Do Colleges Favor Students from Lower-Achieving High Schools? Evidence from the University of California

Sam M. Young*

June 14, 2017

Abstract

In addition to race and socioeconomic status, colleges may also consider the achievement level of a student's high school during admissions, perhaps favoring students from lower-achieving (disadvantaged) schools for equity or diversity reasons. This paper uses data from the University of California (UC) to estimate the direction and magnitude of how a school's overall achievement level affects admissions for otherwise-similar applicants. All else equal, I find mean school achievement to be negatively associated with probability of admission at five out of eight UC campuses between 2001 and 2006. For an average campus, attending a one-decile higher-achieving school associates with a 0.8 percentage point decrease in admission probability. This marginal effect appears to be somewhat larger at the lowest-achieving high schools.

Keywords: school choice; human capital; educational economics; college admissions; postsecondary education

JEL codes: I21, I24, I28

*Department of Economics, University of California, San Diego (smyoung@ucsd.edu). I am grateful to Kate Antonovics for providing data, and Tongshan Chang and Susannah McCormick at the UC Office of the President for follow-up assistance. I thank Kate Antonovics, Julian Betts, Gordon Dahl, Sally Sadoff, Krislert Samphantharak, and especially Julie Cullen for their invaluable guidance and feedback. I also thank Eli Berman, Prashant Bharadwaj, Mae Brown, Melissa Famulari, Mike Furchtgott, Matt Goldman, Roger Gordon, Sam Krumholz, Shanthi Manian, Ben Miller, Brigitte Roth Tran, Larry Schmidt, Ling Shao, Jeff Shrader, Igor Vaynman, and other seminar participants at UC San Diego for helpful comments. Any errors are my own.

1 Introduction

In addition to race and socioeconomic status, the high school that a student attends is another important dimension of an advantaged versus disadvantaged past. Coursework, peers, and overall resources vary across schools, giving rise to vastly different experiences between students of different secondary schools. Many of these differences can be systematically approximated using a school's overall level of achievement. Generally speaking, the higher a school's average level of achievement, the more resources and advantages it typically offers.

In light of this, it is possible that colleges may incorporate the overall achievement or "quality" of a student's school into their admissions decisions. Many colleges, for instance, value equity or diversity along a variety of dimensions, and may wish to increase the share of students from disadvantaged schools. School quality may also help put students' past achievements into context, thereby conveying additional information about student ability.

Whether driven by diversity or other reasons, examples can be found in which school quality directly enters into admissions. For instance, section 51.805 of the Texas Education Code lists 18 factors that determine admissions for students who do not otherwise gain automatic admission. These include "the financial status of the applicant's school district" and "the performance level of the applicant's school." Similarly, colleges in the University of California system follow a set of 14 admissions criteria designed to evaluate students relative to their life circumstances. Coming from a "disadvantaged social or educational environment" is listed among various personal background factors taken into account. To help achieve this goal, at least two UC campuses used (and displayed on their websites) admissions rubrics giving explicit bonuses to students from lower-achieving schools.

Moreover, these bonuses occur even net of the admissions consideration of class rank — which is otherwise often discussed as a way in which the high school affects admissions. Outside of class rank, there has been little academic attention paid to the direct use of school quality in admissions, and little is known about its magnitude. This paper attempts to help fill this gap.

Using administrative data from the University of California, I estimate the admission rules of UC colleges between 2001 and 2006, paying particular attention to how a school's average achievement level affects admission probability net of all other observed information. This analysis is done from a partial equilibrium perspective, focusing on the admissions stage itself, after students have already completed high school and submitted their applications. Absent any barriers to identification, the effect of interest answers "How much more or less likely is an otherwise-identical applicant (ex post) to be admitted when coming from a higher-achieving school?"[1]

[1] To clarify, 'ex post' means the comparison of interest is between students who are identical *after* having gone through high school, and who both decided to apply to a given college. Except for their high school, these students look identical from a college's perspective. These students may not have been identical 'ex ante,' or before entering

The administrative data allow me to control for many essential determinants of college admissions, including high school GPA, SAT scores, intended major, and socioeconomic status. However, these data ultimately make up only a subset of the information observed during actual admissions. A general threat to identification is that, among the remaining items that are observed *during admissions but not in the data*, there may be other factors that vary systematically with school quality. Effects of these factors may be incorrectly attributed to school quality, resulting in biased parameter estimates.

Taking results at face value, I find school quality to be negatively associated with probability of admission at five out of eight UC campuses between 2001 and 2006 while controlling for other factors — including status within the UC's system-level percent plan. Aggregating across campuses, the average association is about -0.8 percentage points for a one-decile increase in high school achievement, or roughly equivalent to a 0.04 decrease in high school GPA. The estimated marginal effect appears to be somewhat larger at the lowest-achieving schools. Back-of-the-envelope calculations suggest that this admissions practice increases the share of attending students from below the median high school by 14 percent (from 24.7 to 28.2 percent), while decreasing the share from above the median high school by 5 percent.

Acknowledging the possibility of omitted variables bias, these results provide an initial estimate, or a starting point, for an effect that has had little past benchmark for its magnitude. This magnitude matters from a policy perspective, since it ultimately affects students' access to postsecondary education. Any effects on either the extensive or intensive margin of college attendance may have long-term impacts on students' earnings and other life outcomes.[2] Admissions practices may also affect students' incentives for which high schools to attend, with both partial and general equilibrium impacts on students.[3] While these latter topics fall outside the scope of the current piece, they provide reasons to study the topic at hand and gain a more complete understanding of how high schools affect postsecondary education access.

The rest of the paper proceeds as follows. Section 2 reviews background and related literature. Section 3 provides a conceptual and empirical framework. Section 4 describes the administrative data used. Section 5 presents empirical results. Section 6 concludes and offers additional discussion.

high school.

[2]The evidence on returns to college brand name or selectivity is slightly mixed, but more commonly indicates an effect than no effect. Brewer, Eide and Ehrenberg (1999), Behrman, Rosenzweig and Taubman (1996), Hoekstra (2009), and Black and Smith (2006) are among the papers that find significant effects, while Dale and Krueger (2002; 2011) have two notable exceptions that find no significant effect.

[3]Evidence from Texas suggests that its top 10% policy — in which students ranking in the top 10% of their class are guaranteed admission to any state university — induced some students to attend lower-achieving schools (Cullen, Long and Reback, 2013; Cortes and Friedson, 2014). This behavior may have mixed impacts on other students. Students who "trade down" may provide a positive peer externality on raw achievement, but may also confer a negative class rank externality by displacing incumbent students from the top 10%.

2 Background and Related Literature

In the broadest sense, this paper fits within the overall study of steps colleges take to increase access for students from disadvantaged groups. Such efforts have been common in the U.S. since at least the 1960s, and in recent decades have expanded to include a widening set of personal background factors. This can be partly seen in Long (2007). Beginning in the 1990s, several states began moving away from (and even banning) traditional race-based affirmative action at public universities. Subsequently, many of these states began considering a broader set of contextual factors, such as family income, the high school setting, and other aspects of hardship.

On the role of the high school, there currently exist two segments of the literature that explore how high schools affect access to postsecondary education. Each of these literatures, however, approaches the issue of access from a slightly different angle than the current piece. As such, neither literature is entirely parallel, nor provides definitive answers to the key question of this paper.

In one literature, a handful of papers examine the relationship between a high school's overall achievement and college attendance. From this group of studies, there is evidence that attending a higher-achieving school increases rates of college attendance (Deming et al., 2014; Berkowitz and Hoekstra, 2011), and even the selectivity of the college attended (Berkowitz and Hoekstra, 2011). These analyses, however, are generally interested in ex ante effects — comparing students who are identical *before* entering high school. Intermediate effects of the high school upon other aspects of a student's record — including the decision of whether to apply to college — are thus included as part of the overall effect. So even as these studies have found beneficial effects from attending a higher-achieving school, it is difficult to disentangle how colleges themselves respond to school quality (once students apply) from the effects of these other intermediate outcomes.

A second literature focuses on how a school's achievement affects admissions *via class rank*. To the extent that peer achievement matters greatly for class rank, it is sometimes argued that class rank preferences implicitly favor students from lower-achieving schools.[4] Such arguments are often raised in the context of percent plans, in which students ranking in the top $x\%$ of their high school class are guaranteed admission to at least one state university.[5]

By construction, these plans offer at least one form of access to students from disadvantaged schools. However, in practice they may have only limited reach. An important point is that their

[4] The influence of peer achievement on class rank is sometimes called the "frog pond" effect. Espenshade, Hale and Chung (2005) study this phenomenon in the context of college admissions. There is also evidence from lotteries that attending a higher-achieving school may lower class rank. See, for example, Cullen, Jacob and Levitt (2006).

[5] Percent plans began in the late 1990s, with Texas, Florida, and California being the most prominent states with such a plan in place. Most academic studies of these plans have looked at Texas, where the percent plan applies at individual campuses. This, however, is often not the case in other settings. It is thus somewhat unclear how the results from Texas generalize to other settings.

net effects depend on how students inside the top $x\%$ fare *without* the percent plan in place. In the case of Texas, where the percent threshold occurs at the top 10%, students within the top 10% had already been admitted with *near* certainty prior to the percent plan (Tienda et al., 2003), even to flagship campuses. Conditional on applying, admission outcomes typically remained unchanged for students inside these plans' direct reach.[6]

With this being the case, there is no obvious reason why school quality cannot enter more directly into the admissions process. In fact, both the Texas Education Code and the University of California's list of admissions criteria indicate that this sometimes occurs even in the presence of a percent plan. What remains missing is that there has been little formal study of this more direct usage. Thus, the overall ways in which high schools affect access to colleges are not fully understood.

2.1 UC-Specific Background

For the University of California specifically, Figures 1 and 2 provide some additional insights.

Figure 1 shows a list of 14 admissions criteria that all UC campuses use (University of California, 2012), with the aim of evaluating students' merits relative to the context of their life circumstances. Most relevant for this paper are items 5, 7, and 13, which involve students' high schools.

Item 5 pertains to Eligibility in the Local Context (ELC), or the UC's system-level percent plan. Beginning with the Fall 2001 entering cohort, originally students ranking in the top 4% of their high school class were guaranteed admission to at least one UC campus, though not a campus of the student's choosing.[7] Compared to Texas, the guaranteed admission is considerably weaker, as it does not apply to individual campuses. Nevertheless, it offers one form of access in a way that takes the high school into account.

