JEL: I24, I2

Are Students Dropping Out or Dragging Out the College Experience?

The Roles of Socioeconomic Status and Academic Background

ABSTRACT

Disadvantaged students are substantially less likely to complete a college degree in six years than more advantaged students. The majority of the race/ethnicity differential and 20-35% of the family income and parental education differential is explained by academic background. However, 36% of those without a degree are still enrolled. When taking such persistence into account, we find Hispanics are less likely to have graduated because they are more likely to drag out the college experience, not because they have dropped out. On the other hand, first generation college students appear to be at greater risk of dropping out, rather than persisting.

Leslie S. Stratton
Professor
Department of Economics
Virginia Commonwealth University
IZA Bonn, Germany

James N. Wetzel
Professor
Department of Economics
Virginia Commonwealth University

Are Students Dropping Out or Dragging Out the College Experience?

The Roles of Socioeconomic Status and Academic Background

I. INTRODUCTION

Substantial differences in achievement by socioeconomic status have been documented throughout the educational structure from K-12 through higher education and are a frequent subject of discussion in the public policy arena. The No Child Left Behind Act of 2001 places substantial pressure on K-12 educators to eliminate such differences. College admissions offices have been encouraged to support equal opportunity/access policies for decades. More recently, colleges are being pushed to improve student outcomes, with public financial support becoming contingent on performance. However, it is critical to control for academic background in order to distinguish the differential impact of socioeconomic status as compared to academic background on college outcomes. If academic background is important, then policy makers could focus further attention and resources on K-12 education in order to improve not only K-12 education itself but also college outcomes. If socioeconomic status remains important after controlling for academic background, then policy changes aimed at supporting those at risk at the college level, as identified by socioeconomic status, may be necessary to increase college success.

Furthermore, the traditional approach to measuring 'success' in college by relying only on graduation rates may be misleading. Graduation rates vary depending upon the time period of the analysis. Data that were once assessed using four-year graduation rates as the goal are now commonly assessed using six-year graduation rates, but even this extended measure fails to capture the substantial persistence observed amongst those who have not graduated. We find

that 36% of those who have not graduated at the end of six years are still enrolled. These students are not necessarily "failures"; they may simply be taking longer to graduate. It is important from both a research and a public policy perspective that statistical analysis take into consideration not only degree receipt but also enrollment status when last observed.

We perform such a statistical analysis using the 1996-2001 Beginning Postsecondary Survey. These data comprise a national sample of undergraduates whose enrollment status is observed for six years even as they move from one institution to another. Restricting the analysis to those initially enrolled at four-year institutions, we find that controlling for academic preparation/ability substantially reduces the gap in graduation rates between less and more advantaged socioeconomic groups, particularly for African Americans and somewhat less so for first generation college students. There remains a significant six to 11 percentage point differential in graduation rates for students from lower income and less educated households. More importantly, we also find that those who are still enrolled six years following matriculation are substantially different from both those who are no longer enrolled as well as those who have graduated, and that the marginal impact of socioeconomic status on persistence differs across the population. For example, being Hispanic is associated with greater persistence, whereas being a first generation college student is associated with a higher probability of non-enrollment at the six year mark.

II. LITERATURE REVIEW

A substantial body of research addresses the decision to attend college. Much of it is based on Becker's (1964) model of education as an investment in human capital. According to this theory, individuals pursue a college degree if the expected net present value associated with

doing so is positive. If one focuses narrowly on financial aspects, the benefits are the increased financial earnings of a college graduate relative to those of a high school graduate and the costs are the direct costs such as tuition and books as well as the indirect costs in the form of foregone earnings while in college. Taking a broader perspective, benefits include the various psychic and social benefits associated with college attendance and costs include the time away from family responsibilities as well as the sacrifice of leisure time to class attendance and to study time.

Initial enrollment differences, also referred to as "access", by socioeconomic status, have declined over the last decades but remain substantial. Socioeconomic status is captured here by race/ethnicity, family income, and parental education. The College Board (2010) reported that in 2007 on average 67.2% of high school graduates were enrolled in a two- or four-year college immediately after graduation. The comparable figure for Hispanics was 60.9%, for African Americans 55.6%, for those from low income households 55.0%, and for those whose parents had completed no more than high school 50.9%. If some student subgroup, such as Hispanics, is "under-represented," the converse subgroup (non-Hispanics) is necessarily overrepresented relative to the population average.

Differences by socioeconomic status would be consistent with theory if the costs and/or benefits varied systematically for these different populations. Barrow and Rouse (2005) find no differences in the returns to education by race/ethnicity. Kane (1994) explains aggregate trends in college enrollment for African Americans during the 1980's as a function of changes in college costs and parental background. He is unable to control for academic ability but posits that ability is closely linked to family background. Cameron and Heckman (2001) find substantial differences in school achievement at age 15 for men by race/ethnicity but, in further analysis, find that controlling for long-term family income, academic ability, and family

background (including parental education) explains all these substantial differences and more. They conclude that African American and Hispanic men are actually more likely to complete high school and attend college than white men if these control factors are taken into account. Carneiro and Heckman (2002) find that the family income-college enrollment relation is primarily driven by pre-enrollment differences by family income in ability. Vignoles and Powdthavee (2009) report that income differences in college attendance in the UK are entirely explained by academic background. Belley and Lochner (2007), however, report that, particularly for the less able, income has become a more important factor in driving college enrollment in the U.S., and Dynarski (2003) documents a relation between financial aid and enrollment as well as completion that suggests a link to income. Clearly then it is important to include controls for all measures of disadvantage – socioeconomic and academic – in order to accurately gauge the importance of each factor individually. Unfortunately the data necessary to do so can be difficult to obtain, particularly when looking at the educational attainment of older individuals (Deming and Dynarski 2009).

