logistic regression with standard errors adjusted for clustering at the IHE level. Variables included in the estimation were: ethnicity, LEP status, gender, economically disadvantaged status, high school attendance rate, dual enrollment participation, diploma type, chemistry and Algebra II SOL participation, Algebra I and English SOL scale scores, AP participation category, and private-non-profit status.

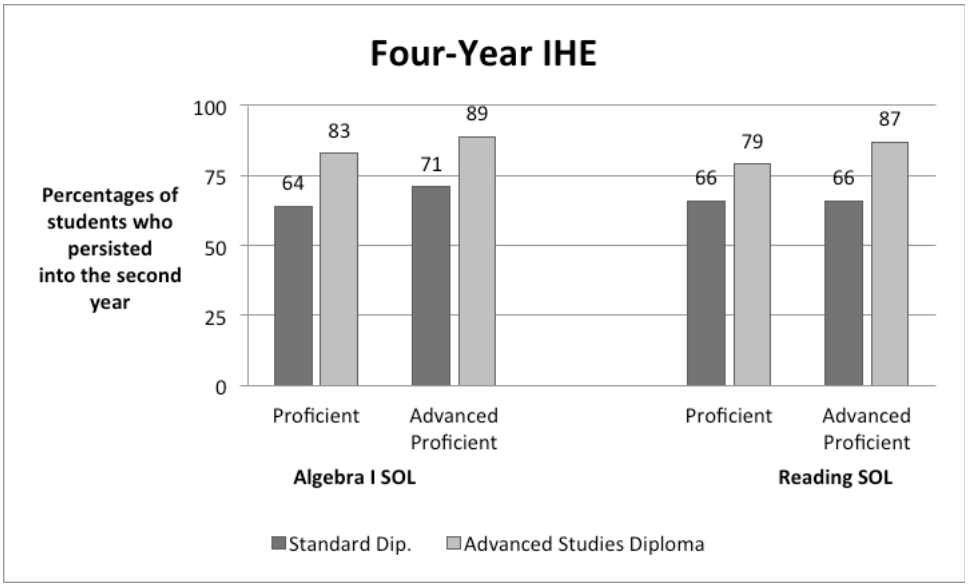
SOL performance and second year persistence

One research objective pertained to how performance on Virginia SOL tests was associated with second year persistence in two- and four-year IHE in Virginia. **Figure 16** shows differences in persistence rates between students earning proficient and advanced proficient scores on state assessments in two- and four-year institutions. Higher SOL scores were associated with increased persistence at both two- and four-year institutions, but the differences were greater in four-year institutions. For example, students earning advanced proficient were 9 percentage points more likely than students earning proficient scores to persist in a four-year institution (89 percent versus 80 percent based on the Algebra I SOL, and 85 percent versus 76 percent based on the Reading SOL). In two-year institutions, the corresponding differences in persistence rates were 6 and 5 percentage points respectively. Students with advanced proficient scores on Algebra I had higher persistence rates than students earning advanced proficient on the Reading SOL. It is not clear whether this is a function of greater predictive power of mathematics achievement or a function of test characteristics. For example, the difference could be a result of different test difficulty between the Reading and Algebra I tests. If the Reading SOL is an easier test relative to the Algebra I test, then students scoring advanced proficient on reading would include students who were less well prepared relative to the students scoring advanced proficient on the Algebra I test.

To identify how much of the differences in persistence rates were associated strictly with differences in SOL scores versus differences in other student characteristics, we created

Figure 18. Second-year persistence rates, by Diploma status, SOL proficiency, and IHE type





Source. SCHEV enrollment file and VDOE graduate records.

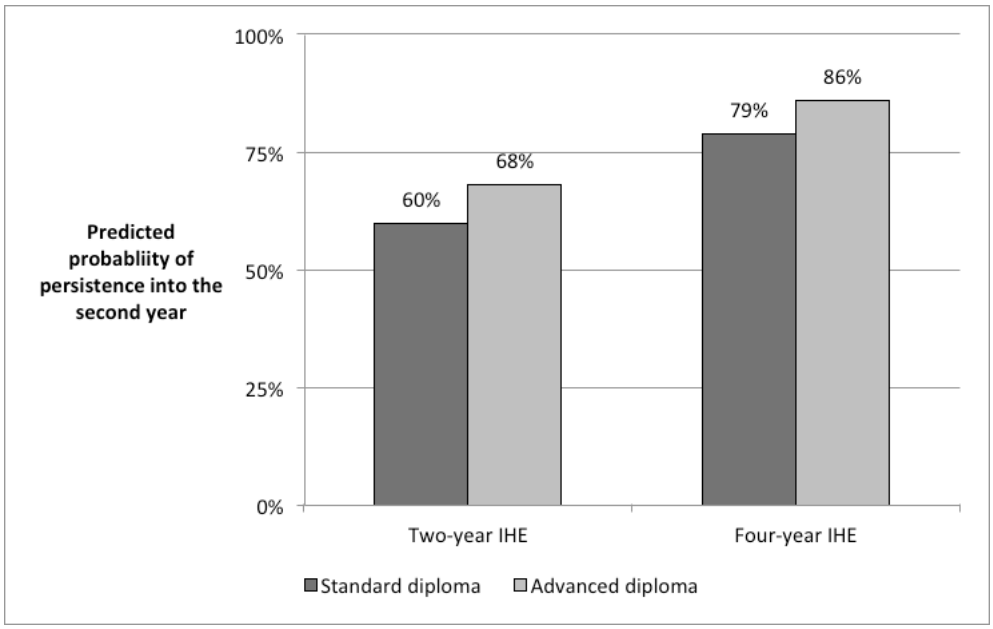
a statistical model to isolate the association between each student characteristic and the probability that a student would persist into the second year. Because influences on persistence may have different effects in two- and four-year institutions, we built separate models for each type of institution.[14]

At two-year IHE, neither the Algebra I SOL nor English SOL scores were statistically significant predictors of persistence. This means that the differences in persistence rates shown in Figure 16 between students earning proficient and advanced proficient scores in two-year institutions were likely attributable not to students' higher SOL scores, but to the other ways in which the two groups of students differed. For example, students earning advanced proficient scores were more likely to earn an Advanced Studies diploma than were students merely scoring proficient.

By contrast, the analysis for four-year institutions did show statistically significant

[14] *Private and public four-year institutions were analyzed in the same statistical model.*

Figure 19. Predicted probability of persistence into second year for Advanced Studies and Standard diploma recipients, by IHE type



Source. SCHEV enrollment records and VDOE graduate file.

