# Caste- and Gender-Based Affirmative Action and College Quality in India

**Surendrakumar Bagde**
Indian Administrative Services

**Dennis Epple**
Carnegie Mellon University and NBER

**Lowell Taylor**
Carnegie Mellon University, NORC and NBER

August 8, 2015

We are grateful to David Card, Greg Caetano, Josh Kinsler, Kevin Lang, Karthik Muralidharan, Ronni Pavan, Michael Peress, Jesse Rothstein, Karl Scholz, and Chris Taber for helpful comments, and thank seminar participants at the CES/ifo Conference on Economics of Education, Delhi School of Economics, Indian School of Business, Indian Statistical Institute, the 7th Annual Conference on Indian Economic Growth and Development, Human Capital Working Group Conference on Measuring and Interpreting Inequality, National Bureau of Economic Research, and the following universities: Berkeley, Boston, Davis, Rochester, Stanford, Washington, Northwestern, Case Western, Carnegie Mellon, and Wisconsin.

**Abstract:** Public policy in modern Indian features affirmative action programs intended to reduce inequality that stems from a centuries-old caste structure and history of disparate treatment by gender. We study one such program—an admissions policy that fixes percentage quotas, common across more than 200 engineering colleges, for disadvantaged castes and for women. We find the program increases attendance of targeted students, particularly at relatively higher-quality colleges. An important concern is that affirmative action might harm intended beneficiaries by placing them in academic programs for which they are ill-prepared. Our analysis of achievement and graduation reveals no evidence of adverse impacts.

Many societies face the difficult public policy challenge of dealing with an historical legacy of discrimination and exclusion based on racial, ethnic, gender, or hereditary categories. One policy response is to implement affirmative action policies which explicitly favor historically disadvantaged groups. The hope is to level the playing field in the short term and to affect a longer-term transformation whereby society eventually no longer needs affirmative action.[1]

Affirmative action policies enacted in almost any situation generate controversy because they alter the allocation of scarce resources. In the case of higher education, which we study here, the preferential admission status granted to one student can result in the exclusion of some other student from a particular college. In addition, there are concerns about the effectiveness of affirmative action in higher education even for intended beneficiaries. At issue is the possibility that affirmative action can harm targeted students by placing them in academic situations for which they are poorly suited, creating a "mismatch." The argument is that students with weak preparation may fare poorly in the challenging environment of selective colleges; these students would be better off in institutions where their academic preparedness more closely matches that of their peers. As we discuss below, empirical evaluation of this issue has proved difficult.

Our paper contributes to the extant literature by examining the college matriculation and academic success of students in more than 200 private non-profit engineering colleges in a large State in India, where affirmative action is established policy. For this analysis we have assembled unique data, which include extensive information about a large number of applicants to many colleges. We know each student's caste and gender, and eligibility for preferential admission under the established rules of a gender- and caste-

[1] As Harry Blackmun famously suggested, in his legal opinion on affirmative action in higher education in the U.S., "I yield to no one in my earnest hope that the time will come when an affirmative action program is unnecessary and is, in truth, only a relic of the past" (*Regents of the University of California vs. Bakke,* 1978).

based affirmative action program. Our data also include student performance on standardized entrance examinations, as well as prior performance on standardized high-school completion examinations. Finally, we have two important measures of progression in college: *test scores*, collected after the first year in college, and *on-time graduation* (i.e., graduation at the end of the fourth year).

The design of the affirmative action program, combined with the availability of rich data, provide us with three important research advantages relative to most previous studies:

First, because policy establishes the uniform application of affirmative action, we are able to make a precise determination of how affirmative action policies affect student admission and college choice. Such a determination is often difficult in other settings in which affirmative action has been studied. For example, in the U.S. each university or college has its own admission policy, often based on factors that are difficult to quantify, such as student essays and interviews, weight placed on extracurricular activities, recommendation letters, and so forth. Identifying beneficiaries of preferential treatment is not possible, so instead imprecise inferences are drawn (typically using information about race or ethnicity, and variables such as prior academic performance and experience of economic hardship). In contrast, in our setting students are admitted through a fully transparent common admission system laid out by the State government. Colleges have discretion in the admission for up to 20 percent of their available seats, but the remaining 80 percent of seats are allocated via a common admission system. The affirmative action policies mandate specified proportions among these seats that must be set aside for students on the basis of gender and caste. We know each applicant's caste or social group, gender, and college, and we know the extent to which the applicant's choice priority is affected by the affirmative action policy.

Second, in many evaluations of affirmative action, credible comparison of academic outcomes across students is difficult. For example, subjective grading standards vary across faculty members, departments, and institutions, so grades and graduation rates are not comparable across colleges. Standardized examinations, such as the GRE, are typically taken only by students seeking admission for advanced degrees. Our setting provides a resolution of this problem. We have matriculation data for all engineering colleges that participate in the centralized allocation system. Of these colleges, 215 are affiliated; they share a common curriculum and administer externally-graded examinations to all students after the first year. This enables us to compare academic progress across colleges. We also have data on on-time graduation, which provides us with a second useful measure for evaluating the impact of the affirmative action program.

Third, in virtually all academic settings applicants are not obliged to accept an offer of admission. Among the college applicants we study, for example, those who qualify for a seat may choose not to attend any engineering college in the state, opting instead to attend some other academic institution, or to not attend any college. Because we have high school and entry examination scores for the entire applicant pool, not just matriculants, we can study factors that influence attendance decisions.

Affirmative action can potentially benefit members of targeted groups by enabling some applicants to attend college who would otherwise not have been admitted, and also by increasing priority in college selection. We document a substantial impact along the first dimension; affirmative action increases college attendance among targeted students, with effects that are proportionally greatest for the men and women in the most disadvantaged castes. As for the second dimension, we find that improved priority in college selection allows targeted students to attend higher-quality colleges, which in turn increases first-year achievement. Our analysis suggests that affirmative action policies have little impact on on-time graduation. In short, we find no evidence of "mismatch" that harms intended beneficiaries.

Our paper is organized as follows. In Section I we discuss affirmative action, with particular attention to the Indian case. Section II describes the data. In Section III we estimate models of matriculation, achievement, and graduation, and in Section IV we discuss results. Section V concludes.

## I. Affirmative Action

Affirmative action policies seek to increase diversity among those selected for productive or developmental opportunities—jobs, slots in school, military positions, government contracts, etc.—often as a means of ameliorating a legacy of discrimination and marginalization by society. Affirmative action entails some form of preferential treatment to a disadvantaged group, identified by gender, religious affiliation, ethnicity, race, and/or caste.

