Revised: August 4, 2009

The Determinants of Mismatch Between Students and Colleges

Eleanor Dillon
Department of Economics
University of Michigan
ewdillon@umich.edu

Jeffrey Smith
Department of Economics
University of Michigan
NBER and IZA
econjeff@umich.edu

Abstract: We expect that when new students enroll in college the most able and accomplished high school graduates will go to the top colleges. In practice, we find many relatively weak students at competitive schools and even more high-ability students at relatively low quality schools. How do students end up "mismatched" with the quality of their college? We develop and estimate a model of how students and their families make the decision of which, if any, college they will attend. We look separately at the types of students who end up over-qualified and under-qualified for their college, with particular attention to lack of information and financial constraints during the college application process. Using data on a very recent cohort of college applicants we find that while both information and finances play a role in predicting which students are most likely to end up poorly matched with their college, lack of information about college is the more important constraint.

* We thank Dan Black, Sarah Turner, Bill Johnson, and the participants at an AEA session for helpful comments and suggestions. Of course, all remaining errors and omissions are our own.

I. Introduction

The process by which students and schools work together to sort applicants into different colleges has important implications for students, colleges, and the economy as a whole. Differences in college quality can have a major effect on a student's experience during and after college, in terms of the gains that students make in college and in their probability of completing their degree in a timely fashion. In consequence, the match between student and college characteristics also affects efficiently the huge investments[1] that schools, governments, and private organizations make in college education in work to grow the supply of college educated labor in the economy. This sorting process is also interesting in its own right. Understanding how students choose among colleges gives us a broader understanding of the economic and social factors that influence the college choice than we can get by looking only at the extensive choice of whether to enroll in college at all. Quantifying the effect of varying tuition costs on college choice also makes explicit the social cost of state tuition walls for public universities.

We develop an informal model of how students and their families make the multi-faceted decision of which, if any, college they will attend, considering both economic constraints, such as varying tuition costs, and the differences in information and guidance available to different students. We expect that some students are better informed about the types of colleges they might attend, the consequences of attending different qualities of college, and about their own abilities relative to other college students. This type of information constraint could affect mismatch in either direction, leading to students who are either over-qualified or under-qualified for the school they attend. Students may also

[1] $143 billion during the 2007-08 academic year according to the College Board's *Trends in Student Aid 2008*.

be more or less able to pay for the more expensive colleges. These financial constraints should be more important in predicting a student's probability of being over-qualified for her college.

We estimate the relative importance of these factors in predicting the quality of the match between student and college characteristics using data on the very recent cohort of high school graduates surveyed in the National Longitudinal Survey of Youth 1997. We define the match between students and schools as the gap between the percentile of the student's ability and the percentile of the college's quality, with the assumption that effective sorting will put the most able students at the best colleges. We consider over-qualification and under-qualification separately since it seems likely that they are different phenomena.

While we find some evidence that both finances and information can cause students to attend a college with which they are "poorly matched," information constraints appear to play a stronger role. This is true for both stronger students at weaker schools and weaker students at strong schools. The information constraints are quite symmetric, influencing over-qualification and under-qualification in roughly the same way. As predicted, financial constraints are more important in predicting over-qualification.

The next section reviews the literature related to college match. We present our model of college choice in Section III. In Sections IV and V we discuss our data and our methods of estimating ability and college quality. Section VI presents our results and Section VII concludes.

II. Related Literature

Previous economic research has indirectly addressed the determinants of mismatch between student ability and college quality. Griffith and Rothstein (2007) find that students are more likely to apply to a selective college if they are more able, but also if they live near one, are female or Asian, attended a private school, have highly-educated parents, or live in a zip code with high average income. Brand and Xie (2007) and others in the sociology literature have stressed that the college decision is not a straightforward balancing of costs and returns and is heavily influenced by social and economic background. We move beyond the existing literature by looking explicitly at the predictors of the quality of the match between student ability and college quality, rather than the predictors of attending a selective school, and by allowing for different determinants of over and under-qualification.

Our work is motivated by a number of papers which find evidence that students who are over or under-qualified for their school may be less likely to graduate and more likely to transfer or take extra time to complete their degree. Light and Strayer (2000) find that U.S. students who are mismatched, in either direction, with the first school they attend are less likely to complete a degree at that school, either because they transfer schools or because they drop out. Among U.K students, Arulampalam et al (2005) find that students who are about even with or more able than their peers are less likely to transfer or drop out of university than students who are weaker than their peers. While they do not address mismatch directly, Bound et al (2007) find that, controlling for student characteristics, college quality affects time to degree. In contrast, Alon and

Tienda (2005) found that attending a selective school did not increase time to degree on average, but their estimates did not entirely control for the students at more selective schools being more able.

College mismatch has the potential to lead to higher or lower wages after graduation. If workers receive an extra wage premium for attending a high quality college, as was found by Black and Smith (2004), then a student who attends a school for which she is over-qualified may forego some of her potential returns to attending college. For the same reason, students may benefit by attending a higher quality school, even if they are under-qualified for it. However, the evidence on the importance of college quality on future earnings is mixed. Dale and Krueger (2002), for one, find very little effect of college quality on future earnings once they control for student ability, since more able students generally attend better schools. Arcidiacono (2004) points out that the major choice has a much larger effect on later earnings than college choice. If individual characteristics are the main determinant of earnings than students should focus attending a college for which they are at least well qualified, so that they can make sure of demonstrating their ability clearly through good grades.

Finally, this paper fits into a larger study of imperfect sorting in other markets. For example, William Johnson (2006 and 2008) finds evidence that search costs prevent workers from sorting perfectly into jobs with different combinations of hours and compensation. There is also a large literature on the existence and limits of Tiebout sorting among neighborhoods.

III. The College Choice Process and Mismatched Outcomes

We define college mismatch as a situation where a student attends a college of a quality that does not match her abilities. We use a multi-faceted measure of quality that includes characteristics of the student body, but also of the faculty and of the school. This definition is broader than the one used frequently in the literature that looks only at the student body, comparing the test score of each student to the average at her school. While we prefer this more comprehensive measure of student and college characteristics, we also estimate determinants of mismatch between a student's SAT score and the average SAT score at her school and our main findings are the same. Our definition is also not dependent on any consequences from mismatch, for example looking at the student's grades in college in comparison to her peers. Rather, we look at mismatch as any deviation from a perfect sorting where the strongest students go to the top schools and on down the line.

Mismatch is not necessarily a bad thing for the student or the school. A student may get better grades at a weaker school, sending a stronger signal of ability, if not of material mastered, on the job market. A weak student at a high quality school may benefit from more rigorous classes, even if she struggles with them, and from better advising and academic support. Given the large numbers of students who end up severely mismatched with their schools, it seems likely that there is at least a perception on the part of students and their parents that some types of mismatch may be good for the student.