Note. Estimated marginal effects derived from logistic regression with standard errors adjusted for clustering at the IHE level. Variables included in the estimation were: ethnicity, LEP status, gender, economically disadvantaged status, high school attendance rate, dual enrollment participation, chemistry and Algebra II SOL indicators, Algebra I and English SOL scale scores, AP participation category, diploma type, and private-non-profit status.

associations between persistence and SOL scores when other student characteristics were held constant (**Figure 17**). For example, an otherwise typical individual scoring 400 on the Reading SOL had close to an 80 percent chance of returning in the second year, whereas typical students scoring 500 had an 84 percent chance of returning (a 4 percentage point increase). Having higher Algebra I scores gave students a slightly greater incremental advantage compared to having higher reading scores, as shown by the slightly steeper line for Algebra I in the figure.[15]

Diploma type and second year persistence

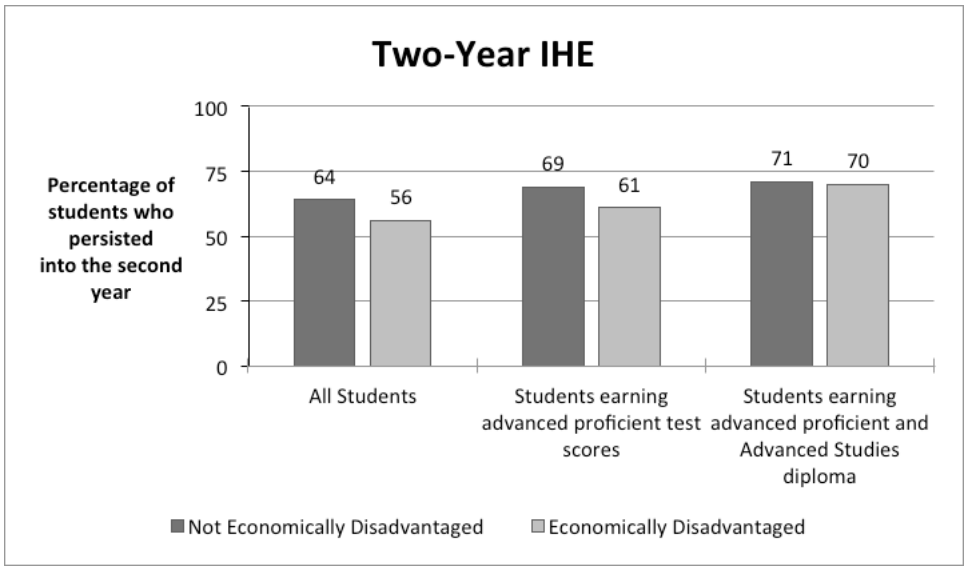
Figure 18 shows disaggregated second-year persistence rates in two- and four-year institutions by two achievement indicators: proficiency level on the SOL and diploma type (Standard or Advanced Studies).[16] As shown in Figure 18, there was a strong association for students enrolled in both two- and four-year institutions between receipt of an Advanced Studies diploma and students' likelihood of persisting into the second-year of college. Persistence rates were 11 to 21 percentage points higher for students earning Advanced Studies diplomas compared to students earning Standard diplomas, with significant variability in the size of the advantage for Advanced Studies diploma earners associated with IHE type and proficiency levels on SOL tests.

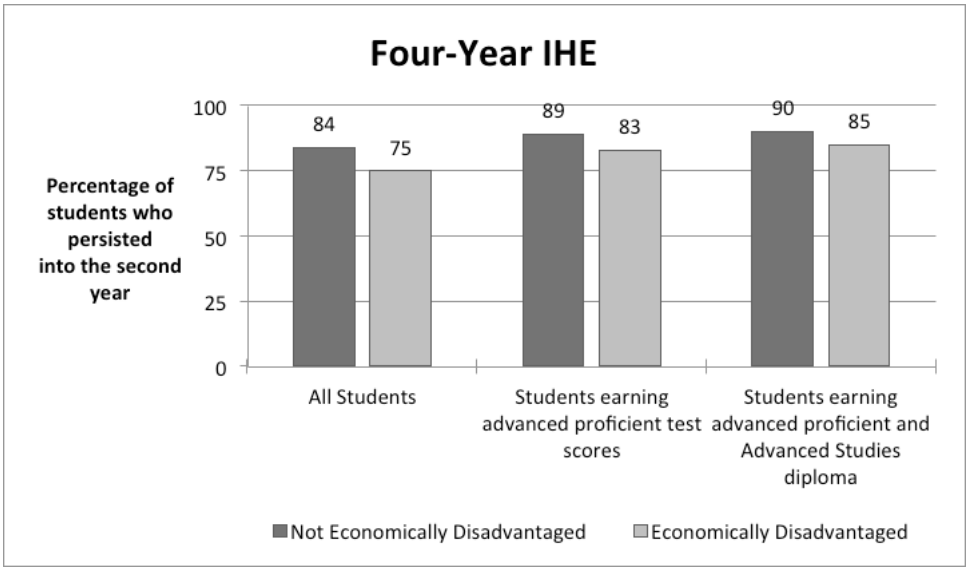
In two-year institutions, while diploma type was clearly associated with students' persistence, proficiency levels on SOL tests were not predictive of persistence once diploma type is taken into account. This is consistent with our earlier finding that SOL scores were not effective predictors of persistence in two-year institutions.

[15] *This slight difference was statistically significant (unlikely to be due to pure chance), as confirmed by a Wald test (Chi-square statistic=6.92). However, the slight difference in the incremental contribution of higher Algebra I scores versus higher Reading scores may not be of much practical significance.*

[16] *Private-non-profit schools are included in the four-year IHE category.*

Figure 20. Second year persistence rates in two- and four-year higher education institutions, by Algebra I SOL achievement level, economic disadvantage status, and diploma type



