

College Selectivity and Degree Completion

Scott Heil

CUNY Office of Institutional Research and Assessment

Liza Reisel

Institute for Social Research

Paul Attewell

*Graduate Center of the City University of
New York*

How much of a difference does it make whether a student of a given academic ability enters a more or a less selective four-year college? Some studies claim that attending a more academically selective college markedly improves one's graduation prospects. Others report the reverse: an advantage from attending an institution where one's own skills exceed most other students. Using multilevel models and propensity score matching methods to reduce selection bias, we find that selectivity does not have an independent effect on graduation. Instead, we find relatively small positive effects on graduation from attending a college with higher tuition costs. We also find no evidence that students not attending highly selective colleges suffer reduced chances of graduation, all else being equal.

KEYWORDS: college selectivity, graduation, selection bias, propensity score matching, tuition

SCOTT HEIL is director of analysis and reporting at the CUNY Office of Institutional Research and Assessment, 555 West 57th St., #1240, New York, NY 10019; e-mail: scott.heil@cuny.edu. His research interests include social stratification, higher education, the history of education, and quantitative methods.

LIZA REISEL is senior research fellow at the Institute for Social Research in Oslo, Norway. Her areas of specialization are disparities in education and the labor market as well as comparative policy research. Her most recent research focuses on gender segregation in education and the labor market.

PAUL ATTEWELL holds the title of Distinguished Professor of Sociology & Urban Education at the Graduate Center campus of the City University of New York. His research focuses on the intersection of social stratification and education, examining the academic progress of low-income undergraduates at non-elite colleges. In recent research funded by the Gates Foundation, his team has undertaken randomized control trials of interventions aimed at increasing academic momentum and improving graduation rates of undergraduates at community colleges.

American colleges can be arrayed along a spectrum of selectivity, from those that have few requirements other than the high school diploma to those that scrutinize academic records and admit only a small fraction from a pool of highly accomplished applicants. What is less clear is how much of a difference it makes whether a student of a given ability enters a more or a less selective college. Does college selectivity affect an undergraduate's likelihood of graduation?

This selectivity question is often framed in terms of the “match” between students and their college. College selectivity and the matching of students with institutions have been linked to several policy issues. For example, affirmative action admissions policies for minority students are said to result in underqualification—as measured by standardized test scores and/or high school grades—leading some commentators to claim that this would undermine the academic progress of affirmative action beneficiaries (Sander & Taylor, 2012; Sowell, 2003; Thernstrom & Thernstrom, 1997). Others have raised the opposite concern: that substantial numbers of economically disadvantaged students are overqualified, in that they attend colleges that are less selective than those students' academic scores would merit. They are therefore denied the benefits that a more selective college might bring (Bowen, Chingos, & McPherson, 2009). These otherwise opposite arguments are built upon a shared assumption: that the academic selectivity of a student's college is consequential for the student's academic progress. However, it has been difficult to establish the independent, causal effect of college selectivity on student outcomes (cf. Black & Smith, 2004; Cohodes & Goodman, 2012). Is it the institution itself that succeeds in getting its students through to degree completion, or is the apparent selectivity effect merely a consequence of the quality of the students entering the institution? Figure 1 provides a synopsis of how this question might play out for a hypothetical student faced with three choices.

Using nationally representative longitudinal data, this article examines whether college selectivity has a substantial effect on a student's chances of graduating with a baccalaureate degree over and above the student's personal attributes, and it estimates the size of that college selectivity effect. Because self-sorting and institutional selection are so central to college placement, we employ statistical techniques that reduce selection bias prior to estimating the effects of college academic selectivity using multilevel models. This combination of propensity score weighting and multilevel modeling allows us to better measure the effect of college selectivity on graduation separate from students' own characteristics. In simple terms, we adjust for students' individual likelihoods of attending a college of a particular level of selectivity, and then we compare the extent to which individual- and college-level predictors are associated with the student's chances of graduating. We also consider the asymmetry of effects, separately estimating the advantage derived from attending a more selective college and the

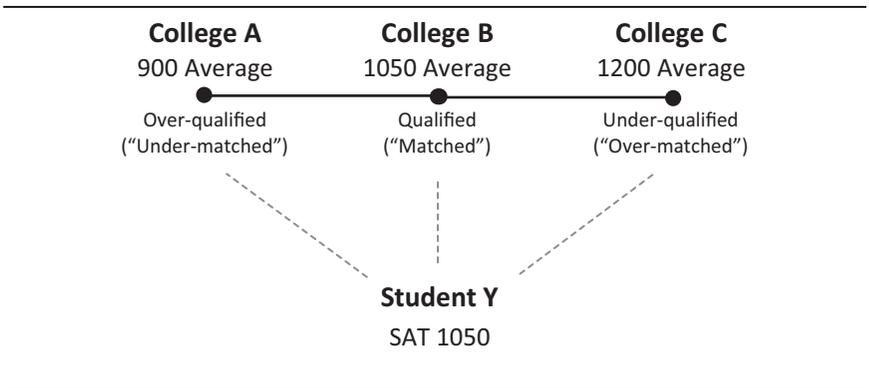


Figure 1. Three hypothetical college choices for a student with a combined math and verbal SAT score of 1050.

disadvantage suffered by students who enter less selective institutions. These two effects may not be the same, as explained in the following. In addition, we determine whether the effect of selectivity is the same across the institutional spectrum. Finally, the mechanisms underlying these effects on graduation are examined: Is it the SAT/ACT selectivity of a college that makes a difference in graduation rates? Or are other characteristics of the college more important?

Theoretical Perspectives and Prior Research

Past studies indicate that a sizeable share of students in U.S. higher education attend colleges at which their academic preparation is markedly higher or lower than the institutional average. Based on the 1997 cohort of the National Longitudinal Survey of Youth, Dillon and Smith (2009) computed the number of students with ASVAB scores more than 20 points different from their college’s average; by that measure, they found that roughly a quarter of undergraduates are overqualified and another quarter are underqualified. Using statewide data from several states, Bowen et al. (2009) found, based on GPA and test scores, that 40% of students were overqualified for the institution they attended and that Black students and those of lower socioeconomic status were much more likely to be overqualified than White and affluent ones.