Even if access to higher education were independent of socioeconomic status, educational attainment may not be. Enrollment in college does not guarantee graduation. The College Board (2010) reports that on average only 56.1% of those entering college, even with the intent of earning a bachelor's degree, graduate within six years. While these 2007 figures reflect an increase from 1997, they clearly demonstrate that degree attainment is far from universal.[1] Furthermore, for African Americans the comparable figure is 40.5%, while for Hispanics it is 46.8%. Recall that these measures are contingent upon enrollment and so are already reflective of a selective population.

While theory suggests that only students who expect the benefits of pursuing a college degree to exceed the costs enroll in college, this does not preclude dropout. Expectations may change as students obtain new information about their expected returns.[2] This new information could be relative to their academic ability, the cost of college, or their likely future returns. Students coming from less advantaged households may be more likely to update their expected net benefits as they may have had less accurate information when making the decision to enroll in the first place. As educational attainment is lower for these populations in general, their knowledge of the process and its net benefits is likely less accurate.

There is a substantial literature pertaining to graduation in the education field,[3] less so in economics. Often these studies employ data on students from only a single institution (see for example, DesJardin, Ahlburg, and McCall 1999), even though Adelman (2006) finds that as many as 60% of all undergraduates attend multiple institutions. Some notable exceptions include Adelman (2006) who uses NELS data, Cameron and Heckman (2001) who use NLSY data, and Cragg (2009) who uses the data employed here. Each demonstrates the importance of controlling for academic background. Adelman in particular argues that controls for test scores, high school grades, and high school curriculum are all important and jointly dominate the impact of socioeconomic status.

Graduation in these analyses is typically modeled as a binary outcome occurring within a fixed time frame. Those who have not graduated within that time frame are treated as a homogenous population. Work in the persistence literature suggests that this assumption may be unwarranted. In examining student persistence from the first to the second year of college, Stratton, O'Toole, Wetzel (2008) find significant differences within the population of non-persisters (those not enrolled one year following matriculation) between those who reenroll

within the subsequent 12 months and those who do not. If all degree recipients completed their requirements within a fixed period of time, measuring success using only degree receipt would fully capture the variable of interest. However, students seem to be taking longer and longer to complete their requirements. Attention these days is focused on six-year graduation rates. These six-year numbers are the rates that four-year institutions are required to provide under the 1990 Student Right-to-Know Act. Even following students for only six years may not be sufficient to identify all 'successful' undergraduates. We address this censoring by using information on enrollment six years following matriculation to distinguish between individuals who are still enrolled in college (persisters) and those who are not (non-persisters), while also identifying degree recipients.

Focusing on populations that have been historically underrepresented at postsecondary institutions, we contribute to the literature (1) by expanding the set of six-year college outcomes to recognize not just those who have completed their degree, but also those who are still persisting in their studies, and (2) by using a representative national data sample of younger college students that follows students as they move between institutions and includes detailed information on respondents' test scores, high school grades, and high school curriculum.

III. ANALYSIS FRAMEWORK

Standard analyses of six-year college outcomes use a logit model to distinguish between those who graduate and those who do not. We begin by estimating such a simple logit controlling only for gender, race, ethnicity, parental education, household income, age, unemployment rate, and marital and parental status. We use these results to estimate the marginal impact of socioeconomic status as measured by race, ethnicity, parental education, and

income on graduation probabilities. These marginal results tell us the impact of each factor, ceteris paribus. We then add controls for academic background/ability and recalculate the marginal impact of socioeconomic status to determine the degree to which socioeconomic status rather than academic preparedness influences graduation rates. Finally, we estimate a specification that controls for a broad array of additional covariates sometimes included in attrition and/or graduation studies to assess the impact these other controls have on observed marginal effects by socioeconomic status. These three steps mimic those employed by Vignoles and Powdthavee (2009) to analyze persistence in the UK.

We then take an important, additional step to call attention to the distinction between persistence and dropout. Specifically, we expand the traditional analysis to further distinguish between those who are enrolled in the last term and those who are not. This analysis requires estimation of a multinomial logit specification. The application is much like that in Stratton, O'Toole, and Wetzel (2008) who use a multinomial logit specification to distinguish between continued enrollment, stopout, and dropout in the first year of college. Thus, the same specifications estimated for the simple logit are rerun for the richer multinomial logit specification to calculate the marginal impact socioeconomic status has upon this three-fold and much more meaningful measure of college outcomes. This analysis will allow us to determine whether some less advantaged populations might have lower graduation rates, not because they are no longer engaged but rather because they are taking longer to graduate.

IV. DATA

The data employed in this analysis come from the restricted access 1996-2001 Beginning Postsecondary Survey (BPS) collected by the National Center for Educational Statistics (NCES)

of the Department of Education. These data constitute a nationally representative sample of students who first matriculated to a postsecondary institution in the 1995-1996 academic year. We restrict our analysis to those individuals with enrollment information through spring 2001 so that we have adequate time to track progress. Given the focus on academic programs culminating in a Baccalaureate degree, enrollment at less than two-year institutions and other institutions which are not likely to offer academic credit (such as beauty, training, and trade schools) is ignored. Some of those initially attending a two-year school are seeking a Baccalaureate degree. However, due to the unobserved and heterogeneous goals of this population, we follow common practice and further restrict our analysis to those in the sample who initially enrolled at a four-year institution. Subsequent enrollment at a two-year institution is recognized. These restrictions yield a sample of 6190 individuals.

Information on academic preparation and student ability is critical for this analysis. These data are missing for a substantial fraction of older students and those not from the United States. As a result, students from abroad and students age 23 and above are excluded from the analysis. A handful of individuals are excluded due to missing age or other characteristics of interest. These restrictions leave a final estimation sample of about 5820 individuals.[4] Sample statistics for this population are reported in Table 1. All the results reported here utilize the BPS longitudinal weights so as to replicate a nationally representative sample; all statistical estimates are corrected for the BPS's complex survey design.