Even with ELC, however, items 7 and 13 hint at further consideration of students' high schools. Item 7 calls for students' academic performance to be evaluated (at least in part) relative to the peers at their school, while item 13 lists a "disadvantaged social or educational environment" as a specific hardship for UC colleges to take into account. Item 13 in particular hints that there may be a direct bonus associated with attending a lower-achieving school, ceteris paribus. At least two examples can be found in which this was the case; for many years during the 2000s, both UC Davis and UC San Diego published detailed admission rubrics on their websites, indicating such a bonus.

[6]It is still possible that the policy may have affected access by inducing changes to application behavior. The total number of applicant-sending high schools increased after the top 10% policy (Long, Saenz and Tienda, 2010), and top 10% students appeared to focus their applications more heavily toward their preferred campuses (Long, Saenz and Tienda, 2010; Long and Tienda, 2010). These channels, however, are considered "general equilibrium" from the perspective of the current piece, which focuses only on the admissions stage, and takes applicant behavior as given.

[7]The eventual campus of admission is not required to be among those to which students originally applied. In this way, students cannot functionally select the campus of admission by only applying to their preferred campus.

UC San Diego's rubric is displayed in Figure 2.[8] Up through Fall 2011 admissions, UC San Diego used this formula to score students on an overall level of quality, taking into account both individual achievement and background/contextual factors.[9] Applicants whose overall score exceeded some threshold were subsequently offered admission.

Looking at the formula, bonuses can be seen for various forms of hardship or disadvantage. These include socioeconomic factors, such as low family income or being in the first generation to attend college, but also (separately) the achievement level of a student's school. Within step II, one sees a bonus of 300 points — worth 300 points of combined SAT score or 0.3 points of high school GPA — given to students from the bottom 40% of California public high schools. Furthermore, this bonus was not mutually exclusive from the consideration of ELC. While ELC status did not guarantee admission (to the specific campus), ELC received its own bonus that was also worth 300 points.

3 Conceptual and Empirical Framework

Using the UC San Diego example as motivation, consider a stylized model of college admissions. Suppose that, at a given college, students are rated on an overall level of desirability (A_i^*) based on the contents of their application. Applicants whose desirability exceeds some cutoff c are then offered admission. This can be represented with the following system:

$$A_i^* = \beta Q_i + x_{1i}'\gamma_1 + x_{2i}'\gamma_2 + u_i \tag{1}$$

$$Admit_i = 1\{A_i^* > c\} \tag{2}$$

where i is an index for students who applied, Q_i denotes the average achievement level or "quality" of a student's high school, x_1 and x_2 denote all other items contained in the application, and u_i denotes an idiosyncratic error, perhaps from the specific admissions officers assigned to each student. Let x_1 contain items of the application that are also available to the researcher (in a dataset), and x_2 contain the remaining items that are missing from the data.

β represents the consideration given to school quality, on the margin, as its own piece of information. Holding fixed all other aspects of the application, β is the degree to which a students is viewed more or less favorably when coming from a slightly higher-achieving school. The goal of this paper is to estimate $\hat{\beta}$ and its corresponding average marginal effect on admission probability across a set of institutions.

[8] The particular screenshots displayed were retrieved July 2, 2015 using an archived version of UC San Diego's admissions website. The source URL is http://web.archive.org/web/20041205230635/http://admissions.ucsd.edu/dev3/info/comreview.html.

[9] UC Davis' formula was similar, but with slightly different numbers. In recent years, many UC campuses have begun switching to a system of 'holistic review,' which no longer uses a set formula, but similarly tries to combine the many aspects of a student's application into a single review score.

For the case where u_i is normally distributed with unit variance, a student's probability of admission, conditional on applying, is $\Phi(\beta Q_i + x'_{1i}\gamma_1 + x'_{2i}\gamma_2)$. With access to the same information used in admissions, β, γ_1, and γ_2 can be consistently estimated using a probit regression, recovering the complete admissions rule. The fact that students sort nonrandomly to high schools does not otherwise prevent identification.

In practice, a challenge arises from the fact that datasets are typically incomplete. That is, they do not contain all items observed during admissions. Some aspects of the application, such as the transcript or personal statements, have many nuances that are difficult to fully capture. When only a portion of this information is available, a practical version of Equation 1 can be written as

$$A_i^* = \beta Q_i + x'_{1i}\gamma_1 + \eta_i, \tag{1'}$$

where $\eta_i = x'_{2i}\gamma_2 + u_i$ makes up a composite error term from the researcher's perspective. When the researcher now estimates

$$P(Admit_i = 1) = \Phi(\beta Q_i + x'_{1i}\gamma_1), \tag{3}$$

$\hat{\beta}$ and its corresponding marginal effect may be biased in either direction from the truth, depending on what items are in x_2 and how they correlate with Q and x_1. Omitted variables bias is thus a principal threat to identification.

The next section gives a more specific sense of the degree to which omitted variables are a concern, describing the administrative data and what variables are present versus missing.

4 Data

This paper uses administrative data on freshman undergraduate applicants to the University of California between 2001 and 2006. The data are gathered by the UC Office of the President (UCOP), which maintains a master database of all applicants to the UC system. In total, 460,704 students appear during these years applying to eight different UC campuses.[10]

The dataset includes individual-level data on students' UC applications and admission outcomes, high school GPA, SAT scores, ELC status, race, and parental income and education. In addition, there is also a summary measure of achievement for the high school attended, known as the Academic Performance Index (API). This measure is recorded at decile level in the data, with the highest-achieving schools making up the 10th decile. API decile is subsequently used as the value of Q_i when estimating Equation 3.

[10]Overall there are nine UC campuses with undergraduates. The current analysis omits UC Merced, which had only two admission cohorts during this time frame (Fall 2005 and 2006).

Prior to decile conversion, API is originally computed using each school's average performance on state standardized tests.[11] Specifically, each year a school's average test scores are mapped to a numeric value between 200 and 1000. These values are computed annually by the California Department of Education, and within California are often used to describe school-level achievement. API is not computed for private or out-of-state schools, due to the measure's origin from the California Public Schools Accountability Act of 1999. Moreover, students from those schools may also face a different set of admissions criteria. To keep things cohesive, all subsequent analysis restricts attention to the 320,721 applicants (about 70% of the total) from California public high schools.

Table 1 summarizes the data for this group of students, with additional information in Tables A2 and A3 of the appendix.

A few key patterns can be seen across API deciles. First, the vast majority of applicants come from the state's higher-achieving schools.[12] About 41 percent of applicants come from API deciles 9–10, or the top 20% of public high schools. Only 10 percent of applicants come from the bottom two API deciles, with an additional 13 percent from deciles 3–4. Overall, this translates to an average API decile of about 7.0 among applicants, and between 6.6 and 7.4 at each campus.

Second, API decile shows a pronounced positive correlation with socioeconomic status. Whether looking at race, parental income, or parental education, applicants from higher API deciles tend to come from more advantaged backgrounds.

Third, students' past academic achievements vary systematically between API deciles. Students from higher API deciles have higher average SAT scores and (to a lesser extent) high school GPA, while also taking more Advanced Placement courses on average. Yet the students from *lower* API deciles are somewhat more positively selected from within their schools. Looking at the rates of ELC eligibility, about 30 percent of applicants from the lowest-achieving schools (API deciles 1–2) rank within the top 4% of their class, compared to 12 percent from the highest-achieving schools.

Having summarized the dataset as a whole, it is worth noting a few limitations about the UCOP data. While the dataset contains many of the most essential determinants of admissions, there are undoubtedly many additional items in students' applications that are not present. For instance, there is no information about personal statements, extracurricular achievements, the exact transcript, or class rank (other than ELC status). Admission effects stemming from these items cannot be directly controlled during estimation, and become possible sources of omitted variables

[11] Until 2014, API used results of the Standardized Testing and Reporting (STAR) Program and the California High School Exit Exam. STAR was then replaced by a different testing program, though that does not affect the admission years in this paper. For high school students, STAR covered English-language arts, mathematics, history, and science. Additional details about API can be found in California Department California Department of Education (2012).

[12] This can be seen visually in a histogram in Figure 3.

bias.

Another limitation is that, aside from API decile, no additional information is provided about students' high schools. School name and location are not stated, nor are schools given unique identifiers. (These and other steps have been taken to guard student anonymity.) This prevents school fixed effects from being used during estimation, and also prevents instrumenting for API decile, since most instruments require knowledge about location.

A final limitation is that, for anonymity purposes, some variables are blurred from their exact values and are instead reported within a bin range. The most notable of these are high school GPA and SAT scores. Much of this issue can ultimately be resolved using an additional measure of student achievement, known as the Index score. Section A.1 of the appendix discusses this measure in more detail.

5 Results

Tables 2–4 show the marginal effects from estimating Equation 3 using various specifications. Before presenting these results, I note that standard errors may be somewhat understated in these tables, due to an inability to cluster error terms at the desired level.[13] Appendix A.2 discusses this issue in more detail, and computes the degree of within-cluster error correlation needed to make results insignificant.

Table 2 shows results with varying sets of controls. The top panel shows results with all campuses pooled, while the bottom panel shows the estimated marginal effect of API decile at each campus. For brevity, I primarily comment on the pooled marginal effect estimate, but these patterns are also roughly true at individual campuses.

Pooled, the estimated marginal effect appears as large as -0.021 when only controlling for the most essential student achievement measures and year. That is, attending a higher-achieving school by one decile associates with a 2.1 percentage point decrease in admission probability in column 1. Additional academic controls other than ELC (column 2) do not affect this estimate by much.

Column 3 next adds controls for race, parental income, and parental education — all of which are strongly correlated with API decile, but also bring many advantages/disadvantages of their own.[14] By all indications, UC colleges actively take these contextual factors into account. Once controlling for these demographics, the estimated marginal effect of API decile shrinks noticeably

[13]Intuitively, error terms should be grouped at the high school level due to various similarities in experience between students of the same school. This clustering cannot currently be achieved, since individual high schools are not indicated in the data.

[14]Under California Proposition 209, race cannot be explicitly considered during UC admissions. However, race is ultimately included in these regressions since it may affect other aspects of a student's experience that do enter into admissions.

to -0.013, but remains significant.