A variety of research has posited that being underqualified on measures of academic performance or aptitude relative to one’s peers is disadvantageous while attending a less selective institution in which one is relatively overqualified is associated with positive educational outcomes. Variations on this idea—sometimes called the “frog pond hypothesis” or the “big fish

little pond effect”—have appeared in research on college selectivity and subsequent career choice (Cole & Barber, 2003; Davis, 1966), high school selectivity and subsequent college admissions (Attewell, 2001), and bar exam performance among beneficiaries of affirmative action policies (Sander, 2004). Analyzing a large international sample of 10th graders, Marsh (2005) and Marsh and Hau (2003) linked academic selectivity to student test scores via the student’s “academic self-concept,” a construct based on students’ responses to a series of questions about how competent and effective they perceive themselves in several academic areas. They found that students who attended the most selective high schools, and therefore work alongside other high-performing pupils, had a lower academic self-concept and subsequently performed worse on standardized tests than otherwise similar students who attended less selective schools. Though not linked to institutional selectivity per se, Van Laar and Sidanius (2001) describe an analogous general theory of how low status students within an institution may engage in “self-protecting” tactics that could undermine their academic performance.

A related strand in the literature argues that when the average skill level of classmates greatly exceeds his or her own, a less prepared student may struggle academically to maintain the expected standard of performance, which in turn lowers his or her chances of graduating (Sander & Taylor, 2012; Sowell, 2003; Thernstrom & Thernstrom, 1997). This thesis appears in research on affirmative action policies that result in racial minority beneficiaries having lower academic qualifications or test scores than most of their fellow undergraduates, summarized in Sander and Taylor (2012). Chang, Cerna, Han, and Sæenz (2008) reported a more nuanced finding that such an effect might differ for underrepresented minorities (URMs) according to the ethnic makeup of the college they attend; among science majors, they found that college selectivity was negatively associated with persistence of URM students at predominantly White and Hispanic colleges, but it was positive for persistence at Historically Black Colleges and Universities (HBCUs).

Other work, by contrast, has reached nearly the opposite conclusion. Using population data from the North Carolina public university system, Bowen et al. (2009) measured a 15 percentage-point shortfall in unadjusted degree completion rates by overqualified students; in multivariate models with controls, this shrank to 10 points. Bowen and Bok (1998) examined a similar question using a sample of highly selective private colleges. They found that graduation rates were higher in more selective colleges and that Black students who attended a more selective college within this already selective group of colleges have higher graduation rates than otherwise similar Black students who attended less selective colleges. Also analyzing a cohort of students at elite colleges, Small and Winship (2007) produced a result similar to that of Bowen and Bok (1998). They contrasted Whites and Blacks using multilevel models and estimated the effects of eight

different institutional factors on graduation. They found that every 100 SAT point increase in a college's selectivity improved a Black student's probability of graduation by about 6 percentage points and a White student's probability by 3 percentage points, a statistically significant effect. Of the institutional dimensions tested, only SAT selectivity was significant. They interpreted the SAT effect as stemming from positive peer group effects among students at highly selective colleges. In student samples from a wider range of institutions, Alon and Tienda (2005), Long (2008), and Melguizo (2008) added statistical adjustments for selection bias and found positive effects of college selectivity on graduation, in some cases specifically among URM students, although Melguizo noted that the measured effect of college academic selectivity was reduced upon adding a variable to account for selection bias. Others have measured positive effects of undergraduate college selectivity on graduate school attainment (Zhang, 2005) and wages (Black & Smith, 2004).

Light and Strayer (2000) identified a general negative effect of mismatch. After modeling for selection bias, they found that an extreme difference—in either direction—between a student's academic test scores and the institutional average was associated with lower probability of college graduation, notably for the academically weakest students who attended the most selective colleges.

A final body of literature challenges notions about the relationship between college selectivity and student outcomes. Drawing on a national data set, Adelman (1999) argued that there was no difference between "highly selective" and "selective" colleges in terms of their entrants' chances of graduating; he maintained that the meaningful distinction was between nonselective and any form of selectivity. Dale and Krueger (2002) went further to conclude that there was no independent effect of college selectivity on students' subsequent wages. Using both an elite college data set and a nationally representative one, they modeled wages after graduation from college on a rich set of individual- and college-level covariates after matching samples of similar students who attended different colleges. While they initially found an association between college SAT and student outcomes—with greater benefits for lower income students—once they tested their complete model, they observed that college tuition was more predictive of graduates' earnings than academic selectivity, which was not significant in the model once tuition was added. They also noted that the average SAT scores of the colleges a student applied to—but did not attend—were more predictive than those of the college he or she actually attended. Dale and Krueger (2011) applied the same model to a more recent national data set and obtained largely the same finding. Espenshade and Radford (2009) began by replicating Bowen and Bok (1998) using the identical data set. But after they added to the model several student-level predictors, such as high school

GPA, immigrant generation, home ownership, and employment during college, the college selectivity effect was no longer statistically significant.

To summarize: Despite multiple studies and considerable methodological sophistication, the research literature on college selectivity and college completion offers contradictory hypotheses and reports conflicting findings. To some extent, this inconsistency may be attributable to differences in data and method: Some studies focus on elite colleges while others analyze the full range of colleges. Some studies only track students who remain at one college for their entire undergraduate career, while others follow students from college to college. Some focus on Black students while others estimate effects across all races. Analytical techniques also differ: Some of the better known studies had very limited controls for student academic preparation, and the majority of the studies reviewed previously did not make allowances for selection bias.

In this article, we attempt to address some of these shortfalls by combining propensity score matching techniques with multilevel models in order to more directly address selection bias. We use a rich set of individual- and college-level predictors and estimate the relationship between college selectivity and graduation probabilities for a nationally representative sample of students enrolled in four-year colleges. We ask whether attending a more selective college increases a student's probability to graduate once selection bias is adequately accounted for. In other words, do highly selective colleges have higher graduation rates simply because of their more highly qualified student body, or is there something about the colleges themselves that would benefit any student attending the college regardless of their own qualifications?