A large literature studies the economic properties of such policies, with an eye toward understanding their impacts in markets and improving their effectiveness for accomplishing desired social objectives.[2] Fryer and Loury (2005), for example, provide an excellent discussion of the tradeoffs inherent in affirmative action policies. Fryer, Loury, and Yuret (2007) study the welfare economics of affirmative action policies, including those that subsidize the skills development of disadvantaged groups. Fang and Moro's (2011) overview includes a discussion of the role affirmative action plays in markets where statistical discrimination is important. Theoretical work by Hickman (2009) finds beneficial incentive effects on college preparation of having individuals compete within their demographic group rather than compete with members of all groups. Recent work by Hafalir, Yenmz, and Yildirim (2013) demonstrates that there are attractive theoretical properties of affirmative action policies based on "minority reserves"—the class of policies used by the Indian colleges we study.

*A. Studies in the U.S. Context*

As for empirical work, a fair amount of research on affirmative action focuses on the U.S. context. That literature shows, first of all, that affirmative action clearly does sway the admission process in the U.S.—increasing the probability that African American students are admitted to elite colleges. For example, Bowen and Bok (1998) find that at each 50-point SAT interval the probability of college admission in elite schools is considerably higher for black students than for comparable white students, so that a race-neutral admission policy would substantially reduce the overall probability of admission for black college applicants to top schools.[3] Long (2004) estimates that accepted minority students in "elite colleges" (colleges in the top decile) would be reduced by 27 percent if preferential admissions for minority students in the U.S. were eliminated. Epple, *et al.* (2008) likewise find that race-neutral policy would

[2] The economics of affirmative action is developed in such work as Lundberg and Startz (1983), Coate and Loury (1993), Moro and Norman (2003), and Fryer and Loury (2005). Epple, Romano, and Sieg (2002, 2006, 2008) study admission and financial aid policies in higher education, and college and applicant preferences regarding socio-economic and racial diversity.

[3] Bowen and Bok (1998) note that their claim is consistent with the observed outcome for the University of California, Berkeley, when it switched from a race-sensitive to race-neutral admission policy, as that policy change reduced admission rates for black applicants from 49 percent to 16 percent. See Card and Krueger (2005) and Hinrichs (2012) for more evidence along these lines. The legal status of affirmative action in public institutions is currently in flux; the U.S. Supreme Court is now considering an important case, *Fisher v University of Texas*.

markedly reduce attendance of minority students in upper-tier colleges, and Howell (2010) finds advantages in admission for black and Hispanic applicants among the "most selective colleges."

As for the estimated impact of affirmative action policies on intended beneficiaries, results vary. Some studies suggest that there is little reason to be concerned about mismatch in U.S. higher education. For instance, Bok and Bowen (1998) find that within each SAT score interval, graduation rates for black students are positively correlated with college selectivity, and Alon and Tienda (2005) find that students generally benefit from attending more-selective colleges, with minority student gaining more, compared to white students, from attending a "most selective institution." Fischer and Massey (2007) similarly find no adverse impact of affirmative action for college students.

On the other hand, Sander (2004) argues that preferences in law school admission harm black students. Affirmative action leads black students into selective schools, he argues, where their grades suffer, leading to poor performance on the bar exam. A sequence of papers—Ayres and Brooks (2005), Ho (2005), Chambers, *et al.* (2005), and Barnes (2007)—evaluates this claim. Rothstein and Yoon (2008) provide a good discussion about the key issue: decisions about the construction of counterfactuals matter a great deal here. Under some reasonable choices the evidence of mismatch disappears. These papers underscore the difficulty of evaluating mismatch in the context of U.S. higher education.

Rothstein and Yoon (2008) also provide a useful discussion about mechanisms that might cause mismatch for intended beneficiaries of affirmative action. In the college environment we study, students all pursue the same broad course of study, engineering, and in that respect the environment is similar to law schools, in which students are preparing for the same qualifying exams (bar exams) and same profession. Mismatch does not occur in such circumstances because a beneficiary is taking the wrong coursework *per se* but, as Rothstein and Yoon argue, can occur for a student who is placed with generally-stronger classmates if materials and assessments are more rigorous than would be optimal for the student,[4] or if the

[4] For example, Duflo, Dupas, and Kremer (2011) found that tracking Kenyan children into classes on the basis of initial achievement improved educational outcomes, plausibly because it allowed teachers to teach students at a more appropriate level.

student finds it demoralizing to be in an environment in which peers are academically stronger and in which "social distance" from other students is high. On the other hand, increased priority in college choice afforded by affirmative action allows a targeted student to choose a higher-quality college, or at least one for which the student is better suited, and this could conceivably improve student performance. Careful empirical work is necessary to determine which of these forces is more important.

Of course in almost every affirmative action program, including the one we study, the policy serves merely to expand the choices of targeted students, so mismatch results only for students who make unwise educational decisions. This observation receives careful treatment in the recent work of Arcidiacono, *et al.* (2011). In their model, when intended beneficiaries of affirmative action make rational choices, they can be made worse off by affirmative action only when colleges offering preferential admissions have private information about the post-enrollment educational process.[5]

Even if an affirmative action program benefits a targeted student, e.g., by creating the opportunity to attend a school with higher-quality instruction and better resources, this might create a different kind of mismatch. A beneficiary might receive only a small gain from attending her school of choice, while the non-targeted student she displaces suffers a substantial decline in value-added. In our work below, we assess the impact of affirmative action not only on targeted students but also on other students.

*B. The Caste System and Affirmative Action in India*

As we have mentioned, India is an important setting for the study of affirmative action in higher education. The Indian caste system divides society into closed hereditary groups or castes (Shah, *et al.*, 2006). The numerous castes in India have been rather carefully classified, graded inequality being a fundamental principle of the system.[6]

---

[5] Arcidiacono, *et al.* (2011) examine evidence from one institution, Duke University, and find that Duke does indeed have relevant information that the student does not. While this finding cannot be used to establish that mismatch occurs at Duke, it does indicate that the possibility of mismatch exists.

[6] Economists have studied a variety of aspects related to the caste system. Akerlof (1976) famously provides thoughts about how a rigid caste structure could persist as a long-run equilibrium. Among many other examples, Munshi and Rosenzweig (2006) study how the caste system shapes career choices for Indian men and women, and Anderson (2011) studies the possibility that the caste structure leads to a breakdown in trade involving water rights at the village level. Each of these papers provides a valuable discussion of the nature and economic consequences of the caste system.