Column 4 adds dummies for whether students applied to each specific UC campus, and the mean SAT score of attending students among applied campuses. Borrowing somewhat from Dale and Krueger (2002; 2011), students' decisions about which colleges to apply to contain important information about their overall quality — in many cases incorporating information from items in x_2 (e.g. extracurriculars) that colleges see but are not in the data. These additional variables help control for otherwise-unobserved aspects of student quality. While these variables themselves are usually significant, suggesting they do convey additional information about students, the average marginal effect for API decile remains mostly unchanged at -0.012.

Finally, column 5 adds Eligibility in the Local Context, or a dummy for ranking within the top 4% of one's high school class. This guarantees eligible students admission to at least one UC campus, but not a campus of their choosing.

While ELC is not legally binding for individual campuses, most UC campuses still appear to respond to ELC status. Across all campuses, ELC is associated with an 11 percentage point increase in admission probability, on average, even after controlling for various other measures of achievement. This has important implications on access for students from lower-achieving schools. Since many applicants from lower API deciles are ELC eligible (up to 30 percent from the lowest-achieving schools), the consideration toward ELC also provides a degree of access to students from these schools. Yet this does not appear to be a complete substitute for more direct use of school quality. After controlling for ELC in column 5 of Table 2, the marginal effect of API decile shrinks considerably at some campuses, and by about a third in the pooled case compared to column 4. But the pooled marginal effect remains significantly negative, suggesting some additional direct consideration. The final pooled estimate of -0.8 percentage points is, on average, equivalent to about a 3.5 point decrease in Index score, or a 0.04 decrease in high school GPA.

Turning attention to individual campuses, Table 3 shows results for each campus using the full set of controls. In this table and others, campuses are ordered by average SAT score among attending students, so that campuses with similar selectivity appear adjacent to one another.

Campus-specific marginal effects range from $+0.002$ to -0.024, and are significantly negative at five out of eight campuses — including five of the six most selective. While the estimated marginal effect is nominally positive and significant at the remaining three campuses (Irvine, Santa Cruz, and Riverside), it is worth noting that none of these campuses has an estimated marginal effect of more than $+0.002$, or 0.2 percentage points. Moreover, analysis of standard errors (see Appendix A.2) suggests that the positive significance disappears at these campuses under corrected standard errors clustered by high school. Magnitudes are much more pronounced at the five campuses where the marginal effect appears negative.

5.1 Nonlinearity Across API Deciles

Table 3 reports the average marginal effect at each campus as a single parameter. Yet it is possible that the marginal effect is nonlinear across API deciles. For instance, some campuses may view the difference between a 2nd and 3rd decile school (e.g.) to be more or less meaningful than the difference between a 9th and 10th decile school.

To allow for nonlinearity, I also estimate each admissions rule using a specification that includes each API decile individually. These results are displayed in Table 4. For all campuses, the 7th API decile (roughly the average among applicants) is omitted as a baseline. Table cells report associated differences relative to the 7th API decile, while marginal effects of a one-decile change in API can be computed by comparing adjacent estimates.

The main patterns from this specification can be seen in Figure 4. The marginal change in admission probability appears to be nonlinear and generally largest at lower-achieving high schools. While the exact marginal effect varies from decile to decile, and across campuses, results that include API decile using a spline function (see Table A5) indicate that the average marginal effect is generally around -1.5 percentage points in the bottom half of the API distribution, and fairly close to zero in the upper half.

Compared to a counterfactual where all students are treated as coming from the 7th API decile (roughly the average among applicants at each campus), these results suggest that, for an average UC campus, school quality can contribute up to a 7.5 percentage point increase in admission probability for students from the lowest-achieving schools. Relative to those students' counterfactual probability of admission (41.1% across all campuses), this represents an 18 percent proportional increase.[15]

For students from the highest-achieving schools, meanwhile, the decrease in their admission probability appears to be about 0.6 percentage points (or less) at an average campus. This translates to about a 1 percent proportional decrease, relative to those students' counterfactual probability of admission.[16] Thus, whether measured by raw or proportional change, the admissions use of school quality seems to affect students in the lower API deciles much more than students in the upper API deciles.

5.2 Enrollment Composition

Using these results, it is possible to construct back-of-the-envelope estimates of how the (apparent) use of school quality affects the resulting pool of admitted and attending students. These results

[15] Here, the counterfactual probability of admission is computed supposing all admission rules are as estimated in Table 4, but that all students are evaluated as coming from the 7th API decile.

[16] Computing counterfactual admission probability the same way as before, these students have a counterfactual probability of admission of 53.2% across all campuses to which they apply.

are displayed in Tables 5 and 6.

Table 5 first shows the actual, predicted, and counterfactual composition of admitted students by API decile. In all cases, the predicted probability of admission is the probability implied by the regression in Table 4, while the counterfactual probability is computed supposing that all admission rules are as estimated in Table 4, but that all students are treated as coming from the 7th API decile.[17] This generates a predicted and counterfactual probability of admission for each student at each campus to which he/she applies. Summing these probabilities within each API decile and within each campus gives the predicted and counterfactual compositions displayed in Table 5.

For enrollment composition, I next combine a student's probability of admission with the probability of enrollment *conditional on admission*, at a given campus, to estimate the overall probability that each student enrolls. Mathematically,

$$\hat{E}(Enroll_i) = \hat{P}(Admit_i = 1) \cdot \hat{P}(Enroll_i = 1 | Admit_i = 1), \quad (4)$$

with the counterfactual version replacing $\hat{P}(Admit_i = 1)$ with $\hat{P}(Admit_i = 1)_{counter} = \Phi(x_{1i}'\hat{\gamma}_1)$. To obtain $\hat{P}(Enroll_i = 1|Admit_i = 1)$, I estimate a probit regression of enrollment among the admitted students at each campus, using the same explanatory variables as when modeling admissions in Table 4. These conditional enrollment patterns are then assumed true of all applicants to that campus, including those who were not admitted.[18] Summing the resulting values of $\hat{E}(Enroll_i = 1)$ within each API decile and within each campus gives the predicted and counterfactual compositions displayed in Table 6.

Looking at the numbers in Tables 5 and 6, one sees that the predicted admission and enrollment shares (Panel B) generally come quite close to matching their actual compositions. The most common error is that the model sometimes overestimates the share of attending or admitted students from the very top API deciles. Correspondingly, comparisons between the actual and counterfactual shares in Tables 5 and 6 may somewhat overstate the impact on students in the upper API deciles.

Looking first at the shares of admits, between 2001 and 2006, about 27.0 percent of admits came from API deciles 1–5, or high schools below the California median. Under the constructed counterfactual, this share may have been only 25.0 percent if not for the admissions consideration given to school quality. At least according to this exercise, this admissions practice increases the

[17] To be precise, the counterfactual probability of admission replaces each student's API decile with 7 and recomputes the probability of admission given by the probit function. Mathematically, $\hat{P}(Admit_i = 1)_{counter} = \Phi(x_{1i}'\hat{\gamma}_1)$, where the estimated effect due to API decile has been nullified. Depending on students' proximity to the admission margin, this may affect some students' admission probability more than others, even within the same API decile and campus.

[18] Table A6 displays the results of these conditional enrollment or "yield" equations. The yield equations can only be estimated on the pool of admitted students, since the researcher never observes whether students who were not admitted would have enrolled if given the chance.

share of admits from API deciles 1–5 by about 8 percent (25.0 to 27.0 percent), while decreasing the share of admits from deciles 6–10 by about 3 percent (75.0 to 73.0 percent). As with the estimated marginal effects earlier, the biggest gains (in a proportional sense) appear to accrue at the lowest API deciles. While only about 8.6 percent of admits came from the API deciles 1–2, this is about a 13 percent gain over the estimated counterfactual of 7.6 percent.

Once factoring in students' predicted probability of enrollment, the effects described earlier become somewhat more pronounced among students who ultimately attend. According to Table 6, about 28.2 percent of attending students between 2001 and 2006 came from the API deciles 1–5. This is about 14 percent higher than the estimated counterfactual of 24.7 percent. For students from the very lowest-achieving schools (API deciles 1–2), the proportional increase is about 26 percent (7.0 to 8.8 percent). Meanwhile, the decrease in the share of attending students from API deciles 6–10 is about 5 percent (75.3 to 71.8 percent).

5.3 Possible Explanations: Academic Ability Versus Other Factors

Preferences for students from lower-achieving high schools, ceteris paribus, may be driven by either academic or non-academic reasons. On one hand, many colleges have preferences for a diverse student body. But school quality may also convey additional information about students' ability, (e.g.) by helping to put students' past achievement into context.

Using college performance measures from the UCOP data, it may be possible to test which of these explanations has more empirical support. If the admissions behavior is driven primarily by academic considerations, one generally expects to see school quality predict college GPA in the same direction as its admissions use. To see whether this is the case, Table 7 regresses students' cumulative college GPA on API decile and other observables from the admissions equation.[19]

The top panel shows results using ordinary least squares (OLS). Here, although API decile is often a negative predictor of admission probability, it positively predicts college GPA at each campus. This is consistent with a number of other studies of how high school characteristics predict college grades.[20]

However, selection may play an important role in these OLS results. College GPA, at a given campus, is only observed for students who ultimately enroll — itself a subset of students who are first admitted. College GPA results obtained on this subsample may not be true of the original pool of applicants. Making matters worse, the ways in which students are admitted and enroll may

[19] I use cumulative college GPA, rather than first-year GPA, in case the influence of high schools upon college performance weakens over time. However, results using first-year college GPA are very similar to those in Table 7. Results using bachelor's degree completion as the outcome are included in the appendix.

[20] Some examples include Betts and Morell (1999), Black et al. (2015), and Rothstein (2004). These papers often do not include a school's overall achievement directly as a regressor, but their results are broadly consistent with a positive relationship.

itself create correlations between API decile and unobservables, among enrollees, even if no such correlation existed among all applicants.[21]

To try to correct for selection, panels B and C estimate the college GPA equation using a Heckman (1979) correction. The two panels differ in how they model the selection equation. Panel B uses no exclusion restrictions, relying only the nonlinearity of the inverse Mills ratio for identification, while panel C uses admission offers to other UC campuses as selection shifters.