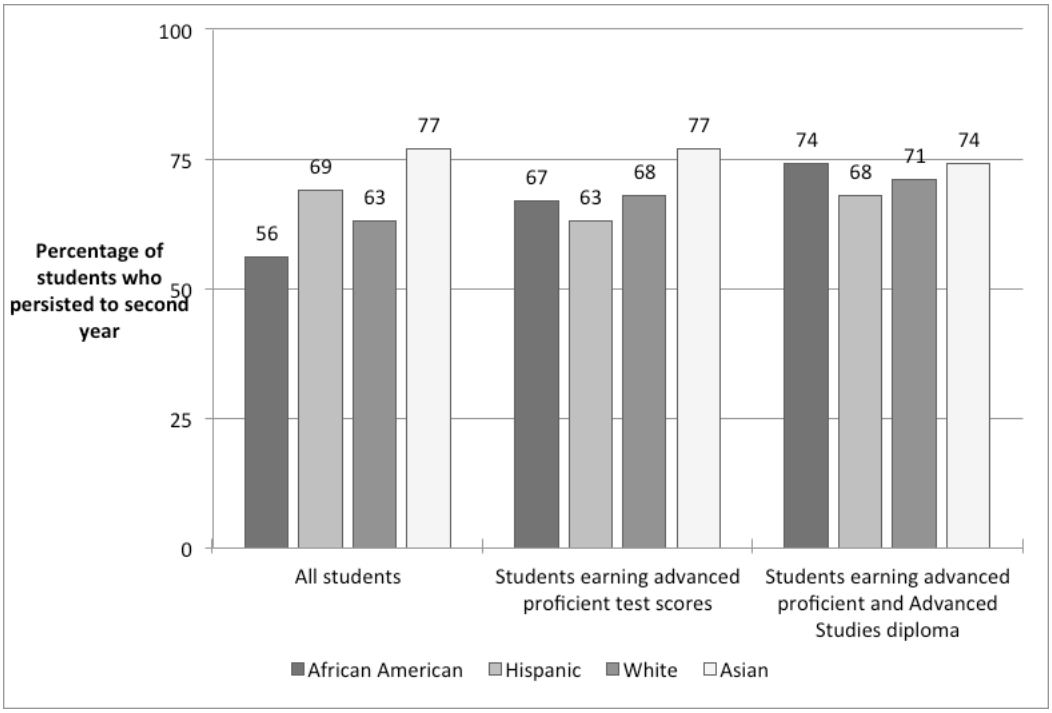


Source. Second year persistence data taken from the linked SCHEV enrollment file and the VDOE graduate records. Algebra I SOL score, economic disadvantage status, and diploma type from VDOE graduate records.

For example, among Advanced Studies diploma recipients, there were no differences in second year persistence rates (71 percent compared to 71 percent) between students who scored advanced proficient or proficient on the Algebra I SOL test. Among students enrolled in four-year institutions, diploma type and SOL scores were predictive of persistence rates. For example, among Advanced Studies diploma recipients, students who scored advanced proficient on the Reading SOL persisted at an eight percentage point higher rate than students who scored proficient (87 percent versus 79 percent).

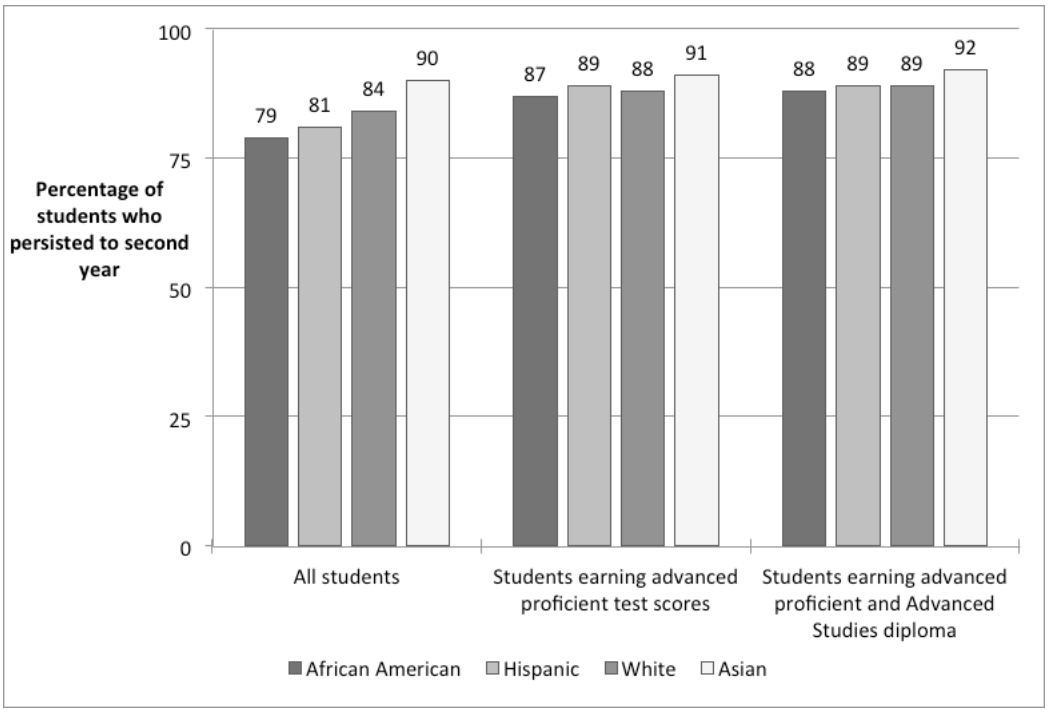
We also used the statistical models described in the previous section to examine the increase in persistence associated with receipt of an Advanced Studies diploma in a hypothetical case in which the Advanced Studies and Standard diploma recipients otherwise had the same demographics and prior SOL scores. **Figure 19** shows the results of this analysis. (The full list of student characteristics controlled for is shown in the note at the bottom of the figure.) In

Figure 21. Second year persistence rates at two-year IHE by Algebra I SOL achievement level, student ethnicity, and diploma type



Source. SCHEV enrollment records and VDOE graduate file.

Figure 22. Second year persistence rates at four-year IHE, by Algebra I SOL achievement level, student ethnicity, and diploma type



Source. SCHEV enrollment records and VDOE graduate file.

two-year institutions, Advanced Studies diploma recipients with state-average characteristics were 8 percentage points more likely (68 percent versus 60 percent) to persist into their second year of college than were their Standard diploma counterparts; for four-year institutions, the difference was 7 percentage points (86 percent versus 79 percent).

5. To what extent did better high school preparation close income and ethnic gaps in second year persistence?

When students were disaggregated by income, economically disadvantaged students persisted at lower rates than non-economically disadvantaged students (56 percent vs. 64 percent in two-year institutions, and 75 percent versus 84 percent in four-year institutions – see **Figure 20**). Likewise, African American and Hispanic students persisted at slightly lower rates than White students in four-year institutions (**Figure 22**), although Hispanics persisted at higher rates than Whites in two-year institutions (**Figure 21**).[17] Nora, Barlow, and Crisp (2005), in a review of literature on factors associated with second year persistence at IHE in addition to original research conducted by the authors, found considerable variation in persistence across student ethnic groups. Surprisingly, however, they found that second year persistence rates were lowest for White students compared to Hispanic and African American students.

