The Role of Socioeconomic Status When Controlling for Academic Background in a Multinomial Logit Model of Six-Year College Outcomes

By

Leslie S. Stratton*
Professor

James N. Wetzel**
Professor

June 2011
Preliminary Draft

* Corresponding Author: Leslie S. Stratton, Department of Economics, Virginia Commonwealth University, 301 W. Main Street., P.O. Box 844000, Richmond, VA 23284-4000. lsstratt@vcu.edu, (804) 828-7141, FAX: (804) 828-9103. Research Fellow at IZA, Bonn, Germany.
** Department of Economics, Virginia Commonwealth University, 301 W. Main Street., P.O. Box 844000, Richmond, VA 23284-4000.

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 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 for helpful suggestions.

The Role of Socioeconomic Status When Controlling for Academic Background in a Multinomial Logit Model of Six-Year College Outcomes

By

Leslie S. Stratton
Professor

James N. Wetzel
Professor

June 2011

ABSTRACT

Socioeconomic status as measured by race, ethnicity, income, and parental education is highly associated with college degree receipt. It is difficult, however, to identify the separate effect of each of these measures given their substantial overlap, and it is difficult to statistically differentiate between the impact of academic background/ability and socioeconomic status as the former information is not always available. We use a national sample of first time undergraduates at 4 year institutions from the 1996-2001 Beginning Postsecondary Survey to shed light on these factors. As we observe that a substantial fraction (36%) of those who have not yet graduated are still actively enrolled at the six year mark, we examine not only graduation but also persistence, using a multinomial logit to model outcome. The results indicate that between 30 and 55% of the raw graduation rate differential observed for those from more disadvantaged backgrounds is attributable to differences in academic preparation/ability. Furthermore persistence and withdrawal represent statistically different outcomes. Hispanics appear on average to be less likely to have graduated after six years because they are substantially likely to still be enrolled, not because they are more likely to have given up. Conversely first generation college students appear to be at greater risk of dropping out.

Keywords: College Outcomes, College Graduation, College Persistence, Academic Background, Socio-economic Status

The Role of Socioeconomic Status When Controlling for Academic Background in a Multinomial Logit Model of Six-Year College Outcomes

Substantial differences in six-year college graduation rates by socioeconomic status and academic background/ability have been documented and are a frequent subject of discussion in the public policy arena. Often, and usually due to data limitations, evidence for the discussion is institution-specific or controls for either socio-economic status or academic background but not for both. Given that socioeconomic status is also associated with differences in K-12 education and thus with academic background, it is critical to control for academic background in order to identify the separate impact of socioeconomic status on college outcomes. From a policy perspective this is important because if socioeconomic status has little impact on college outcomes after controlling for academic background, then intervention in K-12 education will be more cost effective than policy changes at the postsecondary level – and vice versa. Unfortunately, the traditional logit model used to distinguish between individuals who have graduated and those who have not graduated fails to recognize the substantial persistence observed amongst those who have not graduated. We find that 36% of those who have not graduated after 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 perspective and a public policy perspective that statistical analysis take into consideration not only degree receipt but also enrollment status when last observed. Such analysis may, for example, reveal that while those from more disadvantaged backgrounds are less likely to have graduated in six years,

holding academic background constant, they are more likely to persist. Polices to help such persisters proceed faster to the degree may in this case be cost effective.

We perform exactly such an analysis using a national sample of first time undergraduates from the 1996-2001 Beginning Postsecondary Survey. 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. Still there remains a significant 6-9 percentage point differential in graduation rates for less advantaged populations. We also find that those who are still enrolled six years following matriculation are substantially different from both those who are not enrolled and those who have graduated and that the marginal impact of socioeconomic status on persistence differs across the population. Being Hispanic is associated with more persistence whereas being a first generation college student is associated with more non-enrollment.

Review of the Literature

A substantial amount of research has been conducted on college enrollment and year-to-year persistence. Becker (1964) models education as a human capital investment that individuals pursue if the overall expected benefits of doing so outweigh the expected costs. 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. Academic and social matching factors between the student and the institution are likely to affect these broader psychic and social returns. Tinto's theoretical work (for example 1975) on student persistence emphasizes the importance of the institution-individual match from both an academic standpoint and from a social fit or congruence standpoint. Bean's theoretical work (1980) stresses the importance of the student's intentions when enrolling, recognizing that students' motivations may not all be alike.