The process by which students are sorted into schools has several stages and involves choices by both the student and the school. The student first decides which

colleges to apply to, then the colleges decide which students to admit, and finally the student chooses among her offers of admission. We assume the college applicants are rational and forward-looking. Nevertheless, during this process there are several influences that could cause a student to end up at a school that does not match her abilities, including information constraints, financial constraints, and social pressures.

Lack of information on the part of either the student or the school could result in college mismatch. The student may not have complete information about the quality of different colleges, or about how her abilities compare with other college applicants. Both misunderstandings could cause her to apply to an inappropriate mix of schools. Lack of information about college qualities could also cause her to choose a poorly matched school out of the set of schools to which she is accepted. The student's application may also be a poor indication of her true ability, for example if she over- or under-performed on the SAT or if the college admissions team is unsure how harshly her high school grades its students and cannot get a strong ability signal from her GPA. If a college misinterprets the student's ability it may admit her to a school for which she is ill-prepared or reject her from a school that would suit her. We expect that less well-informed students will be more likely to be mismatched with their college in either direction.

In a basic framework where students make the best college match they can subject to their budget, we expect very able lower income students to be less likely to attend the highest quality schools, and therefore be more likely to be over-qualified for the school they do attend. Budget constraints will tend to push students toward schools for which they are over-qualified, since more elite schools tend to be more expensive. These

budget constraints are somewhat offset for strong students by financial aid, which has grown more generous in recent years. In fact, students from relatively low-income families may end up paying less to attend the top private universities than they would to attend a mid-range, but less well-endowed, public university (Avery and Turner 2008). However, students do not know what their final financial aid offers will be when they decide which schools to apply for in the fall. In any case, students may want to stay closer to home to avoid travel heavy travel costs or so that they can live at home instead of paying board. Again, this will tend to increase over-qualification more than under-qualification since the students have an incentive to attend a nearer school even if they are over-qualified for it, but schools have no incentive to accept weaker students just because they live nearby. However, low-income students may be more likely to be under-qualified for their school if schools have begun to practice some affirmative action to admit more low-income students, who are certainly a minority at selective schools. In addition, the state school system can generate mismatch in either direction. Most state schools offer discounted tuition to state residents, making them more affordable than other options. In addition, some state schools have requirements about admitting state residents and may have a lower, or no, admission threshold for local students.

Finally, there are some considerations that may push students away from their closest match that have nothing to do with either information or financial constraints. The college experience is about more than the classes and students' choices may reflect that. Students may choose to go to the same schools that their friends are going to, the school that their parents attended, or any school where they feel they will fit in with the student body, even if that school is not a perfect match for them academically. Social

pressures could generate mismatch in either direction. Strong students may avoid the top schools if they are worried about fitting in with the other students there. Students who feel pressure to attend elite schools may also have help getting into those schools, possibly from legacy, or more generally because parents who want their children at the best schools can help them with SAT prep, tutoring, and extra college advising. Student athletes could also end up at a school for which they are under-qualified academically because they were recruited or made their choice partially based on their athletic ability. When these individual preferences influence college choice we will observe positive or negative measured mismatch, but students may still be at the school that is best for them in a broader sense. Since we have no information about individual preferences for schools they will simply show up as error in our estimates.

IV. The Data

We use the National Longitudinal Survey of Youth 1997 Cohort (NLSY97) data, which allows us to study a very recent cohort of college students. This survey, a second generation of the extensively used NLSY79, covers a group of American youth born between 1980 and 1984. The first interview was in 1997, with follow-up interviews each year since. The majority of the sample graduated high school and made their college choice between 1999 and 2002. 84 percent of the un-weighted sample graduated high school or got a GED. Of these high school graduates, 42 percent attended a four-year college. We focus on the 3,182 respondents who have attended a four-year college in the United States. We exclude two-year schools from our analysis because of the difficulties

of comparing quality across different kinds of schools.[2] Some students are excluded from the multivariate analysis because they attended colleges for which we do not have measures of quality, did not take the ASVAB, or are missing other covariates. We discuss the construction of our sample in more detail in the data appendix. We are left with around 1,700 observations for our multivariate analyses.

The NLSY97 sample includes both a representative cross-section of this generation of Americans and an over-sample of black and Hispanic students. We combine these samples in our analyses. We use probability of inclusion weights to combine the two samples, and also to control for differing sampling and response rates in different regions and across age, gender, race, and ethnicity groups.

The NLSY97 survey has an extremely rich set of information about family background, region and neighborhood characteristics, and characteristics of the student.[3] Specifically, almost all of the survey participants took the Armed Forces Vocational Aptitude Battery (ASVAB) test, which is usually administered to applicants to the U.S. military. This test has twelve components, covering both the sorts of skills measured by the SAT such as arithmetic, vocabulary, and reading comprehension and other skills such as electronics knowledge and spatial reasoning. The ASVAB test score offers a somewhat richer measure of ability in high school than the SAT or ACT score, and should be less influenced by variation in studying effort and preparation, since there was nothing riding on this test for the NLSY participants.[4] The ASVAB score is also useful

[2] See Reynolds (2009) for an analysis of the choice between starting at a two-year of four-year college.
[3] We use both the public use NLSY97 data and the restricted Geocode data, which contains more information about the respondent's location and about schools attended.
[4] The ASVAB test is not a straightforward measure of "innate" ability because it includes the influences and training that the student has had up to the point she takes the test. See Neal and Johnson (1996) for a more thorough discussion of what the ASVAB test is measuring. We consider demonstrated ability in high school to be the relevant variable because it captures what students bring to the college application process,

because it gives us a measure of ability that is potentially relevant to college performance but not observed by colleges while they are making their admissions decisions. We can therefore capture some of the college mismatch generated by incomplete information on the part of the colleges.

All survey participants took the ASVAB between 1997 and 1998, so their age when they took the test varies and most participants were younger than the larger population taking the test. In addition, the ASVAB is a computer adaptive test, meaning that test takers are asked different questions over the course of each section based on their responses to early questions. The score for each section reported by the NLSY is calculated based on both the number of questions answered correctly and the difficulty of those questions estimated from an earlier sample of test takers. We take the first principal component factor across the 12 section scores as our raw measure of ability.[5] We then calculate each respondent's percentile within the sample of college-bound NLSY97 respondents who took the test at the same age, weighted by probability of inclusion in the sample.

We construct a multifaceted index of college quality that combines student characteristics, college and faculty characteristics, and measures of students' revealed preferences over schools. For college quality we merge data from the U.S. Department of Education's Integrated Post-Secondary Data System (IPEDS) and U.S. News and World Report with the colleges listed in the NLSY97 dataset. The components of our

---

without the variation in college preparation that influences the SAT. Neal and Johnson also summarize evidence that, unlike the SAT, the ASVAB test show no signs of racial bias.