Methods

Data

We analyzed the Beginning Postsecondary Students (BPS) Longitudinal Study data set, which is available from the U.S. National Center for Education Statistics (NCES) through a restricted data license (Wine, Heuer, Wheelless, Francis, & Franklin, 2002). The 1996/2001 waves of this study followed a nationally representative sample of first-time U.S. undergraduates for six years, including transfers from one institution to another. The data set includes both students who entered college directly after high school and those who delayed entry for any number of years, provided that the start of the cohort period (fall 1995) was their first attempt in higher education. Its two-stage sampling design is suited for an analysis of institutional effects because it contains observations on several students with potentially different background characteristics, per institution. We limited the sample to entrants to four-year colleges (excluding for-profits) who participated in

the final (2001) wave of the study. Because we were interested in modeling reasonably timely graduation from college, the six academic years represented in this data set made for an appropriate window to measure baccalaureate degree completion. We then combined the BPS student-level data with college characteristics drawn from the Integrated Postsecondary Education Data System (IPEDS) database. The latter contains information provided by college administrators about the SAT distribution of the college, the socio-demographic characteristics of the student body, as well as tuition and cost measures. The BPS student data are provided with weights that adjust for sampling and differences in response rates and attrition; we used these in all models.

Variables

The dependent variable is a dummy variable indicating whether a student had graduated with a bachelor's degree within six years of entering college. Student-level predictors included age, race/ethnicity, gender, independent for purposes of federal aid, being a single parent, having dependents, married, parents' income and income squared, parents' education, math curriculum in high school, high school GPA, student's combined SAT score, whether the student considers himself or herself to be primarily an employee who is going to college or a student who is working to pay expenses, several measures of financial aid, paid work hours while in college, first-term part-time enrollment, and delayed college entry.

From the descriptive statistics table in the appendix in the online journal, we see that 67% of the sample graduated in 6 years ($N = 5,480$). If we consider the socioeconomic background variables, we see that there is quite a lot of variation across the students in the sample. Only 2% of the students in the sample have parents with less than a high school diploma, but the rest of the sample is relatively evenly distributed across the remaining four parental education categories (high school diploma, some college, BA degree, or MA or more). About a quarter of the students in the sample work 1 to 15 hours a week and about a quarter work 15 to 30 hours a week. As much as 11% of the students in the sample work more than 30 hours a week while in college. Eighty-six percent of the students applied for financial aid but on average did not receive very large amounts in federal grants and loans. The largest amount of aid is concentrated in the "Other aid" category, which includes merit aid and other aid awarded according to criteria defined by individual states or cities or at the discretion of each institution. Most students in the sample have calculus or precalculus as their highest high school math course, and almost half the sample (45%) received mostly A grades in high school.

Among the sample of colleges ($N = 420$), slightly more than half are public colleges. The average college has 10% non-Hispanic Black students, 5%

Asian students, 5% Hispanic students, and less than 1% of Native American students. The average full-time in-state tuition fee is slightly below \$8,000 a year, varying from \$1,300 to \$23,000. (See appendix in the online journal for further details.)

The college-level predictors were reported for the first college that a student attended. These included: mean SAT score for the college; percentage of the student body who were Black, Hispanic, Asian, or Native American; annual tuition and fees; and a dummy variable for public/private sector. The average SAT score for the college is the main predictor of interest in the analyses described in the following. It was derived from IPEDS by averaging the 25th and 75th percentile combined SAT score for each college. After reviewing the selectivity ranking of institutions in the data set on this variable, we found a few examples of colleges that were less selective according to a classification typology provided in the BPS data set but reported average SAT test scores in the top quartile. It is plausible that these were colleges that did not require the test for admission but reported a value on IPEDS. In those cases, when an alternative value from Barron's *Profiles of American Colleges* or similar reliable sources was available, we chose the lower number. Finally, we used multiple imputation based on a range of college characteristics available in the IPEDS data set to recover missing average college SAT values for the few cases that had no college SAT scores listed either on IPEDS or in the Barron's guide.

Modeling Strategy

Because the BPS data set contains multiple students from the same colleges, the nested structure of the sample violates the independent observations assumption of typical logistic regression. Multilevel models are appropriate for this sample design, with student-level covariates represented as Level 1 and institution-level covariates as Level 2. We estimated all multilevel models using the HLM7 software with a random intercept for each college, and standard errors were adjusted for the clustering of students within colleges (Raudenbush, Bryk, Cheong, Congdon, & Du Toit, 2004). Predictors were grand-mean centered. Since our outcome of interest, earning a bachelor's degree, is binary, the models were specified as multilevel logistic regressions, which are part of the family of hierarchical generalized linear models (HGLMs).

Next we provide equations for the set of models we specified (with the 35 level 1 terms represented in the matrix X), following the hierarchical linear modeling (HLM) notation convention:

$$\text{Prob}(\text{ANYBA}_{ij}=1|\beta_j)=\phi_{ij}$$

$$\log[\phi_{ij}/(1-\phi_{ij})]=\eta_{ij}$$

$$\text{Model I: } \eta_{ij} = \gamma_{00} + \gamma_{01} \times \text{SAT}_j + u_{0j}$$

$$\text{Model II: } \eta_{ij} = \gamma_{00} + \gamma_{01} \times \text{SAT}_j + \gamma_{10-350} \times X_{ij} + u_{0j}$$

$$\begin{aligned} \text{Model III: } \eta_{ij} = & \gamma_{00} + \gamma_{01} \times \text{SAT}_j + \gamma_{02} \times \text{PUBLIC}_j + \gamma_{03} \times \text{PCTBLACK}_j \\ & + \gamma_{04} \times \text{PCTNATAM}_j + \gamma_{05} \times \text{PCTASIAN}_j + \gamma_{06} \times \text{PCTHISP}_j \\ & + \gamma_{07} \times \text{TUITION}_j + \gamma_{10-350} \times X_{ij} + u_{0j} \end{aligned}$$

Selection Bias and Propensity Scores

Scholars want to estimate the effect of college selectivity upon degree completion. Recognizing the potential for spurious causation or confounds, conventional regression analysis adds covariates into predictive models as controls. However, control variables do not resolve another potential problem: selection bias. If we consider our variable of central interest as a “treatment”—a dichotomy such as low versus higher college selectivity—we find that on average, treated individuals differ from untreated individuals on many covariates and background variables. They may on average have different SAT scores, different family socioeconomic status (SES) scores, and so on. In other words, the treatment variable is correlated with many observed and unobserved personal characteristics, which implies that the coefficient for the treatment variable in a conventional regression model reflects not only the causal effect of treatment itself but also the influence of those correlated factors. Simply adding covariates as controls to a regression does not avoid this problem of selection bias.