Taking scores on the Algebra I SOL into account, students who scored advanced proficient had higher persistence rates in most cases than did the general student population (middle sets of bars in Figures 20-22, compared with the "all students" bars on the left). In four-year institutions, income and ethnic persistence gaps also narrowed for students scoring at the advanced proficient level. For example, the gap between advantaged and disadvantaged students who scored advanced proficient was 6 percentage points (89 percent versus 83 percent) compared with 9 percentage points (84 percent versus 75 percent) for all non-disadvantaged students compared to all disadvantaged students (Figure 20). A similar reduction in ethnic gaps can be seen in Figure 22.

However, scoring advanced proficient alone did not reduce the gap in persistence between advantaged and disadvantaged students in two-year institutions (64 percent versus 56 percent for all students, and 69 percent versus 61 percent for advanced proficient students in Figure 20). The measured persistence rate in two-year institutions was actually lower (63 percent versus 69 percent) for Hispanic students scoring advanced proficient than for all Hispanic students (Figure 21). It is possible that this difference is a result of better prepared Hispanic students transferring to four-year institutions, data not available for this study.

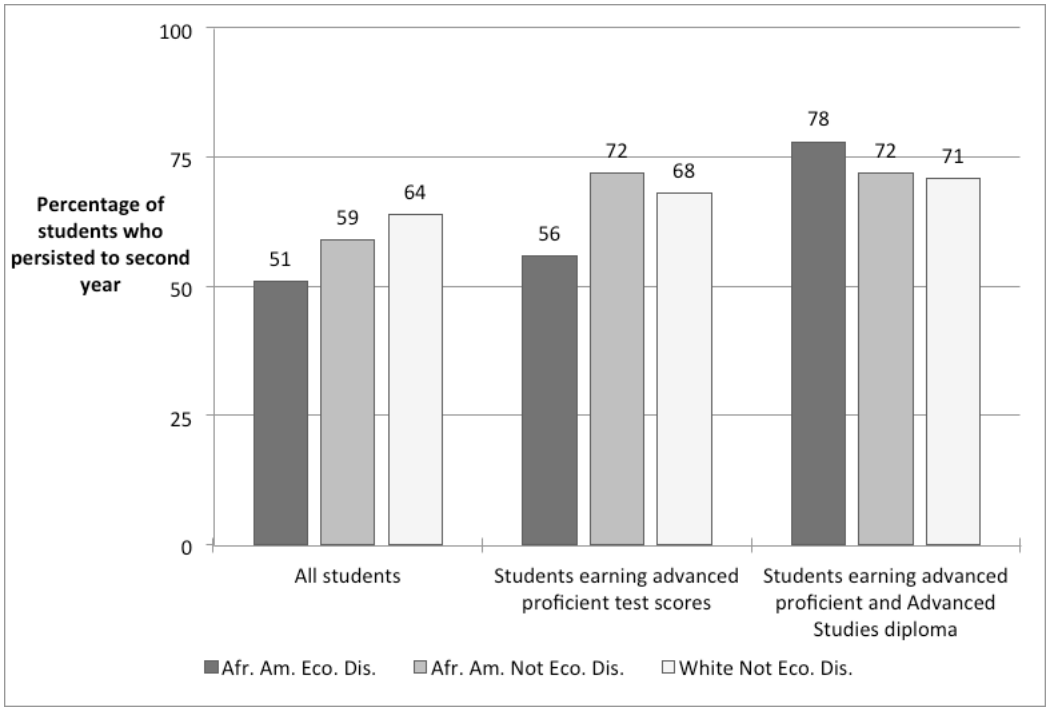
Adding an Advanced Studies diploma for students who also scored advanced proficient in Algebra I (right set of bars in Figures 20-22, compared with the middle set of bars in the three figures) boosted the persistence rates slightly in four-year institutions, and substantially in two-year institutions for certain groups (economically disadvantaged students in Figure 20 and African American students in Figure 21). However, Asian students in this group had lower persistence rates in two-year institutions, possibly a result of increased transfer rates to four-year institutions.

These results indicate that when economically disadvantaged students reach similar rates of academic achievement as their non-economically disadvantaged peers, as measured by proficiency in Algebra I and receipt of an Advanced Studies diploma, the gap in persistence is eliminated in two-year institutions and reduced substantially in four-year institutions.

As a further look at the association between SOL proficiency and income and ethnic gaps, **Figures 23 and 24** show the comparison of persistence rates of economically disadvantaged and non-economically disadvantaged African American students with non-economically

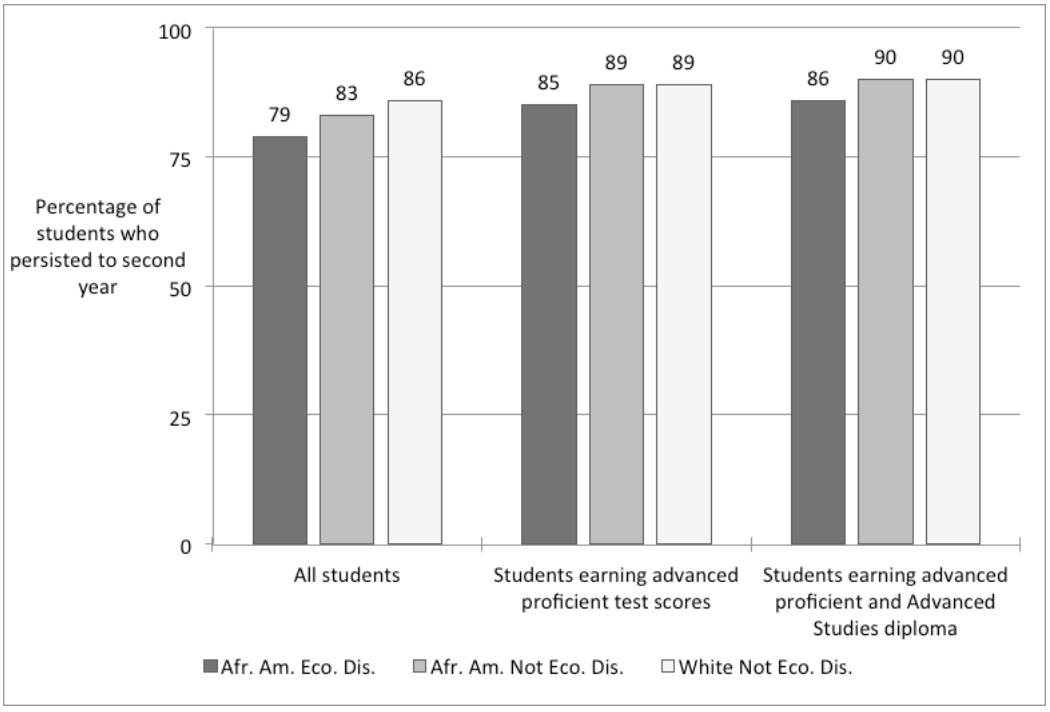
[17] *The lower recorded persistence rate of White students may be due to those students having a higher rate of transfer to four-year institutions.*