[5] Cawley, Heckman, and Vytacil (2001) and Black and Smith (2006) found that the second principal components of the ASVAB score is also relevant in determining later earnings in the NLSY 1979 sample. For our purposes, we need a single measure of ability. The first factor is by far the most important and explains 62 percent of the variation in scores.

college quality index are mean SAT score of entering students, percent of applicants rejected, freshman retention rate, average faculty salary, and faculty-student ratio. We use the first principal component factor across these five measures of quality as our quality index, following Black and Smith (2004). We then calculate the school's quality percentile across all four-year institutions in the United States included in the IPEDS, weighted by student body size.

There are several sources of potential measurement error in our estimates of ability, quality, and college match. An important limitation is that we observe college quality at the school level. In practice, individual departments within a college may be better or worse than the average quality of that college. If a strong student who plans to be a physicist attends a school of medium quality as we measure it, but that school has a top-rate physics program, then we will errantly consider that student over-qualified for her school. Likewise, if an aspiring English major enrolls at a top engineering school we will observe her as well-matched or even under-qualified, when in fact that school may not offer strong training in her area of interest. Additionally, while an index across several dimensions of college quality improves on a single measure of quality there is still some measurement error in college quality (Black and Smith 2006). Finally, the ASVAB score is likely to be an imperfect measure of ability. While the ASVAB includes a richer variety of tests than most standardized tests it still does not capture all the abilities that make for a strong college student. Even if it did measure all relevant abilities, the score from a single ASVAB test would be an imperfect measure of ability because some students will perform above or below their usual level on any given day.

These sources of error will make our results less precise, biasing our estimated relationships toward zero.

Because we weight the quality percentile by student body size, a school in the nth percentile is the school that a student in the nth percentile would attend if you ranked students by quality of school attended. Therefore, if students sorted into schools based purely on ability and school quality, a student in the nth ability percentile would attend a school in the nth quality percentile and mismatch, defined as the different in ability percentile and quality percentile, would be equal to zero for all students.

Gaps in this type of a priori match are quite common. Table 1 gives the joint distribution of student ability and college quality. Students are concentrated along the diagonal, which indicates a good match, but there are also a substantial number of mismatched students. Many previous discussions of mismatch have been framed by a discussion of affirmative action, and have therefore focused on students who seem under-qualified for their schools, but we find that strong students at weak schools are at least as common. The gap between the ability percentage of students and the quality percentile of the college they attend has a roughly normal distribution, shown in figure 1.

V. Understanding the college choice

We use several indirect measures of the student's information about colleges. Parents' education and the share of adults in the student's census district that have college degrees partially capture the student's access to information and role models for college attendance. We also consider whether or not the student lived within a Metropolitan Statistical area, since students from sparsely populated areas may not know many other

students of similar abilities to share information from and are less likely to be targeted by recruitment programs. Finally, we include the quality of the state flagship college and whether the student grew up within fifty miles of a college with which she would be well matched[6] as measures of whether students are familiar with a range of college options, including a good match. The quality of the state flagship also captures whether students have a low-cost way to attend a relatively high quality college. Students with a strong state school face less of a tradeoff between college quality and travel and tuition costs. This effect is likely to be greatest in Central and Western states, where public colleges and universities were established early in the state's development and often remain the premier schools in the state, and least relevant in the Northeast, where there are many excellent private colleges even in states with lackluster public university networks.[7] Greater exposure to college graduates reduces students' uncertainty about the qualities of different schools and about how they compare to other college applicants, lowering the probability that these students will be either over-qualified or under-qualified for their school.

Unfortunately, many of these measures of information may also capture the social pressures and preferences that influence college choice. Students growing up in a neighborhood where many people go on to their local college may prefer to join their friends there, even if that school is not an ideal fit. In contrast, growing up surrounded by college graduates, particularly if many of those graduates went to selective schools, may put pressure on students to go to a top school themselves, and may also give them special

[6] Defined as a school whose quality percentile is within 20 points of the student's ability percentile.
[7] Goldin and Katz (2008) discuss the development of private and public universities in the U.S. and their relative selectivity in different regions.

advantages in applying to those schools.[8] In this case, students with more educated parents, more college graduates in their neighborhood, or higher quality state schools will be more likely to go to high-quality colleges, regardless of their abilities. These students will be more likely to be under-qualified and less likely to be over-qualified for their schools.

Looking at the distribution of student characteristics across school quality and mismatch categories, more educated parents appear to guide their children toward a better match fit between their abilities and college quality. Table 2 shows that the students at higher quality colleges tend to have more educated parents, but this pattern does not control for the possibility that these children also have higher ability. Table 3 looks specifically at match quality, comparing the average characteristics of well matched, very over-qualified, and very under-qualified students. We consider students to be very over or under qualified if there is a greater than 20 percentile point gap between their ability percentile and the quality percentile of the first school they attend. These cutoffs assign about a quarter of the sample to each mismatch category.[9] Table 3 shows that well-matched students have, on average, more educated parents than either over-qualified or under-qualified students. This U-shaped pattern is consistent with our theory of how information would affect match quality. This pattern does not show up in the share of college graduates in the neighborhood or the quality of the state flagship. Even controlling for ability, students at higher quality schools live in states with higher-quality flagship schools and have a higher percentage of college graduates in their neighborhood. This monotonic pattern means those students whose college quality percentiles exceed

[8] See, for example, Kane and Avery (2004) on the differences in college expectations across socio-economic groups.
[9] As discussed in more detail in the next section, our results are robust to changes in these cutoffs.

their ability percentiles have the highest average state flagship school quality and concentration of college graduates in their neighborhood, followed by the well-matched students, then the over-qualified students. This pattern is consistent with the idea that cultural standards can influence college choices regardless of individual ability.

Students facing stricter budget constraints should be less likely to attend a top-quality school, even if they are very able, because those schools tend to be the most expensive. To capture the student's budget constraint we consider their family income and wealth. We focus on income and wealth in 1997, somewhat before most students finished high school, because we have the most complete financial information in the first year of the survey. Income and wealth are both good, but potentially incomplete measure of the family's ability to pay for college. If additional information about the family's permanent income is captured in parent's education and neighborhood characteristics then these variables will pick up some of the student's financial constraint as well as information constraints.

The pattern of college choice by income and wealth is quite consistent with the prediction that constrained students will be more likely to be over-qualified. As evident in Table 2, college-bound students are much better off financially than students who do not continue to college and students are the higher quality schools are much better off than those at lower quality schools. The average family income of students in the top quality quartile of colleges is almost twice the average of students in the bottom quartile. Consistent with the theory, over-qualified students have substantially lower average family wealth and income than either well-matched or under-qualified students.