The set of castes known as Scheduled Castes (SC) are at the bottom of the caste hierarchy, and traditionally have suffered the most discrimination in terms of social exclusion and restricted access to educational opportunities; indeed SCs (formerly known as "untouchables") are outside the traditional four-fold division of Hindu society. Also, there are tribal communities, known as Scheduled Tribes (ST), who have lifestyle and religious practices quite distinct from mainstream Indian society (Deshpande, 2005). STs often live in remote and inaccessible places, making access to education difficult. Hnatkovska, Lahiri, and Paul (2012) document large differences in wages and consumption of SC/STs and non-SC/STs. They also document that over the past two decades there has been some convergence, much of which is accounted for by improvements in education among SC/STs.

In addition to the disadvantaged SC/STs, certain socially and educationally disadvantaged communities or castes are officially designated as Backward Classes (BC). This designation is determined mainly by the extent of educational "backwardness," position within the hierarchy of castes, and the traditional occupations within these classes. In the State in which our colleges are located, the BCs are further divided into four distinct groups: BC-A, BC-B, BC-C, and BC-D.

The Indian constitution, implemented in 1950, mandates affirmative action for SC and ST groups. Affirmative action has subsequently been extended by law to other disadvantaged castes, with variation across the States of India. Affirmative action operates in such matters as employment in the public sector, recruitment into civil services, and in education. While a variety of initiatives are intended to improve access to secondary and higher education for disadvantaged groups, the most important aspect of affirmative action in higher education is the reservation of seats for students from these social groups.

The extent of the quota system is substantial, though the Supreme Court of India has ruled that allotments to disadvantaged castes cannot exceed 50 percent. In the State we study seats are reserved as follows: 15 percent for SC, 6 percent for ST, 7 percent each for BC-A and BC-D, 10 percent for BC-B, and 1 percent for BC-C. Castes not accorded quotas are treated as a single group, designated Open.

In addition, in India, as in many countries, women are traditionally underrepresented in higher education. Affirmative action policy is implemented separately for men and women in the State we study.

One third of seats within each caste group, including Open, are reserved for women.

A small recent literature provides evidence about the impact of affirmative action in Indian higher education—mostly focusing on men. In the Indian context there seems to be little doubt that affirmative action does affect college admission, increasing opportunity for SC, ST, and BC students.[7] Importantly, it appears that the policy of targeting intended beneficiaries by caste (particularly SC and ST) has the intended effect of generally targeting students from disadvantaged families, i.e., does not simply result in prosperous lower-caste students displacing lower-income higher-caste students (see Bertrand, Hanna, and Mullainathan, 2010, and also Frisancho Robles and Krishna, 2012).

As for the impact of affirmative action on students themselves, the literature is quite sparse. In one recent study, Kochar (2010) shows that in one college (characterized as one of the "most selective colleges") affirmative action increases the variance of ability within classrooms, which in turn is detrimental to student learning. Frisancho Robles and Krishna (2012) similarly examine data from an "elite" college, and show that SC/ST students generally fall behind same-major peers in terms of grades.[8] Both of these papers, then, provide reason for concern about mismatch in elite Indian institutions.

Bertrand, Hanna, and Mullainathan (2010) estimate positive labor market returns to attending an engineering college among both lower-caste groups (SC/BCs) and upper-caste groups. As the authors suggest, "This contradicts the extreme view that seats in engineering colleges are 'wasted' on lower-caste-candidates." Due to small sample sizes, however, the causal estimates of benefits of attending an engineering college are accompanied by large standard errors, and difference in benefits across castes are not statistically significant.

Our objective is to build on the existing research to understand the effects of affirmative action. As

[7] See, e.g., Darity, Deshpande, and Weisskopf (2011), who draw parallels between the Indian and U.S. cases, and evaluate how affirmative action programs might affect admission prospects for potential beneficiaries in the two societies. These authors also provide reference to the earlier literature. (For research that gives a sense of challenges in the Indian educational system at the pre-college level, see Muralidharan and Sundararaman, 2011).

[8] As the authors of both papers note, they do not have access to entry examination scores, which creates challenges for controlling for pre-college preparation. The Kochar (2010) and Frisancho Robles and Krishna (2012) studies use GPA as their measure of academic achievement. Both institutions being studied seek to maintain uniformity across faculty and subjects, making GPA a valuable indicator of academic progression. As we have argued, though, there is additional merit in having an independently-measured metric of progress.

we have mentioned, the institutional design of affirmative action, combined with the extensive data we have for a large number of students, give us a substantial advantage over research addressing similar questions in the U.S. context. In comparison to recent analyses of affirmative action in Indian higher education, our work benefits from the use of data for a large number of students from many colleges (rather than one college only) as well as an external examination process, which has an advantage over internal assessments, which might vary across colleges and which might be subject to caste- or gender-based bias.

**II. Data**

We have gained access to data for 215 engineering colleges in a major State (with a population of more than 80 million). Priority for admission to the engineering colleges is based on the student's rank on a common entry examination and on caste and gender. In particular, the highest ranked individual in a caste/gender group can choose any college. Then the second highest chooses. This process continues, with choice among colleges limited to those having seats available for the individual's caste and gender at the time that the individual chooses. Open seats of men are allocated first based on overall entry exam score. Males of disadvantaged castes, and women from any caste, who qualify for a seat in a college without invoking caste/gender priority may take a seat on this round. A seat obtained in this way does not reduce the quota of the individual's caste/gender. Moreover, if such an individual subsequently "moves up" to another college by invoking caste/gender priority, the vacated seat may be taken by another member of the same caste/gender without reducing the quota of the individual's caste/gender.

Administrative records include the entry examination scores used in the college admission process, and also each student's gender and caste. To these records we have matched individuals' high school records from the State, which provide the scores received on the standardized examination taken at the end of high school. The starting point of our analysis is 131,290 individuals who took the State's entry examination for admission to engineering college and were ranked.[9] In matching individuals by name and identification number to high school records, we achieved a match rate of 89%, giving us a sample of

[9] Men in the Open and BC castes with very low scores (below 40 out of 160 questions) on the entry examination—approximately 7,000 individuals—are deemed ineligible for an engineering college and are not given a rank.

116,192 of individuals—80,771 men and 35,421 women for whom we have scores for both the high school exam and the entry exam (with components chemistry, math, and physics).

