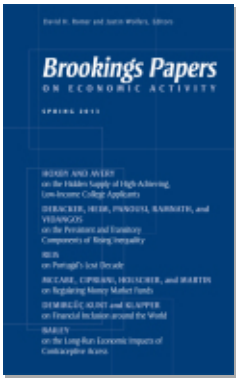
The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students

Caroline Hoxby, Christopher Avery



Brookings Papers on Economic Activity, Spring 2013, pp. 1-65 (Article)

Published by Brookings Institution Press
DOI: 10.1353/eca.2013.0000

CAROLINE HOXBY
Stanford University

CHRISTOPHER AVERY
Harvard Kennedy School

The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students

ABSTRACT We show that the vast majority of low-income high achievers do not apply to any selective college. This is despite the fact that selective institutions typically cost them *less,* owing to generous financial aid, than the two-year and nonselective four-year institutions to which they actually apply. Moreover, low-income high achievers have no reason to believe they will fail at selective institutions since those who do apply are admitted and graduate at high rates. We demonstrate that low-income high achievers' application behavior differs greatly from that of their high-income counterparts with similar achievement. The latter generally follow experts' advice to apply to several "peer," a few "reach," and a couple of "safety" colleges. We separate low-income high achievers into those whose application behavior is similar to that of their high-income counterparts ("achievement-typical") and those who apply to no selective institutions ("income-typical"). We show that income-typical students are not more disadvantaged than the achievement-typical students. However, in contrast to the achievement-typical students, income-typical students come from districts too small to support selective public high schools, are not in a critical mass of fellow high achievers, and are unlikely to encounter a teacher who attended a selective college. We demonstrate that widely used policies—college admissions recruiting, campus visits, college mentoring programs—are likely to be ineffective with income-typical students. We suggest that effective policies must depend less on geographic concentration of high achievers.

In this study we show that a large number—probably the vast majority—of very high-achieving students from low-income families do not apply to a selective college or university.[1] This is in contrast to students with the same test scores and grades who come from high-income backgrounds: they are overwhelmingly likely to apply to a college whose median student has achievement much like their own. This gap is puzzling because we find that the subset of high-achieving, low-income students who do apply to selective institutions are just as likely to enroll and progress toward a degree at the same pace as high-income students with equivalent test scores and grades. Added to the puzzle is the fact that very selective institutions not only offer students much richer instructional, extracurricular, and other resources, but also offer high-achieving, low-income students so much financial aid that these students would often pay *less* to attend a selective institution than the far less selective or nonselective postsecondary institutions that most of them do attend.

We attempt to unravel this puzzle by characterizing low-income, very high achieving students in the U.S. using a rich array of data, including individual-level data on every student who takes one of the two college assessments, the ACT and the SAT. We divide the low-income, very high-achieving students into those who apply similarly to their high-income counterparts ("achievement-typical" behavior) and those who apply in a very dissimilar manner ("income-typical" behavior). We do this because we are interested in why some low-income high achievers appear to base their college-going on their achievement, whereas others base it on their income. We find that income-typical students are fairly isolated from other high achievers, both in terms of geography and in terms of the high schools they attend. In fact, their lack of concentration is such that many traditional strategies for informing high-achieving students about college—for

1. Hereafter, "low-income" and "high-income" mean, respectively, the bottom and top quartiles of the income distribution of families with a child who is a high school senior. "High-achieving" refers to a student who scores at or above the 90th percentile on the ACT comprehensive or the SAT I (math and verbal) and who has a high school grade point average of A- or above. This is approximately 4 percent of U.S. high school students. When we say "selective college" in a generic way, we refer to colleges and universities that are in the categories from "Very Competitive Plus" to "Most Competitive" in Barron's *Profiles of American Colleges.* There were 236 such colleges in the 2008 edition. Together, these colleges have enrollments equal to 2.8 times the number of students who scored at or above the 90th percentile on the ACT or the SAT I. Later in the paper, we are much more specific about colleges' selectivity: we define schools that are "reach," "peer," and "safety" for an individual student, based on a comparison between that student's college aptitude test scores and the median aptitude test scores of students enrolled at the school.

instance, college admissions staff visiting high schools, or after-school programs that provide mentoring—would be prohibitively expensive. We also show that income-typical students have a negligible probability of meeting a teacher, high school counselor, or schoolmate from an older cohort who attended a selective college.

In contrast, we show that achievement-typical students are *highly* concentrated. Some of these low-income students attend a small number of "feeder" high schools that contain a critical mass of high achievers. Some feeder schools admit students on the basis of an exam or previous grades; others are magnet schools; still others contain a subpopulation of low-income students in a student body that is generally affluent. Since these high schools are nearly all located in the largest school districts of very large metropolitan areas (not even in medium-size metropolitan areas), their students are far from representative of high-achieving, low-income students in general. Moreover, we show evidence that suggests that these schools may be "tapped out"—that their students are already so intensively recruited by selective colleges that further recruitment may merely shift students among similar, selective colleges, and not cause students to change their college-going behavior in more fundamental ways.

The evidence that we present is descriptive, not causal. This is an important distinction. For instance, we cannot assert that a high-achieving, low-income student would act like an achievement-typical student rather than an income-typical student if he or she were moved to a large metropolitan area with a high school that practices selective admission. Moreover, we do not assert that income-typical students would have higher welfare if they applied to college in the same way that achievement-typical and high-income high achievers do. We leave such causal tests for related studies in which we conduct randomized, controlled interventions. Nevertheless, our descriptive evidence makes three important contributions. First, it documents that the number of low-income high achievers is much greater than college admissions staff generally believe. Since admissions staff see only the students who apply, they very reasonably underestimate the number who exist. Second, our evidence suggests hypotheses for why so many low-income high achievers apply to colleges in a manner that may not be in their best interest, and that is certainly different from what similarly high-achieving, high-income students do. Most of our hypotheses are related to the idea that income-typical students—despite being intelligent, literate, and on colleges' search lists (that is, the lists to which selective colleges mail brochures)—lack information or encouragement that achievement-typical students have because they are part of local,

critical masses of high achievers. Third, our descriptive evidence allows us to explain why some traditional interventions are unlikely to change the situation and allows us to identify other interventions that could plausibly do so.

Our previous work (Avery, Hoxby, and others 2006) was perhaps the first to identify the phenomena described in this paper, but there is now a small literature on the topic of "undermatching." We especially note the work of William Bowen, Matthew Chingos, and Michael McPherson (2009), Eleanor Dillon and Jeffrey Smith (2012), and Amanda Pallais (2009). Relative to those studies, our study's strengths are its comprehensiveness (we analyze the entire population of high-achieving students, not a sample); our complete characterization of each U.S. high school, including its history of sending its students to college; our ability to map students to their exact high schools and neighborhoods (this allows us to investigate exactly what they experience); and our use of accurate administrative data to identify students' aptitude, application behavior, college enrollment, and on-time degree completion. The sheer comprehensiveness and accuracy of our data are what allow us to test key hypotheses about why some high-achieving, low-income students are income-typical and others are achievement-typical. Our data also allow us to assess which interventions might plausibly (and cost-effectively) alter such behavior.

The paper is organized as follows. In the next section we present some background on college policies directed toward low-income high achievers. In section II we describe our data sources. In section III we present a descriptive portrait of very high-achieving U.S. students—their family incomes, parents' education, race, ethnicity, and geography. In section IV we show that high-achieving students' college application behavior differs greatly by family income. We also show that, conditional on applying to a college, students' enrollment, college grades, and degree receipt do *not* differ by family income (among students with similar incoming qualifications). In section V we divide low-income high achievers into achievement-typical and income-typical groups. We then compare factors that might affect the college application behavior of these groups. In section VI we consider several interventions commonly directed toward low-income high achievers, and we demonstrate that they are likely to be cost-prohibitive for income-typical students. To drive the point home, we contrast colleges' difficulty in identifying low-income high achievers with their ease in identifying top athletes. In section VII we conclude by discussing which hypotheses we have eliminated and which still need testing, and we speculate on the sort of interventions that could plausibly test

whether income-typical students' welfare would be greater if they were better informed.

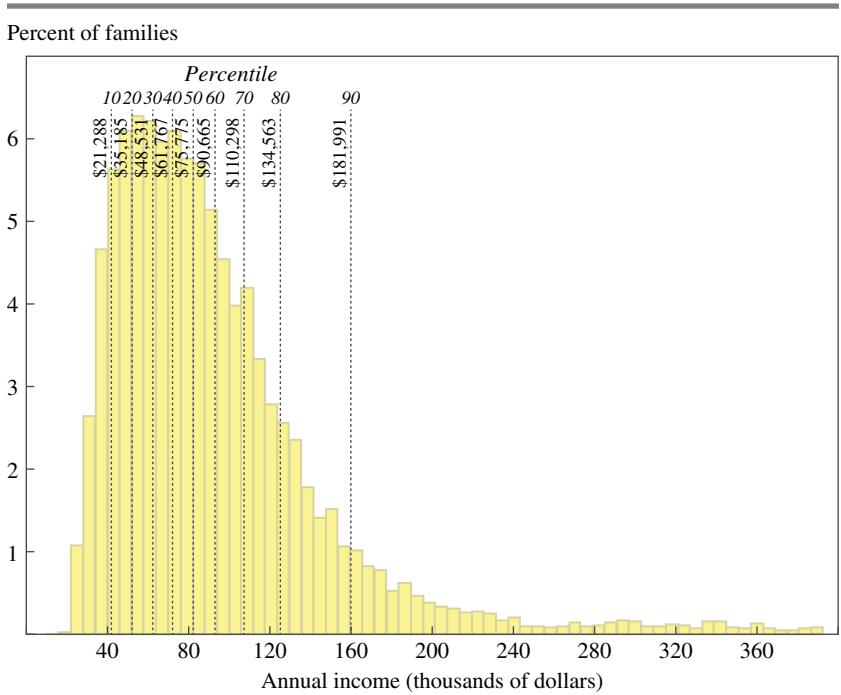
I. Background on College Policies Directed toward Low-Income High Achievers

Many students from low-income families have poor college outcomes: they do not attend college, they drop out before attaining a degree, they earn so few credits each term that they cannot graduate even in 1.5 times the "correct" time to degree, or they attend institutions with such poor resources that even when they do graduate, they earn much less than the median college graduate. These poor college outcomes are often attributed to low-income students being less academically prepared for college and less able to pay for college. These are certainly valid concerns. As we show later, high-income students (those from families in the top income quartile) are in fact much more likely to be high achievers at the end of high school than are low-income students. Nevertheless, some low-income students *are* very high achievers: at the end of high school, they have grades and college assessment scores that put them in the top 10 percent of students who take one of the ACT or SAT college assessment exams or, equivalently, the top 4 percent of all U.S. secondary school students.

High-achieving, low-income students are considered very desirable by selective colleges, private and public, which are eager to make their student bodies socioeconomically diverse without enrolling students who are unprepared for their demanding curricula. The ultimate evidence of colleges' eagerness is their financial aid policies, which, as we shall show, are very generous toward such students. However, we have also observed this eagerness personally among hundreds of college leaders and their admissions staff. Many spend considerable amounts on recruiting the low-income students who do apply and on (not always successful) programs designed to increase their numbers of low-income applicants. There are many reasons for selective institutions to prefer socioeconomic diversity. These include, to name just a few, a deep respect for merit regardless of need; the fact that students whose lives were transformed by highly aided college education tend to be the most generous donors if they do become rich; a belief that a diverse student body makes instruction and research more productive; and pressure from society.

In recent years, selective schools' aid for low-income high achievers has become so generous that such students' out-of-pocket costs of attendance are zero at the nation's most competitive schools, and small at other very

Figure 1. Distribution of Annual Family Income, Families with a Child in 12th Grade, 2008



Source: 2008 American Community Survey.

selective schools. Figure 1 shows the distribution of annual income in 2008 for families with a child in the 12th grade—a good indicator for a family having a child of college-going age in the next year. The 20th percentile of this distribution was $35,185. Table 1 shows the out-of-pocket costs (including loans) such a student would have experienced in the 2009–10 school year at a variety of selective and nonselective institutions. The table is organized based on institutions' selectivity as classified by Barron's *Profiles of American Colleges:* most competitive, very competitive, competitive 4-year institutions, nonselective 4-year institutions, and (nonselective by definition) community colleges and other 2-year institutions. Table 1 also shows the colleges' comprehensive cost for a student who needs *no* financial aid (the "sticker price") and their instructional expenditure per student. What the table reveals is that a low-income student who can gain admission to one of the most selective colleges in the U.S. can expect to pay *less* to attend a very selective college with maximum instructional

Table 1. College Costs and Resources, by Selectivity of College[a]
Dollars per year

Selectivity	*Out-of-pocket cost for student at 20th %ile of family income*	*Comprehensive cost (cost before financial aid)*	*Average instructional expenditure per student*
Most competitive	6,754	45,540	27,001
Highly competitive plus	13,755	38,603	13,732
Highly competitive	17,437	35,811	12,163
Very competitive plus	15,977	31,591	9,605
Very competitive	23,813	29,173	8,300
Competitive plus	23,552	27,436	6,970
Competitive	19,400	24,166	6,542
Less competitive	26,335	26,262	5,359
Some or no selectivity, 4-year	18,981	16,638	5,119
Private 2-year	14,852	17,822	6,796
Public 2-year	7,573	10,543	4,991
For-profit 2-year	18,486	21,456	3,257

Sources: Barron's *Profiles of American Colleges* and authors' calculations using the colleges' own net cost calculators and data from the Integrated Postsecondary Education Data System (IPEDS), National Center for Education Statistics.

a. All costs include tuition and room and board. Out-of-pocket costs include loans. At the very competitive level and above, the net cost data were gathered by the authors for the 2009–10 school year. For all other institutions, net cost estimates are based on the institution's published net cost calculator for the year closest to 2009–10, but never later than 2011–12. These published net costs are then reduced to approximate 2009–10 levels using the institution's own figures for room and board and tuition net of aid, from IPEDS, for the relevant years. Instructional expenditure data are from IPEDS.

expenditure than to attend a nonselective 4-year college or 2-year institution. In short, the table demonstrates the strong financial commitment that selective colleges have made toward becoming affordable to low-income students.[2]

In related work (Avery, Hoxby, and others 2006), we analyze Harvard University's introduction of zero costs for students with annual family incomes of $40,000 and below starting in 2005. (Harvard is a relevant option for the students we analyze in this paper.) Harvard's policy was

2. Note that such a student's out-of-pocket costs (including loans), in absolute terms, peak at private colleges of middling to low selectivity. This is because these colleges have little in the way of endowment with which to subsidize low-income students and receive no funding from their state government (as public colleges do) with which to subsidize students. Moreover, the most selective colleges spend substantially more on each student's education than is paid by even those students who receive no financial aid (Hoxby, 2009). Thus, when a low-income student attends a very selective college, he or she gets not only financial aid but also the subsidy received by every student there.

quickly imitated or outdone by the institutions with which it most competes: Yale, Princeton, Stanford, and some others. All such institutions subsequently raised the bar on what they considered to be a low enough income to merit zero costs, to the point where even students from families with income above the U.S. median can often attend such institutions for free. Although less well endowed institutions followed suit to a lesser extent (usually by setting the bar for zero costs at a lower family income than the aforementioned institutions did), the result was very low costs for low-income students at selective institutions, as table 1 shows.

In our other work we show that Harvard's policy change had very little effect—at least, very little *immediate* effect—on the income composition of its entering class. We estimate that it increased the number of low-income students by approximately 20, in a class of more than 1,600 (Avery, Hoxby, and others 2006, table 1, top row). Interestingly, this very modest effect was not a surprise to many college admissions staff. They explained that there was a small pool of low-income high achievers who were already "fully tapped," so that additional aid and recruiting could do little except shift them among institutions that were fairly similar. Put another way, they believed that the *overall* pool of high-achieving, low-income students was inelastic. Many felt that they had already tried every means open to them for recruiting low-income students: guaranteeing need-blind admission,[3] disproportionately visiting high schools with large numbers of free-lunch-eligible students,[4] sending special letters to high achievers who live in high-poverty ZIP codes,[5] maintaining strong relationships with guidance counselors who reliably direct low-income

3. In order to guarantee low-income students that they are at no disadvantage in admissions, many colleges maintain "Chinese Walls" between their admissions and financial aid offices. Consequently, many schools can only *precisely* identify low-income students once they have been admitted. However, admissions officers target recruiting by analyzing applicants' essays, their teachers' letters of recommendation, their parents' education, and their attendance at an "underresourced" high school.

4. Even highly endowed colleges cannot afford to have their admissions staff personally visit many more than 100 high schools a year, and there were more than 20,000 public and more than 8,000 private high schools nationwide in the school year relevant to our study.