Finally, we consider the effects on mismatch of student demographics. Affirmative action programs should lead to black and Hispanic students being more likely to be under-qualified for their schools, based on their measured ability. Consistent with this prediction, under-qualified students are more likely to be black or Hispanic than either well-matched or over-qualified students, seen in Table 3. We also consider sex and the student's age when he or she graduated high school, although our model has no predictions for the effect of age or sex on match quality.

VI. Multivariate Analysis

We estimate the probability that a student will be substantially over- or under-qualified for the first college she attends using probit models. We consider the effect of variables related to budget constraints, information and social constraints, and demographics. In our baseline specification, presented in the first two columns of Table 4, we again consider students poorly matched with their school if there is a greater than 20 percentile point gap between their ability percentile and their school's quality percentile. Models using other cutoffs, presented in Table 6, suggest that our results are quite robust to the choice of cutoff.

The estimated effects of exposure to college graduates on mismatch suggest that the cultural influences of growing up with more or less contact with college graduates dominate the additional information gained from these interactions. Consistent with the correlations shown in tables 2 and 3, students with higher quality state flagships and a higher share of college graduates in their neighborhood are more likely to be under-qualified and less likely to be over-qualified for their college. These effects are quite

symmetric. Increasing the quality of a student's state flagship by one percentile raises their probability of being under-qualified and lowers their probability of being over-qualified by 0.5 percent on average. Living within fifty miles of a suitable school has an important effect: a 15% reduction in the probability of being overqualified for one's college. Proximity is less important for predicting which students will be under-qualified for college, consistent with the fact that students without a well-matched college nearby are far more likely to have only a school of lower quality nearby than to have only higher quality schools nearby.

In the univariate correlations, the students of more educated mothers were less likely to be either over- or under-qualified, suggesting that they were better informed when choosing their colleges. However, controlling for other factors, the children of college graduates are less likely to be over-qualified but more likely to be under-qualified than the children of mothers with just a high school degree. This pattern is consistent with the cultural standards argument, but not with greater information leading to a better match.

Ability, as measured by the student's ASVAB score percentile, has a strong mechanical effect on the probability of mismatch. Very able students will have fewer schools for which they will be under-qualified and many schools for which they will be over-qualified. This effect is apparent in the estimates. However, a second measure of ability, high school GPA, has the opposite effect on mismatch. Students with higher ASVAB scores are less likely to be under-qualified for their colleges, but controlling for ASVAB score, students with higher GPAs in high school are more likely to be under-qualified for their schools. The reverse pattern is true for the probability of being over-

qualified. These results suggest that the incomplete information colleges have about their applicants' abilities contributes to college mismatch. Controlling for ASVAB-measure ability, which colleges do not see, students with lower grade point averages, which colleges do see, are more likely to end up at a school for which they are over-qualified, based on the performance on the ASVAB.[10]

We use family wealth, which is more closely related to permanent income, as our measure of financial constraint in our baseline specification. Consistent with the univariate analyses in tables 2 and 3, the wealthiest students are almost 7 percent less likely, on average, to be over qualified for their college. Interestingly, the least wealth are also less likely to be over qualified and more likely to be under qualified for their college than students in the middle of the wealth distribution. This pattern suggests that colleges may be practicing some affirmative action based on economic background. Estimates using income instead of wealth as a measure of ability to pay for college are presented in Table 5 and are very similar to the baseline specification.

Once we control for ability and socio-economic background, demographics play a fairly small role in predicting the quality of the college match. We find no evidence that affirmative action programs are raising the probability that black or Hispanic students will be under-qualified for their school using ASVAB score as our measure of ability. In fact, black students are, on average, 4% less likely to be under-qualified. However, this result is only weakly significant and may not hold at all schools. Krueger, Rothstein, and Turner (2006) find evidence of affirmative action, higher probability of admission conditional on test scores for minorities, mainly at the most selective schools. In Table 7 we consider interactions between sex and race and between demographics and ability, but

[10] The same pattern emerges if we use SAT score as the measure of college-observed ability.

still find little evidence of racial or ethnic affirmative action. ASVAB scores have very similar effects on probability of mismatch across sex, race, and ethnicity groups.

We also consider an alternative specification of match quality based only on the student's SAT score relative to the average SAT score of the incoming class at her college, presented in the third and fourth columns of Tables 4 and 5. This specification relies only on measures of ability observed by the colleges, so it does not capture all the mismatch that arises because colleges have imperfect information about the true ability of applicants and may also mask some of the students' lack of information about their own abilities relative to other college applicants. Additionally, the SAT score already embodies some of the guidance students have about applying for college if this information leads them to put extra effort into preparing for the SAT or ACT exams.

Using this measure of mismatch, both higher ASVAB scores and higher GPAs make students more likely to be over-qualified and less likely to be under-qualified, the mechanical relationship between ability and match that we would expect. The pattern that less wealthy students are more likely to be under qualified for their school is even more apparent looking at SAT match quality. However, students from wealthy families are no longer less likely to be under qualified using this measure, which we expect if part of the way wealthier parents ensure that their children will not end up at weaker schools is by investing in making their standardized test scores as strong as possible. Similarly, the share of adults in the neighborhood with college degrees and the quality of the state flagship have smaller effects using this specification, suggesting that these factors may influence mismatch through the student's preparation for college, including their

preparation for taking the SAT. Using this alternative measure of mismatch also reverses some of our conclusions about affirmative action.

VII. Conclusions

Looking only at simple correlations between student characteristics and match quality, student knowledge of colleges, social influences, and budget constraints all play a role in predicting which students will be substantially over or under qualified for their schools. When we control for the interactions between these effects, imperfect information on the part of both students and colleges emerges as the most important factor. As predicted by our model, financial constraints are more important in predicting over-qualification, which makes sense if low-income able students choose not to attend the most selective schools because they are more expensive. Interestingly, the least wealthy students are no more likely to be overqualified for their colleges than the wealthiest students. Rather, the students in the middle of the wealth distribution are the more likely to be at a lower quality college than their ability suggests they could aim for.

Students who are substantially under-qualified for their school may struggle through their classes and are less likely to complete their degree and more likely to take extra time to complete their degree. In addition to being costly for the students, this attrition and delayed graduation is expensive for the schools, governments, and organizations that subsidize post-secondary education and costly for the economy as a whole if the demand for college-educated labor is not fully met. On the other side, students who end up over-qualified for their college get less rigorous training than they might during their time in college. This may lead to lower earnings once they enter the

market and is again an inefficient use of educational resources, since some of our most able students are not being pushed to expand their knowledge and skills. Our results suggest that policy makers aiming to reduce these inefficiencies should focus on outreach efforts that make sure all students are well informed about their options for college, their chances of being admitted to different schools, and what they can expect to pay to attend them.