Table 1 provides summary statistics. Because affirmative action policy treats men and women differently, we provide summary statistics for men and women separately—men in Panel A and women in Panel B. Row 1 in each panel gives samples sizes. Rows 2 and 3 given mean performance on our two pre-college examinations: the three-hour entry examination for the engineering colleges (the "entry exam score"), and an intensive comprehensive twelve-day examination, approximately 42 hours of exam time, taken in high school (the "high school score").[10]

As is evident from Table 1, there are substantial differences in performance on these examinations among students of different castes. For men and women alike, the ordering of average scores on the entry exam and the high school exam conforms to the commonly accepted ordering of the degree of disadvantage of the caste groups: Scheduled Tribes (ST) and Scheduled Castes (SC) are considered the most disadvantaged, and of the BC classes, BC-A and BC-C are considered more disadvantaged than BC-B and BC-D. Open category students score best. Generally, women and men have similar exam scores.

As we have just noted, priority for college admission and choice is strictly determined by a ranking of students by caste and gender on the entry exam. Two features of this priority ranking deserve emphasis. First, on the entry examination, scores take on 160 discrete values. With more than 130,000 students taking the examination, there are necessarily an enormous number of ties. Ties are broken using the following algorithm: For students with the same score on the entry exam, ties are broken based on the score on the mathematics portion of the exam. For students who have the same score on both of those criteria, ties are broken by score on the physics portion of the exam. For students who have the same scores on all three of these criteria, ties are broken based on the high school examination score. Finally, if any students are still tied, the older candidate is given the higher rank. This idiosyncratic tie-breaking procedure is advantageous for our analysis; it aids in distinguishing the effect on college achievement of priority in choice of college

[10] In the State we study, students take standardized exams in grades 11 and 12. Hence, for the "high school score," we use combined test scores from the $11^{th}$ and $12^{th}$ grade. In each year, exams are spread over about two weeks.

from the effect of a student's own ability.[11]

Second, student priority is by rank on the entry exam *and* by caste and gender. To illustrate the impact of this affirmative action policy, we construct two rank variables: "Entry exam rank" simply ranks students from top to bottom, irrespective of caste and gender, and is normalized to the unit interval, with 1 indicating the best-ranked student and 0 the worst-ranked admitted student. In contrast, "effective rank" reflects an ordering based on the actual priority of college access, given established policy. Students of different caste/gender who have the same effective rank thus have the same priority in selecting colleges. The construction of this variable involves a series of mechanical steps described in Appendix A. For a student whose position improves due to affirmative action, effective rank will be higher than entry exam rank. The higher relative mean values of effective rank for students eligible for affirmative action show that these students have higher priority of access to colleges than comparable Open students.

The fourth row of Panel A in Table 1 shows that among men the ordering of the average *exam rank* across castes is the same as that of the mean entry exam (not surprisingly), while the fifth row shows that the ordering of average *effective rank* is roughly in reverse order. Thus, among men ST students have the highest average effective rank and Open students have the lowest. Panel B of Table 1 shows comparable statistics for women. Here again, women in disadvantaged castes have relatively low average entry exam rank, but fare well with respect to effective rank.

Figure 1 illustrates how exam rank and effective rank are related for caste/gender groups. As with entry exam rank, effective rank is normalized to have 1 as the highest value and 0 as the lowest value. We see that the value 0 pertains now only for the lowest-scoring Open-group man. As a consequence of affirmative action, effective rank is greater than entry exam rank across the board for students in all other gender-caste groups. Consider, for instance, applicants scoring at the 40$^{th}$ percentile of the entry

[11] We were able to verify that assigned ranks correspond to these rules. To give one example, 1501 students scored exactly 70 on the entry exam, of whom 926 were Open students. For these students the first tie-breaker was the score on the math exam, but this still left many ties. For instance, six students scored 32 on the math exam, and of these six the physics scores were as follows: one 17, two 16s, one 13, and two 12s. Thus for four of these students, tie-breaking relied on high school score. For instance, the two students who scored 16 on physics had high school scores of 872 and 757 respectively, and thus the former student was given the higher rank.

examination (i.e., an entry exam rank of 0.40). A male ST applicant with this entry exam performance has an effective rank greater than 0.80, while an ST woman has an effective rank of approximately 0.90. In contrast, an Open-caste male with this same score would has an effective rank of only 0.35.

Of the 116,192 ranked applicants for whom we have data, 53,374 elected to attend an engineering college in the State—enrolling in one of 245 colleges, of which 215 affiliated with a single university. Our sample of matriculants is comprised of the 42,914 students in the university-affiliated colleges—28,755 men and 14,159 women. For these matriculants we have records that allow us to assess subsequent success along two dimensions: First, students take a set of high-stakes examinations at the end of their first year—seven three-hour subject exams. Student progression to the second year depends on success on these exams. We norm the scores, with a mean of 0 and standard deviation of 1. The first row of Panel A in Table 2 gives mean scores by caste for men, and the first row of Panel B in Table 2 gives corresponding summary statistics for women. Within every caste grouping, women on average perform better than men on the first-year test. This is striking, given that women and men score similarly on high school and entry exams (see Table 1). Within both genders scores are relatively lower for the most disadvantaged castes.[12]

For almost all students—99.3% of men and 99.6% of women—we are also able to match graduation records. Our second outcome measure, on-time graduation, is also summarized in Table 2. As we discuss below, affirmative action gives students priority in selecting both college and the discipline of study. Thus, students who gain priority might select disciplines that differ in difficulty from those of students with lower priority, which could affect graduation rates, and indeed members of disadvantaged castes choose the more competitive majors in somewhat higher proportions than open caste students. The on-time graduation rate is high for women (0.90) and quite high also for men (0.75). Given that students in targeted castes generally

[12] The exams are designed to provide common guidance across the university-affiliated colleges on student progress. There are two complications. First, there is some variation across the exams by college major. Second, the final score—which determines progression to year two in the program—is determined by the university-administered exam (with weight 80%) that is assessed externally and a locally designed component (with weight 20%). For our primary measure of first-year performance, we use both components because both are used to assess student progression. We also demonstrate robustness using the following alternative: More than 90% of students (those in the largest majors) take six subject exams in common, out of a total of seven that are administered by the university. Thus, our robustness analysis uses only students who take these six exams and then uses only scores from those exams to evaluate first-year achievement.

have lower pre-college achievement than Open students and likely have other disadvantages (e.g., poorer parents), it is not surprising that they also have lower rates of on-time graduation. In our analysis below we attempt to assess the extent to which these patterns are the consequence of affirmative action policy.