Figure 23. Second year persistence rates at two-year IHE for subgroups of African American students compared with non-economically disadvantaged White students, by Algebra I SOL performance and diploma type



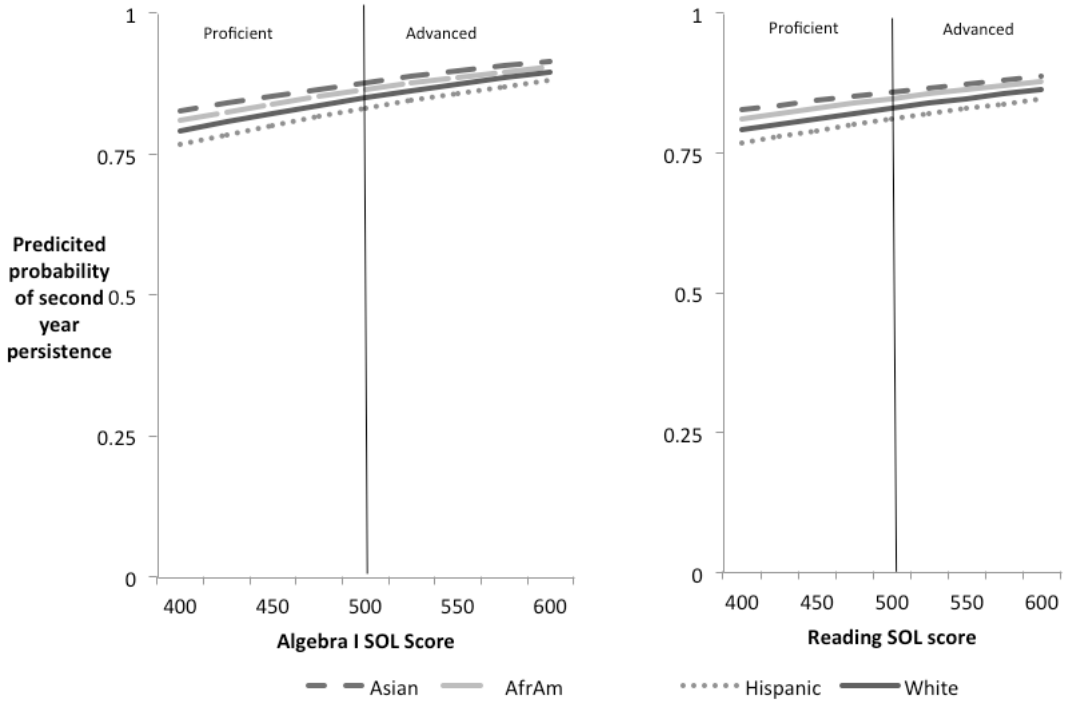
Source. SCHEV enrollment records and VDOE graduate file.

Figure 24. Second year persistence rates at four-year IHE for subgroups of African American students compared with non-economically disadvantaged White students, by Algebra I SOL performance and diploma type



Source. SCHEV enrollment records and VDOE graduate file.

Figure 25. Predicted probabilities of second year persistence in a four-year IHE, by English and Algebra SOL scores



Source. Second year persistence data taken from SCHEV enrollment records and SOL scores taken from VDOE graduate file.

Note. Predicted probabilities derived from logistic regression with standard errors adjusted for clustering at the IHE level. Variables included in the estimation were: ethnicity, LEP status, gender, economically disadvantaged status, high school attendance rate, dual enrollment participation, chemistry and Algebra II SOL participation, Algebra I and English SOL scale scores, AP participation category, diploma type, and private-non-profit status.

disadvantaged White students in two- and four-year institutions. Figure 24 shows that in four-year institutions, there was a 7 percentage point disparity between economically disadvantaged African American and non-economically disadvantaged White students (79 percent versus 86 percent) when achievement variables were not included. This gap declined to 4 percentage points (86 percent versus 90 percent) for students who scored advanced proficient on the Algebra I SOL and earned an Advanced Studies diploma. Persistence gaps in two-year institutions also narrowed among the groups when students performed well academically, with economically disadvantaged but well-prepared African American students more likely to stay in the same two-year institution than are their non-economically disadvantaged White counterparts (Figure 23).

Figure 25 presents statistically modeled persistence rates across ethnic groups across a range of Algebra I and Reading SOL scores at four-year institutions, including both public and private universities.[18] When student achievement on the Algebra I or Reading SOL is equivalent, and other variables such as diploma type are held constant,[19] there are minimal substantive differences in persistence rates between racial/ethnic subgroups. For instance, on the Algebra I

[18] *Results at two-year IHE are not presented, because SOL performance was not significantly associated with the probability of second year retention. Results of this estimation are provided in Appendix D.*

[19] *Our statistical model included a separate predictor for each institution, in effect comparing students to others in the same institution. The inclusion of institution-level fixed effects renders most ethnic group differences statistically insignificant, save for African American students, who are more likely to persist into the second year than are White students, after controlling for other student-level attributes. The Asian-White difference, though larger than the African American-White difference, could be due to chance because of the smaller number of Asian students in the sample.*

SOL, at a score of 550, the estimated probability of the average African American student persisting into the second year is 90 percent, compared to 87 percent for the average Hispanic student and 89 percent for the average White student. Comparable scores on the Reading SOL, however, yield slightly larger gaps in the likelihood of second year persistence, although the differences remain quite small. The average Hispanic student who scored 550 on the Reading SOL was approximately 3 percentage points less likely to persist into the second year than were African American students (84 percent versus 87percent).

One finding from the use of institution-by-institution predictors was that there were institutional differences impacting student persistence, even among comparable students. Put simply: some institutions do a better job retaining students, even after adjusting for the composition of enrollees. It is unclear from this analysis, or from available data, what institutional-level characteristics may create an environment more conducive to student persistence. This finding merits further attention to identify exemplar schools to assess the types of policies and systems in place to retain enrolled students beyond their first year. (Such an analysis would need to account for transfers to four-year institutions as another desirable outcome.)