In any case it is clear that students who begin work towards a college degree expect ex-ante to succeed. The empirical literature on persistence and college success necessarily recognizes that these expectations may not be met. Students with limited means may plan ex-ante to enroll less continuously or to enroll part-time in order to have sufficient time in the labor market to support them financially. Thus lower income students may be less likely to persist and take longer to achieve a degree. Alternatively students may obtain new information following matriculation that changes their expected returns (see Altonji 1993 for a model of such decision making under uncertainty and Manski 1989) and induces them to drop out. The new information could be relative to their academic ability, their fit, or their likely future returns.

A substantial empirical literature has developed linking demographic, familial, education, institutional, and economic data to persistence. Pascarella and Terenzini (1978, 1980) and Kahn and Nauta (2001), among others, contribute to the empirical literature on first year attrition using data from single institutions. More recently the empirical literature has expanded to explore institutional retention and degree persistence later in students' college careers. Herzog (2005) examines first to second year persistence, transfer, and dropout behaviors. Allen, Robbins,

Casillis, and Oh (2008) extend this work to look at second to third year retention and transfer behavior.

A similar literature focuses on degree receipt. Typically researchers model graduation using a logit specification (Montmarquette, Mahseredjian, and Houle 2001; Adelman 2006; Goldrick-Rab and Pfeffer 2007). Institutional type and selectivity have been found to be important, independent of fit, as well. Scott, Bailey and Kienzl (2006) examine the differential graduation rates in public and private institutions while Gansemer-Topf and Schuh (2006) and Cragg (2009) discuss institutional selectivity as it relates to graduation. Kuh et al. (2006) provide a substantial review of this empirical literature.

Of particular interest for this study is research that focuses on historically underserved populations. Some studies highlight racial or ethnic differences (Kane 1994 on African Americans; Nora 1987 and Swail, Cabrera, and Lee 2004 on those of Latino origin; Hu and St. John 2001 and Cameron and Heckman 2001 on minority populations more generally). Other researchers focus attention on first-generation college students (Ishitani 2003, 2006) or on social status (Paulsen and St. John 2002). Titus (2006) evaluates the completion rates of students from lower socio-economic backgrounds especially with reference to the institutions they attend. A related line of research focuses on the role of financial aid and the explicit cost of enrollment on persistence and college outcomes (Long 2004a and 2004b, Baird 2006, and Dynarski 2000 and 2003). Clearly, many of these studies overlap as first-generation and lower income students are also more likely to be racial or ethnic minorities. It is also the case that a number of these studies are based on data from a single institution and/or focus on first year outcomes rather than graduation.

One study of college success that covers multiple institutions and controls for multiple covariates is Adelman (2004). A substantial focus of this analysis is the important role of success in high school. To control for innate ability and academic preparation, Adelman (2004) advocates the direct use of test scores and measures of high school preparation. Unfortunately, this level of background detail is often unavailable to researchers. Nevertheless, its importance is underscored by the fact that Adelman concludes that reported racial and ethnic differences in college success are substantially decreased when adjusting for previous academic preparation. He finds that how well students performed in a high quality high school environment is a more important determinant of college success than race or ethnicity. His focus is, however, on graduation alone.

Work in the persistence literature strongly suggests that non-enrollment may combine heterogeneous populations. Stratton, O'Toole, Wetzel (2008), for example, find that stopouts and dropouts constitute distinct populations when analyzing first year outcomes. 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. Following students for only six years may not be sufficient to clearly identify all 'successful' undergraduates. We address this censoring by using information on enrollment six years following matriculation to distinguish between persisters and non-persisters as well as degree recipients.

The knowledge embedded in these studies represents a substantial increase in what we know currently relative to what we knew several decades ago. Focusing on populations that have been historically underrepresented at postsecondary institutions, we extend this knowledge set (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 includes detailed information on respondents' academic background and ability and that follows students as they move between institutions. Such data are essential to assess the role of socioeconomic status in accounting for college success.

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, 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 academic preparedness rather than socioeconomic status per se influences graduation rates. Finally we estimate a specification that controls for a broad array of additional covariates sometimes included in attrition and/or outcome studies to assess the impact these other controls have on observed marginal effects by socioeconomic status. However, as will be discussed later, some of these may be endogenous.