5. Colleges routinely purchase "search files" from the College Board and ACT that contain names and addresses of students whose test scores fall in certain ranges (and who agree to be "searched"). The colleges can then purchase marketing information on which ZIP codes have low median incomes. The materials they send to students in such ZIP codes typically include, in addition to their usual brochures, a letter describing their financial aid and other programs that support low-income students.

applicants to them,[6] coordinating with or even *running* college mentoring programs for low-income students,[7] paying a third-party organization for a guaranteed minimum number of low-income enrollees,[8] sponsoring campus visits for students from local high schools known to serve low-income families, and personally contacting students whose essays suggest that they might be disadvantaged. Although the admissions staff believed that they might succeed in diversifying their student bodies by poaching from other selective schools or lowering their admissions standards for low-income students, they did not expect additional aid together with more of the same recruiting methods to affect matters much.[9] (Note that the methods we use in this paper to identify low-income students are not available to college admissions staff.)[10]

In this paper, we show that—viewed one way—the admissions staff are correct. The pool of high-achieving, low-income students who apply to selective colleges is small: for every high-achieving, low-income student who applies, there are from 8 to 15 high-achieving, high-income students who apply. Viewed another way, however, the admissions staff are too pessimistic: the vast majority of high-achieving, low-income students do not apply to *any* selective college. There are, in fact, only

6. These schools are informally known as "feeders." Feeder schools are often selective schools (schools that admit students on the basis of exams or similar criteria), magnet schools, or schools that enroll a subpopulation of low-income students despite having most of their students drawn from high-income, highly educated families.

7. Since the vast majority of college mentoring programs rely on students to self-select into their activities, it is unclear whether they identify students who would otherwise be unknown to colleges or merely serve as a channel for students to identify themselves as good college prospects.

8. This practice is controversial. Since the organization may merely be moving low-income students to colleges that pay from colleges that do not, some admissions staff suspect that poaching (not expansion of the pool of low-income applicants) is the reason that the organization can fulfill the guarantees. They suspect that some very selective colleges are able to look good at the expense of others, with little net change in the lives of low-income students. Another controversial aspect is that low-income students who allow themselves to be funneled by the organization do not get to consider the full range of admissions offers they could obtain.

9. In this paragraph, we draw upon personal communications between the authors and many college admissions staff, including those who attend the conferences of the College Board, the Consortium for Financing Higher Education, and the Association of Black Admissions and Financial Aid Officers of the Ivy League and Sister Schools (ABAFAOILSS).

10. Much of the data we use are available only to researchers. Moreover, the analytics involved are far beyond the capacity of the institutional research groups of even the best endowed colleges. We have worked for almost a decade to build the database and analysis that support this paper.

about 2 high-achieving, high-income students for every high-achieving, low-income student in the population. The problem is that most high-achieving, low-income students do not apply to any selective college, so they are invisible to admissions staff. Moreover, we will show that they are unlikely to come to the attention of admissions staff through traditional recruiting channels.

II. Data Sources and Identifying High-Achieving, Low-Income Students

We attempt to identify the vast majority of U.S. students who are very high achieving. Specifically, we are interested in students who are well prepared for college and who would be very likely to be admitted to the majority of selective institutions (if they applied). Thus, as mentioned above, we choose students whose college assessment scores place them in the top 10 percent of test takers based on either the SAT I (combined math and verbal) or the ACT (comprehensive).[11] Since only about 40 percent of U.S. secondary school students take a college assessment, these students are in the top 4 percent of U.S. students. We include in our target group only those students who self-report a grade point average of A- or higher in high school. In practice, this criterion for inclusion hardly matters once we condition on having test scores in the top 10 percent.[12]

Our key data come from the College Board and ACT, both of which supplied us with student-level data on everyone in the high school graduating class of 2008 who took either the ACT or the SAT I.[13] Apart from students'

11. The cutoff is 1300 for combined mathematics and verbal ("Critical Reading") scores on the SAT. The cutoff is 29 for the ACT composite score.

12. We also considered excluding students who had taken no subject tests, since some selective colleges require them. (Subject tests include SAT II tests, Advanced Placement tests, and International Baccalaureate tests.) However, we dropped this criterion for a few reasons. First, many selective colleges do not require subject tests or make admissions offers conditional on a student taking subject tests and passing them. Second, among SAT I takers, few students were excluded by this criterion. Third, ACT comprehensive takers usually take subject tests offered by the College Board or International Baccalaureate. When we attempt to match students between these data sources, errors occur so that at least some of the exclusions are false.

We match students between the ACT comprehensive and the SAT I to ensure that we do not double-count high-achieving students. However, this match is easier than matching the ACT comprehensive takers to College Board subject tests, which students often take as sophomores or juniors in high school.

13. Approximately 2,400,000 students per cohort take a College Board test, and approximately 933,000 students per cohort take the ACT.

test score history, these data sets contain students' high school identifiers, self-reported grades, race and ethnicity, and sex. Validation exercises have shown that students self-report their grades quite accurately to the College Board and ACT (with just a hint of upward bias), probably because students perceive the organizations as playing a semiofficial role in the college application process (Freeberg 1988). The data also contain answers to numerous questions about students' high school activities and their plans for college.

Importantly, the College Board and ACT data contain a full list of the colleges to which students have sent their test scores. Except in rare circumstances, a student cannot complete an application to a selective college without having the College Board or ACT send his or her verified test scores to the college. Thus, score sending is necessary but not sufficient for a completed application. Put another way, score sending may exaggerate but cannot understate the set of selective colleges to which a student applies. Past studies have found that score sending corresponds closely with actual applications to selective colleges (Card and Krueger 2005, Avery and Hoxby 2004). Students who are admitted under an Early Decision or Early Action program often do not apply to colleges other than the one that admitted them early. However, such students typically send scores to all of the schools to which they would have applied had the Early school not admitted them (Avery, Glickman, Hoxby, Metrick 2013). Thus, it is somewhat *better* for our purposes to observe score sending than actual applications: score sending more accurately reveals the set of selective colleges to which the student would have applied. Note, however, that as most 2-year colleges and some nonselective colleges do not require verified ACT or SAT I scores, we do not assume that a student who sends no scores is applying to no postsecondary institutions. Rather, that student is applying to no *selective* institution.

For some of our analyses, we need to know where students actually enrolled and whether they are on track to attain a degree on time (June 2012 for baccalaureate degrees for the class of 2008). We therefore match students to their records at the National Student Clearinghouse, which tracks enrollment and degree receipt. We match all low-income high achievers and a 25 percent random sample of high-income high achievers. We do not match all students for reasons of cost.

The addresses in the data are geocoded for us at the Census block level, the smallest level of Census geography (22 households on average). We match each student to a rich description of his or her neighborhood. The neighborhood's racial composition, sex composition, age

composition, and population density are matched at the block level. Numerous sociodemographic variables are matched at the block group level (556 households on average): several moments of the family income distribution, adults' educational attainment, employment, the occupational distribution, several moments of the house value distribution, and so on. We also merge in income data from the Internal Revenue Service (IRS) at the ZIP code level.

In addition to these data on the graduating class of 2008, we have parallel data for previous cohorts of students who took an ACT or a College Board test. (We have one previous cohort for the ACT and more than 10 previous cohorts for the College Board tests.) We use the previous cohort data in a few ways that will become clear below.

We create a profile of every high school, public and private, in the U.S., using administrative data on enrollment, graduates, basic school characteristics, and sociodemographics. The sources are the Common Core of Data (United States Department of Education 2009a) and the Private School Survey (United States Department of Education 2009b). By summarizing our previous cohort data at the high school level, we also create profiles for each school of their students' usual test scores, application behavior, and college plans. For instance, we know how many students from the high school typically apply to each selective college or to any given group of selective colleges. Finally, we add high schools' test scores, at the subgroup level, for each state's statewide test mandated by the No Child Left Behind Act of 2001. These scores are all standardized to have a zero mean and a standard deviation of 1.

We estimate a student's family income rather than rely on the student's self-reported family income. We do this for a few reasons. First, both the College Board's and the ACT's family income questions provide a series of somewhat wide income "bins" as potential answers. Second, although the College Board's questionnaire appears to elicit unbiased self-reports of family income, students make substantial unsystematic mistakes when their data are compared to their verified data used in financial aid calculations (the CSS Profile data). Third, about 62 percent of students simply do not answer the College Board's family income question. Fourth, although the ACT's questionnaire elicits a high response rate, its question refers to the fact that colleges offer more generous financial aid to students with lower family incomes. This framing apparently induces students to underestimate their family incomes: we find that students often report family incomes that are lower than the 10th percentile of family income in their Census block group.

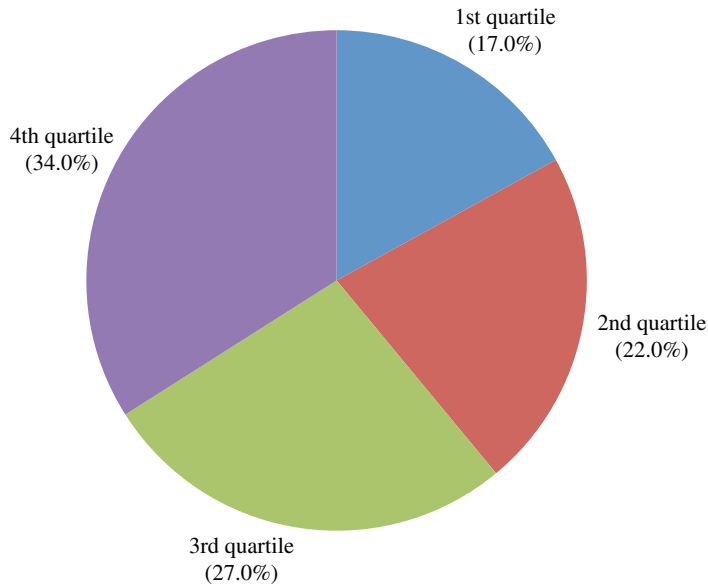
We predict students' family income using all the data we have on previous cohorts of College Board students, matched to their CSS Profile records (data used by financial aid officers to compute grants and loans). That is, using previous cohorts, we regress accurate administrative data on family income using all of our Census variables, the IRS income variables, the high school profile variables, and the student's own race and ethnicity. In practice, the income variables from the Census have the most explanatory power. Our goal is simply to maximize explanatory power, and many of the variables we include are somewhat multicollinear. We choose predicted income cutoffs to minimize Type I error (false positives) in declaring a student to be low-income. Specifically, we choose cutoffs such that, in previous cohorts, only 8 percent of students who are *not* actually in the bottom quartile of the income distribution are predicted to be low-income. We recognize that by minimizing Type I error, we expand Type II error, but it is less worrisome for our exercise if we mistakenly classify a low-income student as middle-income than if we do the reverse. This is because we wish to characterize the college-going behavior of students who are low-income. Since we also find that there are more high-achieving, low-income students than college admissions staff typically believe, we make decisions that will understate rather than overstate the low-income, high-achieving population.

More generally, it is not important for our exercise that our measure of income be precise. What matters for our exercise is that the students we analyze are, in fact, capable of gaining admission at selective colleges—at which time the college's financial aid policies will be implemented. We are confident that the students we analyze are capable of being admitted because we are using the same score data and similar grade data to what the colleges themselves use. Also, we show later that we can accurately predict the colleges at which students enroll, conditioning on the colleges to which they applied. We would not be able to make such accurate predictions if we lacked important achievement and other data that colleges use in their admissions processes.

Hereafter, we describe as low-income any student whose estimated family income is at or below the cutoff for the bottom quartile of the 2008 distribution of incomes among families who had a child in his or her senior year of high school: $41,472.[14] We describe as high-income

14. Since we require microdata to create the relevant distribution, our source for this information is the American Community Survey 2008.

Figure 2. High-Achieving Students, by Family Income Quartile[a]



Source: 2008 American Community Survey and authors' calculations using the combined data set described in the text.

a. High-achieving students are defined as in table 2.

any student whose estimated family income is at or above the cutoff for the top quartile of the same distribution: $120,776. See figure 1 for other percentiles.

III. A Portrait of High-Achieving Students in the U.S.

Who and where are the high-achieving students in the United States? In this section, we briefly characterize them, leaving more detailed analysis of the low-income, high-achieving group for later.

Figure 2 shows that 34 percent of high achievers have estimated family income in the top quartile and 27 percent have estimated family income in the third quartile. That is, high-income families are overrepresented in the high-achieving population. However, 22 percent and 17 percent of high achievers have estimated family incomes in, respectively, the second and bottom quartiles. We estimate that there are at least 25,000 and

Table 2. College Assessment Results of High-Achieving Students, by Family Income[a]

Income quartile	*Average SAT or ACT percentile score among high-achieving students*
First (bottom)	94.1
Second	94.3
Third	94.8
Fourth	95.7

Source: Authors' calculations using data from the ACT, the College Board, IPEDS, and other sources described in the text (hereafter referred to as the "combined data set").

a. High-achieving students are students in 12th grade who have an ACT comprehensive or SAT I (math plus verbal) score at or above the 90th percentile and a high school grade point average of A- or above.

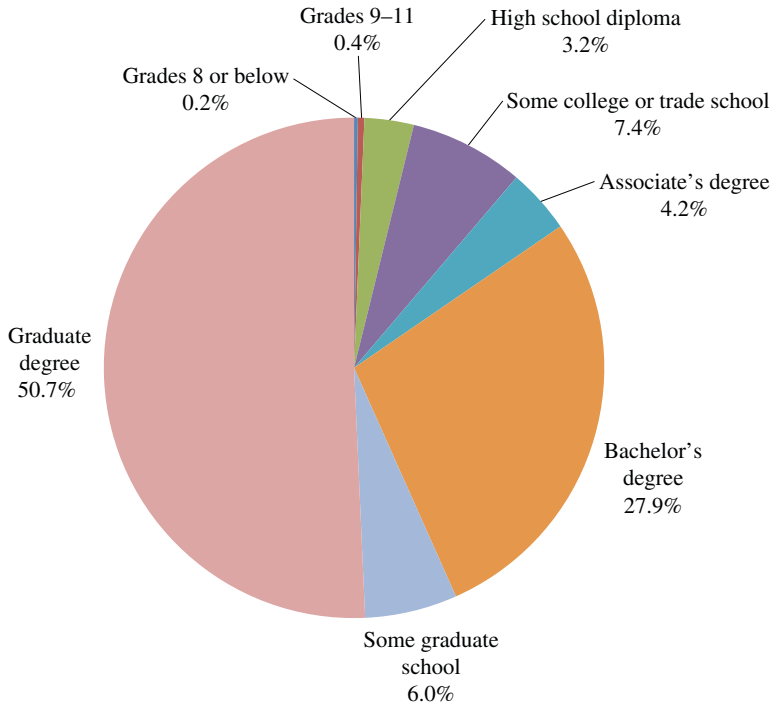
probably about 35,000 low-income high achievers in each cohort in the United States.[15]

Table 2 shows that among high achievers, those who are from higher-income families do have slightly higher college assessment scores, but the difference is small. The average low-income high achiever scores at the 94.1th percentile. The average high-income high achiever scores at the 95.7th percentile.

Data on the parental education of high achievers are unfortunately very incomplete, because ACT takers are not asked to report their parents' education, and 52 percent of SAT I takers fail to answer the question about their parents' education. Moreover, SAT I takers are apparently less likely to report their parents' education when it is low. We base this assessment on the observation that parents' education is more likely to be missing for students who live in Census block groups with low adult education. For what they are worth, however, the data on the parents' education are shown

15. We obtain these numbers by counting the number of high achievers whose estimated family income puts them in the bottom quartile of family income. We subtract a number corresponding to our false positive rate and add a number corresponding to our false negative rate. There are two reasons why this procedure gives us a range rather than an exact number. First, many high achievers appear in both the College Board and ACT data. We cannot definitively eliminate all of the duplicates because their names, addresses, and birthdates often do not exactly match in the two data sets. Eliminating all possible duplicates pushes us toward the lower bound. Second, although our false positive rate is robust to the aid data we use, our false negative rate is not. This is because the false negatives are low-income students who come from block groups where only a small percentage of families have low incomes. Our aid data from such block groups are fairly sparse, and we are therefore not confident about whether we can extrapolate the false negative rate to areas that appear similar but where we have never observed a false negative. Extrapolating pushes us toward the upper bound.