We have one final variable that we use in our empirical work, a proxy for "college quality." While we do not have direct quality metrics for our colleges (faculty qualifications, classroom size, lab resources, etc.), we can form an informative proxy for college quality by assessing the extent to which highly-ranked students choose each college. We do this by constructing a college quality variable equal to the mean entry exam rank among the Open men enrolled in each college. There is substantial variation in this quality proxy. In the highest-quality college this mean is 0.95, while in the lowest-quality college it is only 0.24. There are five women's colleges (99% or more female), and we cannot form our quality measure for these colleges. In analyses that involve college quality we drop the 1484 students who attend these colleges.[13]

## III. College Attendance and Academic Achievement

Affirmative action improves priority in college choice for disadvantaged applicants and may thereby increase college attendance by these students, while also plausibly affecting subsequent academic performance. We study these issues in turn.

### *A. College Attendance*

An applicant to an engineering college in the State we study potentially has three choices: attend an engineering college in the State, attend some other academic institution, or choose the no-college option. We observe only whether the applicant attends an engineering college in the State; presumably the student makes that choice only when it is the highest-utility option among these choices. In Appendix B we develop a simple model that analyzes student choice. Two key predictions follow from our model:

First, the probability of matriculation in a State engineering college is monotonically increasing in effective rank. All else equal, the higher a one's effective rank, the larger will be the choice set of colleges, which increases the value of matriculation. This occurs because student can choose a college of higher

[13] We alternatively formed college quality measured by using Open-caste men *and* women, and then included students in the women's colleges. Nothing important in our analyses changes when we do so.

quality and/or choose a school that better accommodates the student's academic interests or personal circumstances (e.g., attending a school that is closer to home).

Second, conditional on a student's effective rank, the probability of matriculation is an inverted U-shaped function of the student's "latent aptitude" for engineering college. All else equal, we expect relatively low-aptitude students to be less likely to opt for engineering college; these students instead tend to take the no-college options. On the other hand, for very high-aptitude students, aptitude and attendance are inversely related because among these students an increase in aptitude improves the probability of admissions to an Indian Institute of Technology (IIT) or other high-prestige institution and thus *reduces* the probability of matriculation to one of our engineering colleges.

Of course the choice to attend an engineering college depends on costs, which might plausibly vary systematically by caste/gender group.[14] This motivates including caste fixed effects by gender.

Our logic leads us to an empirical model in which the probability of engineering college matriculation for student *i* in caste group *j* depends on that student's caste *j*, effective rank, $r_{ij}$, and variables designed to account for latent ability. Let $I_{ij}$ be our dependent variable—equal to 1if a student attends a State engineering college and 0 otherwise. Our regression is a linear probability model:[15]

(1) $$I_{ij} = \gamma_j + f(r_{ij}) + g(e_{ij}, h_{ij}) + \varepsilon_{ij},$$

where $\gamma_j$ are caste effects, $f(r_{ij})$ is a polynomial in effective rank, and $g(e_{ij}, h_{ij})$ is a polynomial function of the score on the entry exam $e_{ij}$ and high school exam $h_{ij}$ which is intended to serve as a control for latent aptitude. We let $f(r_{ij})$ be a fifth-order polynomial, which allows us to see if $I_{ij}$ is monotonically increasing in effective rank over the entire range of $r_{ij}$, as our model predicts. As for $g(e_{ij}, h_{ij})$, our control for a measure of latent aptitude, we employ a third-order polynomial that includes interactions. College attendance decisions might be quite different for men and women, so we estimate (1) separately by gender.

[14] A tuition waiver is provided to a small number of students—SC and ST students with income below a given threshold—but otherwise tuition is the same for all students. Differences in the opportunity cost and psychic cost of college attendance may create additional variation across individuals in the utility cost of college attendance.
[15] Results were very similar with a probit regression used in an earlier version of the paper.

Because we are using high-order polynomials, direct interpretation of coefficient estimates is difficult, and we thus we report these only in an appendix (see Appendix C). Instead, we focus on a graphical summary which illustrates the estimated relationships.

The first prediction from our theory is that the probability of college attendance is increasing in *priority* of access to college, i.e., we predict a monotone relationship between matriculation and effective rank. We find that effective rank is a powerful predictor of attendance (for both men and women all five coefficients on the effective rank polynomial terms are statistically significant at the 0.01 level), and Figure 2 shows that we observe the expected monotonic relationship for both men and women.[16]

Second, we expect an inverted U-shaped relationship between the attendance index and latent aptitude. As we have noted above, the high school exam score is based on an extensive series of tests, and we therefore expect it to be a quite good measure of aptitude. Also recall that the high school score has very little impact on college admission (mattering only as the last among a series of tie-breakers for students with the same scores on all components of the entry exam). Thus a reasonable way to examine our second prediction is to evaluate the relationship between attendance and high school score, holding exam score and effective rank constant. Figure 3 shows the expected inverted U-shaped relationship for both men and women.[17] This inverted U-shaped pattern is also obtained when we plot the relationship between attendance and entry exam school score while holding high school score and effective rank constant.

*B. Performance on the First-Year Exam*

The first outcome measure we analyze is academic achievement, as indicated by performance on the assessment exam given to students at the end of the first year. Our goal is to provide empirical evidence about the role of preferential admission policies on this outcome among college students.

It is natural in our regression to suppose that exam performance $p_{ij}$ will be closely related to latent ability, so we include a high-order polynomial in the entry exam and high school exam scores $g\left(e_{ij}, h_{ij}\right)$.

---

[16] There are no women with effective rank below 0.20. All men with effective rank below 0.20 are Open-caste, and college attendance is extremely low for these men.

[17] We hold entry exam score constant at the 80th percentile in Figure 5. Results are nearly identical if we use other benchmark percentiles (e.g., the 50th or 90th). We find comparable evidence if we instead use a probit specification.

Even with effective controls for latent ability, we expect that caste and gender might play an additional role in success, because students from differing backgrounds may face differing life circumstance that affect success in college. Thus we also include caste/gender fixed effects. Finally, and most importantly, we are interested in isolating any impact of effective rank on first-year academic achievement. Under the null hypothesis that priority in admission plays no role in academic success, effective rank has no impact on examination performance $p_{ij}$ (assuming, of course, that we have adequately controlled for ability). Alternatively, effective rank might be positively related to performance if priority allows beneficiaries the opportunity to choose a better learning environment, or it might be negatively related if the process of awarding admissions priority creates mismatch for beneficiaries. Furthermore, we want our specification to allow for the possibility that the impact of priority differs across caste/gender groups.