Results in panel B offer somewhat less evidence of API decile being a positive predictor of college GPA, with significance disappearing at four campuses, and the pooled coefficient decreasing by about 60 percent compared to OLS. Yet no campuses show API decile to be a negative predictor of college GPA. Results in panel C, meanwhile, are mostly similar to those obtained using OLS.

These Heckman (1979) specifications have their own challenges to contend with, and ultimately may not fully account for selection.[22] However, all three of the current specifications suggest API decile to be either a neutral or positive predictor of college GPA, ceteris paribus. At least in this sense, it appears the consideration given to school quality during admissions may be motivated more by non-academic considerations. Increasing campus diversity is one possible explanation, though it is also possible that students from lower API deciles are positively selected on leadership, community involvement, or other non-academic contributions — with these benefits outweighing any costs to academic performance from the colleges' perspective.

6 Discussion

This paper examines the direct use of a school's overall achievement or "quality" in college admissions decisions. Though this is potentially a common practice, it so far has received little attention in the literature, and its magnitude is largely unknown. This paper provides an initial attempt to estimate this magnitude across a set of institutions, using administrative data from the University of California.

[21] Cushing and McGarvey (2004) consider the case where students must be above some quality threshold to gain admission, but only students *below* some other quality threshold enroll. This may arise if students apply to many colleges, with differing selectivity, and then enroll at the "best" college to which they are admitted. In such a setting, enrolling students who are positively selected on a particular trait are generally *negatively* selected on unobservables, and vice versa. The end result is that coefficients in OLS regressions of college GPA will be biased in the opposite direction from their true sign, or "toward zero."

Such a result may apply here, as the general setup of Cushing and McGarvey (2004) appears broadly true for UC colleges. Once students are admitted, nearly all traits that positively predict admissions negatively predict enrollment, and vice versa. Students with higher predicted probabilities of admission generally have lower (conditional) rates of enrollment.

[22] In the first case (no exclusion restrictions), the inverse Mills ratio remains *close* to a linear combination of other observables. Identification is often tenuous in such cases (Puhani, 2000). In the second case, although admission offers to other UCs are predictive of enrollment, they also correlate with unobserved student ability, or part of the error term in the college GPA equation.

Controlling for students' high school GPA, SAT scores, race, socioeconomic status, and ELC status, I find a school's overall achievement (measured by API decile) to be negatively associated with probability of admission at five out of eight UC campuses between 2001 and 2006, including five of the six most selective campuses. Averaging across all campuses, a one-decile increase in API is associated with about a 0.8 percentage point decrease in admission probability. This marginal effect appears to be larger at the bottom end of the API distribution, perhaps closer to -1.5 percentage points. Back-of-the-envelope calculations suggest that this admissions practice increases the share of attending students from schools below the California median high school by 14 percent, while decreasing the share from above the median high school by 5 percent.

Due to an inability to observe all items used during admissions, I cannot rule out the possibility of omitted variables bias. These results are thus presented as more of a starting point than a definitive answer for the effect of interest.

This paper is otherwise among the first to consider this particular aspect of how high schools affect access to postsecondary education — an area that is not fully understood. Future work may further explore this area and build upon these results. For instance, even with the current results, it is unclear what overall share of colleges uses a similar admissions practice, and whether there exist systematic differences by region, by selectivity, between public and private colleges, or other dimensions.

Future work may also estimate results from an ex ante perspective. For many parents and policymakers, the greater interest may be in how school quality affects admissions/enrollment *when families still have a chance to move.* While there exists some evidence of positive ex ante effects (Deming et al., 2014; Berkowitz and Hoekstra, 2011; Dynarski et al., 2013), and general findings of strong demand for higher-achieving schools (Black, 1999; Rothstein, 2006), there also appear to be instances of families "trading down" under specific circumstances (Cullen et al., 2013; Cortes and Friedson, 2014). Incentives, as well as families' *perceptions* of incentives (perhaps influenced by salience), may both affect school choice responses in practice. Other authors may investigate these and other possible behavioral responses, and the resulting impacts on families.

References

Behrman, Jere R., Mark R. Rosenzweig, and Paul Taubman. 1996. "College Choice and Wages: Estimates Using Data on Female Twins." *Review of Economics and Statistics*, 78(4): 672–685.

Berkowitz, Daniel, and Mark Hoekstra. 2011. "Does high school quality matter? Evidence from admissions data." *Economics of Education Review*, 30(2): 280–288.

Betts, Julian R., and Darlen Morell. 1999. "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects." *Journal of Human Resources*, 34(2): 268–293.

Black, Dan A., and Jeffrey A. Smith. 2006. "Estimating the Returns to College Quality with Multiple Proxies for Quality." *Journal of Labor Economics*, 24(3): 701–728.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics*, 114(2): 577–599.

Black, Sandra E., Jane Arnold Lincove, Jenna Cullinane, and Rachel Veron. 2015. "Can You Leave High School Behind?" *Economics of Education Review*, 46 52–63.

Brewer, Dominic J., Eric R. Eide, and Ronald G. Ehrenberg. 1999. "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings." *Journal of Human Resources*, 34(1): 104–123.

California Department of Education. 2012. "2011-12 Academic Performance Index Reports Information Guide." Retrieved January 3, 2013, from http://www.cde.ca.gov/ta/ac/ap/documents/infoguide12.pdf.

Cortes, Kalena E., and Andrew I. Friedson. 2014. "Ranking Up by Moving Out: The Effect of the Texas Top 10% Plan on Property Values." *National Tax Journal*, 67(1): 51–76.

Cullen, Julie Berry, Brian A. Jacob, and Steven D. Levitt. 2006. "The effect of school choice on participants: evidence from randomized lotteries." *Econometrica*, 74(5): 1191–1230.

Cullen, Julie Berry, Mark C. Long, and Randall Reback. 2013. "Jockeying for position: strategic high school choice under Texas' top ten percent plan." *Journal of Public Economics*, 97 32–48.

Cushing, Matthew J., and Mary G. McGarvey. 2004. "Sample Selection in Models of Academic Performance." *Economic Inquiry*, 42(2): 319–322.

Dale, Stacy Berg, and Alan B. Krueger. 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, Stacy, and Alan B. Krueger. 2011. "Estimating the Return to College Selectivity over the Career Using Administrative Earning Data." NBER Working Paper 17159.

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2014. "School Choice, School Quality, and Postsecondary Attainment." *American Economic Review*, 104(3): 991–1013.

Dynarski, Susan, Joshua Hyman, and Diane Whitmore Schanzenbach. 2013. "Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion." *Journal of Policy Analysis and Management*, 32(34): 692–717.

Espenshade, Thomas J., Lauren E. Hale, and Chang Y. Chung. 2005. "The Frog Pond Revisited: High School Academic Context, Class Rank, and Elite College Admission." *Sociology of Education*, 78(4): 269–293.

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica*, 47(1): 153–161.

Hoekstra, Mark. 2009. "The effect of attending the flagship state university on earnings: A discontinuity-based approach." *Review of Economics and Statistics*, 91(4): 717–724.

Long, Mark C. 2007. "Affirmative Action and Its Alternatives in Public Universities: What Do We Know?" *Public Administration Review*, 67(2) 315–330.

Long, Mark C., Victor Saenz, and Marta Tienda. 2010. "Policy Transparency and College Enrollment: Did the Texas Top Ten Percent Law Broaden Access to the Public Flagships?" *ANNALS of the American Academy of Political and Social Science*, 627(1): 82–105.

Long, Mark C., and Marta Tienda. 2010. "Changes in Texas universities' applicant pools after the *Hopwood* decision." *Social Science Research*, 39(1): 48–66.

Puhani, Patrick A. 2000. "The Heckman correction for sample selection and its critique." *Journal of Economic Surveys*, 14(1): 1467–6419.

Rothstein, Jesse M. 2004. "College Performance Predictions and the SAT." *Journal of Econometrics*, 121(1-2): 297–317.

Rothstein, Jesse M. 2006. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *American Economic Review*, 96(4): 1333–1350.

Tienda, Marta, Kevin Leicht, Terry Sullivan, Michael Maltese, and Kim Lloyd. 2003. "Closing the Gap? Admissions & Enrollments at the Texas Public Flagships Before and After Affirmative Action." Unpublished.

7 Tables and Figures

1. Academic Grade Point Average (GPA) calculated on all academic courses completed in the subject areas specified by the University's admission requirements (the a-g* subjects), including additional points for completion of University certified honors courses (see 4, below). It is recommended that the maximum value allowed for the GPA shall be 4.0. *An additional subject requirement, labeled "g" Visual & Performing Arts was added in 2002.

2. Scores on the following tests: the Scholastic Assessment Test I or the American College Test. Effective with applicants applying for Fall 2012 and thereafter, the SAT Subject Tests are not required but may be recommended by some programs on some campuses.

3. The number, content of, and performance in courses completed in academic subjects beyond the minimum specified by the University's eligibility requirements.

4. The number of and performance in University approved honors courses, College Board Advanced Placement courses, International Baccalaureate courses, and transferable college courses completed. It is recommended that caution be exercised in order not to assign excessive weight to these courses, especially if considerable weight already has been given in the context of 1, above. Additionally, in recognition of existing differences in availability of these courses among high schools, it is recommended that reviewers assess completion of this coursework against the availability of these courses at the candidate's secondary school.

5. Being identified as eligible in the local context, by being ranked in the top 9% of the class at the end of the junior year, as determined by academic criteria established by the University of California.**Effective with applicants applying for Fall 2012 or thereafter, ELC increased from the top 4% to the top 9%.

6. The quality of the senior year program, as measured by type and number of academic courses (see 3 and 4, above) in progress or planned.

7. The quality of academic performance relative to the educational opportunities available in the applicant's secondary school.

8. Outstanding performance in one or more specific academic subject areas.

9. Outstanding work in one or more special projects in any academic field of study.

10. Recent, marked improvement in academic performance, as demonstrated by academic grade point average and quality of coursework (see 3 and 4, above) completed and in progress, with particular attention being given to the last two years of high school.

Figure 1: Admissions Criteria for University of California Colleges

11. Special talents, achievements, and awards in a particular field, such as in the visual and performing arts, in communication, or in athletic endeavors; special skills, such as demonstrated written and oral proficiency in other languages; special interests, such as intensive study and exploration of other cultures; or experiences that demonstrate unusual promise for leadership, such as significant community service or significant participation in student government; or other significant experiences or achievements that demonstrate the applicant's promise for contributing to the intellectual vitality of a campus.