Statisticians have developed the counterfactual model of causal inference to address the selection into treatment issue. Morgan and Winship (2007) and Reynolds and DesJardins (2009) provide overviews and applications to educational research. The counterfactual strategy attempts to lessen selection bias by constructing a matched sample such that differences on background variables between treated and untreated individuals are minimized. This is analogous to the effects of randomized assignment in an experiment in order to balance experimental and control group subjects on background characteristics. One matching strategy is first to estimate a logistical regression model that predicts who receives treatment, using all available background variables, and to use this logistic regression to calculate the predicted odds of treatment (Morgan & Todd, 2008). Treated and untreated cases are then weighted on their odds of receiving a treatment. Weighting by the odds of treatment has the effect of reducing differences between treated and untreated groups on all other covariates or background

characteristics. It is customary to check for balance both on the odds of treatment and on other covariates and, if needed, respecify the binary regression to improve balance. We went through several rounds of such model refinement for the treatments we present. To gauge the effectiveness of the propensity model at balancing the covariates between the groups, we report average standard biases for our unmatched and matched samples in the following; this statistic measures the difference in standard deviation units between treated and untreated subjects on all measured covariates.

ATT and ATU. The counterfactual model estimates the effect of a treatment upon an outcome (college graduation in our case). Statistically, the technique distinguishes between the “average effect of the treatment on the treated” (ATT) and the “average effect of the treatment on the untreated” (ATU). At first impression, one might expect that the benefit of a treatment for those who are treated should equal the benefit foregone by those who do not receive the treatment. Statisticians, however, show that these two effects are not necessarily equal, so counterfactual analyses undertake two separate matching efforts, one for estimating the ATT and another for the ATU. In the ATT matching, all persons who received the treatment (attended a selective college) are matched with individuals who did not receive the treatment but whose propensity odds suggest that they closely resemble, in academic ability, SES, and other background characteristics, those who received the treatment. The matching algorithm will be unable to match certain individuals who did not receive the treatment and who are very unlike the treated group in background. So the ATT comparison will be between those who went to a selective college and otherwise similar students who did not attend a selective college.

The ATU matching proceeds in a contrasting manner. One begins with students who did *not* attend a selective college and looks for matches among students who did attend selective colleges but in background characteristics are close to those who do *not* attend selective colleges. Thus, the “benchmark” group is different for ATU and ATT estimation, and the individuals constituting the matched groups for ATT and ATU will differ too, although there is considerable overlap, and ultimately the effect size of ATU and ATT may not be the same. The ATT estimate provides a measure of the average effect of going to a selective college among the kinds of students who tend to be in such colleges; the ATU provides a measure of the average effect of attending a selective college among the kinds of students who tend not to go to selective colleges. This distinction can prove important in policy terms.

Limitations

While the present study attempts to measure a nationally representative effect of college selectivity on likelihood of graduation, there are a number

of considerations that we do not address. For one, we did not model nonlinear or extreme levels of divergence between student qualifications and the institutional average; it is possible that the linear and quadratic parameterizations we tested do not capture some other nonlinear pattern that exists for the subset of students whose backgrounds are vastly different from those of most of their peers. For students who attended multiple colleges, as a result of limitations in the data set we have allocated their “institutional effects” to the first four-year college attended; while other researchers have also adopted this convention, it no doubt introduces error into the estimates. In addition, we do not consider whether selectivity criteria other than entrance test scores, including academic and nonacademic criteria, might have a different association with student outcomes. Moreover, for those least selective colleges that do not require the SAT or ACT for admission, it is doubtful whether the reported college average score represents the true population value. As a result, the estimates may be affected by measurement error for such institutions. It is also possible, as suggested by Dale and Krueger (2002), that institutional average test scores alone are too crude a measure to adequately characterize the academic and peer environments of different colleges. A separate issue concerns whether the observation of student qualifications is truncated at the low end; since the lowest scorers on academic tests will tend not to enroll, the models may not be accurate for those students with extremely low scores. Lastly, the findings apply to U.S. higher education in the aggregate and of course cannot predict whether individual students at specific colleges would thrive or not.

Findings

Table 1 presents a summary of the results of three multilevel models calculated for the full range of colleges in the BPS96/01 survey. All models include a random effect for college. Model 1 includes only college SAT selectivity as a predictor; there are no Level 1 controls for individual student attributes. We ran alternative models with college SAT squared and cubed terms. In the large majority of cases, these higher order terms were not statistically significant, and none were of substantive effect, so we have reported the single SAT coefficient here.

Model 2 includes college SAT, and the full individual characteristics listed in online appendix. Model 3 includes college SAT, the individual characteristics listed previously, plus additional college-level variables: a dummy variable indicating that the college is public rather than private, its minority composition, and the college’s annual tuition cost.

These preliminary results in Table 1 do not control for selection; that will be added later. The table reports the logistic coefficient for college-level variables (the main variable of interest being college SAT measured as a standardized z score). The table also reports the p value for the intercept and

Table 1
College SAT Effects in Multilevel Models Predicting Six-Year Graduation

	Coefficient	Individual-Level Controls	<i>p</i> Value	Intercept probability/ Δp
Model 1				
Intercept	.629		<.001	.652
SAT selectivity	.794	No	<.001	.153
Model 2				
Intercept	.866		<.001	.704
SAT selectivity	.281	Yes ^a	<.001	.055
Model 3				
Intercept	.896		<.001	.710
SAT selectivity	.108	Yes	.125	.022
Public college	.148		.459	.030
Percentage Black non-Hispanic	-.001		.836	.000
Percentage Native American	-.039		.008	-.008
Percentage Asian/Pacific Islander	.018		.036	.004
Percentage Hispanic	-.027		.024	-.006
Tuition in thousands	.059		.003	.012
N ^b (Level 1/Level 2)	5,480/420			

Note. Full range of college SATs. Unadjusted for selection effects. BPS 96/01 data.