Detailed personal information is available for every respondent. This includes information on gender, race, ethnicity, and age; state of residence; marital and parental status; and parental education and income. State of residence is used to match the state's 1995 unemployment rate to the sample. Higher unemployment rates imply a lower opportunity cost

associated with college enrollment and may attract a different population of students. Regional dummies are also incorporated. Parental education is identified based on the reported education of the most educated parent, with preference given to parental reports. College degree receipt is the modal response. Almost no student's parents were high school dropouts. We distinguish between those parents with no more than a high school degree, those with some college, and those with a post-graduate degree using dummy variables. First generation college students are variously defined in the literature as either those whose most educated parent has no more than a high school degree or those whose most educated parent has less than a college degree: our specification allows for either definition. A dummy variable is used to identify respondents who declare they are independent of their parents, and income dummies that approximately split the population into quartiles are employed to allow a non-linear income effect. The highest income quartile is treated as the base case.

Academic preparation/ability is captured using a number of different variables, as suggested by Adelman (2006). A dummy variable to indicate high school degree receipt is incorporated to identify graduation and perhaps the character trait 'persistence'. Less than 2% of our sample do not have a degree. A measure of the most advanced math course the student plans to take is included to capture the rigor of the student's high school curriculum. Approximately 11% of the sample fails to report this information. We use a dummy variable to identify these persons and treat Trigonometry as the base case. Alternative specifications using NCES coding of the quality of the student's high school curriculum yield substantially the same results. Standardized SAT test scores and self-reported high school GPA are used to assess individual ability. Again dummy variables are used to identify those with missing values. Students taking the ACT are identified with a dummy variable and their ACT scores converted to SAT scores

using a concordance table published by the College Board (1999). Grades are self-reported, since high school transcripts were not available, and such reports are likely biased upward (more students report an A average than any other outcome). Each of these measures of academic preparation/ability is determined prior to college enrollment. As such this research avoids the endogeneity problem associated with using first year college grades to assess progress towards a degree.

In our final specification, we include information on a wide variety of other factors sometimes incorporated in studies of college outcomes. For example, information on the first institution attended is incorporated at this stage. Specifically, we include controls for institution type (public/private), size, growth rate, and institution selectivity. The Integrated Postsecondary Education Data System (IPEDS) from NCES was used to identify the type, size, and growth rate of the institution. Type and size are commonly included as covariates. The growth rate of the institution over the previous four years is included as a proxy for resource availability. Work by Bound, Lovenheim, and Turner (2010) suggests that students may have difficulty completing their studies at institutions experiencing exceptional enrollment growth. Barron's admissions competitiveness index ratings for 1992 were used to classify institution selectivity (Schmitt 2009). There is substantial evidence that more selective schools have higher success rates all else constant (see, for example, Cragg 2009).

Data on the receipt of financial aid in the first year is also included at this stage. We know which individuals received grants, loans, and/or work-study aid. There are concerns about the accuracy of the reported dollar values. The dollar values also have different implications for enrollment decisions at different institutions given the substantial variation in tuition rates across institutions, as tuition levels affect the unmet need that influences both the receipt of and the

dollar amounts of financial aid. Thus, we follow Hu and St. John (2001) and Johnson (2008) in using dummy variables to take into account financial aid type. The modal respondent used as a base case received some grant aid.

Finally, a dummy variable to identify those who first enrolled in spring 1996 rather than fall 1995 is incorporated. Those not enrolling in fall 1995 may be more marginal students either from an institutional perspective or from a motivational perspective – a factor particularly important in Bean's (1980) model of attrition. Note that all of the variables added in the final specification could be considered endogenous. Institutional characteristics are effectively chosen by the student in deciding to enroll. Financial aid offers are also often institution-specific. Finally, the decision to attend college clearly encompasses the decision of when to attend. Endogenous covariates can bias parameter estimates. However, while such covariates are endogenous as regards the decision to attend itself, our sample is already conditional upon attendance. Given this, one might consider such covariates predetermined for the research issue we address. Thus, while focusing our discussion on our more restricted specifications, we also present results for this expanded specification to assess the sensitivity of our results to the inclusion of such covariates. Behaviors such as stopout and part-time enrollment delay graduation but also represent decisions students make along the way and hence are clearly endogenous with respect to six-year outcomes. To avoid such clear endogeneity, we never include controls for actions taken post-enrollment.

The outcome measures for our analysis are derived using information on Baccalaureate degree receipt and college enrollment at the conclusion of spring 2001. Mimicking previous studies of college outcomes, we construct a simple binary outcome measure to identify those individuals who have graduated as of spring 2001. Column 1 of Table 2 presents average

graduation rates for each of the socioeconomic indicators used in this analysis. These measures are slightly higher than those generally reported as they capture graduation at any institution. The overall fraction of the sample that graduates is 63%. The fraction graduating from the first institution attended (not reported in the table) is 55% - a number that matches the six-year graduation rate calculated using IPEDS data for the 2006 cohort. We find evidence (available upon request) that less advantaged populations are more likely to attend multiple institutions, but no evidence that controlling for this alters the results reported below. Proceeding down Table 2, our sample graduation rates are slightly higher at 66% for whites, and substantially lower at 45% for African Americans and 54% for Hispanics. Graduation rates are lowest for those whose most educated parent has no more than a high school diploma (50%) and highest for those with a parent who has a post-graduate degree (77%). Finally, graduation rates rise from 50% for those with the lowest family income to 76% for those with family incomes of at least $75,000. Raw differences indicate a graduation rate differential of about 21 percentage points for African Americans (66%-45%), ten for Hispanics, 19 for those having the least educated versus college educated parents, and 25 for students from the lowest versus highest income quartiles.

We are also, however, able to distinguish between those who did not graduate but are still enrolled in spring 2001 (henceforth called 'persisters') and those who did not graduate and are not enrolled in spring 2001 (henceforth called the 'not enrolled'). The non-enrollment rate like the graduation rate demonstrates a substantial relation to socioeconomic status (see column 3 of Table 2). While 22% of whites are not enrolled in spring 2001, the fraction of African Americans who are not enrolled is over fifty percent higher at 36.5%. The fraction not enrolling more than doubles across the range of household income and parental education: from less than 13% for parents with post-graduate work to more than 30% for those with no more than a high

school degree and from 14% in the highest income category to 32% in the lowest income category.