References

Alon, Sigal and Marta Tienda. 2005. "Assessing the 'Mismatch' Hypothesis: Differences in College Graduation Rates by Institutional Selectivity." *Sociology of Education* 78(4): 294-315.

Arcidiacono, Peter. 2004. "Ability sorting and the returns to college major." *Journal of Econometrics*, 121: 343 – 375.

Arulampalam, Wiji, Robin Naylor, and Jeremy Smith. 2005. "Effects of In-Class Variation and Student Rank on the Probability of Withdrawal: Cross-Section and Time-Series Analysis for UK University Students." *Economics of Education Review* 24: 251-262.

Avery, Christopher and Sarah Turner. 2008. "Playing the College Application Game: Critical Moves and the Link to Socio-Economic Circumstances." University of Virginia working paper.

Black, Dan and Jeffrey Smith. 2004. "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics*. 121(1): 99-124.

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

Brand, Jennie E. and Yu Xie. 2007. "Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education." California Center for Population Research Working Paper Series 033-07.

Bound, John, Michael Lovenheim, and Sarah E. Turner. 2007. "Understanding the Decrease in College Completion Rates and the Increased Time to the Baccalaureate Degree." PSC Research Report No. 07-626.

Cawley, John, James Heckman, and Edward Vytacil. 2001. "Three Observations on Wages and Measure Cognitive Ability." *Labor Economics*, 8(4): 419-442.

College Board. 2008. *Trends in Student Aid 2008*. New York: College Board Publications.

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," *The Quarterly Journal of Economics*, 117(4): 1491-1527.

Goldin, Claudia and Lawrence Katz. 2008. *The Race between Education and Technology*, Harvard University Press.

Griffith, Amanda and Donna Rothstein. 2007. "Can't Get Here from There: The Decision to Apply to a Selective Institution." Cornell CHERI WP 100.

Johnson, William R. 2006. “Hours Constraints In Market Equilibrium.” University of Virginia working paper.

Johnson, William R. 2008. “Hours Constraints Are Partially Illusory.” University of Virginia working paper.

Kane, Thomas J. and Christopher Avery. 2004. “Student Perceptions of College Opportunities: The Boston COACH Program” in Caroline Hoxby (ed.) *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*. Chicago: University of Chicago Press.

Krueger, Alan, Jesse Rothstein, and Sarah Turner. 2006. “Race, Income, and College in 25 Years: Evaluating Justice O’Connor’s Conjecture.” *American Law and Economics Review*, 8(2): 282–311.

Light, Audrey and Wayne Strayer. 2000. “Determinants of College Completion: School Quality or Student Ability?” *The Journal of Human Resources*, 35(2).

Reynolds, C. Lockwood (2009). "Where to Attend? Estimates of the Effects of Beginning College at a Two-year Institution." Kent State University working paper.

Neal, Derek and William R. Johnson. 1996. “The Role of Premarket Factors in Black-White Wage Differences.” *The Journal of Political Economy*, 104(5): 869-895.

Tables

Table 1: Joint Distribution of Student Ability and College Quality

	College Quality Quartiles				
Ability Quartiles	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile	Total
1st Quartile	8.9 (39.1) [34.6]	7.4 (32.4) [28.2]	4.5 (19.9) [18.3]	2.0 (8.6) [8.4]	(100.0) (N=617.3)
2nd Quartile	6.9 (27.5) [26.8]	7.5 (29.8) [28.5]	6.4 (25.6) [26.0]	4.3 (17.1) [18.3]	(100.0) (N=678.3)
3rd Quartile	5.5 (21.3) [21.2]	6.5 (25.2) [24.7]	7.3 (28.3) [29.4]	6.5 (25.2) [27.7]	(100.0) (N=695.4)
4th Quartile	4.5 (16.9) [17.4]	4.9 (18.4) [18.6]	6.5 (24.5) [26.3]	10.7 (40.2) [45.6]	(100.0) (N=718.6)
Total	[100.0] [N=697.8]	[100.0] [N=709.4]	[100.0] [N=668.9]	[100.0] [N=633.5]	100.0 N=2,709.6

Each cell contains the overall percentage, (the row percentage), and [the column percentage]

Table 2: Average Characteristics of Students by College Choice

	No 4-year College	College Attendees	College quality quartile 1	2	3	4
N	5,480.9	3,503.1	821.3	812.0	796.7	727.9
Average age when R graduated high school	18.9	18.4	18.4	18.5	18.4	18.3
% Male	55.0%	45.6%	44.0%	43.0%	43.8%	47.6%
% Black	18.1%	11.9%	19.1%	11.0%	8.4%	4.9%
% Hispanic	16.1%	7.8%	6.6%	8.4%	7.0%	5.8%
ASVAB percentile	37.6	66.0	59.0	63.1	69.4	78.9
High School GPA	2.6	3.2	3.1	3.2	3.3	3.4
SAT score	872.8	1051.5	972.0	1005.6	1083.9	1168.8
Mother's Highest grade	12.1	14.1	13.5	13.9	14.5	14.9
Father's Highest grade	12.1	14.4	13.7	14.0	14.9	15.4
Quality percentile of state flagship	75.6	74.4	67.5	74.4	76.4	79.5
% in census district with BA	19.0	21.0	18.2	19.8	23.4	23.5
Household income^	$41,774	$71,774	$55,615	$66,701	$76,860	$97,940
Household net worth^	$82,505	$223,068	$124,956	$207,141	$262,676	$350,670
In-state tuition at flagship^	$4,387	$4,502	$4,119	$4,574	$4,507	$4,889

Notes: This table describes the characteristics of students at each college quartile. For example, the third row shows the percent of students attending each college type that are male. Numbers calculated with probability weights to control for sample selection.
^ Dollar amounts are in 1997 dollars.

Table 3: Average Student Characteristics by Match Quality

	College Attendees	Very Under-qualified	Well-matched	Very Over-qualified
N	3,503.1	665.7	1,247.8	796.1
Average age when R started college	18.1	19.0	18.7	18.7
% Male	45.6%	32.9%	45.3%	51.9%
% Black	11.9%	13.6%	12.7%	5.2%
% Hispanic	7.8%	9.3%	6.1%	4.2%
ASVAB percentile	66.0	49.7	67.3	82.1
High School GPA	3.2	3.2	3.2	3.3
SAT score	1,051.5	988.9	1,066.1	1,123.7
Mother's Education*	14.1	14.0	14.4	14.2
Father's Education*	14.4	14.4	14.8	14.4
Quality percentile of state flagship	74.4	79.5	74.4	69.2
% in census district with BA	21.0	23.1	21.1	19.2
Household income^	$71,774	$79,545	$75,627	$65,324
Household net worth^	$223,068	$217,403	$273,202	$201,770
In-state tuition at flagship^	$4,502.2	$4,606.2	$4,547.1	$4,370.7

Notes: This table describes the characteristics of students in each mismatch group. For example, the third row shows the percent of students who are over- under- or well-matched that are male. Numbers calculated with probability weights to control for sample selection. Very over- or under-qualified is defined as a 20-point or greater gap between the student's ASVAB ability percentile and his or her college's quality percentile. These cutoffs assign roughly 25% of the sample to each mismatch category.
^ Dollar amounts are in 1997 dollars.