12. Completion of special projects undertaken either in the context of the high school curriculum or in conjunction with special school events, projects or programs co-sponsored by the school, community organizations, postsecondary educational institutions, other agencies, or private firms, that offer significant evidence of an applicant's special effort and determination or that may indicate special suitability to an academic program on a specific campus.

13. Academic accomplishments in light of the applicant's life experiences and special circumstances. These experiences and circumstances may include, but are not limited to, disabilities, low family income, first generation to attend college, need to work, disadvantaged social or educational environment, difficult personal and family situations or circumstances, refugee status, or veteran status.

14. Location of the applicant's secondary school and residence. These factors shall be considered in order to provide for geographic diversity in the student population and also to account for the wide variety of educational environments existing in California.

Figure 1, continued

Selection Process

Drawing upon the broad guidelines set forth by the Board of Admissions & Relations with Schools (BOARS), the UCSD Faculty Committee on Undergraduate Admissions has approved the following procedures for freshman selection which will be implemented by the Office of Admissions & Relations with Schools.

<u>Step I - Academic Review</u>	Maximum Consideration
Uncapped Grade Point Average (GPA)	4,500
Scores of All Required Exams	3,200
Number of "A-G" Courses Beyond the Minimum	500
<u>Step II - Additional Academic Factors</u>	
Eligibility in the Local Context (ELC)	300
Educational Environment	300
<u>Step III - Socioeconomic Factors</u>	
Low Family Income	300
First Generation College Attendance	300
<u>Step IV – Personal Characteristics and Achievement Factors</u>	
Demonstrated Leadership	300
Special Talents/Achievements/Awards	300
Community and Volunteer Service	300
Participation in Pre-Collegiate/Motivational and Enrichment Programs	300
Special Circumstances/Personal Challenges	500

<u>Step V – Computing a Comprehensive Review Score</u>
Eligible applicants will be assigned a comprehensive review score by totaling the scores from each category listed in steps 1 through 4. Eligible applicants are then ranked based upon that assigned score. Applicants with the strongest combination of academic, personal characteristics and achievement factors will be admitted in sufficient numbers to meet campus enrollment goals.

Inset: Description of Educational Environment

Educational Environment
A disadvantaged educational/school environment indicates that the applicant attends a California high school that is among the 4th or 5th quintile of all California public high schools using the following academic indicators: high school completion rate, percentage of students enrolled in college preparation classes, percentage of students enrolled in Advanced Placement/Honors courses, percentage of students admitted to the UC/CSU, and the percentage of students taking the Scholastic Aptitude Test (SAT) or the American College Test (ACT).

Figure 2: UCSD Admissions Formula Used Through 2011

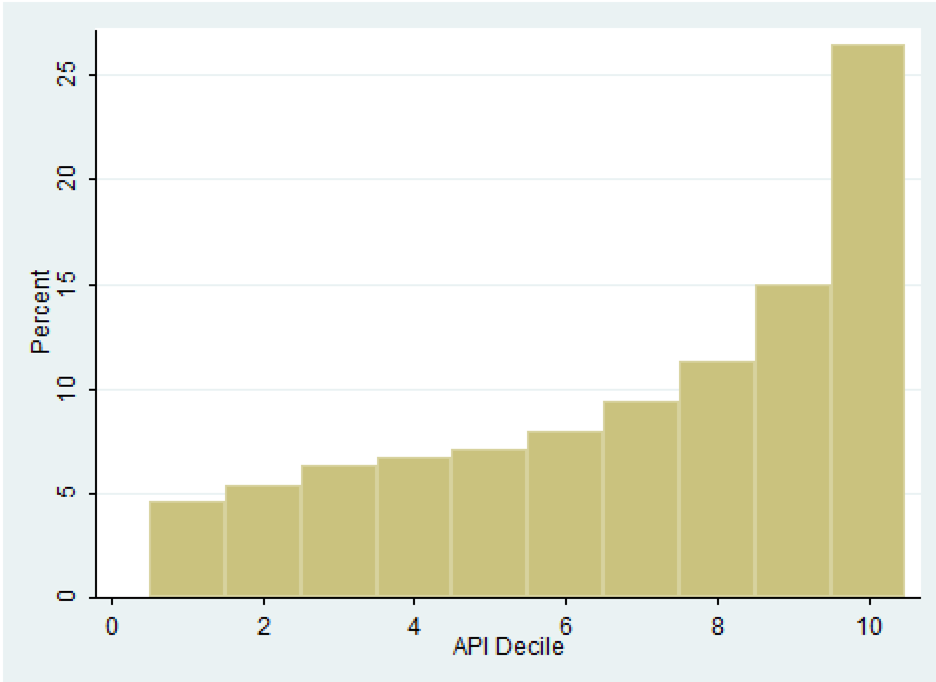

Figure 3: Histogram of API Decile Among Applicants

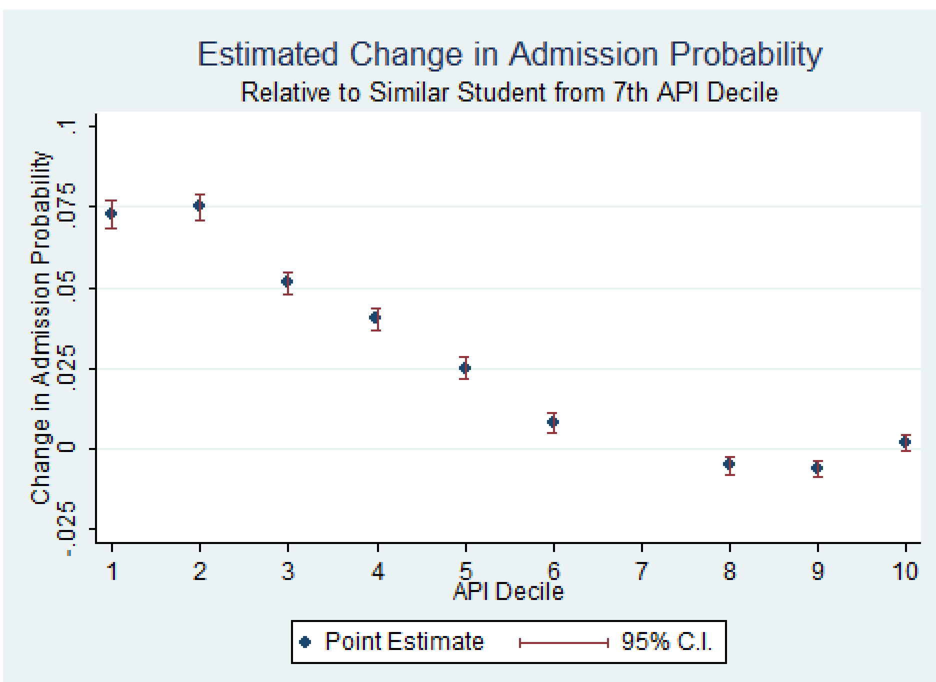

Figure 4: Graphical Depiction of Table 4 Results

Table 1: Descriptive Statistics

	API Deciles					
	1-2	3-4	5-6	7-8	9-10	All
N	31,847	41,695	47,952	66,255	132,972	320,721
Share of Applicants	0.10	0.13	0.15	0.21	0.41	1
<u>Student Achievement Measures</u>						
HS GPA	3.27	3.33	3.40	3.42	3.41	3.38
	(0.50)	(0.49)	(0.47)	(0.46)	(0.45)	(0.47)
SATI Total	982	1071	1132	1173	1235	1161
	(152)	(170)	(164)	(158)	(147)	(175)
AP Courses	6.65	7.07	7.08	7.26	7.39	7.20
	(4.98)	(5.44)	(5.28)	(5.45)	(5.28)	(5.31)
ELC[a]	0.30	0.27	0.25	0.20	0.12	0.19
<u>Socioeconomic Background</u>						
Parental Income Percentile	44.5	56.0	66.7	72.8	78.6	68.4
	(26.2)	(29.3)	(28.9)	(27.0)	(24.8)	(29.1)
Parental Education Percentile	32.3	49.6	61.5	69.6	77.8	65.4
	(29.9)	(32.0)	(29.8)	(26.6)	(22.5)	(30.4)
Underrepresented Minority	0.67	0.40	0.25	0.18	0.09	0.23
White	0.07	0.21	0.35	0.42	0.42	0.35
Asian	0.23	0.33	0.33	0.32	0.40	0.35
Other/Unknown Race	0.04	0.06	0.07	0.08	0.09	0.08
<u>Admission Rates</u>						
Admitted to Berkeley	0.27	0.26	0.26	0.26	0.27	0.27
Admitted to Los Angeles	0.24	0.25	0.24	0.24	0.29	0.26
Admitted to San Diego	0.42	0.45	0.45	0.43	0.45	0.44
Admitted to Irvine	0.44	0.50	0.57	0.59	0.64	0.58
Admitted to Santa Barbara	0.57	0.59	0.60	0.54	0.50	0.54
Admitted to Davis	0.66	0.65	0.62	0.60	0.64	0.63
Admitted to Santa Cruz	0.68	0.75	0.80	0.81	0.82	0.79
Admitted to Riverside	0.78	0.82	0.88	0.89	0.92	0.88

Notes: Standard deviations are in parentheses. Admission rates are conditional on applying.

a — 'ELC' is short for Eligibility in the Local Context, which denotes whether a student ranks in the top 4% of his/her high school class by GPA. These students are guaranteed admission to at least one UC campus, but not a campus of the student's choosing.