With all this in mind, we construct a series of indicator variables as follows: let $z_{if}$ be an indicator variable equal to 1 if person *i* is female, let $z_{im}$ be the corresponding indicator variable for male, and let $z_{ij}$ equal 1 if that person is a member of caste *j* (0 otherwise). Our regression is

$$(2) \qquad p_{ij} = \beta_0 + \beta_f z_{if} + \sum_{j=1}^{6} \beta_j z_{ij} + \sum_{j=1}^{6} \beta_{jf} z_{if} z_{ij} + \alpha_0 r_{ij} + \alpha_f z_{if} r_{ij} + \sum_{j=1}^{6} \alpha_{jm} z_{im} z_{ij} r_{ij} + \sum_{j=1}^{6} \alpha_{jf} z_{if} z_{ij} r_{ij} + g\left(e_{ij}, h_{ij}\right) + \varepsilon_{ij},$$

where, as in (1), $g\left(e_{ij}, h_{ij}\right)$ is a third-order polynomial that includes interactions. The coefficient $\alpha_0$ is the "main effect" of effective rank on performance, and our specification then allows the role of effective rank to differ across caste/gender groups. Under the null hypothesis that *priority in college choice has no impact on performance*, all $\alpha$ coefficients will be 0. We also estimate a variant of regression (2) that includes only effective rank and an interaction of effective rank and the female indicator (i.e., allowing the impact of effective rank to differ by gender but restricting it to be the same across castes within gender).

Identification requires that the effective rank variables not be linearly dependent on the elements of latent aptitude polynomial $g\left(e_{ij}, h_{ij}\right)$. In the variant of our model that has only effective rank and an interaction of effective rank and gender, one potential source of identification is obvious: Figure 1 shows that affirmative action gives rise to wide variation of effective rank for a given level of performance on the entry examination, which is clearly useful for identifying $\alpha_0$ and $\alpha_f$.

When we estimate the variant of our model with all 14 fourteen terms in $r_{ij}$ in equation (2), identification requires variation in effective rank within caste/gender groups (conditional on latent ability). Identification now comes from two sources. First, recall that latent aptitude is measured by $g(e_{ij}, h_{ij})$, a high-order polynomial based on a high school score that includes 42 hours of examinations and the college entry examination. Effective rank, in contrast, is constructed in a specific mechanical way almost exclusively from the entry exam (with the high school exam serving only to break ties in rare cases) and the entry quota system. Effective rank is therefore not linearly dependent on ability $g(e_{ij}, h_{ij})$. Second, identification is enhanced by the idiosyncratic tie-breaking procedure *and* differences across castes in the distribution of the entry exam. Two examples illustrate the combined effects of the latter. For SC and Open students with entry scores in the range from 40 and 45 (out of 160), the range of effective rank is 0.15 for both castes (0.47 to 0.62 for SC and 0.00 to 0.15 for Open). In contrast, for SC and Open students with entry scores from 55 to 60, the range of effective rank among SC students is 0.07 (0.83 to 0.90) while among Open students the range of effective rank is almost twice that amount, 0.13 (0.44 to 0.57).

By construction, effective rank introduces *no* information about ability beyond that contained in entry exam scores. In particular, this implies that the correlation between $r_{ij}$ and unmeasured components of ability remaining in $\varepsilon_{ij}$ in (2) is zero, as required for unbiased estimation of the coefficient of effective rank. We provide strong evidence for the validity of this assumption below.

Estimated coefficients and standard errors for our regression are provided in the first two columns of results in Table 3. Our estimate of the coefficient on effective rank $\alpha_0$ is positive and highly statistically significant. As for the interactions of effective rank and caste/gender indicator variables, only one is statistically significant at the 0.05 level. Indeed we cannot reject the hypothesis that the impact of effective rank is the same for all caste/gender groups (the F-stat = 1.31 for the hypothesis, $\alpha_f = \alpha_{1m} = \cdots = \alpha_{6f} = 0$, with a p-value of 0.20). Thus it appears that increased priority in admission results in stronger first-year academic achievement in a similar fashion across caste/gender groups.[18] We find no evidence of mismatch, as would be indicated by a negative relationship between effective rank and academic achievement.

The second regression reported in Table 3 retains only effective rank and the interaction of effective rank with the female indicator variable. We retain the gender interaction because, as we have seen (in Table 2), women fare better than men in college, and we thus want our empirical specification to allow estimated impacts of admission policies to vary by gender. We find that the impact of effective rank on first-year performance remains positive, and is slightly lower for women than for men. Women tend to outperform men on the first-year exam; this is reflected in the positive female fixed effect, which more than offsets the lower coefficient on effective rank.

As noted above, a key assumption of our analysis is that effective rank is not correlated with unmeasured components of ability in the error term in our achievement equation when we control for latent ability using a flexible function of entry exam score. This implies that our estimate of the coefficient of effective rank is unbiased even if we remove from our regression all six terms involving the *high school score* in our latent ability construct. These six terms are highly jointly significant (p-value $< 0.001$), and dropping them reduces the $R^2$ substantially (from 0.61 to 0.49). Hence dropping these six terms introduces a large unmeasured component of ability into the error term. Nonetheless, when we remove these terms from the regression and re-estimate, the estimated coefficient on effective rank is 0.561, very close to the estimate of 0.542 in regression (2) of Table 3 (with the coefficient on the interaction with female no longer statistically significant). This degree of robustness is strong evidence of the validity of our identifying assumption. We report results in regression (3) in Table 3.

As mentioned in footnote 13, some variation across first-year exam performance may be due to differences across the subject exams by college major and variation in the 20% portion of the score that is locally designed and graded. Fortunately, more than 90% of students (those in the largest majors, including

[18] To explore robustness, we include entry exam rank as an additional regressor to the first model in Table 3. The coefficient of this variable is small (-0.06) and insignificant (p = 0.85) while the coefficient rank of effective rank is negligibly affected (0.66) and remains highly significant (p = 0.009).

those that focus on electrical, electronic, and computer engineering) take six subject exams in common, out of a total of seven that are administered by the university. Thus, an alternative way to proceed is to use *only* students who take these six exams and then use *only* scores from those exams—all of which are externally designed, administer, and graded—to evaluate first-year achievement. Results, reported in regression (4) of Table 3, are very similar to regression (2), which uses the full first-year exam.

*C. Affirmative Action and College Quality*

Regressions given in Table 3 might reasonably be deemed "reduced form." A plausible underlying mechanism is that higher priority in college choice (i.e., a higher effective rank) allows a student to attend a college of higher quality, which in turn this leads to stronger academic achievement. We investigate this mechanism using an IV approach.[19]

We begin with a first stage in which the "quality" of the college a student attends depends on the student's caste/gender, latent ability and effective rank. Our measure of college quality ($q_{ij}$) is the mean entrance exam rank of Open men at the college. This variable is a measure of the academic ability of non-targeted peers, and under the assumption that Open men with high choice priority generally select colleges with positive attributes (well qualified faculty, good laboratory facilities, strong alumni networks, etc.), this measure is also a reasonable proxy for college quality more generally. We anticipate that effective rank will be a good predictor of this college quality measure.