Table 4: Probit Estimates: Very Over-qualified and Very Under-qualified

	ASVAB-College Quality Match		SAT Score Match	
	Under-qualified	Over-qualified	Under-qualified	Over-qualified
Male	-0.017 (0.018)	-0.029 (0.021)	0.046* (0.024)	-0.009 (0.024)
Black	-0.057** (0.023)	-0.021 (0.032)	-0.015 (0.030)	0.085* (0.045)
Hispanic	-0.032 (0.029)	-0.01 (0.038)	-0.013 (0.040)	-0.075 (0.047)
Age started college	0.006 (0.007)	0.004 (0.008)	0.016* (0.009)	-0.007 (0.011)
ASVAB percentile	-0.008** (0.000)	0.008** (0.000)	-0.007** (0.001)	0.007** (0.001)
High school GPA	0.002** (0.000)	-0.002** (0.000)	0.000 (0.000)	0.001** (0.000)
Region of the U.S.				
Northeast	0.136** (0.034)	-0.09** (0.028)	0.007 (0.034)	0.004 (0.035)
South	0.003 (0.027)	-0.025 (0.029)	-0.028 (0.03)	-0.003 (0.034)
West	0.05 (0.031)	-0.047 (0.032)	-0.049 (0.033)	-0.005 (0.039)
Wealth Quartile				
1	0.063* (0.034)	-0.069** (0.034)	0.135** (0.05)	0.007 (0.046)
2	0.003 (0.025)	-0.032 (0.026)	0.068** (0.031)	-0.027 (0.032)
4	0.012 (0.022)	-0.067** (0.023)	0.034 (0.028)	-0.005 (0.027)
In-state tuition at flagship	-0.008 (0.01)	0.012 (0.012)	-0.011 (0.012)	0.012 (0.014)
Mother's educ.				
HS dropout	0.021 (0.035)	-0.033 (0.042)	0.078 (0.049)	-0.063 (0.048)
Some college	0.024 (0.025)	-0.031 (0.025)	0.015 (0.028)	-0.026 (0.03)
College graduate	0.085** (0.028)	-0.064** (0.025)	0.012 (0.03)	-0.059** (0.03)
More than college	0.043 (0.031)	-0.097** (0.028)	0.043 (0.037)	-0.046 (0.034)
Quality pctl of state flagship	0.005** (0.001)	-0.004** (0.001)	0.001 (0.001)	-0.001 (0.001)
% adults in district with BA	0.005** (0.001)	-0.006** (0.001)	-0.001 (0.001)	0.000 (0.002)
Has well-matched school within 50 miles	0.001 (0.025)	-0.156** (0.025)	-0.003 (0.029)	-0.059** (0.029)
Lives outside MSA	-0.041 (0.026)	-0.027 (0.028)	-0.009 (0.030)	0.035 (0.032)
N	1742	1742	1298	1298
Pseudo R2	0.305	0.276	0.255	0.265

Notes: Bold indicates significantly different from zero with 5% confidence, italics with 10%. Parameters are mean derivatives.

Table 5: Probit Estimates: Using Income Quintiles Instead of Wealth Quartiles

	ASVAB-College Quality Match		SAT Score Match	
	Under-qualified	Over-qualified	Under-qualified	Over-qualified
Male	-0.009 (0.018)	-0.026 (0.021)	0.047* (0.024)	-0.008 (0.024)
Black	-0.046* (0.024)	-0.011 (0.033)	-0.008 (0.031)	0.086* (0.046)
Hispanic	-0.04 (0.029)	-0.022 (0.039)	-0.03 (0.04)	-0.09* (0.049)
Age started college	0.009 (0.007)	0.006 (0.008)	0.017* (0.009)	-0.008 (0.011)
ASVAB percentile	-0.008** (0)	0.008** (0)	-0.007** (0.001)	0.008** (0.001)
High school GPA	0.002** (0)	-0.002** (0)	0 (0)	0.001* (0.001)
Region of the U.S.				
Northeast	0.137** (0.034)	-0.082** (0.03)	0.002 (0.034)	-0.007 (0.035)
South	0.007 (0.027)	-0.021 (0.03)	-0.019 (0.031)	-0.021 (0.034)
West	0.059* (0.032)	-0.04 (0.033)	-0.032 (0.035)	-0.01 (0.04)
Income quintile				
1	0.057 (0.035)	-0.050 (0.035)	0.049 (0.042)	-0.036 (0.046)
2	0.03 (0.034)	-0.059* (0.032)	0.005 (0.038)	-0.099** (0.035)
4	0.005 (0.027)	-0.003 (0.027)	-0.022 (0.03)	-0.046 (0.03)
5	0.057** (0.029)	-0.088** (0.026)	0.039 (0.034)	-0.068** (0.03)
In-state tuition at flagship	-0.005 (0.01)	0.011 (0.012)	-0.007 (0.013)	0.005 (0.014)
Mother's educ.				
HS dropout	0.015 (0.036)	-0.032 (0.044)	0.108** (0.053)	-0.066 (0.049)
Some college	0.034 (0.026)	-0.017 (0.026)	0.027 (0.029)	-0.027 (0.03)
College graduate	0.068** (0.029)	-0.055** (0.026)	0.017 (0.031)	-0.058* (0.031)
More than college	0.029 (0.031)	-0.075** (0.031)	0.023 (0.038)	-0.041 (0.036)
Quality pctl of state flagship	0.005** (0.001)	-0.004** (0.001)	0.001 (0.001)	-0.001 (0.001)
% adults in district with BA	0.005** (0.001)	-0.006** (0.001)	-0.001 (0.002)	0 (0.002)
Has well-matched school within 50 miles	-0.003 (0.025)	-0.157** (0.025)	-0.004 (0.029)	-0.046 (0.03)
Lives outside MSA	-0.047* (0.026)	-0.023 (0.028)	-0.015 (0.03)	0.029 (0.033)
N	1684	1684	1252	1252
Pseudo R2	0.3067	0.2758	0.2632	0.2697

Notes: ** indicates significantly different from zero with 5% confidence, * with 10%. Parameters are mean derivatives.