Table 2: Probit Average Marginal Effects with Varying Controls

Panel A: Pooled Regression

	(1)	(2)	(3)	(4)	(5)
API Decile (10=Highest-Achieving)	-0.021***	-0.020***	-0.013***	-0.012***	-0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ELC Status					0.110***
					(0.001)
Observations	1,195,656	1,195,656	1,195,656	1,195,656	1,195,656
Pseudo R-squared	0.477	0.498	0.510	0.520	0.527

Panel B: Marginal Effect of API Decile by Campus

	(1)	(2)	(3)	(4)	(5)
Berkeley	-0.020***	-0.019***	-0.013***	-0.013***	-0.009***
Los Angeles	-0.022***	-0.021***	-0.013***	-0.013***	-0.011***
San Diego	-0.029***	-0.029***	-0.018***	-0.017***	-0.011***
Irvine	-0.010***	-0.009***	-0.008***	-0.007***	0.001***
Santa Barbara	-0.039***	-0.038***	-0.028***	-0.026***	-0.024***
Davis	-0.025***	-0.025***	-0.014***	-0.012***	-0.008***
Santa Cruz	-0.005***	-0.004***	0.000	0.000	0.001*
Riverside	0.001***	0.001***	0.001***	0.002***	0.002***
Application Year (3-Year Bin)	Y	Y	Y	Y	Y
HSGPA and SATI[a]	Y	Y	Y	Y	Y
Other Academic Controls[b]		Y	Y	Y	Y
Race, Parental Income and Ed.			Y	Y	Y
UC Campuses Applied[c]				Y	Y

Notes: The dependent variable is admission, conditional on having applied. The pooled regression uses separate intercepts for each campus, but otherwise a single set of coefficients. To maintain the full sample across specifications, explanatory variables add an indicator for "missing value" when needed. Except for income percentile, this is typically only a small percentage of cases. Robust standard errors are in parentheses, clustered by student in the pooled regression. *** p<0.01, ** p<0.05, * p<0.1.

a – Includes dummies for bin ranges of high school GPA and SAT Math and Verbal scores, along with an additional measure known as the Index score, which is a linear combination of high school GPA and SAT scores. Index score is included using a cubic polynomial.
b – Includes SATII Writing and Third Subject scores, the number of Advanced Placement courses taken, and intended college major.
c – Includes dummies for whether a student applied to each UC campus, along with the mean SAT score of attending students among campuses applied.

Table 3: Probit Average Marginal Effects by Campus

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Decile	-0.009***	-0.011***	-0.011***	0.001***	-0.024***	-0.008***	0.001*	0.002***	-0.008***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0002)
ELC Status	0.092***	0.032***	0.156***	0.298***	0.131***	0.202***	0.063***	0.052***	0.110***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.005)	(0.007)	(0.004)	(0.001)
Observations	148,435	185,619	183,958	161,765	163,038	139,695	103,702	109,444	1,195,656
Pseudo R-squared	0.453	0.439	0.594	0.583	0.519	0.476	0.505	0.478	0.527
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517

Notes: The dependent variable is admission, conditional on having applied. All regressions include the full set of controls from Table 2. The pooled regression uses separate intercepts for each campus, but otherwise a single set of coefficients. Robust standard errors are in parentheses, clustered by student in the pooled regression. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: Probit Average Marginal Effects with Each API Decile

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
1st API Decile	0.094***	0.113***	0.069***	0.031***	0.110***	0.104***	0.007	-0.014***	0.073***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	(0.002)
2nd API Decile	0.076***	0.119***	0.081***	0.019***	0.134***	0.115***	0.009	-0.005	0.075***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.004)	(0.002)
3rd API Decile	0.037***	0.078***	0.069***	0.003	0.106***	0.083***	0.005	-0.011***	0.051***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.002)
4th API Decile	0.034***	0.049***	0.051***	-0.005	0.095***	0.059***	0.007	-0.006	0.040***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.002)
5th API Decile	0.019***	0.018***	0.033***	0.005	0.069***	0.030***	0.000	0.000	0.025***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.002)
6th API Decile	0.005	-0.001	0.011***	0.003	0.038***	0.003	0.001	0.005	0.008***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)
7th API Decile	—	—	—	—	—	—	—	—	—
8th API Decile	-0.007*	-0.006*	-0.009***	0.010***	-0.022***	-0.000	-0.002	-0.001	-0.005***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.001)
9th API Decile	-0.010***	-0.005	-0.015***	0.015***	-0.046***	-0.001	0.011***	0.009***	-0.006***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.001)
10th API Decile	-0.004	0.004	-0.010***	0.020***	-0.058***	0.030***	0.012***	0.005*	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.001)
ELC Status	0.094***	0.036***	0.155***	0.293***	0.132***	0.197***	0.062***	0.052***	0.110***
	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)	(0.005)	(0.007)	(0.004)	(0.001)
Observations	148,435	185,619	183,958	161,765	163,038	139,695	103,702	109,444	1,195,656
Pseudo R-squared	0.454	0.442	0.595	0.584	0.520	0.480	0.505	0.478	0.528
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517
P-value for Linearity[a]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000

Notes: The 7th API decile is omitted as a baseline. All regressions include the full set of controls from Table 2. Robust standard errors are in parentheses, clustered by student in the pooled regression. *** p<0.01, ** p<0.05, * p<0.1.

a — Reports the *p*-value on an *F* test that the differences between probit coefficients for consecutive API deciles are all equidistant.

Table 5: Actual, Predicted, and Counterfactual Admission by API Decile (Percentage Shares)

Panel A: Actual Admissions

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	4.6	4.4	3.0	3.5	4.2	3.6	3.7	5.0	3.9
2	4.9	5.4	4.2	4.5	5.0	4.2	3.8	6.2	4.7
3	5.5	6.0	5.5	5.4	5.8	5.2	4.9	7.0	5.7
4	6.2	6.1	6.1	5.7	6.4	6.0	5.5	7.2	6.2
5	6.3	5.6	6.5	6.3	6.9	6.5	6.1	7.3	6.5
6	7.2	6.1	7.4	7.0	7.8	8.3	7.9	7.4	7.5
7	8.6	7.7	8.9	8.6	9.1	8.5	9.6	9.2	8.9
8	10.4	9.6	10.8	10.6	11.2	10.7	12.2	10.6	10.8
9	14.5	14.6	15.2	16.5	14.9	13.8	15.2	15.0	15.0
10	31.9	34.4	32.3	31.9	28.7	33.2	31.2	25.2	30.7

Panel B: Predicted Admissions

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	4.6	4.3	3.0	3.5	4.2	3.6	3.7	4.9	3.9
2	4.8	5.3	4.2	4.5	4.9	4.2	3.7	6.2	4.7
3	5.4	6.0	5.5	5.3	5.8	5.1	4.9	7.0	5.7
4	6.2	6.1	6.1	5.7	6.4	6.0	5.4	7.2	6.2
5	6.3	5.7	6.5	6.3	6.8	6.5	6.1	7.3	6.5
6	7.2	6.1	7.4	7.0	7.8	8.3	7.9	7.4	7.5
7	8.6	7.8	8.9	8.6	9.2	8.5	9.6	9.2	8.9
8	10.4	9.6	10.8	10.6	11.2	10.8	12.2	10.6	10.9
9	14.5	14.6	15.2	16.5	14.9	13.8	15.2	15.1	15.1
10	31.9	34.4	32.3	32.0	28.8	33.2	31.2	25.2	30.8

Panel C: Counterfactual Admissions

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	3.0	2.7	2.6	3.4	3.3	3.1	3.7	5.1	3.4
2	3.6	3.3	3.5	4.4	3.8	3.6	3.7	6.2	4.2
3	4.8	4.6	4.8	5.4	4.7	4.7	4.9	7.1	5.2
4	5.6	5.3	5.6	5.9	5.4	5.7	5.4	7.2	5.8
5	6.1	5.6	6.1	6.4	6.0	6.5	6.1	7.3	6.4
6	7.4	6.7	7.3	7.1	7.3	8.6	7.9	7.3	7.5
7	9.0	8.4	9.1	8.8	9.2	8.9	9.7	9.2	9.1
8	11.1	10.7	11.3	10.7	11.7	11.3	12.3	10.6	11.2
9	15.7	16.1	16.1	16.5	16.3	14.5	15.2	14.9	15.6
10	33.7	36.6	33.7	31.6	32.2	33.2	31.0	25.1	31.6

Notes: Cells display the overall percentage across admission cohorts 2001 to 2006. Predicted admission probabilities are generated by the regression in Table 4. Counterfactual admission probabilities are estimated using those same regression results, but supposing all students are treated as coming from the 7th API decile.

Table 6: Actual, Predicted, and Counterfactual Enrollment by API Decile (Percentage Shares)

Panel A: Actual Enrollment

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	5.0	4.9	2.2	2.9	4.5	3.5	3.4	5.1	3.9
2	4.8	6.2	4.1	3.4	5.1	4.3	3.1	8.3	4.9
3	5.5	7.2	5.6	4.5	6.1	5.8	4.4	9.1	6.0
4	6.2	6.8	6.3	5.0	6.6	7.2	4.7	8.6	6.5
5	6.3	5.9	6.8	6.3	7.3	7.4	6.4	8.9	6.9
6	7.8	6.0	7.8	6.6	8.4	9.8	8.9	7.5	7.8
7	8.5	7.6	9.4	8.8	9.9	9.2	10.7	9.7	9.2
8	10.4	9.6	11.4	11.3	12.3	11.5	14.5	10.4	11.3
9	13.8	15.0	15.7	19.6	16.3	12.9	16.1	13.9	15.4
10	31.7	30.7	30.8	31.6	23.4	28.4	27.7	18.4	28.0

Panel B: Predicted Enrollment

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	5.1	4.9	2.2	2.8	4.4	3.4	3.4	5.1	3.9
2	4.8	6.1	4.0	3.4	5.0	4.3	3.1	8.2	4.9
3	5.5	7.1	5.5	4.5	6.1	5.7	4.3	9.0	6.0
4	6.2	6.6	6.2	4.9	6.6	7.1	4.8	8.7	6.4
5	6.3	5.8	6.7	6.3	7.2	7.3	6.5	9.0	6.9
6	7.8	5.9	7.6	6.6	8.5	9.8	8.9	7.6	7.8
7	8.5	7.6	9.4	8.9	10.1	9.2	10.7	9.8	9.2
8	10.3	9.6	11.4	11.3	12.4	11.5	14.4	10.5	11.3
9	13.8	15.1	15.7	19.6	16.4	13.0	16.2	13.9	15.4
10	31.5	31.3	31.1	31.7	23.5	28.7	27.7	18.3	28.2

Panel C: Counterfactual Enrollment

Decile	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside	Overall UC
1	3.2	2.8	1.7	2.6	3.1	2.8	3.4	5.2	3.1
2	3.4	3.5	3.1	3.3	3.3	3.5	3.1	8.3	3.9
3	4.8	5.1	4.5	4.6	4.3	4.9	4.3	9.2	5.2
4	5.5	5.6	5.3	5.2	5.0	6.5	4.8	8.8	5.8
5	6.1	5.8	6.1	6.4	5.9	7.3	6.5	9.0	6.6
6	8.0	6.6	7.6	6.8	7.6	10.4	9.0	7.5	7.9
7	9.0	8.4	9.6	9.2	10.1	9.9	10.8	9.7	9.6
8	11.2	11.1	12.1	11.3	13.3	12.3	14.6	10.5	12.0
9	15.2	17.2	17.0	19.4	18.9	14.0	16.0	13.7	16.5
10	33.7	33.9	33.1	31.0	28.5	28.4	27.4	18.1	29.4

Notes: Cells display the overall percentage across admission cohorts 2001 to 2006. Each student's probability of enrollment is computed as the product of admission probability and probability of enrollment conditional on admission. Conditional enrollment probability is estimated using a probit regression of enrollment on the same explanatory variables in Table 4, among the admitted students at each campus. Predicted and counterfactual admission probabilities are computed the same way as described in Table 5.