Nevertheless, these data indicate that persistence at the six year mark is widespread. The first row of column 2 indicates that 13% of the entire sample is continuing to work towards a degree, meaning that 36% (13/(13+23)) of those who have not graduated are persisting. Results are similar when we define persistence as enrollment at any time in the last academic year, with persistence rising to about 40% of non-enrollment.[5] The fraction persisting is furthermore usually higher for those from less advantaged socioeconomic backgrounds as 19% of African Americans and 17% of those with the lowest household income are still enrolled. Thus, there is evidence that the lower graduation rate observed for less advantaged populations six years following matriculation may be partially explained by their higher persistence and partially offset by higher subsequent graduation rates.

These raw statistics suggest that researchers who lump all non-graduates into one category for statistical analysis may be using an oversimplified outcome measure that underestimates long-term college success. While the BPS does not follow these students beyond their sixth year, we can look at those who were persisting at the end of their fifth year and see how they progressed in the following year. Of those who were enrolled but did not graduate in the final term of their fifth year, 26% had graduated and 52% were still enrolled at the end of year six. If the progression from year five to year six is any indication of future trends, many of those classified as persisting in year six may well complete their baccalaureate degree within a year or two.

V. RESULTS

The parameter estimates for the key socioeconomic variables we obtain from simple logit models of graduation are reported in Table 3. Other parameter estimates are available upon request. A positive coefficient indicates an increased probability of graduating. The first column presents results for the model that controls only for basic demographic characteristics. The second column provides results when also controlling for academic preparation/ability, while the third column controls for the broadest array of covariates.

As the magnitude of any effect is difficult to infer from the parameter estimates in a logit model, numerical marginal effects are reported below the coefficient estimates.[6] In nonlinear specifications such as a logit, marginal effects will differ depending upon the location of the observation in the probability distribution. Marginal effects will be larger in the center of the distribution as a movement of β in either direction will capture a larger population. Thus, it is important to select a base case for analysis that holds approximately constant the baseline probabilities. As our primary interest is in identifying the relation between socioeconomic status and college outcomes, we maintain as a base case a single, white, non-Hispanic, childless, 17 year old male from New England and a residence with a sample average unemployment rate, with a college educated parent, and an annual household income greater than $75,000 – an individual from a distinctly advantaged socioeconomic background. Academic preparation and ability are assumed to be approximately modal with the highest expected level of math being trigonometry, high school GPA being between a B and an A-, and SAT test scores falling between 800 and 1100, all for respondents with a high school degree. When including the most inclusive set of covariates, the respondent is assumed to attend a public college of average selectivity that has consistently fewer than 5,000 students; to receive some grant aid; and to begin college in the fall term. The predicted probability of graduating for an individual with

these characteristics ranges from 73.4% for the base model, to 75.6% for the model controlling for academic preparation/ability, to 72.8% for the most inclusive model – thus the location in the distribution, which is so important for the interpretation of logit results, is approximately constant and the marginal effects can be reasonably compared across specifications.

The basic specification (column 1) illustrates significant differences by socioeconomic status. Focusing on the marginal effects, African Americans are 13% less likely to graduate than Whites; Hispanics are 7% less likely to graduate than non-Hispanics; first generation college students are about 11 to 14% (depending on the definition) less likely to graduate than students whose most educated parent has a college degree; and those from the lower half of the income distribution are 9-11% less likely to graduate than those from the highest income quartile, holding all else equal. These differences are somewhat smaller than the raw differentials observed in Table 2 where differences between, for example, the African American and White graduation rates do not control for ethnicity, parental education, or household income, but the differences vary by population. Thus, the difference is on the order of 20-30% lower for Hispanics and first generation college students; 35% lower for African Americans; and about 60% lower for the lowest income quartile. Income in particular is a lot less important when jointly controlling for other basic demographic characteristics and conditioning on initial enrollment, even when not taking into account measures relating to ability.

The marginal impact of socioeconomic status on graduation is, however, further reduced when controlling for academic preparation/ability (column 2). The decrease is on the order of 66% for African Americans and 45% for Hispanics. The decrease is somewhat smaller for those from the bottom half of the income distribution (18-36%) and for first generation college students (16-25%). All of these changes are greater than a standard deviation in magnitude.

Only one marginal effect remains as high as ten percentage points after controlling for academic preparation/ability, whereas previously four of six were larger than ten percentage points. Overall, the impact of high school preparation/ability is both significant and substantial. The marginal impact (not reported but available upon request) of moving either from the lowest level of math (algebra/geometry) to calculus or from a combined SAT test score of less than 800 to a combined SAT test score of more than 1100 is on the order of nine to ten percentage points. The marginal impact associated with reporting a high school GPA of A versus B- or lower is even larger at 30 percentage points! Student performance in high school is a strong proxy for student success in college – much more so than socioeconomic status.

Including the commonly used, but possibly endogenous, covariates (column 3) has only a modest impact. The marginal effects for race/ethnicity rise by about one percentage point. The marginal effects for parental education rise less. The marginal effects for income rise more. That the marginal effect of being in the lowest income quartile changes the most (2.6 percentage points) is likely because this expanded specification includes controls for financial aid receipt. However, none of these differences are over one standard deviation in magnitude and so none are statistically significant. Thus, focusing on the less inclusive specification yields the same results and avoids any taint of endogeneity.

Numerical marginal effects from the multinomial logit specification are reported in Table 4 for each specification and for each outcome. The first row indicates the predicted probability given base case characteristics. Again, these probabilities need to be similar across specifications in order to allow comparison of the marginal effects across specifications. The predicted probability of graduating ranges from 73.2% to 75.7%; the predicted probability of still being enrolled ranges from 9.6% to 11%; and the predicted probability of not being enrolled

ranges from 14.7% to 16.6%. These are all of relatively comparable magnitude. Not surprisingly, the predicted marginal impact of each characteristic on the probability of graduating itself, using the multinomial logit specification, is almost exactly that generated by the logit specification. Thus, we focus our discussion on the other outcomes.