Table 6: Robustness Checks for ASVAB-College Quality Mismatch Cutoffs

	More than 10 percentile gap		More than 30 percentile gap	
	Under-qualified	Over-qualified	Under-qualified	Over-qualified
Male	0.000 (0.021)	-0.014 (0.023)	-0.006 (0.016)	-0.031* (0.018)
Black	-0.002 (0.029)	-0.041 (0.034)	-0.051** (0.017)	-0.037 (0.03)
Hispanic	-0.009 (0.035)	-0.021 (0.04)	-0.052** (0.02)	0.025 (0.036)
Age started college	0.003 (0.008)	0.003 (0.009)	0.012** (0.006)	0.005 (0.007)
ASVAB percentile	-0.009** (0.000)	0.008** (0.001)	-0.006** (0.000)	0.007** (0.000)
High school GPA	0.003** (0.000)	-0.002** (0.000)	0.002** (0.000)	-0.002** (0.000)
Region of the U.S.				
Northeast	0.123** (0.034)	-0.146** (0.031)	0.113** (0.031)	-0.101** (0.022)
South	-0.033 (0.03)	-0.032 (0.032)	-0.02 (0.023)	-0.003 (0.027)
West	0.025 (0.034)	-0.084** (0.034)	0.029 (0.027)	-0.022 (0.029)
Wealth Quartile				
1	0.037 (0.037)	-0.056 (0.039)	0.085** (0.033)	0.006 (0.035)
2	0.035 (0.029)	-0.046 (0.029)	0.016 (0.023)	0.017 (0.025)
4	0.023 (0.026)	-0.063** (0.026)	0.015 (0.021)	-0.033 (0.021)
In-state tuition at flagship	-0.004 (0.012)	0.005 (0.013)	-0.016* (0.009)	0.009 (0.011)
Mother's educ.				
HS dropout	0.069 (0.043)	-0.068 (0.044)	-0.001 (0.03)	-0.046 (0.037)
Some college	0.042 (0.028)	-0.015 (0.028)	0.022 (0.023)	-0.031 (0.022)
College graduate	0.111** (0.03)	-0.098** (0.028)	0.082** (0.027)	-0.058** (0.022)
More than college	0.061* (0.034)	-0.108** (0.032)	0.039 (0.028)	-0.102** (0.023)
Quality pctl of state flagship	0.005** (0.001)	-0.004** (0.001)	0.004** (0.001)	-0.002** (0.001)
% adults in district with BA	0.005** (0.001)	-0.007** (0.001)	0.004** (0.001)	-0.004** (0.001)
Has well-matched school within 50 miles	0.031 (0.028)	-0.114** (0.029)	0.004 (0.023)	-0.111** (0.021)
Lives outside MSA	-0.006 (0.03)	-0.023 (0.030)	-0.023 (0.022)	0.012 (0.025)
N	1742	1742	1742	1742
Pseudo R2	0.2638	0.2403	0.3052	0.2975

Notes: ** indicates significantly different from zero with 5% confidence, * with 10%. Parameters are mean derivatives.

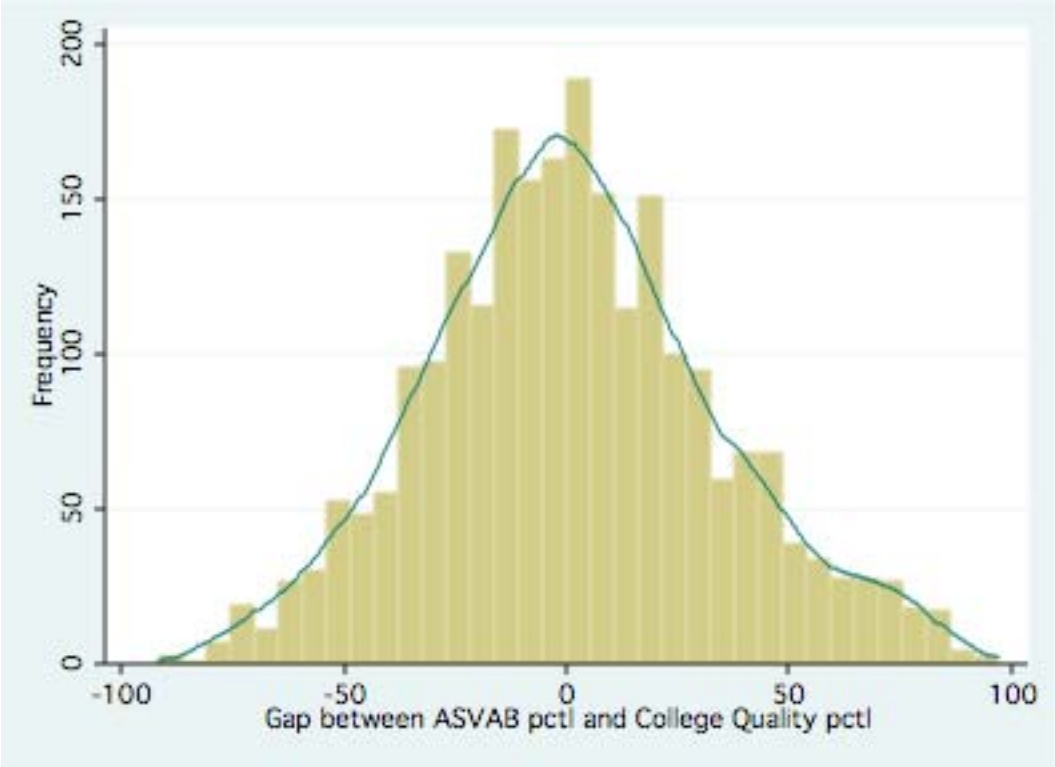
Table 7: Probit Estimates: ASVAB-College Quality Match with Interactions