Figure 3. High-Achieving Students, by Parents' Educational Attainment[a]



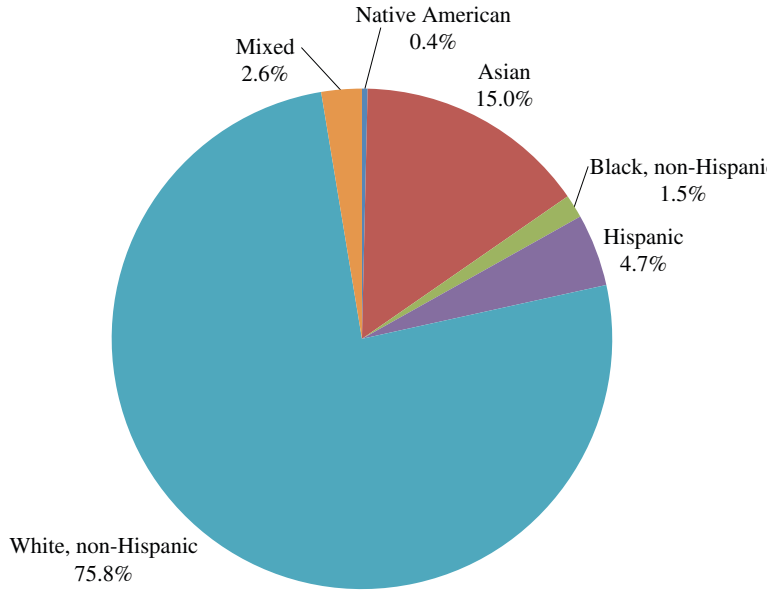
Source: Authors' calculations using the combined data set described in the text.

a. Parents' educational attainment is the highest level attained by either parent. Percentages are of those high-achieving students (defined as in table 2) who took a College Board test and answered the question about parents' education (61 percent of high achievers declined to answer; the ACT questionnaire does not include a similar question).

in figure 3.[16] More precisely, we show the greater of the father's reported educational attainment and the mother's reported educational attainment. Of students who report their parents' education, 50.7 percent say that at least one parent has a graduate degree, 27.9 percent say that at least one parent has a baccalaureate degree, and another 6 percent cite "some graduate school" (but no degree); 11.6 percent claim that at least one parent has an associate's degree or "some college or trade school" (but no degree), and only 3.8 percent report neither parent having more than a high school

16. We do not attempt to correct these data for biases because we do not have verified data on parents' education that we could use to estimate the errors accurately. This is in contrast to family incomes, where we do have a source of verified data (the CSS Profile).

Figure 4. High-Achieving Students, by Race and Ethnicity[a]



Source: Authors' calculations using the combined data set described in the text.

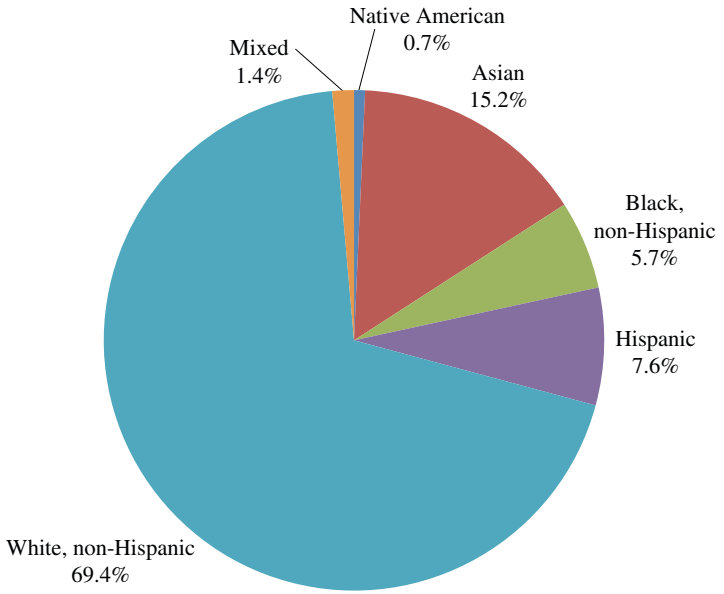
a. Percentages are of those high-achieving students (defined as in table 2) who took an ACT or a College Board test and answered the question about their race or ethnicity (2.1 percent of high achievers declined to answer).

diploma. Perhaps the most interesting thing about the parents' education data is that they seem to indicate that high achievers are reluctant to report that they have poorly educated parents. This is in contrast to the family income data from the same College Board questionnaire. Many students did not answer the income question, but those who did answered it in an unbiased (albeit fairly inaccurate) way.

Figure 4 displays information on high achievers' race and ethnicity, which 98 percent of students voluntarily report on the ACT or the College Board questionnaire. Of all high achievers, 75.8 percent say that they are white non-Hispanic, and another 15.0 percent say that they are Asian. The remaining 9.2 percent of high achievers are associated with an underrepresented minority,[17] either Hispanic (4.7 percent), black non-Hispanic

17. "Underrepresented minority" is the term of art in college admissions. Notably, it excludes Asians.

Figure 5. High-Achieving, Low-Income Students, by Race and Ethnicity[a]

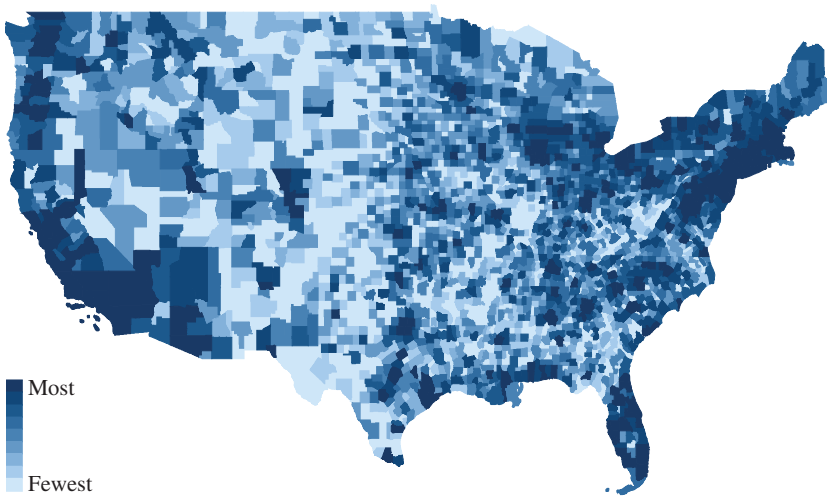


Source: Authors' calculations using the combined data set described in the text.
a. Percentages are of those high-achieving students (defined as in table 2) from bottom-quartile-income families who took an ACT or a College Board test and answered the question about their race or ethnicity (2.1 percent of all high achievers declined to answer).

(1.5 percent), Native American (0.4 percent), or mixed race/ethnicity (2.6 percent). If we focus on *low-income* high achievers only (figure 5), we see that 15.4 percent are underrepresented minorities. Interestingly, the entire increase in this share comes out of the percentage who are white. Asians make up 15.2 percent of low-income high achievers, almost identical to their share of all high achievers.

A key takeaway from figure 5 is that a student's being an underrepresented minority is not a good proxy for his or her being low-income. Thus, if a college wants its student body to exhibit *income* diversity commensurate with the income diversity among high achievers, it cannot possibly attain this goal simply by recruiting students who are underrepresented minorities. If admissions staff do most of their outreach to low-income students by visiting schools that are largely Hispanic and black, the staff should realize that this strategy may lead to a student body that is diverse on specific racial and ethnic dimensions but that is not diverse in terms of family income.

Figure 6. Numbers of High-Achieving Students, by County[a]

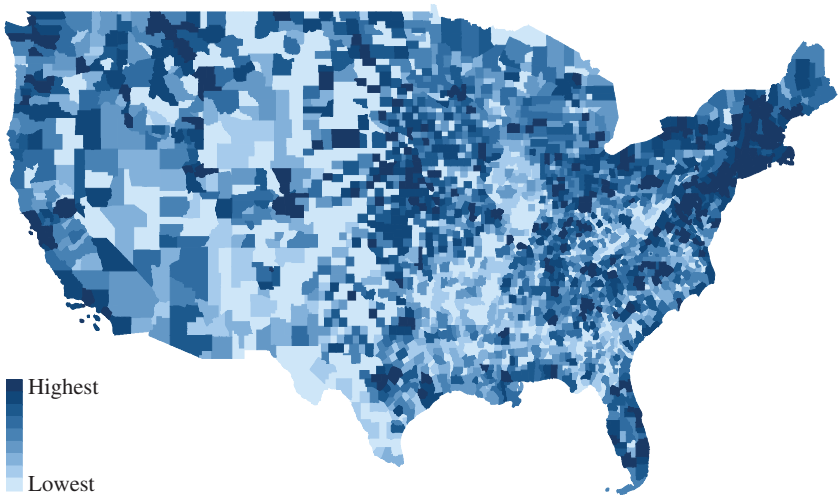


Source: Authors' calculations using the combined data set described in the text.
a. Counties are ranked by the absolute number of high achievers (as defined in table 2) living in the county in 2008 and then grouped into deciles; counties are then shaded according to their decile.

Figure 6 is a choropleth map showing the number of high-achieving students in each county of the United States. Counties are an imperfect unit of observation because some are large in land area and some are small. Nevertheless, they are the most consistent political units in the United States.[18] The darker is the county's coloring, the more high-achieving students it contains. What the map demonstrates is that critical masses of high-achieving students are most likely to be found in the urban counties in southern New England (Massachusetts, Connecticut, Rhode Island), the Mid-Atlantic (New York, New Jersey, eastern Pennsylvania), southern Florida, and coastal California from the Bay Area to San Diego. The other critical masses are more scattered, but a person familiar with U.S. geography can pick out Chicago (especially), Houston, Dallas-Fort Worth, Atlanta, and some smaller cities. In short, if one's goal were to visit every county where one could gather at least 100 high achievers, one could concentrate entirely on a limited number of cities on the East and West Coasts and a few cities in between.

18. That is, the size and scope of municipalities, school districts, and other jurisdictions are far *less* consistent than those of counties.

Figure 7. Shares of All 17-Year-Olds Who Are High-Achieving, by County[a]



Source: Authors' calculations using the combined data set described in the text.

a. Counties are ranked by the number of high achievers (as defined in table 2) living in the county in 2008 divided by the number of 17-year-olds living in the county, and then grouped into deciles; counties are then shaded according to their decile.

Some part of the above statement is due to the fact that high-income, highly educated parents are somewhat concentrated in the aforementioned areas, and such parents, as we have shown, are somewhat more likely to have high-achieving children. However, some part of the above statement is due purely to population density. That is, even if children in all counties were equally likely to be high-achieving, there would still be critical masses of them in densely populated counties, and vice versa. The choropleth map in figure 7 illustrates the role of population density by showing the number of high-achieving students *per 17-year-old* in each county. The darker a county is, the higher is its decile on this relative measure. The map makes it clear that this relative measure is far less concentrated than the absolute measure that favors densely populated counties. In fact, one can see a belt of counties that tend to produce high achievers running from Minnesota and the Dakotas south through Missouri and Kansas. A good number of counties in Appalachia, Indiana, and the West outside of coastal California also tend to produce high achievers. In short, if one's goal were to meet a nationally representative sample of high achievers, one's trip could not be concentrated on a limited number of counties on the coasts and a few cities in between.

IV. College Applications, Enrollment, and Degree Receipt among High-Achieving Students in the U.S.

In this section, we analyze the college application choices, enrollment decisions, and on-time degree receipt of high-achieving students in the United States, paying attention to how low-income students' behavior does or does not differ from that of high-income students. Because colleges in the United States are so varied and large in number, we characterize them by the college assessment score of their median student, expressed as a percentile of the national college assessment test score distribution. This statistic, although admittedly insufficient to describe colleges fully, has important qualities. First, it is probably the single best, simple indicator of selectivity—much better than a college's admissions rate, for instance (Avery, Glickman, Hoxby, and Metrick, 2013). Second, when an expert college counselor advises students on how to choose a portfolio of schools to which to apply, he or she usually tells students to apply to a few schools that are a "reach," four or more schools that are "peer" or "match," and one or more schools that are "safe." Similar advice is widely available on the Internet sites of college advising organizations with a strong reputation, including the College Board and the ACT. Expert college counselors use schools' median test scores to define "reach" schools (typically, those whose median score is more than 5 percentiles above the student's own), "peer" schools (typically, those where the school's median score is within 5 percentiles of the student's own), and "safety" schools (typically, those whose median score is 5 to 15 percentiles below the student's own).[19] Naturally, the exact cutoffs for these categories vary from expert to expert, and high-achieving students are often advised to apply to their state's public flagship university, even if it falls below the safety zone.[20] High-achieving students are generally advised to apply to at least eight schools.

19. Experts also advise students to look at the high school grade point average that is typical of a college's students. However, such grade-based categories are not terribly relevant to high-achieving students because selective colleges vary so much more on the basis of college aptitude test scores than on the basis of high school grades.

20. State flagship universities are something of a special case. On the one hand, they vary widely in selectivity. On the other hand, even flagships with low overall selectivity can create opportunities (formal or informal) for their highest-aptitude students to get an education oriented to students with their level of achievement. These opportunities may include research jobs and taking courses primarily intended for doctoral students.

IV.A. College Application Behavior: A Graphical Analysis

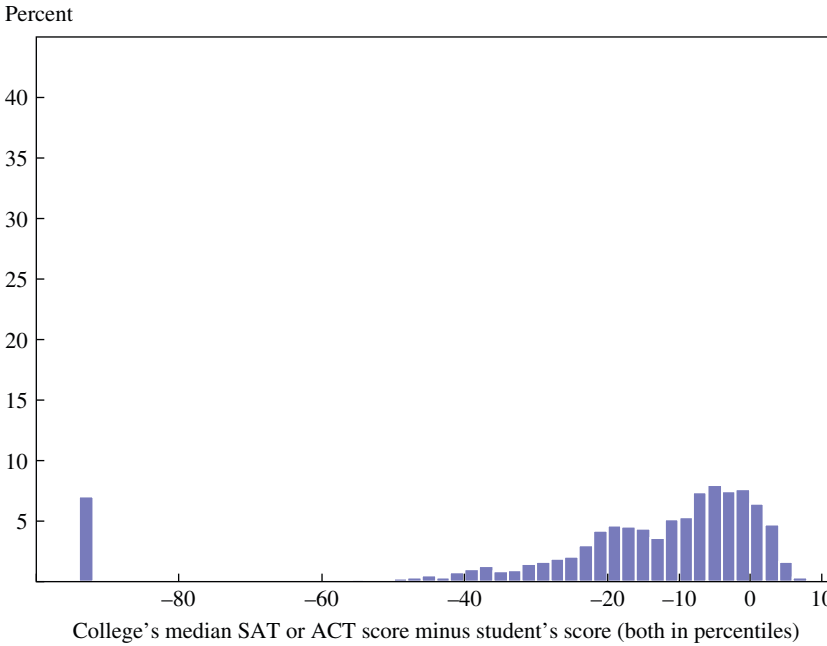
In this subsection, we provide graphical evidence of what students' application portfolios look like. This presentation is somewhat informal but useful for fixing ideas and defining categories before we move to the formal econometric analysis in the next subsection. In what follows, an "application" is defined as sending a test score to a college.[21]

Figure 8 is a histogram of the application portfolios of high-income students. It is important to understand how this and subsequent figures are constructed. On the horizontal axis is the difference between the applied-to college's median test score and the student's own score, in percentiles. Thus, if an application is located at zero, the student is applying to a peer school whose median student has exactly the same score. An application at, say, +8 is a reach, and an application at, say, −13 is a safety. Since nonselective colleges do not require their students to take college assessments (and thus do not report a median student score), an application to a nonselective school is placed at −94, which is zero minus the average percentile score of high-achieving students in the data. It is not obvious where to place applications to nonselective schools, but −94 has the advantage that such applications cannot be mistaken for applications to a school that is selective but that sets a very low bar.

Each student is given a weight of 1 in the histogram, and this weight is split evenly over that student's applications. This is to ensure that the histogram does not overrepresent the behavior of students who apply to more schools, since, after all, each student will enroll at just one (initially at least). Thus, if a student puts all of his or her eggs in one basket and applies to a single +8 school, that student's full weight of 1 will show up in the +8 bar. If a student applies to one +8 school, one +6 school, one +4 school, and so on down to one −8 school, one 9th of that student's weight will show up in each of the relevant bars. Note that each bar is 2 percentiles wide.

21. As noted above, a student may often apply to a nonselective college without sending scores, although a good number of students send scores to them for apparently no reason (the first few sends are free) or for placement purposes (that is, to avoid being placed in lower-level or even remedial courses). If we match students to their enrollment records in the National Student Clearinghouse, we can add to their set of applications any nonselective school in which they enrolled without sending scores. This does not change the figures much, although it does systemically raise the bar for nonselective applications. We do add applications in this way for the analysis in the second half of this section, but it makes too little difference here to be worthwhile, especially as we would then have to show figures for a sample of the students, rather than the population of them.