Conclusion and next steps

VDOE was interested in investigating the associations between indicators of high school achievement and student success in college. This report describes the results of the first series of analyses conducted by the TSP. The report focused on associations between student achievement in high school and enrollment and persistence in Virginia's two- and four-year IHE and follows a series of VDOE analyses that focused on associations between high school achievement and enrollment in four-year IHE nationwide. Several measures of high school achievement were used in the analyses, including student scores on state end-of-course assessments in reading and Algebra I; participation in high school chemistry and Algebra II; type of diploma or other high school credential earned; and participation in college-level courses while in high school. Additionally, several demographic and student characteristics were used throughout the analyses. In college, data on enrollment in credit-bearing and developmental (non-credit-bearing) mathematics and English courses, and persistence into the second year in the same institution were used as outcome variables.

In summary, this study showed the following:

- Indicators of academic achievement that were the focus of this study, SOL end-of-course tests in Algebra I and reading, and diploma type, were associated with enrollment in college.

- Students who earned Advanced Studies diplomas had higher postsecondary enrollment rates and lower rates of placement in developmental education courses in college when compared to students who earned Virginia's Standard diplomas. For high school graduates of 2008, Virginia's minimum course requirements for the Advanced Studies diploma compared to the Standard diploma included one additional course credit in each of mathematics, laboratory science, history and social science, and at least three credits in foreign language.

- Earning advanced proficient scores on Virginia's end-of-course reading and Algebra I assessments was also associated with lower rates of enrollment in developmental education. Developmental course enrollment rates were lowest when students earned Advanced Studies diplomas and advanced proficient scores on these tests.

- Achievement gaps in enrollment and persistence in postsecondary education were clearly evident for minority students and those from economically disadvantaged families.

However, when these students reached high achievement levels—by earning Advanced Studies diplomas and advanced proficient scores on state assessments—the gaps in enrollment and persistence were substantially reduced or completely eliminated. The public education community must address the achievement gaps in the PreK-12 education system in order to eliminate achievement gaps in higher education.

These results represent the first phase of a two-phased research project aimed at understanding associations between high school achievement and postsecondary enrollment and persistence. Phase II of this project is focused on understanding associations between additional achievement variables in high school, including end-of-course Algebra II and writing assessments, and course performance in college.

APPENDIX A:
Match Procedure and Diagrams

In this section, we describe the content and structure of each file provided to the TSP. In addition, we describe how we linked the files from each data source to assemble the database used in the analyses.

The master file (VDOE File) provided by the VDOE contained de-identified student-level data including SOL test scores and demographic attributes of students (gender, ethnicity, economic disadvantage status, and date of birth). In addition, the file included a host of academic characteristics, including LEP and dual enrollment status, AP course- and exam-taking history, and diploma type. This file contained records for all graduates and non-graduate completers (students who earned GEDs and certificates of program completion) from several graduating cohorts, including the 2006, 2007, 2008, and 2009 graduating years, with a total of 334,174 observations. Only students who graduated in the spring of 2008 were retained, reducing the file to 86,312 observations. In addition, the file contained a binary indicator denoting whether a student enrolled in a Virginia postsecondary institution as a first-time entering freshman in the fall of 2008. Whether or not a student enrolled in a Virginia postsecondary institution was determined using a probabilistic matching process between VDOE and SCHEV records that matches an estimated 83 percent of all Virginia high school graduates enrolled in postsecondary institutions in Virginia. Students who were not identified as enrolled were eliminated, yielding 31,434 unique student observations.

SCHEV generated a cohort file of first-time entering freshmen at Virginia's IHE for the 2008-2009 school year (SCHEV Cohort File). The file contained 43,811 non-unique records; 11,114 of these records could not be matched to VDOE records and were eliminated, yielding a file containing 32,697 non-unique student enrollment records. A retention flag was also included in the file, representing whether a student returned to the institution in the subsequent school year (fall 2009).[20] These records included only those first-time freshman enrolled in a Virginia two- or four-year IHE who were identified as being from Virginia.

A total of 166 non-unique student records were duplicated in the SCHEV Cohort File. The duplicated records were identified and eliminated in the following manner. Students with duplicated records for institution number, second year persistence status, and unique student identifier were reduced to a single observation. This eliminated the 83 duplicated records, producing a file with 32,614 non-unique student records. Approximately 66 records were duplicated (33 unique students) due to concurrent enrollments at multiple IHE. These records were retained to ensure an accurate link to semester and course enrollment records discussed below.

These two files (VDOE File and the SCHEV Cohort File) were linked using the de-identified unique student identifier common to both data sets to create a cohort enrollment file with a matching success rate of 96.5 percent (Table A1). A number (1,148) of students within the SCHEV cohort file were not found in the VDOE spring 2008 graduating cohort file. This discrepancy arose for two reasons. First, approximately 178 students did not graduate in the spring 2008 cohort which was the graduating cohort of interest for the project. Second, the remaining unmatched students were those included in the SCHEV cohort file but who did not have a graduation or completer record in the VDOE file. These records were removed from the SCHEV cohort file since data on the covariates of substantive interest (namely, secondary outcomes) were not available. The final merged file contained 31,466 non-unique student records. Again, students concurrently enrolled at multiple postsecondary institutions remained in the file.

[20] *The retention flag provided by SCHEV only indicates students who returned to the same institution in the subsequent fall semester. This is an important limitation, since 2-year institutions, for instance, with high rates of transfers to 4-year year IHE will have systematically lower retention rates.*

The VDOE also provided the TSP with two de-identified data files generated by the SCHEV. One file included institution and de-identified student-level data, reported at the semester level, such as institution type, student enrollment status, first and second academic major, financial aid recipient status, degree type, and demographic characteristics. Students' records were duplicated because the SCHEV file was collected at the institution-semester level. Thus, a student enrolled in both the fall and spring semesters would be represented twice in the merged file. This file included 85,407 non-unique student records. Of these, 21,950 were eliminated due to a missing unique student identifier, reducing the file to 63,457 records. Approximately 132 additional records were eliminated that had identical values on each variable in the data set. An additional 22 records were eliminated for students who were concurrently enrolled in the same type of institution in the same semester (two-year or four-year IHE). Last, for students concurrently enrolled in different types of institutions in the same semester (19 unique students, with a total of 38 duplicated records), all enrollment records were retained. The final file contained 63,303 unique student records.

This file was linked to the merged VDOE/SCHEV cohort file using several student-level variables (Table A2). Additional variables were added to the merge to avoid erroneous merges: thus, students were linked not only using their unique identifier, but also using birth month, birth year, and gender. The match success rate was 96.5 percent, generating a file of 61,098 students. Approximately 2,241 student records that were only in the SCHEV semester enrollment file were eliminated, because no information on high school outcomes was available.