^aModel 2 and Model 3 in this and the following tables summarize the SAT effects from multilevel models that contain 35 additional Level 1 covariates. Space precludes our listing the coefficients for all the individual-level covariates. The complete output for all models is available upon request from the authors.

^bAll *N*s are rounded to the nearest 10 in compliance with IES restricted data regulations.

the college level estimates and a delta *p* statistic. For each model, the delta *p* value of the intercept should be interpreted as the conditional probability of graduating for an average student attending the average college in the sample. The delta *p* of college SAT can be interpreted as the change in graduation probability associated with one standard deviation change in college SAT, when other covariates are held at their mean values.

Model 1 shows a strong statistically significant relationship between college selectivity and graduation rate, when no individual student characteristics are controlled. For every one standard deviation increase in college SAT, the graduation probability increases by 15.3 percentage points. When plotted, the relationship is represented in Figure 2 (the solid dark line), and it is this that gives an initial impression that college selectivity is strongly associated with higher graduation rates, consistent with institutional data sets such as the IPEDS. However, the picture will change considerably when we add individual Level 1 predictors and other Level 2 predictors.

Model 2 in Table 1 adds 35 individual student characteristics to the multilevel model. The college selectivity coefficient is still significant, but it

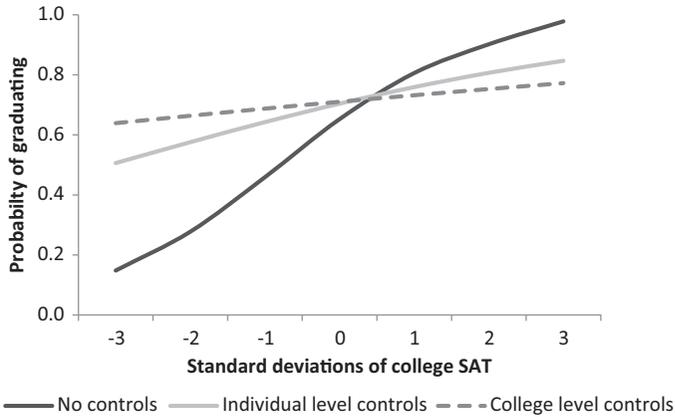


Figure 2. Relationship between college SAT and individuals' chances of graduating.

drops to roughly one third of its previous size. Now a one standard deviation increase in college SAT is associated with a 5.5 percentage point increase in the probability of graduation (graphed as the solid grey line in Figure 2).

Model 3 in Table 1 adds more institutional-level characteristics. In this third model, the college SAT coefficient shrinks to less than half its size in Model 2 and does not attain statistical significance. The delta p indicates that for one standard deviation increase in college SAT, a student's probability of graduating increases by 2.2 percentage points, a quite small effect that is not significantly different from zero (graphed as the dashed line in Figure 2).

Thus far, our analyses have not considered selection bias, nor have they considered whether the effect of college SAT differs at different points along the SAT spectrum. To address these issues, we defined two different contrasts or "treatments": first, T1, attending a mid or high selectivity college rather than a relatively unselective college; second, T2, contrasting students attending highly selective colleges versus a mid selectivity college (excluding the lowest selectivity colleges). Low, mid, and high selectivity colleges were defined by dividing average college SAT scores into thirds of the distribution of average SAT scores across colleges. The cut-off for T1 was an average college SAT score of 1016, and the cut-off for T2 was an average college SAT score of 1133. These cut-offs are very similar to those reported in the Bowen et al. (2009, p.15) study, where the least selective state colleges ("SEL B's") are reported to have average SAT scores up to 1058 points, and

their “flagship institutions” are reported to have average SAT scores from 1125 points and up.

For each treatment, propensity score matching was then used to develop weights that reduced the differences on observed covariates between “treated” and “untreated” samples in T1 and T2 separately for ATT estimation and ATU estimation. There were therefore four sets of propensity weights in all. The balance of covariates showed a greatly reduced bias on observables after applying the propensity weights for each of these combinations.¹ These propensity weights were combined with sampling weights and used in a set of multilevel models.

Table 2 summarizes HLM models after adjustment for selection using the propensity score method described previously, reporting the ATT for both treatments. It answers the question: How much does the average student who attends a more selective college benefit by that attendance? The results are relatively similar for the two treatments or contrasts, suggesting that the average college SAT effect does not change very much when moving from low to high selectivity colleges. The raw selectivity effect on graduation is 7 percentage points for one standard deviation increase in college SAT when moving from low selectivity to mid or high selectivity (T1). The raw selectivity effect on graduation is 8.6 percentage points for a standard deviation increase in college SAT within the more selective groups of colleges (T2). This drops to 5.9 and 6.7 percentage points, respectively, after the other individual-level covariates are added and drops to 2.2 and 1.7 percentage points, respectively, after institutional predictors other than SAT are added in Model 3. The college SAT estimates in Model 3 are not statistically significant at the .05 level. We interpret this as meaning that there may at most be a small independent college selectivity effect—about a 2 percentage point increase in graduation chances associated with attending a one standard deviation more selective institution—but the population parameter may be zero.

At the same time, we see in Model 3 that college tuition level is significantly associated with graduation. The estimates indicate that adjusted for selection and net of individual-level variables and several other college-level variables including SAT selectivity, a \$1,000 increase in annual full-time in-state tuition is associated with a small increase in graduation rates (1.5 and 1 percentage points, respectively, for T1 and T2). In addition, we see a small graduation advantage associated with attending a college with a higher percentage of Asian students and a small disadvantage associated with attending a college with a higher percentage of Hispanic students.