The results clearly indicate that the factors distinguishing non-enrollment from graduation and those distinguishing persistence from graduation are significantly different (p-value 0.00 for all specifications). Non-enrollment and persistence are different outcomes, and policy makers should address these behaviors separately in acting to improve college outcomes.

Looking at the results from the basic specification, there are striking differences in the predicted distribution of non-graduates by socioeconomic status. Holding all else constant, the marginal effect of being Hispanic is over twice as great on persistence (5.3%) as it is on non-enrollment (1.9%). Conversely, the marginal impact of being a first generation college student on non-enrollment is distinctly larger (11-12%) than on persistence (0 to 2.6%). African Americans and those from the lowest income strata have a somewhat higher marginal probability of persisting but also a larger relative chance of not enrolling. Overall, it appears that Hispanics who have not graduated in six years may not have given up but may be on the slow road to graduation while first generation college students may be gone for good.

The results for Hispanics and first generation college students are robust across specifications, albeit with somewhat smaller and less significant marginal effects as more controls are added. As was the case with the simple logit, the marginal effect of belonging to the lowest income quartile on graduating is much smaller when controlling for academic background. This decrease stems largely from a reduction in the marginal impact on non-enrollment. The marginal effect of coming from a low income household on the probability of

persisting diminishes only slightly. Controlling for the largest set of covariates, including first year financial aid type, again increases the marginal effect of income on graduation and non-enrollment. To see if this effect could be driven by differential first year financial aid by income, interactions between income and aid type were incorporated in the specification. These terms were neither jointly nor individually significant. The marginal impact of being African American on outcome probabilities declines precipitously after controlling for academic ability. Even though the marginal effect of being African American on non-enrollment remains greater than the marginal effect on persistence, neither impact is statistically significant at conventional levels.

Educational background continues to have the same large marginal impact on the probability of graduating that it had using the simple binary outcome measure. Thus, the marginal impact associated with the rigor of the student's curriculum and with the student's test score is on the order of nine to ten percentage points, while the marginal impact associated with a change from the lowest to the highest high school GPA is around 30 percentage points. Of greater interest is how the marginal impact of academic background differs for those who have not graduated. While each of our measures appears to have a statistically significant marginal effect on both persistence and non-enrollment, there are some differences. The predicted marginal effect of high school GPA is about three times as large upon non-enrollment as upon persistence. Jointly the coefficients to our high school GPA measures are significantly different for non-enrollment than for persistence at the one percent level. By contrast, test score measures appear to have a greater marginal impact on persistence than on non-enrollment. While those with both higher and lower than median test scores have significantly different probabilities of persisting, only those with lower test scores are significantly more likely not to be enrolled.

Finally, while the rigor of the high school curriculum has three times the predicted marginal impact on non-enrollment as on persistence, these effects are not significantly different from zero. Thus, the multinomial logit specification highlights some differences in how academic background is associated with six-year persistence and dropout behaviors. These differences further emphasize the importance of utilizing a multinominal logit versus the traditional binominal logit to evaluate college outcomes.

To test the robustness of our results and to see if any patterns arise using different observation windows, we reran the analysis (1) coding respondents enrolled at any point during the sixth year as persisters and (2) using fifth year (Spring 2000) outcomes (results available upon request). Obviously, a smaller fraction has graduated in five years as compared to six (58% versus 63%). While 20% were still enrolled in spring 2000 (year five), 16% were enrolled at some point during the 2000-2001 academic year, and 13% were still enrolled in spring 2001. The fraction classified as having withdrawn is relatively stable, ranging from 22% in year 5 to 23% in year six. This stability arises because most of those classified as withdrawals have not been enrolled for three years and 40% have not been enrolled for four years. The majority are long term dropouts. Reestimating the multinomial logit model with these alternative definitions of the dependent variable does not substantially change our results. If anything they show that academic background explains a greater share of the graduation rate differential at the five than at the six year cutoff. This result may be due to the fact that as students persist, their high school record matters less.

We also tested for interaction effects between race/ethnicity and income/parental background. No significant interaction was identified. From this we can infer that the effects of low income or first generation status are not different by race or ethnicity.

VI. CONCLUSION & DISCUSSION

Lower socioeconomic status has long been associated with a failure to complete college. In this study we make two primary contributions. First, we examine the relation between commonly used socioeconomic factors and graduation. In doing so, we are able to include a broader array of controls for individual academic preparation/ability than is typically possible. This approach allows us to assess the impact of socioeconomic status on college outcomes, holding academic background constant. Second, and significantly different from prior studies, we distinguish between those non-graduates who are still enrolled six years following matriculation and those who are not still enrolled. Standard logit analysis treats all non-graduates the same, and hence in some sense as failures. We find that 36% of those who had not graduated in six years were still enrolled when last observed. Persistence at the six year point is substantial. Furthermore, our results indicate that persistence and non-persistence are statistically distinct outcomes. Evidence from those persisting at the five year mark suggests a good fraction of those still enrolled after six years may in fact go on to graduate. Thus, persisters may just be taking longer to graduate. If students from more disadvantaged backgrounds are disproportionately likely to persist, significant differences in graduation rates by socioeconomic status may disappear over time.

Using a national sample of first time undergraduates matriculating in 1995/96, we find that simply jointly controlling for basic demographic (primarily socioeconomic) characteristics explains a substantial fraction of the raw graduation rate differences reported by socioeconomic status. There is a lot of overlap in terms of income, parental education, race, and ethnicity. Still, the differences in graduation rates remain substantial, typically over ten percentage points.