A second SCHEV file included course enrollment data at Virginia IHE. In addition, the file contained course name and number, subject area, grade, enrollment status, and whether the course was non-credit bearing. The file contained only mathematics and English course enrollment records for the 2008-2009 school year, with a total of 103,472 non-unique records. Due to a missing unique student identifier, 27,175 records were dropped. An additional 142 students had extraneous duplicated course records that were eliminated. Last, within each semester, several students were repeated, within the same institutions, course number, course area (i.e., English or mathematics) with different course grade assignments. Within each semester, for these repeated courses, the course record with the highest reported grade was retained. Repeating this data reduction process across each semester reduced the file to 75,848 non-unique student records. The analytic files included student records that were duplicated both within semesters and between, because students may have enrolled in multiple courses in either the fall or the spring semester. Moreover, students had a unique record for each course enrollment record. Therefore, students who enrolled in mathematics in the fall 2008 semester and an English course in the spring 2009 semester had two enrollment records.

The course enrollment file was linked to the collated Cohort File (Table A3). Eighty-two percent of the Cohort File was successfully linked to the course enrollment records contained within the course enrollment file. There were two primary sources of unsuccessful merges between the cohort dataset, which contained IHE enrollment information, and the course enrollment dataset. First, 2,511 student records from the course enrollment file were not contained within the original VDOE Cohort file. Second, 13,458 students in the Cohort File did not enroll in an English or mathematics course during the 2008-2009 academic year. As a result, these students were not reported in the course enrollment file because only course enrollment records for English or mathematics were included for this analysis. The final collated data set included 86,795 non-unique student records.

For the analyses of developmental course enrollment patterns addressed in Questions 1[21], 2, and 3, course enrollment records across semesters captured in the final IHE enrollment and course enrollment file were collapsed down to a single record of developmental

[21] *Because the question pertains to the composition of two unique student groups (those who enrolled in a postsecondary institution and those who did not), a supplementary data set was assembled that appended the non-enrolled students that were originally dropped from the VDOE Cohort File.*

or non-developmental course enrollment. For instance, a student may have enrolled in multiple sequential or repeated developmental mathematics courses in the fall and spring semesters. For each student in the original VDOE Cohort File (n=31,434), an array of variables were created to capture the number and types of developmental courses a student took.

Table A1. Stage 1 merge, VDOE File to SCHEV Cohort File

	VDOE File (n=31,434)	SCHEV Cohort File (n=32,614)
Linking variable	Unique student identifier	Unique student identifier
Match success rate	96.48% (n=31,466)	
Output File name	VDOE/SCHEV Cohort File	

Table A2. Second-stage merge, VDOE/SCHEV Cohort File to Course Enrollment File

	VDOE/SCHEV Cohort File (n=31,466)	Course Enrollment File (n=63,303)
Linking variables	Unique student identifier Postsecondary school code Birth month Birth year Gender	Unique student identifier Postsecondary school code Birth month Birth year Gender
SCHEV match success rate	96.46% (n=61,098)	
Output File name	Cohort File	

Table A3. Third-stage merge, Cohort File to Course Enrollment File

	Cohort File (n=61,098)	Course Enrollment File (n=75,848)
Linking variables	Unique student identifier Postsecondary school code Reporting period (semester)	Unique student identifier Postsecondary school code Reporting period (semester)
Match success rate	82.1% (n=73,337)*	
Output File name	VDOE/SCHEV/Course Enrollment Cohort File	

**Records from the master file cohort file were retained from the merge procedure. The number of matched records were 73,337, but the total number of records, including those from the master file without a matched pair in the Course Enrollment file, was 86,795.*

APPENDIX B:
Technical Material for Multivariate Analyses Section

The outcomes of interest were whether an individual student enrolled in a developmental course or persisted into the second year. Thus, the dependent variable in each analysis was binary, assuming values of "1" for a positive outcome and "0" for a negative outcome. Limited dependent variables necessitated econometric techniques that adjust for the non-continuous and non-linear structure of the dependent variable. Adopting standard ordinary least squares (OLS) regression techniques with a limited dependent variable introduces numerous statistical violations and jeopardizes the researchers' ability to extract sound inferences from the statistical results. To avert these dangers, maximum likelihood estimation (MLE) techniques (i.e., logistic regression) were used to produce estimates of the impact of student characteristics on the likelihood of either developmental course enrollment or second year persistence. These logits were then converted to predicted probabilities to ease the interpretation of the results.

An additional econometric challenge was introduced by the clustering of students in different IHE. This non-independence violates standard assumptions of the independence (normally, identically, and independently distributed) observations within a given sample, suppressing the standard errors and generating potentially fallacious inferences from the statistical analyses.[22] Logistic regression with clustered standard errors at the IHE level was used throughout to account for this non-independence. This approach was adopted over a multilevel model primarily for computational simplicity.

[22] *For readable, but technical, discussions of this issue, see Primo, Jacobsmeier, and Milyo (2007) and Arceneaux and Nickerson (2009). Both demonstrate that both multilevel approaches and clustered standard errors produce equivalent standard errors.*

APPENDIX C:

Multivariate Logistic Regression Estimates For Research Objective I

Table C1. Logistic regression estimates for English and mathematics DC enrollment

		English			mathematics		
		Odds Ratio+		SE	Odds Ratio+		SE
Demographics	Unidentified	0.63	*	0.15	0.49	***	0.10
	Am. Indian	0.52	*	0.18	0.56		0.21
	Asian	1.09		0.20	1.06		0.22
	Afr. Am.	0.82		0.23	0.87		0.23
	Hispanic	2.01	**	0.68	1.30	*	0.20
	White (reference group)						
	LEP in high school (vs. not LEP)	1.40	**	0.19	0.74		0.19
	Male (vs. female)	0.83	***	0.05	0.76	***	0.07
	Eco. Disad. (vs. not Eco. Disad)	1.19		0.14	0.96		0.09
Diploma type	Standard	2.44	***	0.35	2.31	***	0.30
	Disabilities	2.32	***	0.60	1.00		0.33
	GED	0.86		0.17	2.04	***	0.43
	Advanced (reference group)						
High school academic outcomes							
	High school attendance rate	1.00		0.00	0.99		0.00
	Chemistry SOL (vs. no Chemistry SOL)	0.88		0.12	0.71	***	0.05
	Algebra II SOL (vs. no Algebra II SOL)	0.85		0.12	0.68	***	0.08
	Algebra I scaled score	0.99	***	0.00	0.99	***	0.00
	English scaled score	0.98	***	0.00	0.99	***	0.00
AP and dual credit participation							
	Enrolled	0.44	***	0.11	0.65	**	0.12
	Enrolled and took exam	0.33	***	0.15	0.28	***	0.08
	Took exam	0.37	*	0.20	0.10	***	0.07
	No AP participation (reference group)						
	Dual enrollment (vs. no dual enrollment)	0.45	***	0.12	0.63	**	0.13
	Wald Chi^2	1438.23	***		950.98	***	
	n	27,985			27,985		