	Under-qualified	Over-qualified
Male	-0.049 (0.034)	0.035 (0.052)
Black	-0.134** (0.027)	-0.109** (0.050)
Hispanic	-0.101** (0.035)	-0.106 (0.067)
Male* Black	0.108 (0.075)	0.009 (0.015)
Male*Hispanic	0.058 (0.037)	-0.009 (0.008)
Age started college	0.005 (0.007)	0.004 (0.008)
ASVAB percentile	-0.008** (0.000)	0.008** (0.001)
Black*ASVAB percentile	0.003 (0.003)	0.001 (0.002)
Hispanic*ASVAB percentile	0.002 (0.002)	0.001 (0.002)
Male*ASVAB percentile	0.001 (0.001)	-0.001 (0.001)
High school GPA	0.002** (0.000)	-0.002** (0.000)
Region of the U.S.		
Northeast	0.137** (0.034)	-0.09** (0.028)
South	0.009 (0.027)	-0.024 (0.03)
West	0.05 (0.031)	-0.047 (0.032)
Wealth Quartile		
1	0.066* (0.034)	-0.069** (0.035)
2	0.002 (0.025)	-0.033 (0.026)
4	0.011 (0.023)	-0.067** (0.023)
In-state tuition at flagship	-0.007 (0.010)	0.012 (0.012)
Mother's educ.		
HS dropout	0.021 (0.034)	-0.033 (0.043)
Some college	0.023 (0.025)	-0.032 (0.025)
College graduate	0.085** (0.029)	-0.067** (0.025)
More than college	0.045 (0.031)	-0.098** (0.028)
Quality pctl of state flagship	0.005** (0.001)	-0.004** (0.001)
% adults in district with BA	0.005** (0.001)	-0.006** (0.001)
Has well-matched school within 50 miles	0.002 (0.025)	-0.154** (0.025)
Lives outside MSA	-0.040 (0.026)	-0.026 (0.028)
N	1742	1742
Pseudo R2	0.311	0.278

Notes: ** indicates significantly different from zero with 5% confidence, * with 10%. Parameters are mean derivatives.

Figures

Figure 1: Distribution of College Mismatch



Mismatch defined as student ability percentile - college quality percentile. Histogram includes estimated kernel density distribution.

Data Appendix

We use data from the National Longitudinal Survey of Youth 1997 (NLSY97). This data set contains a very rich set of variables collected in annual interviews with 8,984 American youths. The first survey was conducted in 1997, when the respondents were 13 to 17 years old, and follow-up interviews have been conducted every year since then. The NLSY97 sample contains a representative sample of American youth and an over-sample of black and Hispanic youths. We use observations from both groups of respondents, using the weights developed by the survey collectors based on the inverse probability of being included in the survey.

We consider the college choices of respondents who attended a college during the sample period and for whom the first college attended was a four-year college, about a third of the full sample. Our estimates are based on the 2,385 respondents who went to a college for which we have all the components of our college quality measure and who took the ASVAB test, which we use to measure ability. Appendix Table 1 gives details on the construction of our sample.

Almost all the youths in the sample took the Armed Services Vocational Aptitude Battery (ASVAB) test. This test has 12 sections, covering the same topics as the SAT, arithmetic, vocabulary, and reading comp, but also other topics such as electronics knowledge and spatial reasoning. We estimate common factor loadings across the respondents' scores on all 12 sections. Following Black and Smith (2004) and Cawley, Heckman, and Vytacil (2001), we use the first principal component as our raw measure of student ability. To calculate an ability percentile for each respondent who went to college we rank their raw ability measure among the other respondents who reported

attending a four-year college and who were the same age when they took the ASVAB test. Therefore, this percentile indicates their rank *among college goers who were the same age when they took the ASVAB test*, not among the population as a whole.

We calculate a multi-faceted measure of college quality using the same process as in Black and Smith (2006) and Cawley, Heckman, and Vytacil (2001). We use data from the 2007 Integrated Postsecondary Education Data System (IPEDS) and from the 2009 US News and World Report College Rankings (which is based on data collected in 2008). The college quality index combines measures of peer quality: average SAT score of the incoming class, the resources of the school: faculty/student ratios and average faculty salaries, and "voting with your feet" measures of how students and their families assess the school: the share of applicants that are rejected and the freshman retention rate. SAT scores are from US News where available and IPEDS otherwise. All other variables are from IPEDS. We calculate factor loadings across these five quality measures and construct the first component factor.

Our final measure of college quality is that school's percentile among all four-year colleges in the IPEDS database, weighted by student body size. We weight by college size so that if students sorted into schools perfectly by ability and quality the students in the x ability percentile would be at schools in the x quality percentile. In this case, our measure of mismatch, student ability percentile minus college quality percentile, would be zero for all students.

Appendix Table 2 gives details for the construction of the other independent variables in our estimation.

Appendix Table 1: The Sample

Total Observations	8,984
Did not complete high school	1,462
Graduated HS or got GED	7,522
Attended a four-year college	3,182
Total college attendees with q-index	2,812
Of quality quartile 1	*654*
Of quality quartile 2	*634*
Of quality quartile 3	*620*
Of quality quartile 4	*565*
With q-index and ability measure	2,385

College characteristics are for first four-year college attended. We lose some additional observations in the multivariate analysis because other covariates are missing.

Appendix Table 2: Description of Independent Variables

Variable	Description
Male	Dummy variable equal to 1 if the respondent is male, 0 otherwise
Black	Dummy variable equal to 1 if the respondent lists black as a racial category, 0 otherwise
Non-black Hispanic	Dummy variable equal to 1 if the respondent lists Hispanic as an ethnic category and doesn't list black as a racial category
ASVAB percentile	Described in the data appendix
High School GPA	Collected from the respondent's high school transcript and standardized to a 4-point scale weighted by Carnegie credits
Region of the U.S.	Indicator variables for Census region where the respondent lived at the end of high school
SAT score	The combined score on the math and verbal section of the SAT (max score 1600), collected from the respondent's high school transcript. In cases where the student took only the ACT those scores are converted to the SAT using the method described on the ACT's website.
Mother's Education	The respondent's mother's self-reported highest grade completed. This measure is for the mother the respondent lived with in high school, whether that mother was biological, step, or adoptive.
Father's Education	The self-reported highest grade completed by the father the respondent lived with in high school.
Quality percentile of state flagship	Quality percentile, as described in the data appendix, of the flagship state university in the state where the respondent lived at the end of high school
% in census district with BA	The share of the adult (over 25) population that has a B.A. according to the 1990 Census in the census district where the respondent lived at the end of high school
Household income	Total gross income for the household where the student lived in 1997 in 1997 dollars. Later years are used to fill in missing data as long as the student had not yet graduated high school. Quintile cutoffs are from the Current Population Survey from the relevant year.
Household net wealth	Total household assets less total household liabilities for the household where the student lived in 1997 in 1997 dollars. Later years are used to fill in missing data as long as the student had not yet graduated high school. Quartile cutoffs are calculated within the NLSY sample, using sampling weights.
Has well-matched school within 50 miles	Equal to 1 if the respondent lived within 50 miles of a college with which she would be "well-matched" (college quality percentile within 20 points of her ability percentile) when she graduated high school.
Lives outside MSA	Equal to 1 if the respondent did not live in a Metropolitan Statistical, meaning she lived in a rural area, at the time she graduated high school.
In-state tuition at flagship	Posted tuition (not including any financial aid) for state residents at the flagship state university in the state where the respondent lived at the end of high school, from the 2007 IPEDS, in 1997 dollars.