Adding controls for academic background as measured by test scores, high school grades, and high school curriculum reduces those adjusted graduation rate differences by about two-thirds for African Americans, half for Hispanics, and 20-35% for first generation college students and students from low income households. Thus, the predicted difference by race/ethnicity falls to about four percent and becomes at best marginally statistically significant. Those from the lowest half of the income distribution are predicted to be seven percent less likely to graduate relative to those from the upper income quartile. First generation college students are predicted to be between nine and 11% less likely to graduate than students with more highly educated parents. These income and first generation six-year graduation differentials are substantial and statistically significant. While academic background is substantially and significantly associated with college graduation, controlling for academic background does not eliminate all observed differences in graduation rates by socioeconomic status.

We then extend the standard analysis of college outcomes in a novel way to distinguish among three outcomes: graduation, continued enrollment, and non-enrollment. Using a multinomial logit specification, we find evidence that treating all those who have not graduated as a simple, single population is not statistically appropriate. This more complex analysis reveals significant differences in the marginal impact of socioeconomic status on the probability of persisting. Those of Hispanic descent are significantly more likely to persist than non-Hispanics but are not significantly more likely to stop enrolling. Conversely, first generation college students are significantly more likely to not be enrolled, but not significantly more likely to persist than non-first generation college students. African American students and those from lower income households have higher probabilities of both persisting and not enrolling, as compared to their white and higher income counterparts. Although academic background is an

important predictor of college outcomes, controlling for academic background and other covariates does not substantially change this persistence story.

Equal access to higher education has been a social goal for decades now in the United States. Attention has more recently shifted from access to persistence and degree receipt. These outcomes are important for institutions, educators, and policy makers both because limited resources make time spent in school expensive and because it is success in college, not just access, that will help us achieve greater social equality. Most research on persistence has focused on the early years of the college experience, commonly the first to second year transition. Research on degree receipt has focused on six-year graduation rates. That focus on degree receipt fails to distinguish between persisters and non-persisters at the six-year mark. Our analysis begins to fill that substantial void and suggests that long term persistence is deserving of further attention. The fact that many students who are persisting at the five year mark successfully complete their degree in six years is promising, but data that follow students beyond the six year window are needed to determine if those persisting at the six year point actually do graduate. The higher average persistence rate of the Hispanic population also requires some analysis, particularly if policy makers wish to speed time-to-degree for such students.

A common thread throughout this discussion is the importance of academic preparation in the K-12 years on college success. Colleges work with the raw material they receive, and it is costly for colleges to change those K-12 preparations. From a broad policy perspective, improvements in K-12 education are important not only as the United States needs to be more competitive in the global economy, but also because such improvements should improve college success. Success is important especially for those subgroups in society that have been historically underrepresented in college and hence have only slowly advanced up the income and

education social ladder. Policy makers should consider the impact on college graduation and persistence as they evaluate the benefits of improving K-12 education in general. We do, however, continue to observe significant differences in college outcomes by income and family background even after controlling for academic background. This suggests that changes at the college level to help those historically underrepresented may also be in order. Socioeconomic status, due to the luck of the draw at birth, remains a barrier to college completion that needs to be eliminated, if one believes in equal opportunity for all.

Table 1
Sample Means
(% except where noted)

	Mean	Std. Dev.
Basic Specification		
Female	0.550	0.498
White	0.776	0.417
African American	0.109	0.311
Other race	0.115	0.320
Hispanic	0.083	0.276
Parental Education		
High school	0.305	0.012
Some college	0.124	0.329
College	0.251	0.434
Post-graduate	0.264	0.441
Missing	0.055	0.229
Family Income		
Independent	0.028	0.166
Income ($000s)	60.648	54.651
< $25,000	0.224	0.417
$25-$50,000	0.262	0.440
$50-$75,000	0.245	0.430
>= $75,000	0.269	0.443
Age - 17	1.412	0.756
Ever married male	0.004	0.063
Ever married female	0.007	0.083
Father	0.004	0.061
Mother	0.010	0.101
Unemployment rate in state of residence	5.494	1.194
Measures of Academic Preparation/Ability		
No high school diploma	0.011	0.103
Highest level of math:		
Algebra II or less	0.229	0.420
Trigonometry	0.163	0.370
Pre-calculus	0.230	0.421
Calculus	0.259	0.438
Missing	0.119	0.324
Standardized Test Information		

SAT score of 800-	0.186	0.389
SAT score of 800-1000	0.468	0.499
SAT score of 1100+	0.317	0.465
Took ACT test	0.306	0.461
Missing test score	0.029	0.169
High school GPA		
B- or lower	0.088	0.283
B- to B	0.142	0.349
B to A-	0.270	0.444
A- or higher	0.384	0.486
Missing	0.117	0.322
Other Covariates		
Public institution	0.642	0.479
Barron's Admissions Competitiveness Index 1992		
Less selective	0.259	0.438
Moderately selective	0.412	
Very selective	0.328	0.470
Growth in FTE undergraduates (1992-1996 average)		
Negative growth (-1%-/year)	0.310	0.462
No growth	0.410	0.492
Positive growth (1%+/year	0.280	0.449
Institution size		
Number of undergraduates	10398	8630
< 5,000	0.346	0.476
5-10,000	0.237	0.425
10-20,000	0.278	0.448
> 20,000	0.139	0.346
Began in the Spring not Fall term	0.043	0.005
Financial Aid		
Received a loan	0.497	0.500
Received a grant	0.621	0.485
Received work study	0.166	0.372
Number of Observations	~5820	

Eight regional dummies are also incorporated in each specification.