Notes:

*+The odds ratio represents how much more likely students with a given characteristic or level of academic of preparation is to enroll in a developmental course compared to baseline students in the reference group. For example, students who earned a Standard Diploma were more than two times (2.4) more likely than their peers who earned an Advanced Diploma to enroll in a developmental mathematics course. When odds ratios are close to one, this indicates that the differences in developmental course enrollment between the two groups are statistically indistinguishable. *p < .10; **p < .05; ***p < .01*

APPENDIX D:

Multivariate Logistic Regression Estimates For Research Objective II

Table D1. Logistic regression estimates for second year persistence in a two-year or four-year IHE

		2-year IHE			4-year IHE		
		Odds Ratio+		SE	Odds Ratio		SE
Demographics	Unidentified	0.99		0.16	1.08		0.22
	Am. Indian	1.48		0.67	0.75		0.25
	Asian	1.76	***	0.14	1.26	*	0.18
	Afr. Am.	0.83	***	0.06	1.19	*	0.12
	Hispanic	1.32	***	0.07	0.90		0.17
	White (reference group)						
	LEP in high school (vs. not LEP)	1.78	***	0.17	1.41	*	0.23
	Male (vs. female)	0.79	***	0.03	0.81	***	0.05
	Eco. Disad. (vs. not Eco. Disad)	0.80	***	0.04	0.79	***	0.06
Diploma type	Standard	0.70	***	0.04	0.63	***	0.04
	Disabilities	0.62	***	0.09	0.52		0.65
	GED	0.44	**	0.16	1.61		1.45
	Advanced (reference group)						
High school academic outcomes							
	High school attendance rate	1.04	***	0.01	1.06	***	0.01
	Chemistry SOL (vs. no Chemistry SOL)	1.19	***	0.08	1.29	***	0.13
	Algebra II SOL (vs. no Algebra II SOL)	1.15	**	0.08	0.96		0.13
	Algebra I scale score	1.00		0.00	1.00	***	0.00
	English scale score	1.00		0.00	1.00	***	0.00
AP and dual credit participation							
	Enrolled	1.01		0.11	1.15	*	0.09
	Enrolled and took exam	1.08		0.08	1.48	***	0.10
	Took exam	0.45		0.28	1.46	*	0.35
	No AP participation (reference group)						
	Dual enrollment (vs. no dual enrollment)	0.92		0.07	0.92		0.07
Institution type	Private IHE (vs. public IHE)				0.56	***	0.06
	Wald Chi^2	39022.43	***		482.05	***	
	n	11,168			16,796		

*Notes:+The odds ratio represents how much more likely a student with a given characteristic or level of academic preparation is to persist into the second year of college in the same institution they were in during the first year. When odds ratios are close to one, this indicates that the differences in persistence between the two groups are statistically indistinguishable.*p < .10; **p < .05; ***p < .01*

APPENDIX E:

Predictive Accuracy of the Algebra I And Algebra II SOL: Classification Table Results

To assess improvements in predictive accuracy associated with the addition of Algebra I and Algebra II SOL performance, we created a series of classification tables. Classification tables can be conceptualized as simply a 2 X 2 cross-tabulation, with the number of observed positive and negative outcomes (here, enrolled and did not enroll in a DC) on one axis, and the number of predicted positive and negative outcomes (again, the number who enrolled and did not enroll in a developmental course) on the other axis. Outcomes are classified as positive (enrolled) if the estimated probability for a given student is greater than 50 percent.

Improvements in predictive accuracy produced from additional variables in the statistical model are determined by calculating the improvement in the number of outcomes correctly predicted compared to the null model. The null model represents the overall proportion of positive outcomes (enrolled students). As a more concrete example, in the data that were linked from public high schools to Virginia's IHE, 18 percent of students who graduated or completed high school and were enrolled in a Virginia IHE were also enrolled in a mathematics DC in the 2008-2009 academic year. If we assumed that no students had enrolled in a DC, we would have gotten 82 percent correct without any explanatory variables (Table E1, Null model). However, the inclusion of Algebra II SOL improves this classification accuracy by approximately five percent (Table E1), while the addition of Algebra I yields very little improvement in the predictive accuracy of the model.

Table E1. Classification accuracy from the logistic regression model predicting mathematics DC enrollment

	Percentage correctly classified	Percentage improvement from null model
Null model	81.7%	NA
Algebra I SOL	81.3%	-.49%
Algebra II SOL and Algebra I SOL	86.1%	5.4%

APPENDIX F:
References

Arceneaux, K., & Nickerson, D. W. (2009). Modeling Certainty with Clustered Data: A Comparison of Methods. *Political Analysis, 17*(2), 177-190.

Balfanz, R., & Byrnes, V. (2006). Closing the Mathematics Achievement Gap in High-Poverty Middle Schools: Enablers and Constraints. *Journal of Education for Students Placed at Risk, 11*(2), 143-159.

Beaton, A. E., Mullis, I. V., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years.* Boston, MA: TIMSS Study Center.

Byrnes, R. B. (2006). Closing the mathematics achievement gap in high poverty middle schools: Enablers and Constraints. *Journal of Education for Students Placed at Risk, 11*, 143-159.

Gottfried, M. A. (2010). Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal, 47*, 1-32.

Jonas, D. L. (2011). *Highlights of progress in meeting Board of Education objectives.* Presentation to the Virginia Board of Education, April 27, 2011. Available at: http://www.doe.virginia.gov/boe/meetings/2011/04_apr/agenda_items/item_a.pdf.

Lamdin, D. J. (1996). Evidence of student attendance as an independent variable in education production functions. *Journal of Educational Research, 89*, 155-162.

Nichols, J. D. (2003). Prediction Indicators for Students Failing the State of Indiana High School Graduation Exam. *Preventing School Failure, 47*, 112-120.

Primo, D. M., Jacobsmeier, M. L., & Milyo, J. (2007). Estimating the Impact of State Policies and Institutions with Mixed-Level Data. *State Politics and Politics Quarterly, 7*(Winter), 446-459.

State Council of Higher Education for Virginia. (2009). The four-year college pipeline in Virginia. Presentation to the Virginia P-16 Council, October 6, 2009, Richmond, VA.

Virginia Department of Education (2010). Virginia's College and Career Readiness Initiative. Richmond, VA: Virginia Department of Education. Downloaded January 15, 2010: http://www.doe.virginia.gov/instruction/college_career_readiness/resources/introductory_briefing.pdf.