Figure 8. Distribution of High-Achieving, High-Income Students' College Applications, by Student-College Match[a]



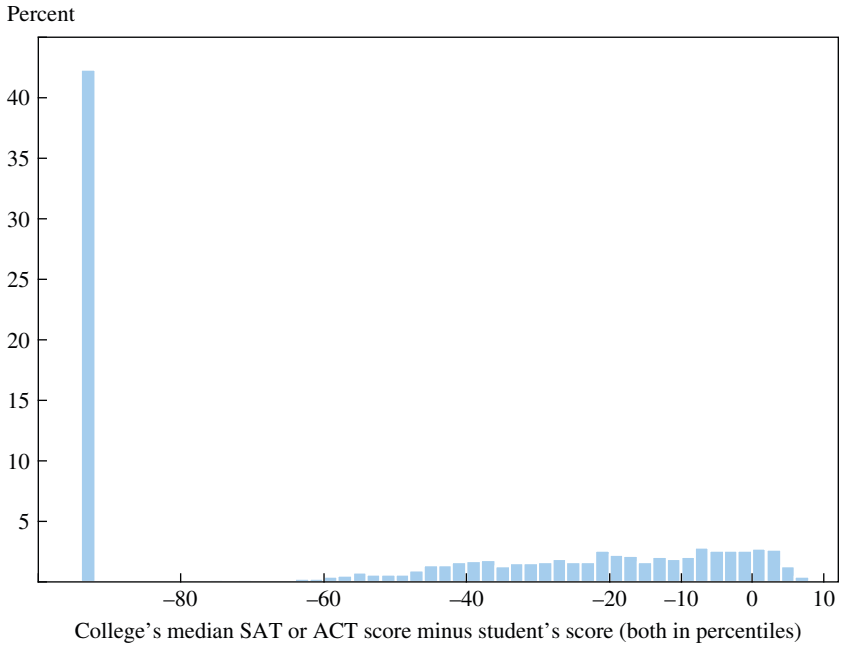
Source: Authors' calculations using the combined data set described in the text.

a. The heights of the bars are determined as follows. Each high-income high achiever is assigned a weight of 1, which is divided equally among his or her college applications. Each application is then placed in the vertical bar associated with the difference between the college's median test score and the applicant's score. The weights of all applications in a bar are added up, and the resulting sum is the height of the bar. The figure thus depicts the distribution of these students' applications in the aggregate among the colleges to which they applied. See the text for further details. High-achieving students are defined as in table 2; high-income students are those in the top quartile of the family income distribution (figure 1).

Figure 8 shows that high-income students largely follow the advice of expert counselors. The bulk of their applications are made to peer schools. They apply to some reach schools as well, but they are mechanically limited in the extent to which they can do this: there are *no* reach schools for slightly more than half of the high-achieving students we study.[22] High-income high achievers also apply fairly frequently to safety schools.

22. For instance, consider a student whose own scores put him or her at the 94th percentile. In order to apply to a reach school, he or she would need to apply to a school whose median student scored at the 99th percentile. There are no such schools—or at least no schools that admit to having such a high median score.

Figure 9. Distribution of High-Achieving, Low-Income Students' College Applications, by Student-College Match[a]



Source: Authors' calculations using the combined data set described in the text.
a. The figure is constructed in a manner analogous to figure 8.

Although not shown in the figure, it is noteworthy that many such students apply to their state's flagship university. These schools vary greatly in selectivity, so that some such applications are in the safe range, but other applications to flagships appear far more safe than anyone would think necessary. For instance, an application by a high achiever to a flagship with a median score at the 50th percentile would end up at –40 to –50. Nevertheless, applying to these schools may be well-advised (see note 20).

The reader might be surprised to find that high-achieving, high-income students apply to some colleges that are nonselective on academic grounds. However, the schools in question are often specialty schools: music conservatories, art or design schools, drama or performing arts schools, cooking schools, and so on. Some of these are highly selective on nonacademic dimensions.

Figure 9 shows that unlike the high-income high achievers, few low-income high achievers follow the advice of expert counselors. More than

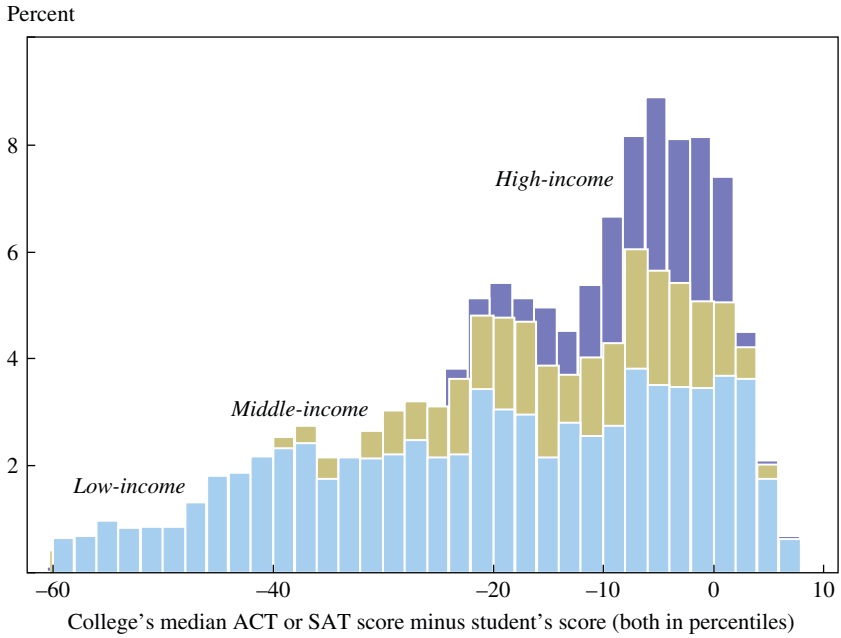
40 percent of the mass in the histogram loads on nonselective schools. (This is an underestimate because scores are not sent to some nonselective schools. If we included every nonselective enrollment as a nonselective application, the nonselective bar on the histogram would rise by 5.1 percentage points.)[23] Moreover, the nonselective colleges to which low-income students apply are rarely of the specialty type mentioned above. They are often local community colleges or local 4-year institutions with meager resources per student and low graduation rates. Much of the height of the nonselective bar is due to the fact that many low-income high achievers apply only to nonselective colleges, or to a nonselective college and a barely selective college.

Figure 10 overlays the histograms for low-income, middle-income, and high-income students who are high-achieving. It cuts off the portion of the histogram that shows nonselective colleges so as to focus on application choices among colleges that are selective to at least some degree. It will be observed that the behavior of the middle-income students (those from families in the two middle quartiles of the family income distribution) is about midway between that of their low- and high-income counterparts. Moreover, even within the subset of applications that are made to selective colleges, high-income students apply much more to peer colleges, and low-income students apply much more to colleges far below the safety level.

Figure 11 contains four panels. The top left-hand panel shows, for all high-achieving, low-income students, the histogram of the most selective college to which each student applied. The top right-hand panel shows the same histogram for high-achieving, high-income students. The bottom left-hand panel shows the histogram for the *second* most selective college to which a low-income student applied (or the most selective, for students who applied to a single college). The bottom right-hand panel shows the same histogram for high-income students. These histograms reveal that the vast majority of high-income high achievers' most selective applications fall within 10 percentiles of their test scores. Their second most selective applications are sent to less competitive, but not much less competitive schools: the vast majority fall between +10 and −15 percentiles. In contrast, low-income high achievers send their most selective applications to the entire

23. We do not treat the sending of no scores as equivalent to applying to no selective institution. The reason is that a student may send no scores because he or she takes both the SAT and the ACT and prefers to send the scores from only one of the two tests. Since we cannot definitively match students across the two data sources (see note 15), we cannot assume that no-score-sending corresponds to no selective applications.

Figure 10. Distribution of All High-Achieving Students' College Applications to Selective Institutions, by Student-College Match[a]



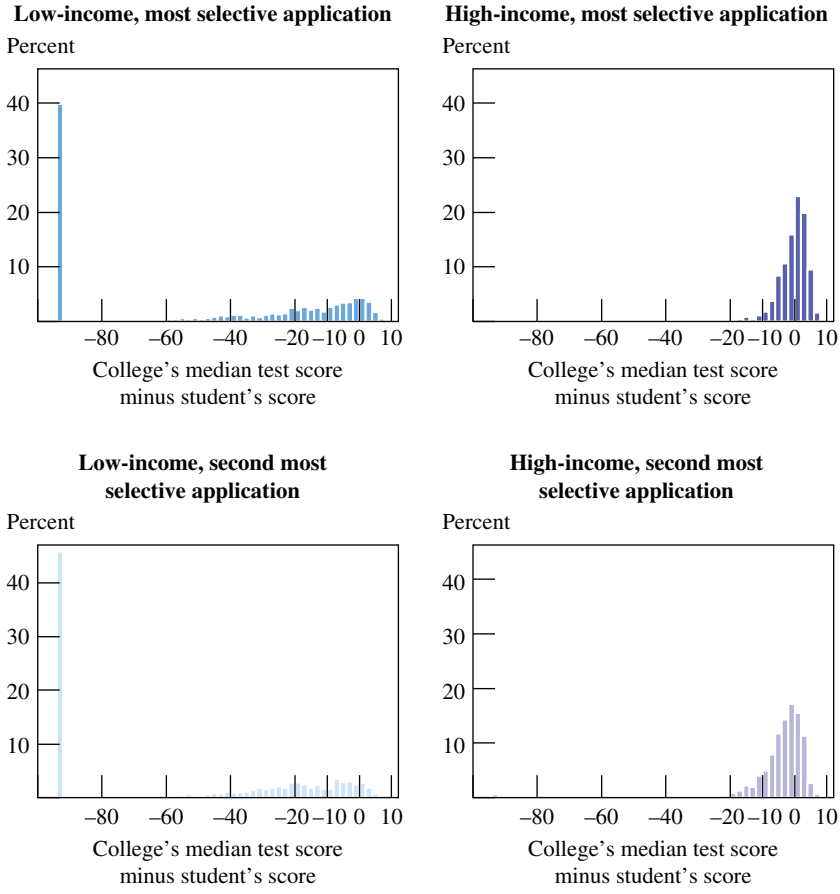
Source: Authors' calculations using the combined data set described in the text.

a. The figure is constructed in a manner analogous to figure 8, truncating the left tail of the distribution.The bars for middle-income students are to be read as extending downward to zero behind the low-income bars, and the high-income behind the middle-income.

range of colleges: nonselective and –60 to +10. Their second most selective applications are, again, to less competitive (but not necessarily much less competitive) schools. All of this suggests that there may be two distinguishable types of low-income high achievers: those who apply much as their high-income counterparts do, and those who apply in a manner that is very different.

In fact, 53 percent of low-income high achievers fit the profile we will hereafter describe as income-typical: they apply to no school whose median score is within 15 percentiles of their own, and they do apply to at least one nonselective college. At the other extreme, 8 percent of low-income high achievers apply in a manner that is similar to what is recommended and to what their high-income counterparts do: they apply to at least one peer college, at least one safety college with a median score not more than 15 percentiles lower than their own, and apply to no nonselective colleges.

Figure 11. Distributions of High-Achieving Students' Most Selective and Second Most Selective College Applications, by Family Income[a]



Source: Authors' calculations using the combined data set described in the text.
a. Each panel is constructed in a manner analogous to figure 8.

We hereafter designate such students as achievement-typical, noting that once a student fits the above criteria, he or she usually applies to several peer colleges, much as high-income students do.

The remaining 39 percent of low-income, high achievers use application strategies that an expert would probably regard as odd. For instance, we see some students apply to only a local nonselective college and one extremely selective and well-known college—Harvard, for instance. No expert would advise such a strategy because the probability of getting into

an extremely selective, well-known college is low if a student applies to just one—even if the student's test scores and grades are typical of the college's students. Moreover, such a strategy reveals that the student is interested in extremely selective institutions yet is not applying to the other schools that are, for most purposes, indistinguishable from the one to which he or she applied. Another strategy that appears is a student applying to a single public college in his or her state that is selective but is much less selective than the state's flagship university. Although about half of these application choices could be motivated by distance from home, the other half cannot because the flagship university is nearer. Another strategy that falls into the idiosyncratic category is a student applying to a single private college outside his or her state that *is* selective, but much less selective and much poorer in resources than the student's private peer colleges would be. Such choices are odd because although the private peer colleges might offer fewer scholarships that are explicitly merit-based, they offer much more generous need-based aid, so that the student would pay less to attend and would enjoy substantially more resources. Furthermore, it is almost never sensible for a low-income student to apply to a single private, selective college: such a student can use competing aid offers to improve the aid package at his or her most preferred college.

We have described a few salient strategies that appear among low-income high achievers who are neither achievement-typical nor income-typical. However, most of these students' portfolios do not evince any pattern that can be readily described. Thus, below we turn to an econometric analysis, in which we can simultaneously consider a large number of factors correlated with students' application choices.

IV.B. College Application Behavior: An Econometric Analysis

In this subsection we assess the factors that are associated with a student's choice of his or her application portfolio, using a conditional logit model in which a student can apply to all colleges in the United States but decides to apply only to some. This model is based on a random utility framework and assumes that the student prefers all colleges to which he or she applies over the colleges to which he or she does not apply. We do not assume anything about the student's preference ordering within the colleges to which he or she applied.[24] Each possible college matched with

24. We considered estimating a rank-ordered logit model (Beggs, Cardell, and Hausman 1981), on the assumption that the order in which the student sent scores to colleges indicates

each student is an observation, and the dependent variable is a binary variable equal to 1 if the student submits an application to the college and zero otherwise.

The explanatory variables we consider are the difference between a school's median test score and the student's own test score if positive, the same difference if negative,[25] an indicator for the school's being nonselective, the distance between the student's home and the school, the square of this distance, an indicator for the school being the most proximate, an indicator for the school being public, an indicator for the school being in-state for the student, an indicator for the school being the flagship university of the student's state of residence, the sticker price of the college, the likely net cost of the college for the student, and the student-oriented resources per student at the college. We fully interact these explanatory variables with indicators for the student being low-income, high-income, or in between. Thus, we estimate separate coefficients for each income group. In the tables we do not show the coefficients for the middle-income group because they nearly always fall between those of the high- and low-income students, but the coefficients are available upon request.

Table 3 shows the results of this estimation. The coefficients are expressed as odds ratios so that a coefficient greater than 1 means that an increase in the covariate is associated with an increase in the probability that the student applies to the school, all other covariates held constant. Based on our graphical analysis, we expect to find very different coefficients for low- and high-income students, and we do.[26] Note that, although it is convenient to describe the coefficients as though they literally revealed

---

the rank order of his or her preference among them. (All colleges to which no application is sent are assumed to generate net utility below the bottom-ranked college.) If we do this, the rank-ordered logit generates fairly similar results, in part because many students do not send scores to more than a few colleges. However, the order of score sending might be a poor proxy for some students' preference orderings because they choose a first batch of colleges to receive their scores before they know what those scores are. Once they learn their scores, they choose a second batch of colleges to receive their scores. At application time, they presumably prefer the second batch to the first.

25. That is, we do not assume that the response of a student to mismatch is symmetric around his or her own test score. A student may only slightly like being at a reach school, for instance, but strongly dislike being at a safety school.

26. In Avery and Hoxby (2004), we found much smaller differences in the behavior of low- and high-income students, but all the students we sampled attended high schools that were at least somewhat reliable feeders. As we will show, the low-income students we sampled were thus very disproportionately what we call "achievement-typical" students who *do* behave fairly similarly to high-income students.

Table 3. Conditional Logit Regressions Explaining High-Achieving Students' College Applications[a]

Factor	*Low-income students*	*High-income students*
College is a peer school[b]	1.015	76.214***
College is a safety school[c]	3.009***	14.895***
College is nonselective	0.748***	1.6e-9***
Tuition before discount (thousands of dollars)	0.865***	1.176***
Average tuition discount (percent)	1.091**	0.925**
Could live at family home (college is <10 miles away)	4.942***	0.810***
Could go home often (college is <120 miles away)	1.556***	1.185***
Distance in miles to college	0.996	0.998
Square of (distance in miles/1,000)	1.056**	1.283***
College is in-state	2.595***	1.206***
College is private	0.838***	1.002
College is for-profit	0.834***	0.012***
Highest degree offered is 2-year	0.925**	0.009***
College is a university	0.997	0.567***
College is a liberal arts college	0.717***	0.973*

Source: Authors' regressions using the combined data set described in the text.

a. Results of a conditional logit estimation in which the dependent variable is an indicator equal to 1 if a high-achieving student applies to the college and zero otherwise. Coefficients are expressed as odds ratios, so that a coefficient greater than 1 means that an increase in the covariate is associated with an increase in the probability that the student applies to a college with the indicated factor, all other covariates held constant. High-achieving students are defined as in table 2. Low- and high-income students are those from families in the bottom and top quartiles of the family income distribution, respectively. Asterisks indicate statistical significance at the *10 percent, **5 percent, or ***1 percent level.

b. The absolute value of the difference between the college's median test score and the student's own is within 5 percentiles.

c. The college's median score is 5 to 15 percentiles below the student's own

preference, they should not be given such a strong interpretation or a causal interpretation. For instance, students might "disfavor" distance not because distance itself generates negative utility but because distant schools have, say, distinct cultures that the student dislikes.