Table 2
Raw Outcomes by Socio-Economic Status

	Six Year Outcome Probabilities		
Sample	Graduate	Still Enrolled	Not Enrolled
Full	63.23	13.36	23.41
Race			
White	65.60	12.33	22.07
African American	44.65	18.80	36.55
Other	64.85	15.11	20.04
Ethnicity			
Non-Hispanic	64.08	12.75	23.17
Hispanic	53.91	20.02	26.07
Parental Education			
≤ High School	50.07	16.58	33.36
Some college	55.53	12.99	31.48
College	69.27	12.51	18.22
Post-graduate	76.97	10.39	12.64
Income			
< $25,000	50.82	17.44	31.73
$25-$50,000	57.52	14.00	28.47
$50-$75,000	66.88	12.76	20.36
≥ $75,000	75.81	9.87	14.32
Number of Observations	~5820		

Table 3
Impact of Socioeconomic Status on Six Year Graduation Rate
Results from a Logit Model

	Base Case		With Academic Preparation/Ability		Largest Set of Covariates	
	Coefficient		Coefficient		Coefficient	
African American	-0.5955	***	-0.2172		-0.2622	*
	(0.1324)		(0.1365)		(0.1452)	
	-13.09%		*-4.22%*		*-5.49%*	
Hispanic	-0.3440	**	-0.2039		-0.2217	
	(0.1385)		(0.1507)		(0.1611)	
	-7.25%		*-3.95%*		*-4.60%*	
Parental Education						
≤ High School	-0.6543	***	-0.5257	***	-0.5071	***
	(0.0783)		(0.0796)		(0.0819)	
	-14.51%		*-10.92%*		*-11.09%*	
Some College	-0.4924	***	-0.4385	***	-0.4357	***
	(0.1350)		(0.1331)		(0.1340)	
	-10.65%		*-8.95%*		*-9.42%*	
Post Graduate	0.2846	**	0.1837		0.1523	
	(0.1192)		(0.1298)		(0.1293)	
	5.20%		*3.23%*		*2.91%*	
Household Income						
< $25,000	-0.4281	***	-0.2975	**	-0.3966	***
	(0.1435)		(0.1395)		(0.1535)	
	-9.16%		*-5.89%*		*-8.51%*	
$25-50,000	-0.4952	***	-0.4305	***	-0.4918	***
	(0.1183)		(0.1188)		(0.1260)	
	-10.72%		*-8.77%*		*-10.73%*	
$50-75,000	-0.2595	*	-0.1790		-0.1806	
	(0.1334)		(0.1350)		(0.1450)	
	-5.38%		*-3.45%*		*-3.72%*	

Standard Errors in parentheses. Marginal effect reported below.

Asterisks indicate significance level: *** 1%, ** 5%, * 10% for a 2-tailed test.

All specifications include controls for gender, other race, independence from parents, age-17, region of residence, the unemployment rate in the state of residence, and gender-specific marital and parental status.

Academic preparation/ability measures include controls for highest math expected in high school, high school GPA, SAT equivalent test scores, and high school degree receipt.

The largest set of covariates includes the type of first year financial aid received; a dummy to identify those who first enter in the spring term; college type (public/private), selectivity, growth rate, and size.

Table 4

Marginal Impact of Socioeconomic Status on Three Six Year Outcomes

Results from a Multinomial Logit Model

	Base Case			With Academic Preparation/Ability			Full Set of Covariates		
	Graduated	Still Enrolled	Not Enrolled	Graduated	Still Enrolled	Not Enrolled	Graduated	Still Enrolled	Not Enrolled
Base Probability	73.35%	10.06%	16.59%	75.72%	9.57%	14.71%	73.17%	10.99%	15.85%
African American	-13.05%	4.30%	8.75%	-4.21%	1.55%	2.66%	-5.55%	2.65%	2.90%
	(0.0000)	(0.0540)	(0.0010)	(0.1350)	(0.3880)	(0.1980)	(0.0960)	(0.3020)	(0.1890)
Hispanic	-7.24%	5.31%	1.93%	-3.84%	3.55%	0.28%	-4.54%	4.08%	0.46%
	(0.0250)	(0.0050)	(0.4250)	(0.1980)	(0.0420)	(0.8970)	(0.1890)	(0.0640)	(0.8470)
Parental Education									
<= High School	-14.71%	2.65%	12.06%	-11.10%	2.05%	9.06%	-11.11%	2.08%	9.04%
	(0.0000)	(0.0450)	(0.0000)	(0.0000)	(0.1090)	(0.0000)	(0.0000)	(0.1780)	(0.0000)
Some College	-10.94%	0.13%	10.81%	-9.31%	-0.05%	9.36%	-9.50%	-0.19%	9.69%
	(0.0010)	(0.9310)	(0.0000)	(0.0040)	(0.9740)	(0.0020)	(0.0070)	(0.9050)	(0.0030)
Post Graduate	5.31%	-1.19%	-4.12%	3.32%	-0.64%	-2.68%	2.95%	-0.42%	-2.52%
	(0.0180)	(0.3010)	(0.0280)	(0.1460)	(0.5670)	(0.1290)	(0.2200)	(0.7470)	(0.1480)
Household Income									
< $25,000	-9.23%	3.57%	5.65%	-5.95%	3.07%	2.88%	-8.64%	3.72%	4.92%
	(0.0020)	(0.0590)	(0.0190)	(0.0290)	(0.0960)	(0.1840)	(0.0040)	(0.1100)	(0.0540)
$25-50,000	-10.75%	2.31%	8.44%	-8.83%	2.01%	6.82%	-10.76%	2.51%	8.25%
	(0.0000)	(0.1530)	(0.0000)	(0.0000)	(0.1860)	(0.0000)	(0.0000)	(0.1760)	(0.0000)
$50-75,000	-5.37%	2.01%	3.36%	-3.49%	1.48%	2.02%	-3.80%	1.48%	2.32%
	(0.0500)	(0.1150)	(0.1830)	(0.1760)	(0.2250)	(0.3690)	(0.1850)	(0.2970)	(0.3480)

P-values in parentheses. The models correspond to those estimated for the logit specification.

The base probability is for a single, childless, 17 year old white, non-Hispanic, non-first generation male with a household income of > $75,000, who lives in New England, in a state with a sample average unemployment rate.