Table 7: Regression Results for Cumulative College GPA

Panel A: Ordinary Least Squares

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Decile	0.017***	0.023***	0.022***	0.023***	0.029***	0.028***	0.024***	0.025***	0.024***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Observations	17,931	20,827	20,558	20,842	18,663	23,223	14,427	17,534	154,005
R-squared	0.275	0.305	0.274	0.220	0.321	0.320	0.219	0.202	0.327

Panel B: Heckman Correction With No Exclusion Restrictions

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Decile	0.002	-0.004	0.003	0.019***	0.024	0.024***	0.024***	0.022***	0.010**
	(0.005)	(0.007)	(0.007)	(0.006)	(0.016)	(0.003)	(0.002)	(0.004)	(0.004)
Inverse Mills Ratio	0.774***	0.956***	1.433***	-0.182	0.162	1.383***	-0.080	0.322	2.474***
	(0.231)	(0.236)	(0.461)	(0.252)	(0.463)	(0.416)	(0.050)	(0.378)	(0.660)
Observations	17,931	20,827	20,558	20,842	18,663	23,223	14,427	17,534	154,005

Panel C: Heckman Correction With Other UC Admission Offers

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Decile	0.020***	0.021***	0.021***	0.025***	0.029***	0.028***	0.024***	0.025***	0.024***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Inverse Mills Ratio	-0.154***	0.069***	0.085***	0.086***	0.022	0.088***	0.062***	0.057**	-0.143***
	(0.032)	(0.017)	(0.010)	(0.013)	(0.016)	(0.015)	(0.019)	(0.022)	(0.012)
Observations	17,931	20,827	20,558	20,842	18,663	23,223	14,427	17,534	154,005

Notes: The dependent variable is cumulative college GPA. All regressions include the full set of controls from Table 2. Heckman results are estimated using the two-step procedure. Panel B uses no additional selection shifters in the Heckman first stage, while Panel C uses dummies of admission offers to other UC campuses. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

A Appendix

A.1 Index Score

To help protect student anonymity, the UCOP dataset does not report students' exact high school GPA or SAT scores, but instead reports them within a bin range. Unfortunately, these bins are somewhat coarser than one would like to fully control for academic achievement. Raw high school GPA is reported in bins of 0.25, while SAT score is reported in bins of 50 points by category (except the top bin, which goes 700–800).

Much of this problem can ultimately be resolved using an additional measure of student achievement, known as the Index score. Index score is a composite measure of student achievement, computed by UCOP as a linear combination of high school GPA and SAT scores. Unlike the individual components, this Index score is recorded precisely in the data. It is thus able to detect fine differences in achievement that are invisible to the GPA and SAT bins. Index score thus serves as the primary academic control in all regressions, and is included using a cubic polynomial for flexibility.

Figure A.1 shows a histogram of Index score among applicants from California public high schools.

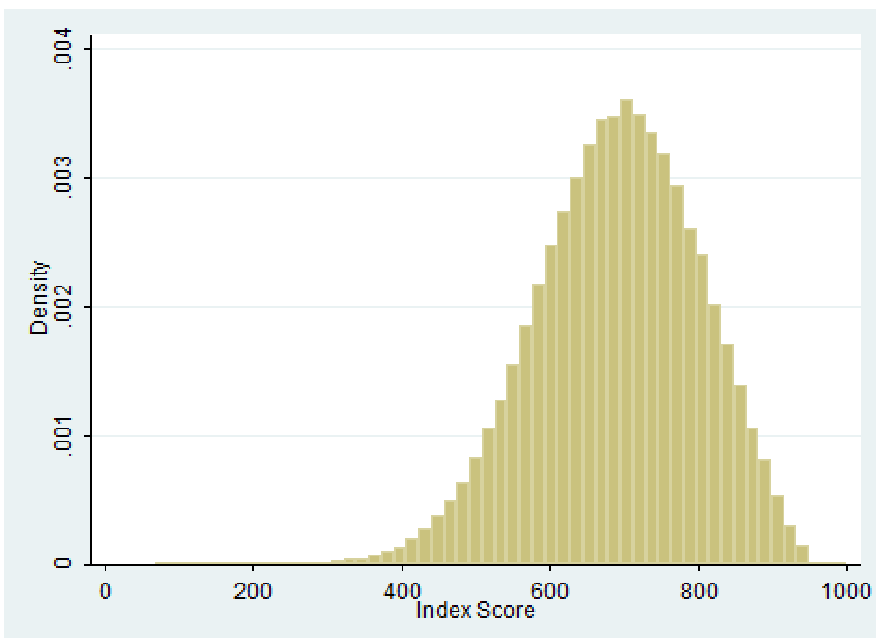

Figure A1: Histogram of Index Score

Index score has a maximum value of 1000, with up to 400 points coming from high school GPA, and 300 points each from SATI Math and Verbal scores. Regressing Index score on continuous approximations (provided by UCOP) of high school GPA and SAT scores, the approximate formula is:

$$\text{Index Score} \approx -165 + 90 \cdot \text{HSGPA} + 0.48 \cdot \text{SAT Math} + 0.48 \cdot \text{SAT Verbal}. \tag{A1}$$

One point to keep in mind, when using this measure, is that not all combinations that achieve the same Index score are necessarily equally effective for admissions. In practice, most UC campuses seem to prefer students with higher GPA and lower SAT score than vice versa. To help address this issue, all regressions include dummies for the bin value of high school GPA and SAT score (by category), along with the cubic polynomial of Index score.

A.2 Standard Errors

Due to various similarities in experience between students of the same high school, intuitively one wishes to group error terms at the high school level. This clustering cannot be achieved, however, since individual high schools are not indicated in the data. Error terms are instead clustered by student in pooled regressions, and currently unclustered for individual campuses. While a possible alternative is to cluster by API decile, that choice is also inappropriate, as ten deciles is too few to invoke asymptotic properties, and also encounters challenges using cluster-bootstrap techniques (Cameron, Gelbach and Miller, 2008).

This section instead computes the degree of within-cluster error correlation needed to make results in Table 3 insignificant based on the Moulton (1986) factor. For the case of simple regression, the OLS standard error should be inflated by a factor of

$$\sqrt{1 + \rho_x \rho_u (var(N_g)/\bar{N}_g + \bar{N}_g - 1)} \tag{A2}$$

where ρ_x denotes within-cluster correlation of x, ρ_u denotes within-cluster error correlation, N_g is the number of observations in cluster g, and $\bar{N}_g$ is the average cluster size. While this result is inexact with more than one regressor, it can still provide a reasonable approximation (Moulton, 1990). Further, although this result is for OLS, here the z statistics on average marginal effects obtained using a probit regression are very similar to those obtained under OLS (see Table A4 for OLS results).

Treating individual high schools as clusters, $\rho_x \approx 1$, as except for high schools right near decile cutoffs, which may finish above or below the cutoff in a given year, API decile is mostly stable from year-to-year. Next, based on supplemental tabulations from UCOP, a total of 1,276 California public high schools had at least one student applying to UC colleges between 2001–2006. Assuming this count also holds true for individual campuses, and that the proportional composition by high school is also constant across campuses, one can plug in approximate values of $\bar{N}_g$ and $var(N_g)$, and thereby compute the minimum value of ρ_u that renders results insignificant.

For example, the average marginal effect for UC Berkeley has a z statistic of -21.60 in Table 3. Significance at the 5% level disappears if standard errors are understated 11.02 times relative to the truth. Under the stated assumptions, UC Berkeley has an average cluster size of 116.33 with a variance of 24,010.8. Significance at the 5% level disappears if $\sqrt{1+\rho_u(24,010.8/116.33+116.33-1)} \geq 11.02$. This occurs when $\rho_u \geq 0.374$. That is, within-high school error correlation of at least 0.374 renders results insignificant for UC Berkeley in Table 3.

Table A1 shows the required value of ρ_u for each campus. As seen, even a small degree of within-high school error correlation makes results insignificant at UC Irvine and UC Riverside. However, these are both campuses where the average marginal effect is currently estimated to be positive. Loss of significance at these campuses does not substantively affect results. For campuses where the estimated average marginal effect is negative, ρ_u needs to be at least 0.276 before significance disappears.

Table A1: Within-High School Error Correlation Thresholds

	Berkeley	Los Angeles	San Diego	Irvine	Santa Barbara	Davis	Santa Cruz	Riverside
z	-21.60	-29.69	-29.63	$+3.55$	-57.85	-18.01	$+1.95$	$+6.06$
Required Inflation on S.E.	11.02	15.15	15.12	1.81	29.52	9.19	N/A	3.09
Average Cluster Size	116.33	145.47	144.17	126.78	122.77	109.48	81.28	85.77
Var(Cluster Size)	24,010.8	37,545.7	36,876.3	28,516.5	28,966.7	21,266.0	11,720.0	13,053.5
Minimum ρ_u	0.374	0.568	0.570	0.007	> 1	0.276	N/A	0.036

Notes: Table shows the required inflation on standard errors for the coefficient on API decile to become insignificant at the 5% level in Table 3. No threshold is computed for UC Santa Cruz, whose z statistic is not significant at the 5% level.