Table 3 reports the ATU. It answers the question: How much of an advantage in graduation prospects does the average student who normally would not attend a more selective college gain by attending a more selective college? Before controls, the selectivity effect on graduation in Model 1 is 7.8 percentage points for one standard deviation increase in college SAT when moving from low selectivity to mid or high selectivity (T1). Within the more

Table 2
College SAT Effects in Multilevel Models of Six-Year Graduation, Average Treatment Effect on the Treated (ATT)

Outcome: BA at 6 Years	Treatment 1: Attending Mid or High Selectivity Versus Low				Treatment 2: Attending High Selectivity Versus Mid Selectivity (omitting low)			
	Coefficient	Individual-Level Controls	<i>p</i> Value	Intercept prob./ Δp	Coefficient	Individual-Level Controls	<i>p</i> Value	Intercept prob./ Δp
Model 1								
Intercept	1.000		<.001	.731	1.471		<.001	.813
SAT selectivity	0.390	No	<.001	.070	0.718	No	<.001	.086
Model 2								
Intercept	0.874		<.001	.706	1.422		<.001	.806
SAT selectivity	0.302	Yes	<.001	.059	0.500	Yes	<.001	.067
Model 3								
Intercept	0.907		<.001	.712	1.451		<.001	.810
SAT selectivity	0.111	Yes	.244	.022	0.111	Yes	.512	.017
Public college	0.266		.312	.051	0.317		.424	.044
Percentage Black non-Hispanic	-0.003		.548	-.001	0.000		.97	.044
Percentage Native American	-0.031		.303	-.006	-0.127		.209	-.020
Percentage Asian/Pacific Islander	0.003		.787	.001	0.033		.004	.005
Percentage Hispanic	-0.016		.176	-.003	-0.067		<.001	-.011
Tuition in thousands	0.076		.002	.015	0.067		.041	.010
<i>N</i> (Level 1/Level 2)	5,400/420				3,620/240			

Note. Propensity score adjusted for selection effects. BPS 96/01 data.

Table 3

College SAT Effects in Multilevel Models of Six-Year Graduation, Average Treatment Effect on the Untreated (ATU)

	Treatment 1: Attending Mid or High Selectivity Versus Low			Treatment 2: Attending High Selectivity Versus Mid Selectivity (omitting low)				
	Coefficient	Individual-Level Controls	<i>p</i> Value	Intercept prob./ Δp	Coefficient	Individual-Level Controls	<i>p</i> Value	Intercept prob./ Δp
Outcome: BA at 6 Years								
Model 1								
Intercept	.142		.077	.535	0.847		<.001	.700
SAT selectivity	.321	No	.001	.078	0.479	No	.027	.090
Model 2								
Intercept	.826		<.001	.696	1.396		<.001	.802
SAT selectivity	.155	Yes	.049	.032	0.293	Yes	.028	.043
Model 3								
Intercept	.901		<.001	.711	1.429		<.001	.807
SAT selectivity	.023	Yes	.782	.005	0.023	Yes	.878	.003
Public college	.213		.409	.042	0.267		.391	.038
Percentage Black non-Hispanic	.000		.89	.000	0.003		.493	.000
Percentage Native American	-.032		.025	-.007	-0.072		.26	-.011
Percentage Asian/Pacific Islander	.010		.328	.002	0.024		.023	.004
Percentage Hispanic	-.021		.034	-.004	-0.070		<.001	-.011
Tuition in thousands	.078		.004	.016	0.062		0.033	.010
<i>N</i> (Level 1/Level 2)	5,400/420				3,620/240			

Note: Propensity score adjusted for selection effects. BPS 96/01 data.

selective groups of colleges, the raw selectivity effect on graduation is 9 percentage points for a standard deviation increase in college SAT (T2). When student characteristics are added in Model 2, the effects of college SAT drop to 3.2 and 4.3 percentage points, respectively, for T1 and T2, and when tuition and other college variables are added, the effects of college SAT become nonsignificant and the delta p values become very small (half a percentage point or less for a standard deviation of college SAT). We interpret this as a clear indication that students who are unlikely to enroll in selective colleges in the first place would not have improved their chances of graduating by enrolling in a more selective college.

However, also the ATU estimates indicate a significant effect of tuition in Model 3, almost identical in size to the ATT estimates. This indicates that among students with a low probability to enter a more highly selective college there is a small graduation benefit to attending a more expensive college. We also see the same minority composition effect within the more selective colleges (T2), as we saw for the ATT in Table 2; a small graduation advantage associated with going to a college with a higher percentage Asian students and a small disadvantage associated with going to a college with a higher percentage Hispanic students. In addition, the ATU estimates matching students who would otherwise attend low selectivity colleges with similar students at mid or high selectivity colleges (T1) show a small negative graduation effect of attending a college with a higher percentage of Hispanics or Native Americans, net of individual-level variables and the other college-level variables.

Discussion

A substantial body of research has examined whether the average academic selectivity of a student's college affects that student's likelihood of graduation. The best known hypothesis argues that *ceteris paribus* a student who attends a more selective college has a much better chance of graduating. Researchers have pointed to a likely mechanism by which this might occur, suggesting that peer group effects in more selective colleges are supportive of retention and academic achievement more generally (Brunello, De Paola, & Scoppa, 2008; Goethals, Winston, & Zimmerman, 1999; Stinebrickner & Stinebrickner, 2006; Winston & Zimmerman, 2004).

However, empirical studies have not always found that college selectivity has a positive effect on graduation, and there are serious methodological limitations in several prior studies (e.g., omitting students who transfer between colleges and failing to address selection bias). This article has tried to address some of those concerns.

We initially found a "raw" college selectivity effect of roughly 15 percentage points in improved graduation rates, similar to that which Bowen et al. (2009) reported. However, when we included a rich set of individual-level student characteristics as controls, we found that this initial effect

of college selectivity was attenuated. When we then added institutional variables beyond college SAT, we found that the college SAT effect was further reduced, primarily due to the effect of a college's tuition cost. This suggests that college selectivity does not have the strong effect on graduation that it has been credited with.