To extend the analysis to account for persistence amongst non-graduates, we further distinguish between those who are enrolled in the last term that they are observed 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. The same specifications estimated for the simple logit are rerun for the multinomial logit to calculate the marginal impact socioeconomic status has upon this richer measure of college outcome. This analysis will allow us to determine whether some less advantaged populations might have lower graduation rates because they are taking longer to graduate, not because they are no longer engaged.

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 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 outside the United States and students age 23 and above were 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 5823 individuals. 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; marital and parental status; and parental education and income. 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. We distinguish between those 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 either as 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. 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 does 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. IPEDS data were 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). Note that these institutional characteristics were effectively chosen by the student in deciding to enroll and

hence may be endogenously determined. Given that concern, we do not include these controls in each specification.

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 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 received some grant aid. Again, these variables were in some sense choice variables for respondents and so may be regarded as endogenously determined. For example, choosing a more expensive school may increase the probability of receiving financial aid or specific types of financial aid. Furthermore, when choosing between similarly costly institutions, a student may select that one which offers more aid.

Finally, dummy variables to control for the region of residence, a dummy variable to identify those who first enrolled in spring 1996 rather than fall 1995, and a measure of the unemployment rate in the respondent's home state are incorporated. Region of residence is included as a general demographic control variable. 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. The unemployment rate may have an effect because college enrollment and participation in the labor market constitute alternative uses of time. High unemployment rates, by making it difficult to

find work, reduce the opportunity cost associated with attending college and thus potentially attract a different college-going population.

While our most expansive specification includes covariates that are endogenous as regards the decision to attend itself, our sample is already conditional upon attendance. Thus, one might consider these covariates predetermined for the research issue we address. We explicitly do not control in any specification for actions taken post-enrollment such as stopout behavior and part-time enrollment. These activities delay graduation but also represent decisions students make along the way and hence are clearly endogenous with respect to six-year outcomes.

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. These measures will be slightly higher than those from single institution studies as they capture graduation at any institution. Column 1 of Table 2 presents average graduation rates for each of the socioeconomic indicators used in this analysis. The overall fraction of the sample that graduates is 63%. 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%), 10 for Hispanics,

19 for those having the least educated versus college educated parents, and 25 for 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% 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.[1] 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

[1] 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 two years of study. On average the enrollment patterns of these individuals are quite different. Nevertheless, we estimate models 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).

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 5 to year 6 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.

RESULTS

The parameter estimates for the key socioeconomic variables obtained from the 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 reports 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 the impact is difficult to infer from the parameter estimates in a logit model, numerical marginal effects are reported below the coefficient estimates.[2] In nonlinear

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

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 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 be from New England with sample average unemployment rate; 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 72.7% for the base model, to 74.4% for the model controlling for academic preparation/ability, to 72.8% for the most inclusive model – thus the location in the distribution is held 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 15% less likely to graduate than Whites; Hispanics are 9% 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 slight for Hispanics (falling from 10% to 9% - a 10% decrease), more substantial for African Americans and first generation college students (on the order of 25-30% lower), and over 60% lower for the lowest income quartile. Income in particular is a lot less important when jointly controlling for other basic demographic characteristics.

The marginal impact of socioeconomic status on graduation is substantially reduced when controlling for academic preparation/ability (column 2). The decrease is on the order of 55% for African Americans, 30% for Hispanics, and 23 to 36% for those from the bottom half of the income distribution. The decline is somewhat smaller at 15 to 29% for first generation college students. All of these changes are over a standard deviation in magnitude. No marginal effect is above 10 percentage points after controlling for academic preparation/ability whereas previously 4 of 6 were above 10 percentage points. 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 10 percentage points. The marginal impact associated with reporting a high school GPA of A versus B- or lower is even larger at 25 percentage points! Student motivation in high school is a strong proxy for student success in college – much more so than socioeconomic status.

Adding more, but possibly endogenous, covariates (column 3) has a modest impact. The marginal effects for first generation and low income students increases by 1 to 2.5 percentage points, while the marginal effect for racial and ethnic groups falls further by 1 percentage point. The different decisions made upon entry to college by different populations appear to have some association with later academic progress. African American and Hispanic students on average enter institutions/accept aid packages that reduce their probability of success. First generation and especially lower income students make college entry decisions that enhance their probability of success. However, none of these differences are over one standard deviation in magnitude and hence are not themselves statistically significant.