We find that high-income students strongly favor reach colleges and disfavor safety colleges (those for which the score difference is negative). Per percentile of difference, this effect is much stronger on the reach side than on the safety side, but recall that high-achieving students can only reach a bit whereas they can apply to *very* safe schools. High-income students strongly dislike nonselective institutions. They also dislike higher net costs but (all else equal) like higher sticker prices. This is probably because higher sticker prices are associated with higher per-student resources, a characteristic they also like. High-income students dislike distance, but the quadratic term indicates that they dislike it only up to a point, after which

they are fairly indifferent. They have a mild preference for in-state schools and their state's flagship university. They do not have a statistically significant preference for publicly controlled schools.

The low-income students exhibit several immediate contrasts. Such students strongly favor nonselective colleges. This was obvious in the graphical evidence. They do not disfavor schools whose median scores are lower than theirs. They slightly disfavor schools with higher sticker prices (recall that these were favored by high-income students) and do not have a preference for net costs that is statistically significantly different from zero. Low-income students do favor schools with higher expenditure per student, but not nearly as much as high-income students do. Distance is strongly disfavored for schools within 100 miles but, thereafter, low-income students are fairly indifferent to it. Low-income students favor in-state schools somewhat more than high-income students do, but low-income students do not exhibit a preference in favor of their state's flagship university. They slightly favor publicly controlled colleges.

Table 4 repeats the estimation but interacts the covariates with indicators for high-income students, middle-income students, low-income achievement-typical students, low-income income-typical students, and other low-income students. The estimated coefficients for achievement-typical students are fairly similar to those for high-income students. It is the income-typical students whose coefficients are strikingly different. Of course, these results are somewhat by design, given the way we categorized low-income students into achievement-typical and income-typical groups. However, the coefficients validate the categorization: achievement-typical students *do* pursue similar application strategies to high-income students. In the next section we assess which factors predict a student being achievement-typical and which predict a student being income-typical.

IV.C. College Enrollment and Progress toward a Degree

In this subsection, we demonstrate that, *conditional on applying to a specific college,* high- and low-income students thereafter behave similarly. There is no statistically significant difference in their probability of enrolling or in their progress toward a degree.

To find the first of these results, we estimate a conditional logit model in which the binary outcome is 1 for the college in which the student initially enrolled and zero for all others. Importantly, we limit the student's choice set to the colleges to which he or she applied. So that the student's enrollment decision is compared to those of students who applied to the same college, we include a fixed effect for each college. We also include

Table 4. Conditional Logit Regressions Explaining Income-Typical and Achievement-Typical Students' College Applications[a]

	Low-income students		
Factor	*Income-typical*[b]	*Achievement-typical*[c]	*High-income students*
College is a peer school	7.21e-8***	87.808***	76.214***
College is a safety school	2.142***	19.817***	14.895***
College is nonselective	0.795***	1.04e-8***	1.6e-9***
Tuition before discount (thousands of dollars)	0.973***	1.004	1.176***
Average tuition discount (percent)	1.000	1.020*	0.925**
Could live at family home (college is <10 miles away)	5.140***	1.477***	0.810***
Could go home often (college is <120 miles away)	1.972***	1.436***	1.185***
Distance in miles to college	0.999	0.999	0.998
Square of (distance in miles/1,000)	1.042*	1.448***	1.283***
College is in-state	4.891***	7.455***	1.206***
College is private	0.662***	0.296***	1.002
College is for-profit	0.806***	0.001***	0.012***
Highest degree offered is 2-year	0.855***	0.016***	0.009***
College is a university	0.956**	0.861***	0.567***
College is a liberal arts college	0.515***	0.167***	0.973*

Source: Authors' regressions using the combined data set described in the text.

a. Results of a conditional logit regression in which the dependent variable is an indicator equal to 1 if a high-achieving student applies to the college and zero otherwise. Coefficients are expressed as odds ratios, so that a coefficient greater than 1 means that an increase in the covariate is associated with an increase in the probability that the student applies to a college with the indicated characteristic, all other covariates held constant. The coefficients for high-income students are repeated from table 3 for ease of comparison. High-achieving students are defined as in table 2. Low- and high-income students are those from families in the bottom and top quartiles of the family income distribution, respectively. Asterisks indicate statistical significance at the *10 percent, **5 percent, or ***1 percent level.

b. Those who apply to no school whose median score is within 15 percentiles of their own and apply to at least one nonselective school.

c. Those who apply to at least one peer college, at least one safety college with a median score not more than 15 percentiles lower than their own, and no nonselective colleges.

interactions between these fixed effects and an indicator for a student's having high or low income. We then test whether each college's high-income or low-income interaction is statistically significantly different from zero. Thus, we test, specifically, whether high- and low-income students who apply to the same college are differentially likely to enroll in it.

We also estimate a variant of this model in which we include an indicator variable for each number of colleges to which the student applied: 1 college, 2 colleges, and so on up to 20 or more colleges. This variant tests whether a high- and a low-income student who apply to the same college

and the same *number* of colleges are differentially likely to enroll in the college in question.

Because so few high-income students apply to nonselective and low-selectivity colleges, many of the high-income × college interactions are dropped by the regression. Therefore, the effect of income on enrolling in such colleges, conditional on having applied, is not identified.

Note that the tests subsume colleges' admissions decisions. That is, if we find that high- and low-income students are equally likely to enroll in a college, conditional on having applied to it and to the same number of colleges, they must be getting treated similarly in the admissions process. Otherwise, they would enroll differentially simply because they had been admitted differentially.[27] Moreover, if we find that high- and low-income students are equally likely to enroll in a college, conditional on having applied to it (regardless of the number of colleges to which they applied), not only must they be treated similarly in the admissions process, but they must also typically apply to the same number of colleges.[28]

Table 5 shows the results from these estimations. The table is organized by colleges' median test scores, with more selective colleges closer to the top. We find that only very small shares of low- and high-income enrollment probabilities (conditional on applying) are statistically significantly different from one another at the 5 percent level. For instance, low-income enrollment probabilities differ from high-income enrollment probabilities in only 4 percent of the colleges that have median scores at the 90th percentile or above. This is about what one would expect from a test at the 5 percent level. The remaining rows of the table contain similar results, all suggesting that low- and high-income students do not enroll differentially, conditional on applying. The results are very similar when the estimation includes an indicator for each number of colleges to which a student applies.

Our test for differential progress toward a degree, conditional on the school at which a student initially enrolled, is constructed in an analogous way. The dependent variable is now the percentage of coursework toward a

27. We can interact additional student characteristics that might affect admission—for instance, race and ethnicity—with colleges' fixed effects. This effectively "soaks up" each college's preferential admissions standards. However, such a specification does not change the estimated coefficients of interest to a noticeable extent, and it makes interpretation slightly harder.

28. This is a somewhat subtle test of whether the achievement-typical students have total application portfolios like those of high-income high achievers.

Table 5. Estimates Showing Whether High-Achieving, Low- and High-Income Applicants Have Different Probabilities of Enrolling in Selective Colleges[a]

	Percent of colleges where low- and high-income students' probabilities of enrolling (conditional on application) are statistically significantly different at the 5 percent level	
College's median test score	*Base specification*[b]	*Base specification plus indicator variables for number of applications sent*[c]
≥90th percentile	4	5
≥80th but <90th percentile	5	5
≥70th but <80th percentile	4	5
≥60th but <70th percentile	3	4
≥50th but <60th percentile	6	5
<50th but college is selective	Not identified[d]	
College is nonselective	Not identified	

Source: Authors' calculations using the combined data set described in the text.

a. Results of a conditional logit estimation in which the dependent variable is an indicator equal to 1 if a high-achieving student enrolls in a particular postsecondary institution and zero otherwise. Each student's choice set is the set of colleges to which he or she applied. High-achieving students are defined as in table 2. Low- and high-income students are those from families in the bottom and top quartiles of the family income distribution, respectively.

b. The only independent variables are indicators for each college interacted with an indicator for whether the student is high- or low-income.

c. Indicator variables for whether the student applied to 1 college, 2 colleges, and so on up to 20 or more colleges are added to the specification in the previous column.

d. Results are not identified for low-selectivity and nonselective colleges because too few high-income students apply to such colleges.

4-year degree that the student appears to have completed as of June 2012.[29] A student who is making on-time progress should have completed 100 percent of his or her coursework by then. We estimate a fixed effect for every college so that students are compared to others who enrolled in the same school. We interact the fixed effects with high- and low-income indicators, and we test whether these interactions are statistically significantly different. Again, the effects for nonselective and low-selectivity colleges are not identified because so few high-income students enroll in them.

The left-hand column of table 6, which is organized in much the same way as table 5, shows the results from this estimation. For selective colleges, we find that only very small shares of colleges have statistically

29. We do not consider progress toward a 2-year degree because virtually none of the high-achieving students reported that a 2-year degree was their educational goal in the descriptive questionnaires that accompany the ACT and SAT I tests.

Table 6. Estimates of Whether Low- and High-Income Students Have Different Probabilities of Persisting at a Selective College, Conditional on Having Enrolled[a]

	Percent of colleges where low- and high-income students' shares of credits earned toward a degree (conditional on enrollment) are statistically significantly different at the 5 percent level	
College's median test score	*Base specification*[b]	*Excluding students from selective and magnet high schools*[c]
≥90th percentile	5	4
≥80th but <90th percentile	4	5
≥70th but <80th percentile	4	5
≥60th but <70th percentile	5	5
≥50th but <60th percentile	4	4
<50th but college is selective	Not identified[d]	
College is nonselective	Not identified	

Source: Authors' calculations using the combined data set described in the text.

a. Results of an ordinary least squares regression in which the dependent variable is the share of credits toward a baccalaureate degree earned by a student by June 2012. Students who do not enroll in a postsecondary institution are not included in the regression. High-achieving students are defined as in table 2. Low- and high-income students are those from families in the bottom and top quartiles of the family income distribution, respectively.

b. The only independent variables are indicators for each college interacted with an indicator for whether the student is high- or low-income.

c. Same specification as in the previous column, but students who attended high schools classified as magnet schools or that select incoming students on the basis of test scores or grades are excluded.

d. Results are not identified for low-selectivity and nonselective colleges because too few high-income students apply to such colleges.

significant differences between the progress of their low- and of their high-income students. For instance, low-income students' progress toward a degree differs from high-income students' progress toward a degree at only 5 percent of the colleges that have median scores at the 90th percentile or above. This is what one would expect from a test at the 5 percent level.

The right-hand column of table 6 reports results of reestimating the model excluding low-income students who attend selective and magnet high schools. The reestimation addresses the possibility that achievement-typical students perform well in college because, although poor, they attended high schools that offer unusually strong preparation. (This is true of some but not most achievement-typical students, as shown below.) We obtain very similar results.

There are two key takeaways from this subsection. First, the *application stage* is where interesting differences appear in the behavior of high-income high achievers and low-income high achievers. If they apply to the

same colleges, their educational paths are similar afterward. Thus, interventions that could make low-income high achievers' college careers look more like those of their high-income counterparts must, as a logical matter, be focused on the application stage or preparation for it. Second, the data do not suggest that low-income students who currently fail to apply to selective colleges and therefore fail to attend one would be rejected or would perform badly if they were admitted and enrolled. Of course, we cannot say that they would do just as well as the low-income students who do apply. One would need to induce low-income students to apply to substantially more selective schools and then estimate causal effects to make such a claim. We do not attempt to do that in this paper.[30] However, we are certainly not struck by evidence that low-income students have poor outcomes when they apply to selective schools.

V. Factors That Predict a Student's Being Achievement-Typical or Income-Typical

In this section, we use simple descriptive statistics to identify some factors that predict whether a low-income student is achievement-typical or income-typical. Our goal in this section is to characterize the two types of low-income students sufficiently well that we can build hypotheses about why they apply to colleges so differently.

Ex ante, our hypotheses fall into three broad categories:

(i) Despite the fact that both income-typical and achievement-typical students have estimated family incomes in the bottom quartile, income-typical students are actually socioeconomically disadvantaged compared to achievement-typical students when we examine their backgrounds more carefully. Given their greater disadvantage, they cannot be expected to behave similarly.

(ii) Income-typical students are likely to be poorly informed about college compared to achievement-typical students.

(iii) Income-typical students are making rational, well-informed choices about college. Their utility from attending nonselective or less selective colleges exceeds the utility they would derive from attending more selective colleges.

30. Hoxby and Turner (2013) implement exactly the causal test needed by inducing income-typical students to apply to substantially more selective institutions. That study finds no evidence that the students thus induced fail to be admitted at normal rates, fail to attain normal grades, or fail to persist at normal levels.

Table 7. Socioeconomic Characteristics of High-Achieving Students[a]

Characteristic	*High-income students*	*Low-income students*	
		Achievement-typical	*Income-typical*
Annual family income (dollars)[b]	157,569	30,475	32,418
Parents' education (years)[c]	18.7	16.0	16.7
Race or ethnicity[d] (percent of total)			
White	74.8	45.1	79.5
Black	2.1	5.2	2.9
Hispanic	5.6	12.6	6.0
Asian	20.5	31.8	7.3

Source: Authors' calculations using the combined data set described in the text.

a. High-achieving students are defined as in table 2. Low- and high-income students are those from families in the bottom and top quartiles of the family income distribution, respectively. Achievement-typical and income-typical students are defined as in table 4.

b. Estimated as described in section II.

c. Highest level of education attained by either parent, as reported by the student. Such self-reporting of parental education is unreliable because students may be more likely not to report if their parents' educational attainment is low.

d. Self-reported.

We can look for evidence of hypotheses in categories (i) and (ii). The hypothesis in category (iii) is inherently untestable, so it is effectively the residual explanation if there is no evidence for other hypotheses. Note that if hypothesis (iii) is the true one, students need not get more utility from attending a nonselective college because it is a good academic match for them. A student might attend a school that is obviously a poor academic match because it enables him or her, say, to look after his or her family. The student might derive sufficient utility from doing this so that his or her college choice is utility maximizing. Cultural and social factors that deter students from applying would also fall under hypothesis (iii). For instance, a student might feel that he or she would enjoy a better social life if he or she attended school with people from a very similar background.

Table 7 reports statistics on several characteristics of the families of high-achieving students that might reveal that income-typical students are truly socioeconomically disadvantaged relative to achievement-typical students. These statistics tend to go the wrong way for hypotheses of type (i). Income-typical students have slightly higher estimated family income than achievement-typical students do. Their (admittedly very flawed) reports of parents' education suggest that income-typical students' parents might have 0.7 years more of education than those of achievement-typical students. Achievement-typical students are more likely to be black or Hispanic, so

Table 8. Socioeconomic Characteristics of the Neighborhoods of High-Achieving Students[a]

		Low-income students	
Characteristic[b]	*High-income students*	*Achievement-typical*	*Income-typical*
Annual family income (dollars)	123,684	32,142	31,767
Adjusted gross income (dollars)[c]	121,448	41.358	37,652
Residents with a B.A. degree			
Number	863	207	
Percent of all adults	66.7	22.0	16.8
Race or ethnicity (percent of total)			
White	86.7	58.2	77.1
Black	2.6	12.8	10.1
Hispanic	4.1	16.9	8.7
Asian	9.2	8.5	2.2

Source: Authors' calculations using the combined data set described in the text.

a. High-achieving students are defined as in table 2. Low- and high-income students are those from families in the bottom and top quartiles of the family income distribution, respectively. Achievement-typical and income-typical students are defined as in table 4.

b. Neighborhoods are census block groups except where noted otherwise.

c. Neighborhood is the ZIP code.

they are presumably more, not less, likely to have experienced discrimination or to expect to experience it in college.