Our main equation is

(3) $$p_{ij} = \beta_0 + \beta_f z_{if} + \sum_{j=1}^{6} \beta_j z_{ij} + \sum_{j=1}^{6} \beta_{jf} z_{if} z_{ij} + \delta q_{ij} + g\left(e_{ij}, h_{ij}\right) + \varepsilon_{ij}.$$

We are instrumenting $q_{ij}$ with effective rank $r_{ij}$. Identification relies on an assumption that effective rank is correlated with our performance measure only through its effect on the quality of the college attended.

Results are reported in Table 4, using our two dependent variables: first, the standardized first-year exam score, and, second, the standardized scores using only tests that are externally administered and graded. From Panel B we see that the first stage works as expected. Effective rank is positively associated

[19] We are grateful to a referee for recommendation of this strategy.

with our college quality measure and is highly statistically significant (the first-stage marginal F-statistics are, respectively, 8.04 and 19.79, quite solid given that we have one instrument).[20] As for the second stage (Panel A), our 2SLS estimates of the coefficient on college quality are positive and statistically significant.[21]

Estimated impacts are quite large; the two regressions in Table 4 imply that a one standard deviation increase in the college quality proxy (0.12) increases first-year test score by, respectively, 0.46 and 0.30 standard deviations. We can link these estimates to affirmative action policy as follows: the first stage of regression (1) implies that a one standard deviation increase in effective rank (0.17) increases a student's college quality by 0.140 × 0.17 = 0.024. Multiplying this increase by the second-stage coefficient on college quality (3.859), the resulting impact on first-year score is an increase of 0.09 standard deviations. A similar exercise for regression (2) indicates that a one standard deviation increase in effective rank yields a 0.10 standard deviation increase in first year score. If instead we estimate this same impact using the reduced-form regression (3) in Table 3, our inference is 0.17 × 0.614 = 0.10, i.e., approximately the same.

Our evidence is consistent with the interpretation that when affirmative action policies improve a student's effective rank, this allows the student to attend a higher-quality school, which in turn serves to improve academic achievement as measured by test performance at the end of the first year.[22]

*D. College Graduation*

Our final analysis focuses on graduation. An important concern raised in the mismatch literature is that intended beneficiaries of affirmative action might find themselves in academic situations for which they are poorly prepared, and thus graduate at lower rates than they would have in the absence of affirmative action policies. While we find above that that affirmative action policies enhance first-year success, this does not guarantee smooth continued progression through college. We can evaluate this issue because for

[20] We also estimated our "first-stage regression" including gender/caste indicator variables interacted with the effective rank variable. We cannot reject that these coefficients are 0 (p-value on the test for equality is 0.41). This is consistent with effective rank having a similar impact on college quality for students in all demographic groups.
[21] As a further robustness check, we estimate this same regression but only for non-OC men (25,977 students), i.e., only potential beneficiaries of affirmative action. The first-stage F-statistic is 18.71, and the coefficient on college quality is virtually unchanged, 2.49 (with a p-value of 0.0056).
[22] This result—that attending a higher-quality school results in stronger performance, as measured in a standardized test—appears in a number other studies, e.g., Pop-Eleches and Urquiola (2013) and other papers cited therein.

most engineering students who completed the first year (99.3% of men and 99.6% women), we also know if the student graduated by the end of the second semester of the fourth year.

Recall from Table 2 that graduation rates are substantially lower for some disadvantaged groups (especially ST, SC, and BC-C men) than for Open men. This fact alone is not persuasive evidence of mismatch. There are many factors other than academic preparation that affect success in college—financial duress, family obligation, or other personal issues—that might be systematically different across caste. Also, it is worth noting that women in all castes receive favorable treatment in admission, and graduation rates are quite high even for women in disadvantaged castes.

Our analysis of college graduation parallels our analysis of first-year achievement. Specifically we estimate a regression that has the same specification as (3), using effective rank as an instrument for college quality, but the dependent variable is on-time graduation (1 for an on-time graduation, 0 otherwise).

Estimates are reported in Table 5. We notice that conditional on latent ability, women graduate at substantially higher rates than men, and this is especially true among some disadvantaged castes. We cannot reject the hypothesis that the impact of college quality on graduation is 0 ($p = 0.81$). To the extent that affirmative action allows students to enroll in higher-quality colleges this seems to neither improve graduation prospects nor harm them.[23] We discuss in this finding in additional detail below.

**IV. Quantifying the Effects of Affirmative Action**

To summarize, our empirical work shows, first of all, that policies serve to substantially increase the priority of college admission (effective rank) of men and women in disadvantaged castes (particularly ST and SC), while reducing priority of Open men. In turn, effective rank is strongly and monotonically related to college attendance. Second, effective rank is positively related to achievement in the first year of college, and this relationship holds for all caste/gender groups. The implication is that affirmative action policies benefit targeted groups while disadvantaging Open men, for whom policies generally reduce effective rank. Third, we find no statistically significant impact of effective rank on graduation.

[23] We also analyze the relationship between effective rank and college graduation in a reduced form regression. We find no statistically significant impact of effective rank on graduation when we control for latent ability.

In this Section, we evaluate the size of these impacts implied by our model.

*A. Effect of Affirmative Action on College Attendance*

Using our college attendance model we predict attendance for each student with and without affirmative action. Table 6 reports predicted changes in attendance by gender for each caste. As this table shows, affirmative action dramatically increases attendance by men and women in disadvantaged castes. To pick one example, among the 1,558 SC women predicted to attend one of the State colleges, 450 would have attended in the absence of the affirmative action program, while the other 1,108 would not have attended.

In Table 7 we show mean *differences* in the predicted attendance rates by caste and gender and by quintiles on the entry examination. We notice three features: First, all disadvantaged-caste men and women benefit from affirmative action, as evidenced by the positive mean difference in attendance for all percentiles of the entry exam rank. Second, the mean differences are negative for Open-caste men and for most Open-caste women, showing the adverse impact of affirmative action on these individuals. Third, the effects of affirmative action tend to be greatest for individuals in middle quintiles of entry exam rank, with effect magnitudes declining as one moves to either the highest or lowest quintile ranges.