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 72.8% to 74.5%; the predicted probability of still being enrolled ranges from 10.5% to 11%; and the predicted probability of not being enrolled ranges from 14.4% to 16.3%. 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.

The contribution of this analysis is in differentiating between persistence and non-enrollment not on revising graduation rate probabilities. As regards this distinction, 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 likely address these behaviors differently 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 (6%) as it is on non-enrollment (2.5%). Conversely, the marginal impact of being a first generation college student on non-enrollment is distinctly larger (11%) than on persistence (0 to 2.5%). African Americans and those from the lowest income strata have more equal marginal effects that are only slightly favor non-enrollment. 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.

These results are robust across specifications, albeit with smaller and less significant marginal effects. As was the case with the simple logit, the marginal effect of income on graduating is larger once one controls for first year financial aid type. The multinomial logit results indicate the larger marginal income effect is offset by non-enrollment rather than persistence. 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.

To test the robustness of our results and to see if any patterns arise using different observation windows, we reran the analysis using (1) sixth year outcomes allowing enrollment at any point during the sixth year and (2) fifth year outcomes to classify respondents as continuing (results available upon request). Obviously, a smaller fraction has graduated in 5 years (58% versus 63%). While 20% were still enrolled in spring 2000 (year 5), 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 6. 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.

CONCLUSION & DISCUSSION

Lower socioeconomic status has long been associated with worse college outcomes. In this study we make two primary contributions. First, we are able to include a broader array of controls for academic preparation/ability than is typically the case, allowing us to identify the impact of socioeconomic status on college outcomes, holding constant academic background. Second, we distinguish between non-graduates who are still enrolled six years following matriculation and those who are not still enrolled. Standard logit analysis with a zero-one dependent variable treats all non-graduates as failures. Our results indicate these are statistically distinct populations and evidence from five year persisters suggests a good fraction of those still enrolled after six years may go on to graduate. Policy makers and institutional researchers should consider these differences as they act to increase graduation rates and promote timely graduation. Use of a national sample of students who are followed as they move between

institutions is also an advantage of this analysis when discussing national graduation rates as a whole.

We find that controlling for basic demographic (primarily socioeconomic) characteristics explains over half of the raw graduation rate differences by income, about a quarter of the raw differences by race and first generation college status, and perhaps 10% of the raw differences by ethnicity. Clearly there is a lot of overlap between these classifications. Controlling for academic background further reduces graduation rate differences by half for African Americans and by 30% or so for Hispanics and students from low income households. Between 15 and 30% of the difference for first generation college students is explained by academic background. Still we observe that African Americans and Hispanics are 6% less likely to graduate, those from the lowest half of the income distribution 7% less likely, and first generation college students 9% less likely to graduate than wealthier, non-first generation, white, non-Hispanics with the same academic background and these graduation differentials are statistically significant.

However, these differentials are based on six year graduation records. The fact that historically the stereotypical college student was expected to complete school in four years and now the norm is to report not four but six year graduation rates suggests that our measure of college 'success' is changing. Some students, particularly those from more disadvantaged backgrounds, may take even longer to graduate. Indeed, we find that 36% of those who had not graduated in six years were still enrolled in the last term they were observed. Persistence at the six year point is substantial. Using a multinomial logit specification to distinguish between those who have graduated, those who are still enrolled, and those who are not still enrolled, we find evidence that each of these states is influenced by different factors and thus that treating all those who have not graduated as a single population is not statistically appropriate.

Further analysis reveals that the marginal impact of socioeconomic status on the probability of persisting differs substantially by socioeconomic indicator. Those of Hispanic descent are significantly more likely to persist than non-Hispanics but are not significantly more likely to be not enrolled. 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 than their white and higher income counterparts. Controlling for academic background and other covariates does not substantially change this 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 issues 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 social equality. Most research on persistence has focused on the early years of the college experience, while research on degree receipt has focused on four or six year outcomes. Our results suggest that persistence continues to be significant even six years following matriculation, and such long term persistence is deserving of 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 identify actual graduation rates for those still persisting at the six year point.