Table 8 reports statistics on several neighborhood factors that are useful for assessing hypotheses of both types (i) and (ii). A person's neighbors reveal something about his or her own socioeconomic disadvantage, but they also reveal something about the information he or she is likely to encounter. The statistics show that income-typical and achievement-typical students live in Census block groups with very similar average family income. However, achievement-typical students' block groups are less white, and more black, Hispanic, and Asian than those of income-typical students. Achievement-typical students also have more baccalaureate degree holders in their block groups, both in absolute number (207 versus 144) and as a share of adults (22.0 percent versus 16.8 percent). This last fact suggests that income-typical students may be less likely to get advice about college from a neighbor with a degree.

Table 9 compares the geography of income-typical and achievement-typical students, and the contrast is striking. Sixty-five percent of achievement-typical students live in the main city of an urban area, whereas only 30 percent of income-typical students do. Even within main city residents, achievement-typical students are much more likely to live in a large

Table 9. Types of Communities Where High-Achieving Students Reside[a]
Percent

		Low-income students	
Community type	*High-income students*	*Achievement-typical*	*Income-typical*
Main city, urban area with population > 250,000	17	26	8
Main city, urban area with population 100,000–250,000	14	21	13
Main city, urban area with population < 100,000	48	18	9
Suburb, urban area with population > 250,000	8	9	9
Suburb, urban area with population 100,000–250,000	0	2	2
Suburb, urban area with population < 100,000	0	4	12
Town, near an urban area	0	5	12
Town, far from an urban area	5	7	13
Rural, near an urban area	6	4	10
Rural, far from an urban area	0	5	10

Source: Authors' calculations using the combined data set described in the text.
a. High-achieving students are defined as in table 2. Low- and high-income students are those from families in the bottom and top quartiles of the family income distribution, respectively. Achievement-typical and income-typical students are defined as in table 4.

urban area (one with population greater than 250,000). Indeed, 70 percent of the achievement-typical students come from just 15 metropolitan areas (out of 334 nationwide): San Francisco, Oakland, Los Angeles, San Diego, Dallas, Houston, Chicago, Cleveland, Pittsburgh, Portland (Maine), Boston, Providence, New York, Philadelphia, and Baltimore.[31]

Only 21 percent of achievement-typical students live in a nonurban area (not necessarily rural, but a town rather than an urban-area suburb). By contrast, 47 percent of income-typical students live in a nonurban area. Put another way, income-typical students tend to be the high achievers who live in counties that have a large number of high achievers *per 17-year-old* (figure 7) but not a large number of achievers in absolute terms (figure 6).

Using administrative data from the U.S. Department of Education, table 10 compares the high schools attended by income-typical and achievement-

31. There were 334 metropolitan statistical areas and primary metropolitan statistical areas in the 2000 Census of Population.

Table 10. Characteristics of High Schools Attended by High-Achieving Students[a]

Characteristic	*High-income students*	*Low-income students*	
		Achievement-typical	*Income-typical*
Number of students per cohort	333	330	241
Type of school (percent of total)			
Regular public school	66	73	86
Magnet school	4	11	0
Independent private school	16	7	3
Catholic or other religious school	15	9	11
Spending per pupil (dollars; public schools only)	15,558	12,975	10,701
Pupil-teacher ratio	16.8	18.3	17.2
Pupil-counselor ratio (public schools only)	307	328	305

Source: Authors' calculations using the combined data set described in the text.

a. High-achieving students are defined as in table 2. Low- and high-income students are those from families in the bottom and top quartiles of the family income distribution, respectively. Achievement-typical and income-typical students are defined as in table 4.

typical students. These statistics should help us to assess these students' academic disadvantages and the amount of college-related information they might obtain at school. Achievement-typical students are considerably more likely to attend a school that is classified as a magnet school or an independent (as opposed to religious) private school. These statistics certainly understate the extent to which the achievement-typical students attend high schools that admit students on the basis of exams or grades. Chester Finn and Jessica Hockett (2012) find that only a small share of such high schools are classified as magnet schools.[32] Spending per pupil at achievement-typical students' public schools is higher, but since facilities and staff costs are often higher in the urban areas where they tend to live, it is unclear whether the higher spending actually gives them an advantage. Pupil-teacher and pupil-counselor ratios are fairly similar for achievement-typical and income-typical students: 18.3 versus 17.2, and 328 versus 305.

Using survey data from the Schools and Staffing Surveys from 1987 to 2007 and data on previous cohorts from the College Board and ACT,

32. Finn and Hockett found most of the selective high schools in their study by word of mouth and by contacting all high schools that were so dissimilar to other schools in their district that they seemed likely to practice selective admissions. Interestingly, many school districts deemphasize the existence of their selective high schools, which can be controversial. This perhaps explains why there was no reasonably accurate list of them before Finn and Hockett (2012).

Table 11. College-Related Characteristics of High Schools Attended by High-Achieving Students[a]

		Low-income students	
Characteristic	*High-income students*	*Achievement-typical*	*Income-typical*
Percent of teachers who graduated from a peer college[b]	8.9	2.9	1.1
Percent of teachers who graduated from a safety college[c]	14.4	7.5	5.0
Number in a typical previous cohort who applied to top 10 U.S. colleges[d]	12.9	7.6	1.6
Number in a typical previous cohort who were admitted to a top 10 U.S. college[d]	12.3	7.4	1.5
Number in a typical previous cohort who enrolled at a top 10 U.S. college[d]	12.3	7.4	1.5
Percent of cohort who are high achievers	17.1	11.2	3.8
Radius to gather 20 high achievers (miles)	2.6	7.7	19.3
Radius to gather 50 high achievers (miles)	4.1	12.2	37.3

Source: Authors' calculations using the combined data set described in the text.

a. High-achieving students are defined as in table 2. Low- and high-income students are those from families in the bottom and top quartiles of the family income distribution, respectively. Achievement-typical and income-typical students are defined as in table 4.

b. A peer college is one where the college's median test score is within 5 percentiles of the score of the average high achiever attending the high school.

c. A safety college is one where the college's median test score is between 5 and 15 percentiles below that of the average high achiever attending the high school.

d. Average over the last 10 years.

table 11 compares college-related factors at the high schools attended by achievement-typical and income-typical students.[33] The first striking statistic in the table is what a tiny share of low-income students' teachers graduated from colleges that would be peer or safety colleges for high-achieving students. Only 1.1 percent of income-typical students' teachers attended peer colleges, and only 5.0 percent attended safety colleges. The shares are larger for achievement-typical students' teachers, but still not large: 2.9 percent from peer colleges and 7.5 percent from safety colleges. Even high-income students do not encounter many teachers with degrees from very selective colleges.

33. We use all of the Schools and Staffing surveys in an attempt to pick up as many high schools as possible, but we nevertheless end up with teacher data for only 34 percent of the high-achieving students we study. We use the survey weights to create statistics that should be nationally representative. For the statistics based on previous cohorts, we use the actual previous cohorts from the College Board but must assume that our one previous cohort from the ACT was representative of the whole previous decade.

Income-typical students attend high schools where just 1.6 students in a typical previous cohort applied to one of the 10 most selective colleges in the United States.[34] In contrast, 7.6 students applied to these colleges in a typical previous cohort of achievement-typical students' schools. Thus, compared to an income-typical student, an achievement-typical student would be much more likely to vicariously experience the process of applying to a very selective college, through an upperclassman. In addition, only 3.8 percent (including the student) of the average income-typical student's high school class, compared with 11.2 percent of the average achievement-typical student's class, are high achievers themselves. Since income-typical students' high schools are, on average, less than two-thirds the size of achievement-typical students' high schools, these low percentages translate into very little school-based contact with other high achievers. The low percentages also suggest that their counselors are unaccustomed to advising students who have opportunities to attend selective colleges.

Of course, one might gather and advise a critical mass of high achievers outside of the high school setting, but the bottom rows of table 11 show that even this is difficult for income-typical students. The radius needed to gather 50 high achievers is 37.3 miles for the average income-typical student, but only 12.2 miles for the average achievement-typical student. Since a college access program cannot expect to get participation from *every* qualified student in the area it covers, the radii shown suggest that most income-typical students cannot be reached by programs that require a critical mass of high achievers to operate at efficient scale.

VI. Thought Experiments: Interventions That Might Inform Income-Typical Students

In this section we consider a few interventions that might affect how informed income-typical students are about their college-going opportunities. We do this because, as shown in the previous section, the data evince no support for hypothesis i (that income-typical students are actually more disadvantaged than achievement-typical ones) so that we are left with hypotheses ii (students are poorly informed) and iii (students are well informed and utility-maximizing). One way to assess hypoth-

34. Arguably, focusing on these colleges overstates the extent of previous cohorts' sophistication about college applications. These colleges are the most likely to show up in odd strategies like applying to one nonselective institution and to Harvard.

esis ii is to consider what information actually reaches or could reach income-typical students. After all, they are low-income high achievers who are apparently desirable applicants. Why should they not, for instance, become informed by their counselors or by traditional college recruitment methods?

VI.A. Traditional Interventions

Colleges often send admissions staff to high schools to recruit high-achieving students. Therefore, consider a thought experiment in which any student who attends a high school that contains at least 20 high-achieving students will have contact with some college admissions staff. (We chose a cutoff of 20 because it is expensive in time and money for admissions staff to visit high schools in which they cannot fill at least a classroom with potential applicants.) If this experiment occurred, 92 percent of high-income high achievers and 66 percent of achievement-typical students would have contact with admissions staff, but only 17 percent of income-typical students would have such contact.

Of course, admissions staff can hold evening or weekend events that students from multiple high schools can attend. Thus, we should also consider what would happen if admissions staff visited every location in the United States where they could gather at least 20 high-achieving students from a 10-mile radius. Such visits would ensure that 94 percent of high-income high achievers and 73 percent of achievement-typical students could meet with admissions staff. But such visits would allow only 21 percent of income-typical students to meet admissions staff.

Clearly, admissions staff visiting students is unlikely to be an effective method of informing income-typical students. What about students visiting colleges? As another thought experiment, consider what would happen if every high-achieving student visited colleges if he or she could reach five peer colleges by traveling 2,000 miles or less. Then 75 percent of high-income high achievers and 71 percent of achievement-typical students would do a college "tour." Only 22 percent of income-typical students would.

In fact, remembering that 70 percent of achievement-typical students are drawn from only 15 urban areas, we note that many of these students need not travel out of town at all to visit one or more selective colleges. Without needing anything other than a subway pass, a New York City student could easily visit Columbia University, Barnard College, New York University, Cooper Union, and at least six other colleges ranked at least "very competitive" by Barron's. A student living in Boston, Chicago,

Los Angeles, Philadelphia, or the San Francisco Bay area would also be spoiled for choice. Even a student from Portland, Maine—an area that might have seemed out of place on our list of 15 urban areas—has Bates, Bowdoin, Colby, and Dartmouth (all very selective institutions) within a modest radius. In fact, we know from colleges' own published materials and communications with their authors that many colleges make great efforts to seek out low-income students from their metropolitan areas. These strategies, although probably successful, fall somewhat under the heading of "searching under the lamppost." That is, many colleges look for low-income students where the *college* is instead of looking for low-income students where the *students* are.

We have already seen that income-typical students are very unlikely to encounter a teacher, counselor, or neighbor who attended a selective college himself or herself. Furthermore, income-typical students' counselors (each of whom typically manages a roster of hundreds of students) cannot be expected to develop expertise about very selective colleges, given the rarity with which they are called upon to advise high achievers. Indeed, at College Board sessions attended by the authors, several counselors reported that when the rare student in their school was qualified to attend very selective colleges, they told him to guide himself or herself by gathering information on the Internet because they themselves lacked expertise. This is despite the fact that counselors who attend College Board sessions are probably more sophisticated and informed than the average counselor.

The logic that makes admissions staff visits ineffective with income-typical students works similarly for after-school or weekend college mentoring programs: programs with sustainable costs are unlikely to reach income-typical students. Of course, college mentoring programs do exist in areas where income-typical students live, but the typical program focuses on motivating students merely to attend college—not on the decisions faced by high-achieving students with many college opportunities. The typical program also does not provide much advice on negotiating the multilayered application process that very selective colleges use.

What about mailing brochures with a specialized letter to students who live in ZIP codes where most families are poor? This strategy might work in the very largest urban areas, particularly if they are densely populated, but it cannot work well outside them. The United States Postal Service defines ZIP codes with the goal of making mail delivery efficient, not with the goal of identifying families with similar incomes. In a place like Manhattan, a ZIP code might be physically small enough to contain families with fairly

uniform socioeconomics. In smaller cities and rural areas, though, the typical ZIP code contains families with diverse incomes, ensuring that mail campaigns targeted to high-poverty ZIP codes systematically fail to reach most low-income students.

VI.B. Novel Interventions

What, then, are some interventions that might inform income-typical students about college and that might overcome the challenge of serving high achievers who are geographically dispersed? First, a college has many more alumni than admissions staff, and alumni are much more broadly distributed geographically than admissions staff. For instance, the anonymous private, very selective university studied by Jonathan Meer and Harvey Rosen (2012) has at least one alumnus or alumna in nearly every U.S. county.[35] Presumably, colleges could give their alumni the names of local students who appear on the search lists of students who are likely qualified for admission. Such alumni-based information interventions could potentially overcome the lack of geographic concentration among income-typical students. The main challenges for such interventions would seem to be the need to coordinate and inform alumni. It would be problematic, for instance, if alumni knew very little about their college's current curriculum or financial aid policies.

Income-typical students are intelligent and able to absorb written material. Thus, other interventions that might affect them would be purely informational ones, whether distributed by mail, online, or through social media. To be effective, however, such interventions must be much better targeted to low-income students than a campaign based on ZIP codes. Also, they cannot simply replicate the content that students already receive in the form of numerous college brochures. The two most obvious deficiencies of these brochures are that they are generic rather than customized to a student's situation (for instance, the student's family finances), and that they have a boosterism that may make it difficult for students to derive information from them. Taking these points to heart, we test several interventions in Hoxby and Sarah Turner (2013) that have the potential to identify *causal* effects of giving low-income students information about their college-going opportunities.

35. Meer and Rosen generously computed the relevant statistics for us.

VI.C. Recruiting Athletes versus Recruiting Low-Income High Achievers

Colleges seem able to identify and recruit students who are top athletes.[36] Should they therefore be able to identify and recruit the vast majority of low-income high achievers? Our analysis suggests that not only is the answer no, but that athletes are the exception that proves the rule.

Regardless of how dispersed they are, it is easy for colleges to identify top athletes. Any top athlete who participates in an individual sport can be easily found on lists of state finalists, often as early as the 10th grade. Most recruited athletes who play team sports also generate statistics (such as rushing yards) that are readily available, or play for a team that participates in state competitions. Even athletes who play only a team sport and whose home team is mediocre can be readily identified by the coaches of the top state teams with whom they compete: "John Smith from High School X is a great running back, even though his team has a mediocre record."[37]

Our conversations with college athletic directors suggest that they use simple, traditional recruiting methods to find athletes. The same methods would not work with low-income high achievers. Again, we emphasize that the data and analytics used in this paper are not available to colleges.

VII. Conclusions

We have demonstrated that the majority of high-achieving, low-income students do not apply to any selective colleges despite apparently being well qualified for admission. These income-typical students exhibit behavior that is typical of students of their *income* rather than typical of students of their *achievement.* There are, however, high-achieving, low-income students who apply to college in much the same way as their high-income counterparts. These achievement-typical students also enroll and persist in college like their high-income counterparts.

36. Many readers of previous drafts have asked us to compare athletes and low-income high achievers, which we are glad to do because the comparison is telling. However, there is *no* evidence that colleges actually identify and recruit most students who have the potential to perform very well in college sports. Our readers tend to assume that this is true, but colleges might, in fact, neglect to recruit many talented athletes.

37. Of course, colleges will likely not identify a student who is a potentially top athlete but only in a team sport and who plays on a weak team that competes only with other weak teams. But arguably that student cannot develop his or her potential in any case.

There are several plausible explanations for income-typical students' behavior:

(i) they cannot afford to attend peer institutions;

(ii) they are actually more disadvantaged than achievement-typical students and therefore behave differently;

(iii) they would fail to be admitted to peer institutions or would fail to thrive at them, were they to apply;

(iv) they are poorly informed about their college-going opportunities;

(v) they have cultural, social, or family issues that make them unwilling to apply to peer institutions, even if they are confident of being admitted and succeeding academically.

We believe we have eliminated explanations (i) and (ii). What is especially striking is that income-typical students pay more to attend less selective colleges than they would pay to attend peer institutions. Our evidence also does not support explanation (iii) since we find that—if they apply—low-income high achievers enroll and persist at the same rates as high-income students with the same test scores. Nevertheless, we cannot definitively test explanation (iii) in this paper. We are mainly left with explanations (iv) and (v), both of which are compatible with the fact that income-typical students are fairly isolated. Hoxby and Turner (2013) rigorously test explanations (iii) and (iv), leaving (v) as the residual explanation.