Using methods that reduce selection bias—caused by the sorting of students into colleges—yielded several additional but consistent findings. First, the raw SAT effect does not seem to vary at different levels of college selectivity, so models that contrast low, medium, and high selectivity parts of the spectrum yield similar estimates. Second, the effect of college selectivity is *not* statistically significant after controlling for individual-level predictors and for other college-level variables, of which tuition cost had the most consistently predictive power, in propensity weighted models. Third, the size of effects differs between ATT and ATU models. The ATT indicates that there could be a small SAT effect on graduation (2.2 percentage points for a standard deviation increase in college SAT), but this does not reach statistical significance. The ATU is much smaller in magnitude and is not significantly different from zero. For comparison, the magnitude of these coefficients is in some of the models smaller than that of the institutional predictors of ethnic composition, notably the percentage of Hispanic students, although the composition predictors themselves are not significant in all models. These findings suggest that college selectivity has at most a small positive impact on graduation chances of students who do attend selective institutions and that students who typically do not attend selective colleges do not suffer lower graduation rates because they attend the less academically selective colleges.

This does not mean that selectivity has no effect, but it calls into question two prominent and competing claims in the recent literature. For the first—that a given student should always prefer the most academically selective school to which he or she can gain admission—our analysis shows that for an individual, the net effect of a moving from a lower to middle tier school, or from a middle to upper tier, although positive, is so small as to be almost unmeasurable. On the other hand, our results lend no support to the claim that “over-match” lowers students' likelihood of graduating. After modeling for self-selection into different levels of college selectivity, we find that outcomes are nominally positive, albeit small and usually not statistically different from zero, for all students regardless of whether they appeared likely or unlikely to attend selective schools as predicted by student background characteristics such as race, gender, socioeconomic status, and pre-college test scores.

Our findings are also contrary to most of the models tested by Long (2008). The present study differs in both modeling strategy and target population (in our case, a full entry cohort as opposed to an age cohort). Among other modeling choices, we separated the measure of average tuition and fees from individual financial aid received. Although Long (2008) cited reduced statistical power in one specification that failed to find evidence

of academic selectivity effects, we did not experience such a loss of sample in our propensity-weighted adjustments. We interpret this difference as another reflection of the correlation between academic selectivity and a variety of other covariates, making specification and sampling choices more consequential.

It should be clear that our results do not preclude that academic selectivity may influence other student outcomes such as future earnings, subjective well-being, and professional networks. Nor does our proxy of test scores for “academic selectivity” exhaust all the forms of admissions selectivity that colleges employ. Through their examination of students’ academic and extracurricular portfolios, many colleges select at least partially on many other characteristics such as community service, written communication skills, course-taking patterns, and cumulative grades, among others. The present analysis does not offer any evidence on those covariates, but future research might benefit from closer measurement and analytical separation of the “selectivity” function that colleges exercise versus other sorts of policies and procedures that may be in place as well as their financial resources. The shorter term college retention literature has tended to investigate more specifics such as categories of student support and interventions, and that level of detail may contain insights for a wider analysis of institutional effects on graduation. The present results simply give some sense of where not to place too much emphasis, namely, the idea that academically selective institutions, as measured by admissions test scores, somehow have a “secret sauce” that gets students to graduate disproportionately relative to their background characteristics.

Our findings come closest to those of Espenshade and Radford (2009). We have added to their study by including all types of four-year colleges, public as well as different types of private colleges, at all ranges of college selectivity. Our finding that college selectivity had no independent relation to the outcome variable once tuition was modeled is also similar to that in Dale and Krueger (2002), only we measured this relationship on the probability of graduating and used a different national data set.

Our findings have important implications for the current policy debate. Some critics use the difference in graduation rates between selective and unselective colleges as evidence that unselective, affordable colleges are underperforming, are of lower quality, or otherwise fail to live up to the standards set by more selective colleges. Our findings suggest that differences in graduation rates are largely driven by the composition of the student body and secondarily by high tuition cost. While we did not explore the mechanisms that might explain the tuition finding, the result is compatible with at least two different hypotheses that have been offered in the literature: (a) Higher tuition may indicate better college resources to promote student success such as more extensive counseling and advisement and (b) higher cost may also incentivize students to graduate because of the large investment that they, and their families, have made in their schooling so that they avert the “moral hazard” of not completing college.

If educational opportunities are to remain widely available regardless of social background, then the colleges that uphold that promise by being less expensive are bound to enroll more nontraditional college students and by definition will not be able to benefit from the consequences of charging high tuition. It is possible that colleges with low graduation rates can make some institutional changes to help their students stay in college and complete their studies. However, as this article shows, merely attending a more selective college does not make much of a difference for a given student's chances of graduating if all else remains the same.

Conclusion

We find at best weak evidence that institutions raise, via academic selectivity, the graduation rates of students who otherwise would have lower chances of graduating. At the same time, we find little support for the hypothesis that academic mismatch has a significant impact on U.S. college completion. Put differently, from the standpoint of an individual student, choosing to enroll at a college whose average admissions test scores are substantially higher or lower does not appear to help or harm her chances of graduating. Thus, other considerations in college selection, such as proximity to family and social supports, favorable financing, the availability of programs and faculty of interest, and personal preferences, might be more salient criteria to inform that decision.

The finding on tuition indicates a need for further research on the institutional and student processes by which college cost is associated with student outcomes. It would also be beneficial to understand more closely how cost is linked to academic selectivity, including the extent and effects of heterogeneity in that relationship. Another timely question is whether the same patterns hold in other sectors of higher education, notably the for-profit sector, which we omitted from our sample but that has grown substantially since the data we analyzed were collected.

These results also have potential ramifications for accountability initiatives at public institutions. Responding to demands by political leaders and other stakeholders, college administrators may be tempted to raise cut-off scores on admissions tests in order to improve their college's performance measurements. Our analysis suggests that any resulting increase will be due to compositional effects from changes in the student body, with a corresponding decrease in college access, rather than any improvement in the effectiveness of the institution per se.

Notes

We would like to thank the three anonymous reviewers for their insightful comments. This research was supported by a grant from the Spencer Foundation.