The base probability for academic preparedness and ability is for an individual who has a high school diploma, expects to complete trigonometry, has an A average in high school, and has an SAT score of 800-1100,

The base probability for the full model is for an individual who receives no financial aid, enters a moderately selective public institution with a constant size of less than 5000 students in the fall term

References

Adelman, C. *The toolbox revisited: Paths to degree completion from high school through college.* Washington, D.C.: U.S. Department of Education, 2006.

Altonji, J. G. "The demand for and return to education when education outcomes are uncertain." *Journal of Labor Economics* 11 (1), 1993, 48-83.

Barrow, L. and C. E. Rouse. "Do Returns to Schooling Differ by Race and Ethnicity?" *American Economic Review*, 95(2), Papers and Proceedings, 2005, 83-7.

Bean, J. P. "Dropouts and turnover. The synthesis and test of a causal model of student attrition." *Research in Higher Education*, 12(2), 1980, 155-87.

Becker, G. *Human capital*. New York: National Bureau of Economic Research, 1964.

Belley, P. and L. Lochner. "The Changing Role of Family Income and Ability in Determining Educational Achievement." National Bureau of Economic Research Working Paper 13527, 2007.

Bound, J., M. F. Lovenheim, and S. Turner. "Why have college completion rates declined? An analysis of changing student preparation and collegiate resources." *American Economic Journal: Applied Economics*, 2(3), 2010, 129-57.

Carneiro, P. and J. J. Heckman. "The Evidence on Credit Constraints in Post-Secondary Schooling." *The Economic Journal*, 112(482), 2002, 705-34.

Cameron, S. V., and J. J. Heckman. "The Dynamics of Educational Attainment for Black, Hispanic, and White Males." *Journal of Political Economy*, 109(3), 2001, 455-99.

The College Board. "Concordance Between SAT I and ACT Scores for Individual Students." Research Notes. Office of Research and Development. RN-07, June 1999.

The College Board. The College Completion Agenda: 2010 Progress Report. 2010.

Cragg, K.M. "Influencing the probability for graduation at four-year institutions: A multi-model analysis." *Research in Higher Education*, 50(4), 2009, 394-413.

Deming, D and S. Dynarski. "Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor." National Bureau of Economic Research Working Paper 15387, 2009.

DesJardin, S.L., D.A. Ahlburg, and B.P. McCall. "An Event History Model of Student Departure." *Economics of Education Review*, 18(3), 1999, 375-90.

Dynarski, S. M. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." *The American Economic Review*, 93(1), 2003, 279-88.

Hu, S., and E. P. St. John. "Student persistence in a public higher education system: Understanding racial and ethnic differences." *Journal of Higher Education*, 72(3), 2001, 265-86.

Johnson, I. "Enrollment, persistence and graduation of in-state students at a public research university: Does high school matter?" *Research in Higher Education*, 49(8), 2008, 776-93.

Kane, T. J. "College entry by blacks since 1970: The role of college costs, family background, and the returns to education." *Journal of Political Economy*, 102(5), 1994, 878-911.

Kuh, G. D., J. Kinzie, J. A. Buckley, B. K. Bridges, and J. C. Hayek. *What matters to student success: A review of the literature*. Final report for the National Postsecondary Education Cooperative and National Center for Education Statistics. Bloomington, IN: Indiana University Center for Postsecondary Research. http://nces.ed.gov/npec/papers.asp. 2006.

Manski, C. F. "Schooling as Experimentation: A Reappraisal of the Postsecondary Dropout Phenomenon." *Economics of Education Review*, 8(4), 1989, 305-12.

Schmitt, C.M. "Documentation for the Restricted-Use NCES-Barron's Admissions Competitiveness Index Data Files: 1972, 1982, 1992, 2004, and 2008." National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C., NCES 2010-330, 2009.

Stratton, L. S., D. M. O'Toole, and J. N. Wetzel. "A multinomial logit model of college attrition that distinguishes between stopout and dropout behavior." *Economics of Education Review*, 27(3), 2008, 319-31.

Vignoles, A. F. and N. Powdthavee. "The Socioeconomic Gap in University Dropouts." *The B.E. Journal of Economic Analysis & Policy*, 9(1), Article 19, 2009.

Footnotes:

* This material is based upon work supported by the Association for Institutional Research, the National Center for Education Statistics, the National Science Foundation, and the National Postsecondary Education Cooperative, under Association for Institutional Research 2010 Grant Number RG10-128. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Association for Institutional Research, the National Center for Education Statistics, the National Science Foundation, or the National Postsecondary Education Cooperative. We thank participants at the 2011 AIR Forum and the University of Cergy Pontoise for helpful suggestions.

Leslie S. Stratton, Professor, Department of Economics, Virginia Commonwealth University, 301 West Main Street, PO Box 844000, Richmond, VA 23284-4000. Phone 804-828-7141, Fax 804-828-9103, E-mail lsstratt@vcu.edu. Research Fellow at IZA, Bonn, Germany.

James N. Wetzel, Professor, Department of Economics, Virginia Commonwealth University, 301 West Main Street, P.O. Box 844000, Richmond, VA 23284-4000. Phone 804-828-7145, Fax 804-828-9103, E-mail jwetzel@vcu.edu.

[1] Bound, Lovenheim, and Turner (2010) report evidence from other sources that completion rates have fallen.

[2] See Altonji (1993) for a model of such decision making under uncertainty and Manski (1989).

[3] See Kuh et al. 2006 for a review

[4] NCES Security restrictions require we round sample sizes to the nearest ten.

[5] To assess the degree to which our results might be sensitive to our definition of persistence, we looked more closely at enrollment records. We find that about 50% of those we classify as not enrolled have enrolled for no more than two years of study in the six years they are observed. They either dropped out, never to return, or floated in and out of college. By comparison, only 3% of those classified here as persisters have completed as few as two years of study. On average the enrollment patterns of these individuals are quite different. Nevertheless, we report below estimates using alternative definitions to test the sensitivity of our results to our chosen definition of persistence and to our chosen window of analysis (six years following matriculation).

[6] Analytic marginal effects are similar and available upon request.