In this paper we have demonstrated that achievement-typical students not only come disproportionately from the central cities of large urban areas but are likely to attend selective, magnet, or other feeder high schools. A majority of achievement-typical students are drawn from only 15 urban areas, each of which has at least one and often several selective colleges. We show that traditional recruiting methods are likely to work better in large, dense urban areas and in the immediate vicinity of the college itself. Probably unintentionally, colleges end up looking for low-income students where the college is, instead of looking for low-income students where the students are. Thus, they recruit the low-income students "under the lamppost" but fail to identify the vast majority of others. We speculate that admissions staff believe that the supply of low-income high achievers is inelastic for two reasons. Many of these students are not on the radar screen because they do not apply. Also, staff spend much of their time informing students who attend high schools that are already so "tapped out" that their efforts merely shift students among colleges but fail to expand the number of low-income, high-achieving applicants.

Even if we knew for certain that income-typical students behaved as they do because they are poorly informed (as opposed to being deterred

by cultural factors), we would not attribute blame to colleges, counselors, or the students themselves. Income-typical students are insufficiently geographically concentrated to be reached, cost-effectively, by traditional methods of informing students about their college opportunities. Their high school counselors cannot be expected to develop expertise about selective colleges when doing so is rarely relevant to their duties, which require them to advise hundreds of students on myriad issues. Low-income high achievers are not necessarily less enterprising than their high-income counterparts; they simply do not have parents or counselors who ensure that they know something about peer institutions.

Our results suggest that interventions likely to affect low-income high achievers' college-going behavior will be ones that do *not* depend, for their efficacy, on the students being concentrated in a limited number of schools or small geographic areas.

ACKNOWLEDGMENTS For inspiration, generous help with data, and numerous useful suggestions, we thank the College Board and ACT. These two organizations' dedication to providing students with well-informed college application advice is the reason that this research exists. We have been especially helped by Connie Betterton, Michael Matthews, Anne Sturvenant, Ryan Williams, and Robert Ziomek. For important comments and suggestions we thank Sarah Turner, Amanda Pallais, Eric Bettinger, Scott Carrell, Charles Clotfelter, Susan Dynarski, Bridget Long, Parag Pathak, Bruce Sacerdote, Douglas Staiger, Jacob Vigdor, and the editors.

References

Avery, Christopher, and Caroline Hoxby. 2004. "Do and Should Financial Aid Packages Affect Students' College Choices?" In *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It,* edited by C. Hoxby. University of Chicago Press.

Avery, Christopher, and Sarah Turner. 2009. "Aid and Application Awareness." Harvard Kennedy School and University of Virginia.

Avery, Christopher, Mark Glickman, Caroline Hoxby, and Andrew Metrick. 2013. "A Revealed Preference Ranking of American Colleges and Universities." *Quarterly Journal of Economics.* 128 (1): 425–467.

Avery, Christopher, Caroline Hoxby, Clement Jackson, Kaitlin Burek, Glenn Pope, and Mridula Raman. 2006. "Cost Should Be No Barrier: An Evaluation of the First Year of Harvard's Financial Aid Initiative." Working Paper no. 12029. Cambridge, Mass.: National Bureau of Economic Research.

Bowen, William, Matthew Chingos, and Michael McPherson. 2009. *Crossing the Finish Line: Completing College at America's Public Universities.* Princeton University Press.

Card, David, and Alan B. Krueger. 2005. "Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence from California and Texas." *Industrial and Labor Relations Review* 58, no. 3: 416–34.

Dillon, Eleanor, and Jeffrey Smith. 2012. "Determinants of Mismatch between Student Ability and College Quality." University of Michigan.

Finn, Chester, and Jessica Hockett. 2012. *Shiny Needles in the Education Haystack: America's Academically-Selective Public High Schools.* Stanford, Calif.: Hoover Institution Press.

Freeberg, Norman. 1988. "Accuracy of Student Reported Information." College Board Report no. 88-5. professionals.collegeboard.com/profdownload/pdf/RR%2088-5.PDF.

Hill, Catherine, Gordon Winston, and Stephanie Boyd. 2005. "Affordability: Family Incomes and Net Prices at Highly Selective Private Colleges and Universities." *Journal of Human Resources* 40, no. 4:769–90.

Hoxby, Caroline. 2009. "The Changing Selectivity of American Colleges and Universities." *Journal of Economic Perspectives* 23, no. 4: 95–118.

Hoxby, Caroline, and Sarah Turner. 2013. "Expanding College Opportunities for Low-Income, High-Achieving Students." SIEPR Discussion Paper no. 12014. Stanford, Calif.: Stanford Institute for Economic Policy Research.

Meer, Jonathan, and Harvey Rosen. 2012. "Does Generosity Beget Generosity? Alumni Giving and Undergraduate Financial Aid." Working Paper no. w17861. Cambridge, Mass.: National Bureau of Economic Research.

Pallais, Amanda. 2009. "Small Differences That Matter: Mistakes in Applying to College." Massachusetts Institute of Technology.

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. 2009. The Common Core of Data. Electronic data for the 2008–09 school year. Downloaded from http://nces.ed.gov/ccd/pubschuniv.asp.

U.S. Department of Education. Institute of Education Sciences. National Center for Education Statistics. 2009. Private School Universe Survey. Electronic data for the 2008–09 school year. Downloaded from http://nces.ed.gov/surveys/pss/pssdata.asp.

Comments and Discussion

COMMENT BY

AMANDA PALLAIS This paper by Caroline Hoxby and Christopher Avery provides a comprehensive analysis of the differences in college application patterns between high-achieving students of differing family incomes. It finds that high-achieving, low-income students apply to substantially different sets of colleges than do their higher-income peers. Over half of the low-income group send SAT or ACT test scores to at least one nonselective college and do not send scores to any college with a median test score within 15 percentiles of their own score. Only 8 percent send scores to a portfolio containing at least one "match" college, one "safety" college, and no nonselective college.

This paper is not the first to note that low-income students apply to different sets of colleges than high-income students (see, for example, Spies 2001, Bowen and others 2005, and Pallais and Turner 2006). However, it is distinguished by its comprehensiveness and the sheer amount of data that allow the authors to fully characterize the application choices of high-achieving students. The paper starts with data on everyone in the high school class of 2008 who took either the ACT or the SAT I. Then it links these students to the colleges they sent scores to, to data on their high schools, and to data on their census block and zip code, as well as to information on whether and where they ultimately enrolled in college and whether they had completed a 4-year degree by 2012.

After showing the differences in application patterns between high- and low-income high achievers, the paper considers the characteristics both of those low-income students whose application behavior is similar to high-income students' (what the authors call "achievement-typical" students) and of those who do not apply to selective institutions ("income-typical" students). Achievement-typical students are more likely to come

from schools and neighborhoods where they could more easily obtain information about colleges (for example, because they are more likely to have teachers who attended selective colleges and friends from earlier cohorts who applied to selective colleges). The paper suggests that many low-income, high-achieving students would actually benefit from attending selective colleges but do not apply, because unlike high-income students, they do not have specific relevant information (for example, about the range of colleges available, colleges' true costs, or the relevant benefits of attending specific colleges).

A closely related explanation for low-income high achievers' distinct application choices is that applying to college or for financial aid is prohibitively difficult for some. For example, they may be less likely to have parents or guidance counselors who can assist them with the application process. This explanation also implies that low-income high-achievers might benefit from attending selective colleges but are failing to apply. However, if the applications themselves are preventing these students from attending selective colleges, simply providing more information without also assisting them in filling out the applications (or simplifying the application process) will not be effective. In the rest of this comment, I summarize some of the existing literature on these two explanations as they relate to low-income students in general, not just high-achievers.[1] This relatively new literature provides many examples in which giving high school students information about colleges or assistance with completing applications affects whether and where students attend college.

A recent paper by Hoxby and Sarah Turner (2013) presents the results of a randomized experiment with several different treatments. In one treatment, they sent high-achieving, low-income students information on colleges' actual net cost.[2] They found that this induced students to apply to more colleges and raised the likelihood both of their applying to a selective college and of their being admitted. (The point estimate also implies that this intervention increased the probability that students attended a selective

1. Throughout the ability distribution, low-income students apply to less selective colleges than their higher-income peers (Pallais 2011) and, conditional on high school performance, are less likely to attend any college (for example, Ellwood and Kane 2000). However, the application barriers that high-achieving students face may be different from those faced by lower-achieving students.

2. As the paper documents, students' net cost of attendance after financial aid often differs substantially from colleges' sticker prices, particularly at selective colleges.

college, but it is not statistically significant.) Another randomized treatment sent students information about suggested application strategies, college graduation rates, and application deadlines. Additionally, it sent students a copy of the Common Application (a standardized application used by many colleges), perhaps making it easier to apply. This treatment also induced students to send more applications and led to their being admitted to more-selective colleges. As a result of this treatment, students attended more-selective colleges.

Hoxby and Turner (2013) also provide evidence that application fees present a barrier to attending selective colleges for high-achieving, low-income students. Although low-income students can obtain fee waivers, the process requires additional paperwork. Randomly selected students in another treatment received fee waiver coupons in the mail and information on where the coupons could be used. Low-income students in this treatment also sent more applications, were admitted to more-selective colleges, and attended more-selective colleges than a randomly selected control group.

Eric Bettinger and others (2012) provide evidence that another aspect of the college application process, the Free Application for Federal Student Aid (FAFSA), is a barrier to low-income students attending college in general. In this project, H&R Block completed the FAFSA for randomly selected students. These students were significantly more likely to attend college than a control group. In contrast, students who received individualized financial aid information and were encouraged to complete the FAFSA on their own were not more likely to attend college than the control group.

Pallais (2012) shows that a small decrease in the cost of sending standardized test scores to colleges can induce low-income students to attend more-selective colleges. Before the fall of 1997, ACT allowed students to send three score reports to colleges for free and charged $6 for each additional score report. Thereafter ACT allowed students to send four score reports for free, with the same marginal cost for additional reports. Before the change, 82 percent of students sent exactly three score reports, while only 3 percent sent four. Afterward, only 10 percent of students sent three score reports, while 74 percent sent four. Both high- and low-income students sent more score reports as a result of the cost change, widened the range of colleges they sent scores to, and sent score reports both to more-selective and to less selective colleges than they would have otherwise. However, only low-income students ended up actually attending more-selective colleges as a result. It could have been the actual decrease

in the cost of the fourth score report that led students to change their behavior: perhaps the $6 was a financial barrier to applying to colleges for low-income students. Alternatively, students could have viewed the number of free score reports as information about the optimal number of score reports to send, interpreting the provision of three (or four) free score reports as reflecting ACT's informed judgment that sending three (or four) score reports was optimal.

Sarena Goodman (2012) and George Bulman (2013) show that inducing students to take a college entrance exam changes their college matriculation outcomes. Goodman (2012) analyzes the effect of mandates in some states that require all high school juniors to take the ACT. She finds that these mandates increased the number of students who took the ACT and did so disproportionately for low-income students. The mandates did not change the overall college attendance rate in these states but did substantially increase the number of students attending selective colleges (which are much more likely to require standardized test scores). Bulman (2013) examines the effect of opening an SAT testing center at a student's own high school. Having such a testing center allows students to take the SAT at their own high school rather than travel to other local schools. He finds that opening such a center increased the probability that a given student took the SAT and the probability that he or she attended a selective college. Both these effects were larger for students attending schools in low-income areas.

Finally, Scott Carrell and Bruce Sacerdote (2012) show that helping high school students navigate the college application process can induce these students to attend college. Their paper presents the results of a randomized intervention targeted at high school students in the winter of their senior year. Eligible students were identified by their guidance counselors as those who were "on the margin" of applying to college: they had expressed interest in applying to college but had made little or no progress in applying. Treated students were chosen at random from this pool. They had their application fees, SAT fees, and ACT fees paid for them and received in-person mentoring by a Dartmouth student. Dartmouth students also helped the students sign up for the SAT or the ACT if they had not already done so, complete essays, complete and file applications, request transcripts and recommendation letters, and start the FAFSA. The mentors sometimes also provided advice on how many and which colleges the students should apply to. Finally, students in the treatment group received $100 for completing their applications. This intervention substantially increased 4-year college going among female students, but

not among men.[3] The intervention also seemed to have larger effects at more-disadvantaged high schools. N3

An important question is whether inducing low-income students to attend college and to attend more-selective colleges actually benefits them. It is hard to answer this question fully without knowing more about students' utility functions or the information they have when making college decisions. However, Hoxby and Avery's paper provides evidence that low-income students actually pay less on average to attend very selective colleges than they would to attend less selective colleges. Moreover, research suggests that low-income students receive particularly high returns from attending college in general (Card 1995) and from attending more-selective colleges (Dale and Krueger 2002, Saavedra 2008). Hoxby and Avery show that low-income students who attend highly selective colleges have graduation rates similar to those of high-income students attending these colleges; thus, low-income students appear to be successful in these selective colleges. Of the studies described above, those that followed the students who were induced by the interventions to attend college or to attend more-selective colleges (Hoxby and Turner 2013, Bettinger and others 2012, Sacerdote and Carrell 2012, and Bulman 2013) all find that these students have high persistence in college. Thus, it seems likely that many low-income students who do not already do so would benefit from attending college and attending more-selective colleges.

REFERENCES FOR THE PALLAIS COMMENT

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *Quarterly Journal of Economics* 1277, no. 3: 1205–42.

Bowen, William G., Martin A. Kurzweil, and Eugene M. Tobin. 2005. *Equity and Excellence in American Higher Education.* University of Virginia Press.

Bulman, George. 2013. "The Effect of Access to College Assessments on Enrollment and Attainment." Working paper. Stanford University.

Card, D. 1995. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." In *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp,* edited by L. N. Christofides, E. K. Grant, and R. Swidinsky. University of Toronto Press.

3. It was not only the $100 payment that increased college going: students receiving the whole intervention experienced large increases in college going relative to a randomly selected group who received only the $100 incentive.

Carrell, Scott E., and Bruce Sacerdote. 2012. "Late Interventions Matter Too: The Case of College Coaching in New Hampshire." Working Paper no. 19031. Cambridge, Mass.: National Bureau of Economic Research.

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, no. 4: 1491–1527.

Ellwood, D. T., and T. J. Kane. 2000. "Who Is Getting a College Education: Family Background and the Growing Gaps in Enrollment." In *Securing the Future: Investing in Children from Birth to College,* edited by Sheldon Danziger and Jane Waldfogel. New York: Russell Sage Foundation.

Goodman, Sarena. 2012. "Learning from the Test: Raising Selective College Enrollment by Providing Information." Working paper. Columbia University and University of California, Berkeley.

Hoxby, Caroline M., and Sarah Turner. 2013 "Expanding College Opportunities for High-Achieving, Low Income Students." SIEPR Discussion Paper no. 12-014. Stanford Institute for Economic Policy Research.

Pallais, Amanda. 2011. "Essays in Labor Economics." Ph.D. dissertation. Massachusetts Institute of Technology.

———. 2012. "Small Differences That Matter: Mistakes in Applying to College." Working paper. Harvard University.

Pallais, Amanda, and Sarah Turner. 2006. "Opportunities for Low Income Students at Top Colleges and Universities: Policy Initiatives and the Distribution of Students." *National Tax Journal* 59, no. 2: 357–86.

Saavedra, Juan E. 2008. "The Returns to College Quality: A Regression Discontinuity Analysis." Working paper. Harvard University.

Spies, Richard R. 2001. "The Effect of Rising Costs on College Choice: The Fourth and Final in a Series of Studies on This Subject." Princeton University Research Report Series no. 117. Princeton University.

COMMENT BY

PARAG A. PATHAK This paper by Caroline Hoxby and Christopher Avery provides convincing evidence of the following fact: a large number of high-achieving, low-income students systematically do not apply to selective colleges or universities. The authors identify two major classes of low-income college applicants. "Income-typical" applicants apply to schools in much the same pattern as do other students in their local area and to no schools whose median scores are similar to their own. "Achievement-typical" applicants apply to schools in much the same pattern as do high-income high achievers, who are mostly from urban areas or have exposure to selective colleges. One noteworthy feature of the low-income, high achieving students in their sample is that most are not underrepresented minorities.

The paper nicely illustrates the importance of descriptive empirical work and the value of a nationally representative data set.

There are some important parallels between this paper and existing work on selective K-12 institutions in the United States. Atila Abdulkadiroglu, Joshua Angrist, and Pathak (2012) study exam high schools, including Boston Latin School and New York City's Stuyvesant High School and Bronx High School of Science. In both Boston and New York, roughly two-thirds of exam school students are eligible for a free or subsidized lunch, an indicator of poverty. Moreover, students enrolled in an exam school are between 1.3 and 1.5 standard deviations ahead of their public school peers on baseline standardized tests. Thus, they are low-income high achievers, but they are a few years away from applying to college.