¹The mean absolute value of the standard bias for T1 was 0.289 before matching and 0.023 and 0.021 after average effect of the treatment on the treated (ATT) weighting and average effect of the treatment on the untreated (ATU) weighting, respectively. The mean absolute value of the standard bias for T2 was 0.248 before matching and 0.021 and 0.039 after ATT weighting and ATU weighting, respectively. Further details about the matching procedure are available from the authors upon request.

References

- Adelman, C. (1999). *Answers in the toolbox: Academic intensity, attendance patterns, and Bachelor's degree attainment*. Washington, DC: U.S. Department of Education.
- Alon, S., & Tienda, M. (2005). Assessing the "mismatch" hypothesis: Differences in college graduation rates by institutional selectivity. *Sociology of Education*, 78(4), 294–315.
- Attewell, P. (2001). The winner-take-all high school: Organizational adaptations to educational stratification. *Sociology of Education*, 74, 267–295.
- Black, D. A., & Smith, J. A. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics*, 121, 99–124.
- Bowen, W. G., & Bok, D. C. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press.
- Bowen, W. G., Chingos, M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.
- Brunello, G., De Paola, M., & Scoppa, V. (2008). *Residential peer effects in higher education: Does the field of study matter?* (Discussion Paper Series No. 3277). Bonn, Germany: IZA (Institute for the Study of Labor). Retrieved from http://www.iza.org/index_html?lang=en&mainframe=http%3A//www.iza.org/en/web/content/publications/papers&topSelect=publications&subSelect=papers
- Chang, M. J., Cerna, O., Han, J., & Sænz, V. (2008). The contradictory roles of institutional status in retaining underrepresented students in biomedical and behavioral science majors. *Review of Higher Education*, 31(4), 433–464.
- Cohodes, S., & Goodman, J. (2012). *First degree earns: The impact of college quality on college completion rates* (No. RWP12-033). Boston, MA: Harvard Kennedy School Faculty Research Working Papers.
- Cole, S., & Barber, E. (2003). *Increasing faculty diversity. The occupational choices of high-achieving minority students*. Cambridge, MA: Harvard University Press.
- Dale, S., & Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics*, 117(4), 1491–1527.
- Dale, S., & Krueger, A. B. (2011). *Estimating the return to college selectivity over the career using administrative earnings data* (Working Paper No. 17159). Cambridge, MA: NBER.
- Davis, J. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology*, 72, 17–31.
- Dillon, E., & Smith, J. (2009). *The determinants of mismatch between students and colleges. Unpublished working paper*. Retrieved from <ftp://ftp.cemfi.es/pdf/papers/Seminar/Dillon%20and%20Smith%20Mismatch%20080309.pdf>

- Espenshade, T. J., & Radford, A. W. (2009). *No longer separate, not yet equal*. Princeton, NJ: Princeton University Press.
- Goethals, G., Winston, G., & Zimmerman, D. (1999). *Students educating students: The emerging role of peer effects in higher education* (Working Paper No. 50). Williams College, MA: Williams Project on the Economics of Higher Education. Retrieved from <http://sites.williams.edu/wpehe/files/2011/06/DP-50.pdf>
- Light, A., & Strayer, W. (2000). Determinants of college completion: School quality or student ability? *The Journal of Human Resources*, 35(2), 299–332.
- Long, M. C. (2008). College quality and early adult outcomes. *Economics of Education Review*, 27, 588–602.
- Marsh, H. W. (2005). Big fish little pond effect on academic self-concept. *German Journal of Educational Psychology*, 19, 119–128.
- Marsh, H. W., & Hau, K. T. (2003). Big fish little pond effect on academic self-concept: A crosscultural (26 country) test of the negative effects of academically selective schools. *American Psychologist*, 58(5), 364–376.
- Melguizo, T. (2008). Quality matters: Assessing the impact of attending more selective institutions on college completion rates of minorities. *Research in Higher Education*, 49(3), 214–236.
- Morgan, S., & Todd, J. (2008). A diagnostic routine for the detection of consequential heterogeneity of effects. *Sociological Methodology*, 38(1), 231–281.
- Morgan, S., & Winship, C. (2007). *Counterfactuals and causal inference*. New York, NY: Cambridge University Press.
- Raudenbush, S., Bryk, A., Cheong, Y., Congdon, R., & Du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- Reynolds, C. L., & DesJardins, S. (2009). The use of matching methods in higher education research: Answering whether attendance at a two-year institution results in differences in educational attainment. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research 24* (pp. 47–97). Dordrecht, Netherlands: Springer.
- Sander, R. H. (2004). A systemic analysis of affirmative action in American law schools. *Stanford Law Review*, 57(2), 367–483.
- Sander, R. H., & Taylor, S., Jr. (2012). *Mismatch: How affirmative action hurts students it's intended to help, and why universities won't admit it*. New York, NY: Basic Books.
- Small, M., & Winship, C. (2007). Black students' graduation from elite colleges: Institutional characteristics and between-institution differences. *Social Science Research*, 36(3), 1257–1275.
- Sowell, T. (2003, February 8). Damaging admissions: Increasing faculty diversity. *Capitalism Magazine*. Retrieved from <http://www.capmag.com/article.asp?ID=2448>
- Stinebrickner, R., & Stinebrickner, T. R. (2006). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics*, 90, 1435–1454.
- Thernstrom, S., & Thernstrom, A. (1997). *America in Black and White: One nation indivisible*. New York, NY: Simon & Schuster.
- Van Laar, C., & Sidanius, J. (2001). Social status and the academic achievement gap: A social dominance perspective. *Social Psychology of Education*, 4, 235–258.
- Wine, J. R., Heuer, E., Wheelless, S. C., Francis, T. L., & Franklin, J. W. (2002). *Beginning Postsecondary Students Longitudinal Study: 1996-2001 (BPS:1996/*

College Selectivity and Degree Completion

- 2001 methodology report* (NCES 2002-171). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Winston, G., & Zimmerman, D. J. (2004). Peer effects in higher education. In C. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 395–424). Chicago, IL: University of Chicago Press.
- Zhang, L. (2005). Advance to graduate education: The effect of college quality and undergraduate majors. *Review of Higher Education*, 28(3), 313–338.

Manuscript received October 28, 2012

Final revision received December 30, 2013

Accepted May 4, 2014