Table A2: Descriptive Statistics, Alternate View

Variable	N	Mean	S.D.	Min	Max
API Decile	320,721	6.99	2.84	1	10
Eligibility in the Local Context	320,721	0.19	0.39	0	1
Index Score	320,443	690	109	67	1000
Raw High School GPA[a]	318,208	3.38	0.47	2.50	3.87
SATI Math[b]	315,420	597	96	400	724
SATI Verbal[b]	315,415	564	98	400	724
SATII Writing[b]	309,509	569	100	400	724
SATII Third Subject Score[b]	290,579	606	105	400	724
Number of AP Courses (in Semesters)	313,749	7.20	5.31	0	90
Parental Income Percentile[c]	255,275	68.4	29.1	6.1	96.3
Parental Education Percentile[d]	306,896	65.4	30.4	2.3	93.7
White	320,721	0.35	0.48	0	1
Asian	320,721	0.35	0.48	0	1
Underrepresented Minority	320,721	0.23	0.42	0	1
Other/Unknown Race	320,721	0.08	0.26	0	1
Science Major	320,391	0.32	0.47	0	1
Social Science Major	320,391	0.15	0.36	0	1
Other Major	320,391	0.53	0.50	0	1
Mean SAT of Applied Campuses	320,391	1205	39	1069	1296
Applied to Berkeley	320,721	0.47	0.50	0	1
Applied to Davis	320,721	0.44	0.50	0	1
Applied to Irvine	320,721	0.51	0.50	0	1
Applied to Los Angeles	320,721	0.58	0.49	0	1
Applied to Riverside	320,721	0.36	0.48	0	1
Applied to San Diego	320,721	0.58	0.49	0	1
Applied to Santa Barbara	320,721	0.51	0.50	0	1
Applied to Santa Cruz	320,721	0.33	0.47	0	1
Cumulative College GPA[e]	154,006	2.99	0.60	1.05	3.95

Notes: This table uses continuous approximations, provided by UCOP, for variables in bins. Values are typically coded at bin midpoints, with some exceptions for top and bottom end bins.

a – Recorded in bins of 0.00-2.99; 3.00-3.24; 3.25-3.49; 3.50-3.74; 3.75-4.00
b – Recorded in bins of 200-449; 450-499; 500-549; 550-599; 600-649; 650-699; 700-800
c – Parental income is originally recorded in bins of $0-9,999; $10,000-19,999; $20,000-29,999; ...; $90,000-99,999; and $100,000+ per year. These have been converted to percentiles using the 2000 U.S. Census.
d – Parental education is originally recorded in bins of no high school; some high school; high school graduate; some college; 2-year college graduate; 4-year college graduate; and postgraduate study. These have been converted to percentiles using the 2000 U.S. Census.
e – Recorded in bins of 0–1.09; 1.10–1.19; 1.20–1.29; ...; 3.90–4.00.

Table A3: Selected Means by UC Campus

Applicants

Campus	N	Admit	Attend	API Decile	HS GPA	SATI Total	ELC	Income Percentile	Underrep. Minority
Berkeley	149,940	0.27	0.12	7.21	3.51	1217	0.26	68.4	0.20
Los Angeles	185,620	0.26	0.11	7.08	3.46	1194	0.24	67.3	0.22
San Diego	185,893	0.44	0.11	7.35	3.45	1197	0.22	70.0	0.18
Irvine	162,732	0.58	0.13	7.04	3.36	1157	0.19	66.3	0.21
Santa Barbara	163,773	0.54	0.11	7.30	3.36	1163	0.15	71.1	0.21
Davis	141,992	0.63	0.16	7.37	3.40	1173	0.17	70.8	0.17
Santa Cruz	105,203	0.79	0.14	7.27	3.28	1142	0.10	69.9	0.21
Riverside	116,222	0.88	0.15	6.64	3.22	1097	0.10	64.4	0.29

Admits

Campus	N	Admit	Attend	API Decile	HS GPA	SATI Total	ELC	Income Percentile	Underrep. Minority
Berkeley	39,964	1	0.45	7.23	3.80	1312	0.60	69.7	0.17
Los Angeles	49,031	1	0.42	7.29	3.75	1309	0.50	68.1	0.17
San Diego	81,451	1	0.25	7.38	3.72	1280	0.44	69.0	0.15
Irvine	94,017	1	0.22	7.38	3.58	1232	0.32	70.9	0.16
Santa Barbara	87,764	1	0.21	7.13	3.59	1220	0.26	70.6	0.20
Davis	88,138	1	0.26	7.36	3.58	1223	0.26	69.9	0.16
Santa Cruz	82,265	1	0.18	7.39	3.41	1169	0.12	71.4	0.19
Riverside	95,903	1	0.18	6.84	3.30	1120	0.11	66.4	0.26

Attending Students

Campus	N	Admit	Attend	API Decile	HS GPA	SATI Total	ELC	Income Percentile	Underrep. Minority
Berkeley	17,966	1	1	7.18	3.77	1289	0.53	68.6	0.16
Los Angeles	20,829	1	1	7.04	3.69	1267	0.41	65.4	0.18
San Diego	20,559	1	1	7.39	3.64	1245	0.32	67.4	0.13
Irvine	20,868	1	1	7.59	3.43	1191	0.17	72.6	0.14
Santa Barbara	18,815	1	1	6.95	3.47	1171	0.15	72.1	0.24
Davis	23,320	1	1	7.11	3.46	1164	0.17	66.8	0.15
Santa Cruz	14,493	1	1	7.40	3.25	1139	0.05	73.1	0.19
Riverside	17,679	1	1	6.34	3.12	1067	0.08	62.6	0.32

Notes: N reports the total number of applicants, admits, or attending students from California public high schools between 2001 and 2006. Individual students appear once at each campus applied or admitted.

Table A4: OLS Admission Results by Campus

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Decile	−0.008***	−0.010***	−0.011***	−0.002***	−0.027***	−0.012***	−0.001	0.001***	−0.008***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0004)	(0.0002)
ELC Status	0.158***	0.074***	0.175***	0.182***	0.042***	0.064***	−0.012***	0.011***	0.108***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Observations	148,439	185,620	183,958	161,768	163,040	139,697	103,707	109,448	1,195,677
R-squared	0.469	0.442	0.579	0.573	0.519	0.480	0.481	0.413	0.515
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517

Notes: Robust standard errors are in parentheses, clustered by student in the pooled regression. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A5: Probit Average Marginal Effects with API Decile as Spline

Panel A: Single Knot at 5.5

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Deciles 1–5	-0.018***	-0.027***	-0.017***	-0.006***	-0.019***	-0.030***	-0.003***	0.003***	-0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
API Deciles 6–10	-0.003***	-0.000	-0.007***	0.006***	-0.026***	0.004***	0.003***	0.001**	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Observations	148,435	185,619	183,958	161,765	163,038	139,695	103,702	109,444	1,195,656
Pseudo R-squared	0.454	0.442	0.594	0.584	0.519	0.479	0.505	0.478	0.527
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517

Panel B: Two Knots at 4 and 7

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
API Deciles 1–4	-0.019***	-0.024***	-0.008***	-0.011***	-0.007***	-0.021***	-0.001	0.003**	-0.013***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
API Deciles 4–7	-0.011***	-0.019***	-0.020***	0.002**	-0.034***	-0.024***	-0.003**	0.002**	-0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
API Deciles 7–10	-0.001	0.005***	-0.002***	0.006***	-0.020***	0.013***	0.005***	0.002*	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Observations	148,435	185,619	183,958	161,765	163,038	139,695	103,702	109,444	1,195,656
Pseudo R-squared	0.454	0.442	0.595	0.584	0.520	0.479	0.505	0.478	0.527
Admit Rate	0.269	0.264	0.443	0.581	0.538	0.631	0.793	0.876	0.517

Notes: Robust standard errors are in parentheses, clustered by student in the pooled regression. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A6: Average Marginal Effects for Probit Regression of Yield, Each API Decile

	(1) Berkeley	(2) Los Angeles	(3) San Diego	(4) Irvine	(5) Santa Barbara	(6) Davis	(7) Santa Cruz	(8) Riverside	(9) Pooled
1st API Decile	-0.040***	-0.065***	-0.109***	-0.063***	-0.056***	-0.115***	-0.021***	-0.049***	-0.068***
	(0.015)	(0.013)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.003)
2nd API Decile	-0.071***	-0.055***	-0.064***	-0.077***	-0.064***	-0.088***	-0.042***	-0.000	-0.053***
	(0.014)	(0.012)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.002)
3rd API Decile	-0.043***	-0.009	-0.049***	-0.056***	-0.042***	-0.067***	-0.029***	0.001	-0.033***
	(0.013)	(0.011)	(0.008)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.002)
4th API Decile	-0.040***	-0.011	-0.038***	-0.044***	-0.040***	-0.031***	-0.029***	-0.004	-0.027***
	(0.013)	(0.011)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.002)
5th API Decile	-0.023*	-0.011	-0.026***	-0.013**	-0.031***	-0.021***	-0.007	0.010	-0.012***
	(0.012)	(0.011)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.002)
6th API Decile	0.014	-0.004	-0.005	-0.014**	-0.017***	0.010	-0.004	-0.010*	-0.004*
	(0.012)	(0.011)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.002)
7th API Decile	—	—	—	—	—	—	—	—	—
8th API Decile	0.005	0.027***	0.011*	0.008	0.006	0.003	0.012**	-0.007	0.006***
	(0.011)	(0.010)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.002)
9th API Decile	-0.008	0.041***	0.006	0.038***	0.020***	-0.012*	-0.006	-0.023***	0.004**
	(0.010)	(0.009)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.002)
10th API Decile	0.001	0.011	0.001	0.013***	-0.020***	-0.006	-0.022***	-0.043***	-0.015***
	(0.009)	(0.008)	(0.006)	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)	(0.002)
ELC Status	-0.037***	-0.041***	-0.025***	-0.042***	-0.039***	-0.004	-0.011**	0.013***	-0.036***
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.001)
Observations	39,964	49,031	81,447	94,014	87,764	88,138	82,264	95,901	618,530
Pseudo R-squared	0.074	0.140	0.121	0.172	0.130	0.145	0.259	0.114	0.124
Mean Dep Var	0.450	0.425	0.252	0.222	0.214	0.265	0.176	0.184	0.250

Notes: The dependent variable is enrollment, conditional on having been admitted. Robust standard errors are in parentheses, clustered by student in the pooled regression. *** p<0.01, ** p<0.05, * p<0.1