The barriers to application for these schools seem, if anything, lower than the barriers faced by low-income, high-achieving students when they apply to college. For instance, the schools in both cities have long histories, they are widely known, and admission requires completing a common application on a standardized timeline. To gauge the extent to which applicants do not apply to a selective exam school in Boston for seventh grade, I estimate linear probability models of application and offers for students, controlling for their baseline test scores and demographics. By comparing the offer probability with the application probability, it is possible to measure the extent to which students who seem likely to obtain an offer at a school are likely to apply.

My table 1 reports the estimates by decile of predicted offer. An important fact shown in the table is that a large fraction of students who would almost certainly be offered admission to one of Boston's exam schools do not apply. Only 75.8 percent of students in the top decile submit an application, even though applicants in this group would be very likely to obtain an offer given their baseline test scores and demographics. That is, for this highest achieving, mostly low-income population, there is roughly a one-quarter gap in the fraction of students who apply to an exam school among those who are most likely to be given an offer. Seen in this light, it is perhaps no longer that surprising that for the more complicated process of applying to a selective college or university, many students who would almost surely be admitted do not apply.

Given the fact that Hoxby and Avery uncover, a natural question is whether it reflects a market failure, which would rationalize some form of policy intervention. There are many aspects to this question, and in what follows I will only touch on a few. First, this paper and other work by Hoxby (2009) provides evidence that more selective colleges provide more

Table 1. Probability of Sixth-Graders Applying to a Boston Exam School, by Decile of Predicted Probability of Receiving an Offer

Decile	*Predicted probability* p *of receiving an offer*[a]	*Probability* p *of applying*
1	0.021	0.085
2	0.042	0.111
3	0.068	0.141
4	0.105	0.178
5	0.160	0.226
6	0.237	0.286
7	0.357	0.361
8	0.520	0.460
9	0.726	0.583
10	0.927	0.758

Source: Author's calculations using data from Abdulkadiroglu, Angrist, and Pathak (2012).

a. Predictions are based on baseline test scores and the following: interactions of application year × baseline decile for both math and English; race; sex; and whether the student is eligible for the free school lunch program. Students classified as having limited English proficiency or as in special education are excluded.

student-oriented resources than less selective colleges, a trend that has increased dramatically in the last two decades (see, for example, figure 2 in Hoxby 2009). This fact, together with the likely scenario that attendance at a selective college will cost low-income students less, because of the college's own financial aid and other scholarship opportunities, seems to suggest that applicants are making suboptimal decisions. Moreover, the evidence that Hoxby and Avery marshal about the tendency of achievement-typical students to come from major urban areas or magnet or independent private schools seems to imply that the income-typical applicants lack adequate information about college, so that reducing application costs (broadly defined) seems likely to boost demand from this population.

There has been progress in making it easier for students to exercise their choice options for K-12 education. In districts that allow a choice of schools to attend, through either open enrollment plans or charters, there has been a recent push toward standard application timelines and common, online application systems. Cities like Denver and New Orleans have recently adopted single-offer coordinated charter and district school admissions schemes, and new assignment mechanisms that make it safe for participants to rank schools truthfully have become increasingly widespread (see, for example, Abdulkadiroglu, Pathak, and Roth 2009, Pathak 2011, Pathak and Sonmez 2008, 2013). The goal of these reforms is to increase access to high-quality educational options for students. Unlike college admissions,

many of these reforms involve changes within existing centralized institutions. However, it seems that decision aids, informational cues, and further guidance would be beneficial in either a decentralized college admissions system or a centralized assignment mechanism.

Second, a perhaps more difficult issue for policy is related to the fact that college admissions is an assignment market. It is possible that low-income, high-achieving students benefit from selective colleges and universities. However, if there are slot constraints at schools, reforms in favor of low-income students would involve a reallocation between these students and other students. Therefore, it seems especially important to investigate whether low-income students benefit more or less from selective education than other students. This may be so either because their fallback options are not as good, or because a selective school education creates social externalities.

Relatively few studies address these questions. Mark Hoekstra (2009) documents earnings effects from attending a flagship state university. Stacy Dale and Alan Krueger (2002, 2011) present "selection-adjusted" evidence on the returns to selective education and find some effects of attending a selective college for students from disadvantaged family backgrounds (as measured by parental education or income). However, identifying the causal effect of selective colleges is a challenging issue. In particular, it seems hard to describe a data generating process under which all selection bias operates through the set of schools one applies to, as in Dale and Krueger's research design. Perhaps just as important, if it were possible to encourage low-income high-achievers to attend selective schools, the relevant margin is probably the switch from a local community college to a lower-end selective college, which may not be the effect identified by Dale and Krueger. On the other hand, the evidence from selective secondary schools seems to point to little benefit to selective education (Bui and others 2011, Abdulkadiroglu and others 2012, Clark 2010). Without a doubt, additional research is needed to measure the benefits of selective education. The additional work by Hoxby and Turner (2013), foreshadowed in this paper, will likely provide valuable evidence on this challenging empirical question.

Finally, the stylized facts documented by Hoxby and Avery indicate another rationale for intervention, one related to issues of equity. Of course, whether equity should be seen as a rationale for policy intervention depends on the objective functions of colleges, and of society more generally. These points are being brought forward in recent policy debates. For example, a possible response to court challenges

of affirmative action is the adoption of broader income-based criteria. Here again an important precedent at the K-12 level seems worth noting. The 2007 Supreme Court ruling in *Parents Involved in Community Schools* v. *Seattle School District No. 1* has been widely seen as outlawing the use of racial preferences in K-12 school admissions. A number of districts throughout the country are experimenting with a redefinition of diversity. For instance, Chicago Public Schools has established a "tier system" for its nine selective high schools. Each census tract in Chicago is given a socioeconomic status score based on median family income, educational attainment, the percentage of single-parent households, the percentage of residents who are homeowners, the percentage that speak a language other than English, and average school performance. At each selective school, 30 percent of seats are reserved to be assigned on the basis of admissions test scores only, and the remaining 70 percent are assigned according to admission test scores within tier. It seems a good bet that school systems will continue to experiment with like-minded policies that seek to redefine diversity for higher education. If so, an encouraging aspect of Hoxby and Avery's finding is that low-income, high-achieving students are plentiful, and if Hoxby and Avery can find them, others might be able to as well.

REFERENCES FOR THE PATHAK COMMENT

Abdulkadiroglu, Atila, Parag A. Pathak, and Alvin E. Roth. 2009. "Strategyproofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." *American Economic Review* 99, no. 5: 1954–78.

Abdulkadiroglu, Atila, Joshua D. Angrist, and Parag A. Pathak. 2012. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." IZA Discussion Paper no. 6790. Bonn, Germany: Institute for the Study of Labor. (Also forthcoming in *Econometrica.*)

Bui, Sa A., Steven G. Craig, and Scott A. Imberman. 2011. "Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Achievement." Working Paper no. 17089. Cambridge, Mass.: National Bureau of Economic Research (May).

Clark, Damon. 2010. "Selective Schools and Academic Achievement." *B.E. Journal of Economic Analysis and Policy: Advances* 10, no. 1: article 9.

Dale, Stacy Berg, and Alan B. Krueger. 2002. "Estimating the Payoff of Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics* 107, no. 4: 1491–1527.

———. 2011. "Estimating the Return to College Selectivity over the Career Using Administrative Earnings Data." Working Paper no. 17159. Cambridge, Mass.: National Bureau of Economic Research (June).

Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *Review of Economics and Statistics* 91, no. 4: 717–24.

Hoxby, Caroline. 2009. "The Changing Selectivity of American Colleges." *Journal of Economic Perspectives* 23, no. 4: 95–118.

Hoxby, Caroline, and Sarah Turner. 2013. "Expanding Opportunities for High-Achieving, Low Income Students." SIEPR Discussion Paper no. 12-014. Stanford Institute for Economic Policy Research.

Pathak, Parag A. 2011. "The Mechanism Design Approach to Student Assignment." *Annual Review of Economics* 3: 513–36.

Pathak, Parag A., and Tayfun Sonmez. 2008. "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." *American Economic Review* 98, no. 4: 1636–52.

———. 2013. "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation." *American Economic Review* 103, no. 1: 80–106.

GENERAL DISCUSSION Donald Kohn wondered what incentives selective colleges have to find and admit high-performing low-income students. Noting that most available diversity statistics measure racial and ethnic diversity and not income diversity, Kohn speculated that admissions officers have less of an incentive to search out high-performing, low-income students who are not members of historically disfavored racial and ethnic groups. He also asked, given the recent drop in male college application and completion rates, whether the missing high-performing, low-income students were predominantly male.

Justin Wolfers mentioned that he had written an op-ed with Betsey Stevenson describing Hoxby and Avery's results, which had led the president of the University of Michigan to inquire about the results and how the university might utilize them. That indicated that at least one selective university was interested in improving its recruitment of high-performing low-income students.

Benjamin Friedman observed that the geographic distribution of the "missing" high-performing, low-income students in the authors' data appeared to be correlated with current political patterns in the United States: relatively larger numbers of such students tend to be found in the "red" (Republican-leaning) states. He thought this correlation might offer some clues for understanding why these students were "missing."

Robert Gordon questioned the assumption in the discussion thus far that improving the situation of high-performing, low-income students would

yield a net gain to society. A student who is admitted to and attends Harvard instead of a Chicago community college almost surely reaps a lifetime benefit, but that student displaces another student from attending Harvard, who then attends a slightly less prestigious school. That displaced student in turn displaces another student at that school, and so on. The sum of these displaced students' lost utilities, Gordon argued, should be taken into account in any net social welfare calculation. Gordon further noted that the shares of black and Hispanic students among the population of high-performing, low-income students are lower than their shares in the overall population.

Raquel Fernández wondered how much additional financial aid is being allocated to high-performing, low-income students and what that might imply for colleges' incentive to admit them. She also asked whether it would it be possible for a high school student's performance information to be automatically made available to selective colleges when the student takes a standardized test.

Citing his own experience as a parent of high school students, Michael Klein proposed that the largest component of the cost of applying to any given college might be the cost of making a campus visit. He wondered how many students attend a college that they had not first visited, and how many high-performing, low-income students are prevented from attending colleges appropriate to their ability by the cost of visiting.

Betsey Stevenson suggested that family issues may help explain why some high-performing, low-income students do not go to the most competitive colleges. For instance, firstborn children might have responsibilities for younger siblings that prevent them from going away to college. In terms of the things over which colleges have more influence, she argued that it might be the perceived, not the actual, cost of a selective college that is deterring applications: many low-income applicants might be poorly informed about the true cost of attending college.

Gary Burtless cited a possible adverse consequence of increasing the number of high-performing, low-income applicants to elite colleges. He recalled reviewing a recent book by Charles Murray that showed that in the 1940s, students at elite universities scored only slightly better, on average, on standardized tests than students at a cross section of Pennsylvania universities. Given that elite universities now accept high-performing students almost exclusively, the distribution of high-performing, low-income students by selectivity of their college is roughly the same today in as the 1940s, but the distribution of other high-performing students has become significantly more concentrated. If that trend were to continue so that all high-performing students ended up attending elite colleges, it

would contribute to what Murray saw as an increasing divide between the elite and the rest in society. Supporting Burtless's point, Gordon noted that until roughly 1960, the majority of Harvard students were from college preparatory schools, but from then on more public high school graduates than college prep graduates attended.

Alexandre Mas observed that colleges generally do a better job of recruiting high-achieving athletes than of recruiting high academic achievers. Could colleges use their athletic recruitment policies as a model for recruiting more high-performing, low-income students? Laurence Ball added that it was absurd to imagine that there are students capable of playing intercollegiate basketball at UCLA or Duke but not applying there and ending up at local community colleges instead.

Ricardo Reis asked how the authors reconciled their finding of large benefits from attending an elite college with those of other studies, such as by Parag Pathak, that find little value added from attending an elite high school. Reis noted that students can apply to elite schools at any of three different levels: high school, undergraduate, or graduate school. But the differences in family income among attendees of elite high schools are less pronounced than those for attendees of undergraduate colleges, and applicants to graduate schools are less tied to their city or state of origin. This suggested to Reis that, of the three levels, only for undergraduate college is geography the binding constraint on elite school attendance for high-performing low-income students.

Willaim Brainard agreed with Stevenson that students face important information constraints in deciding which college to apply to. Even many Yale faculty members and administrators, he reported, cannot accurately state the current cost of a year's tuition at Yale. Brainard suspected that many high-performing students do not apply to elite colleges because they think, mistakenly, that they have a negligible chance of being admitted, or if admitted, being able to afford the cost.

Bradford DeLong observed that the number of undergraduate students at Harvard has roughly tripled since the end of the 19th century, from about 500 then to about 1,600 currently. But the number of qualified individuals who might want to attend Harvard has increased by a factor of at least 10 over the same period. The result has been a substantial increase in selectivity, which means that a student's best strategy is to submit many applications and hope for a good draw from what has become largely a random selection process.

David Romer noted that many colleges like to have students from all 50 states. He wondered how well these colleges do at finding high-achieving

students from small states who do not attend high schools with large numbers of high achievers.

Responding to the discussion, Caroline Hoxby remarked that selective colleges today are quite eager to diversify their student bodies by recruiting more low-income students—so eager that some, regrettably, have been willing to lower their admission standards to do so. These colleges would much prefer to find low-income students who meet their standards and can do the academic work, because such students are less likely to fail to keep up or to become segregated in the easier majors. The problem, Hoxby said, is not in locating high achievers and sending them information: colleges can easily buy lists of high-scoring students from the College Board and ACT; indeed, college applicants today are inundated with brochures, catalogs, application forms, and the like—to say nothing of the vast amount of information now available through the Internet. Part of the problem is that most students have great difficulty navigating this sea of information without help from more knowledgeable people, and such people tend to be scarce in the homes, schools, and neighborhoods of low-income students.

Once a high-achieving, low-income student has applied to a selective college, Hoxby continued, the college will likely make every effort to recruit him or her—offering all-expense-paid visits, for example, or even sending admissions staff to the student's home. But many high-achieving, low-income students, precisely because they lack adequate guidance about the opportunities available to them, never apply. Hence they remain invisible to the selective colleges that are so assiduously looking for them.

Hoxby also addressed the question of whether admitting more low-income students to highly selective colleges simply displaces other students to slightly less selective colleges, thus negating any net social gain. She noted that before the financial crisis, several of the nation's most selective colleges, including Harvard and Stanford, had made plans to expand their freshman classes. Princeton actually did so. In principle, then, the displacement problem could—and may yet—be avoided. A concern, however, has been that higher-income students would end up taking most of the new slots. Thus, the paper's finding that many more low-income high achievers are out there than had been widely believed should encourage these schools to proceed with their expansions.

More broadly, Hoxby thought it a mistake to view college admissions as a zero-sum game in which selective colleges have a fixed quantity of resources to allocate among a fixed number of successful applicants. That might be the case in the short run, she said, but the long run presents a

very different picture. Selective colleges seldom recoup their operational costs from tuition and other upfront revenue—even most of their higher-income students pay no more than about half the true cost of their education. Rather, the books for a given graduating class balance only decades later, as donations from successful—and grateful—alumni flow in. Hoxby added that although in the short run a college might lose more, because of financial aid, on its low-income students than on the average student, often it is the low-income students, mindful of the enormous difference the college has made in their lives, who become the most loyal and generous alumni.

Finally, Hoxby argued that although streamlining the college application process was surely part of the solution, it might inadvisable if by "streamlining" one means reducing the amount of information collected. After all, the aid package that a low-income applicant to a highly selective college receives can be worth nearly half a million dollars. Just as a bank considering a $500,000 small business loan will require extensive information about the borrower's business plan, so it is only reasonable for a college to scrutinize each financial aid applicant carefully, to decide whether he or she is a good investment.

Christopher Avery added, replying to Donald Kohn, that colleges are in fact facing strong pressure to admit more low-income students. One source of that pressure, he said, was the increasing availability of rankings of colleges by their percentage of students who are eligible for Pell grants.

Replying to a point in Amanda Pallais's comment, Avery agreed that it was important for low-income high achievers to apply to many selective colleges rather than a few, because although the average aid package at such colleges is generous, the distribution of aid offers is fairly wide. He also agreed with Pallais that even though the cost of applying to a college is typically negligible relative to the potential long-run benefits of attending, it nevertheless seems to be a behavioral sticking point for many students. That underlined for him the importance of fee waivers for low-income applicants.