Table 1: Sample Means
(% except where noted)

Variables	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
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
Unemployment rate in state of residence	5.494	1.194
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	5823	

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	5823		

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

	Base Case		With Academic Preparation/Ability		Broadest Set of Covariates	
	Coefficient		Coefficient		Coefficient	
African American	-0.6760	***	-0.3328	**	-0.2622	*
	(0.1340)		(0.1356)		(0.1452)	
	-15.16%		-6.83%		-5.49%	
Hispanic	-0.3996	***	-0.2947	**	-0.2217	
	(0.1355)		(0.1479)		(0.1611)	
	-8.59%		-6.00%		-4.60%	
Parental Education						
≤ High School	-0.6198	***	-0.4695	***	-0.5071	***
	(0.0816)		(0.0835)		(0.0819)	
	-13.80%		-9.90%		-11.09%	
Some College	-0.4835	***	-0.4283	***	-0.4357	***
	(0.1326)		(0.1306)		(0.1340)	
	-10.54%		-8.96%		-9.42%	
Post Graduate	0.2743	**	0.1608		0.1523	
	(0.1198)		(0.1256)		(0.1293)	
	5.09%		2.94%		2.91%	
Household Income						
< $25,000	-0.4383	***	-0.2981	**	-0.3966	***
	(0.1428)		(0.1378)		(0.1535)	
	-9.49%		-6.08%		-8.51%	
$25-50,000	-0.4880	***	-0.3957	***	-0.4918	***
	(0.1176)		(0.1184)		(0.1260)	
	-10.65%		-8.23%		-10.73%	
$50-75,000	-0.2404	*	-0.1409		-0.1806	
	(0.1324)		(0.1341)		(0.1450)	
	-5.02%		-2.78%		-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, 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.

Full set of covariates includes region and unemployment rate of state of residence; 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	72.80%	10.85%	16.35%	74.52%	10.55%	14.94%	73.17%	10.99%	15.85%
African American	-15.12%	6.02%	9.10%	-6.78%	3.03%	3.75%	-5.55%	2.65%	2.90%
	(0.0000)	(0.0100)	(0.0010)	(0.0240)	(0.1470)	(0.0910)	(0.0960)	(0.3020)	(0.1890)
Hispanic	-8.77%	6.25%	2.52%	-6.06%	4.83%	1.23%	-4.54%	4.08%	0.46%
	(0.0070)	(0.0050)	(0.3020)	(0.0540)	(0.0200)	(0.5890)	(0.1890)	(0.0640)	(0.8470)
Parental Education									
<= High School	-13.87%	2.49%	11.38%	-9.93%	1.74%	8.19%	-11.11%	2.08%	9.04%
	(0.0000)	(0.0330)	(0.0000)	(0.0000)	(0.1460)	(0.0000)	(0.0000)	(0.1780)	(0.0000)
Some College	-10.64%	0.09%	10.55%	-9.11%	-0.18%	9.29%	-9.50%	-0.19%	9.69%
	(0.0010)	(0.9570)	(0.0000)	(0.0030)	(0.9070)	(0.0010)	(0.0070)	(0.9050)	(0.0030)
Post Graduate	5.13%	-1.15%	-3.98%	2.99%	-0.32%	-2.67%	2.95%	-0.42%	-2.52%
	(0.0220)	(0.3650)	(0.0340)	(0.1940)	(0.8020)	(0.1350)	(0.2200)	(0.7470)	(0.1480)
Household Income									
< $25,000	-9.56%	4.10%	5.46%	-6.14%	3.17%	2.97%	-8.64%	3.72%	4.92%
	(0.0020)	(0.0360)	(0.0250)	(0.0280)	(0.0930)	(0.1710)	(0.0040)	(0.1100)	(0.0540)
$25-50,000	-10.59%	2.65%	7.94%	-8.20%	2.00%	6.21%	-10.76%	2.51%	8.25%
	(0.0000)	(0.1310)	(0.0000)	(0.0010)	(0.2330)	(0.0010)	(0.0000)	(0.1760)	(0.0000)
$50-75,000	-5.03%	2.22%	2.81%	-2.83%	1.38%	1.45%	-3.80%	1.48%	2.32%
	(0.0650)	(0.1150)	(0.2600)	(0.2810)	(0.3220)	(0.5160)	(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 1st generation male with a household income of > $75,000.

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 living in New England, with a sample average unemployment rate, who receives no financial aid, en a moderately selective public institution with a constant size of less than 5000 students in the fall term

References Cited

Adelman, C. (2004). *Principal indicators of student academic histories in postsecondary education, 1972-2000*. Washington, D.C.: U.S. Department of Education.

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

Allen, J., Robbins, S.B., Casillas, A., and Oh, I. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education*, 49(7): 647-664.

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

Baird, K. (2006). Access to college: The role of tuition, financial aid, scholastic preparation and college supply in public college enrollments. *NASFAA Journal of Student Financial Aid*, 36(3): 16-38.

Bean, J. P. (1980). Dropouts and turnover. The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12(2): 155-187.

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

Bound, J., Lovenheim, M.F., Turner, S. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, 2(3): 129-157.

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

Cameron, S. V., and Heckman, J. J. (2001). The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *Journal of Political Economy*, 109(3): 455-499.

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

Dynarski, S. M. (2000). Hope for whom? Financial aid for the middle class and its impact on college attendance. *National Tax Journal*, 53(3): 629-661.

_____. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. *The American Economic Review*, 93(1): 279-288.

Gansemer-Topf, A.M., and Schuh, J.H. (2006). Institutional selectivity and institutional expenditures: Examining organizational factors that contribute to retention and graduation. *Research in Higher Education*, 47(6): 613-642.

Goldrick-Rab, S., and Pfeffer, F.T. (2007). Second chances: Student mobility, institutional differentiation, and stratification in college completion. Paper presented at the American Sociological Association annual meetings.

Herzog, S. (2005). Measuring Determinants of Student Return Vs. Dropout/Stopout Vs. Transfer: A First-to-Second Year Analysis of New Freshmen. *Research in Higher Education*, 46(8): 883-928.

Hu, S., and St. John, E.P. (2001). Student persistence in a public higher education system: Understanding racial and ethnic differences. *Journal of Higher Education*, 72(3): 265-286.

Ishitani, T. T. (2003). A longitudinal approach to assessing attrition behavior among first generation students: Time-varying effects of pre-college characteristics. *Research in Higher Education*, 44(4): 433-449.

_____. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. *Journal of Higher Education*, 77(5): 861-885.

Johnson, I. (2008). Enrollment, persistence and graduation of in-state students at a public research university: Does high school matter? *Research in Higher Education*, 49(8): 776-793.

Kahn, J. H., and Nauta, M. M. (2001). Social-cognitive predictors of first year college persistence: The importance of proximinal assessment. *Research in Higher Education*, 42(6): 633-652.

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

Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., and Hayek, J. C. (2006). *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

Long, B.T. (2004-a). Does the format of an aid program matter? The effect of in-kind tuition subsidies. *Review of Economics and Statistics*, 86(3): 767-782.

_____. (2004-b). How do financial aid policies affect colleges? The institutional impact of the George HOPE Scholarship. *Journal of Human Resources*, 39(4): 1045-1066.

Manski, C. F. (1989). Schooling as Experimentation: A Reappraisal of the Postsecondary Dropout Phenomenon. *Economics of Education Review*, 8(4): 305-312.

Montmarquette, C., Mahseredjian, S., and Houle, R. (2001). The determinants of university dropouts: A bivariate probability model with sample selection. *Economics of Education Review*, 20(5): 475-484.

Nora, A. (1987). Determinants of retention among Chicano college residents: A structural model. *Research in Higher Education*, 26(1): 31-59.

Pascarella, E.T., and Terenzini, P.T. (1978). The relation of students' precollege characteristics and freshman year experience to voluntary attrition. *Research in Higher Education*, 9(4): 3347-3366.

_____. (1980). Predicting freshmen persistence and voluntary dropout decisions from a theoretical model. *Journal of Higher Education*, 51(1): 60-75.

Paulsen M.P., and St. John, E.P. (2002). Social class and college costs. *Journal of Higher Education*, 73(2): 189-236.

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

Scott, M., Bailey, T., and Kienzi, G. (2006). Relative success? Determinants of college graduation rates in public and private colleges in the U.S. *Research in Higher Education*, 47(3): 249-279.

Stratton, L. S., O'Toole, D. M., and Wetzel, J. N. (2008). A multinomial logit model of college attrition that distinguishes between stopout and dropout behavior. *Economics of Education Review*, 27(3): 319-331.

Swail, W. S., Cabrera, A. F., and Lee, C. (2004). *Latino youth and the pathway to college.* Washington, DC: Educational Policy Institute, Inc.

Tinto, V. (1975). Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research*, 45(1): 89-125.

Titus, M.A. (2006). Understanding college degree completion of students with low socioeconomic status: The influence of the institutional financial context. *Research in Higher Education*, 47(4): 371-398.