Even with the attendance gains from affirmative action, the most disadvantaged castes still attend in smaller proportions than their population shares. The quota shares allocated to disadvantaged castes are roughly in proportion to their population shares. For the ST and SC castes, the quota shares are, respectively, 6% and 15%. From Table 2, we see that only 2.5% and 10.8% of male matriculants are from ST and SC castes, well below their allocated shares.[24] The shares of women are well below half for all castes, with the differential between men and women being greatest in the more disadvantaged castes. ST women occupy only 0.9% of shares allocated to women and SC women occupy only 7.7% of these seats.

*B. Effect of Affirmative Action on First-Year Test Scores*

[24] By contrast, the least disadvantaged of the BC castes obtain shares in excess of their allocated quotas. BC-B and BC-D realized shares are, respectively, 18.3% and 14.5%, which are substantially higher than the respective allocated shares, 10% and 7%. As we discuss in Appendix A, the large realized shares relative to quota shares for the BC-B and BC-D castes are a consequence of a controversial regulation, Government Rule 550.

Table 8 provide estimates of the mean effects of affirmative action on first-year academic performance for each caste/gender group, using regression (2) of Table 3. We have seen that ST/SC students have the largest increases in effective rank due to affirmative action policies. These same groups experience the largest estimated increases in achievement due to affirmative action. For ST and SC men, average test-score gains are, respectively, 0.18 and 0.12 standard deviations, and for ST and SC women, they are, respectively, 0.20 and 0.14. These gains are substantial, comparable in magnitude to the striking effects of teacher pay policies obtained by Muralidharan and Sundararaman (2011) in Indian primary schools. Estimated impacts on other disadvantaged castes are somewhat smaller. OC men face *reductions* in effective rank due to affirmative action policies and we thus expect to observe declines in first-year test performance. Average losses are 0.05 standard deviations for OC men. Estimated effects for OC women are very small.

*C. Effect of Affirmative Action on Graduation*

Affirmative action policies in engineering colleges gives priority both in admission and in choice of discipline. Thus students who earned priority might have selected into different disciplines than Open caste men. To get a sense of this issue, in Table 9 we list the major disciplines, ordered according to competitiveness, as measured by high school score.[25] Interestingly, disadvantaged-caste men and women tend to select into the most competitive disciplines (Electronics and Communication Engineering and Computer Science) at slightly higher rates than men admitted in the Open category. There is no evidence that beneficiaries of affirmative action choose easier majors, which might improve on-time graduation.

With this in mind, recall that our regression analysis provides no statistically significant link between effective rank—our measure for priority in admission—and on-time graduation for either men or women. Students from disadvantaged castes choose more competitive majors, but nonetheless we find no compelling evidence that affirmative action reduces the average prospects of graduation for beneficiaries.

*D. Discussion*

[25] Nothing changes if instead we use admission exam scores or effective rank to establish the ordering of disciplines.

Our counterfactual analyses assume that the college choice, achievement, and graduation models would remain unchanged with an alternative system for determining effective rank for college admissions. This is a strong assumption. An alternative ranking system will change the allocation of students across colleges. This reallocation of students can be expected to change peer characteristics within colleges, which may lead to further impacts on student achievement. It may be that our counterfactual analysis understates the adverse effect on disadvantaged castes of eliminating affirmative action. If affirmative action were eliminated, members of disadvantaged castes would have lower priority in college choice. This would tend to concentrate lower-performing disadvantaged students into a subset of colleges that higher-ranked students perceive as less attractive. Thus in the absence of affirmative action, disadvantaged-caste students would almost surely attend colleges with weaker peers than they do with affirmative action. Such colleges might then be less able to attract a strong college faculty.

Fryer, Loury, and Yuret (2007) highlight the importance of affirmative action in incentivizing pre-college effort by members of targeted groups. This in turn can increase application rates and success rates of members of those groups. We see in Table 1 that the proportions of ST and SC men taking the entry examination are far below their proportions in the state we study. The same is true with respect to women applicants, as only 30% of entry exam takers are women. Following the logic of Fryer, Loury and Yuret, these proportions would likely have been even lower in the absence of affirmative action. If so, our counterfactual analysis may understate the benefits of affirmative action for disadvantaged castes, and also many understate the adverse effects of affirmative action on advantaged castes. Having said this, there may also be some positive spillover to students in advantaged castes if affirmative action policies encourage disadvantaged students to improve college preparedness (i.e., make them better peers).

We conducted additional analysis to see if heterogeneity in ability within colleges adversely affected achievement, and we did not find such effects. Having said that, it is possible that concentrating weaker students in the same colleges would have permitted better curriculum targeting (as suggested in Duflo, Dupas, and Kremer, 2011). The preceding discussion makes clear that counterfactual calculations, while informative, should be viewed as suggestive with respect to assessing the magnitude of the overall impact of affirmative action.[26]

## V. Conclusions

Affirmative action in Indian higher education is designed to deal with a difficult issue—societal inequalities that stem from the legacy of a caste system and a history of disparate treatment of women. In this paper we analyze an exceptional assemblage of data for a large number of engineering colleges, which allows us to assess the impact of affirmative action in Indian higher education.

We find that affirmative action policy works largely as intended. The policy clearly increases attendance among targeted students, especially for those in the most disadvantaged groups. In principle, affirmative action policies might harm intended beneficiaries by placing them in academic situations for which they are poorly suited. We find no evidence for this sort of mismatch. To the contrary, increased college choice priority has a positive impact on beneficiaries, at least as indicated by achievement on a comprehensive test taken after the first year of college. The program does not adversely affect average graduation rates despite the fact that disadvantaged-caste students choose competitive majors at higher rates than other students.

Despite the presence of a beneficial affirmative action program, we find large gaps in pre-college preparation, college participation, and college academic performance between the most disadvantaged castes and their more advantaged counterparts. The gaps in participation rates are magnified for women, especially for women from the most disadvantaged castes. Our work also indicates why affirmative action policies generate debate; we find that improved educational outcomes for targeted students come at a cost to those who do not receive preferential treatment.

[26] More generally, while affirmative action works as intended in the large system of engineering colleges we study, this does not rule out the possibility of mismatch in the most elite colleges (see, e.g., Kochar, 2010, and Frisancho Robles and Krishna, 2012).

## References


Akerlof, George (1976), "The Economics of Caste and of the Rat Race and Other Woeful Tales," *Quarterly Journal of Economics,* 90(4), 599-617.

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

Anderson, Siwan (2011), "Caste as an Impediment to Trade," *American Economic Journal: Applied Economics,* 3(1), 239-263.

Arcidiacono Peter, Esteban M. Aucejo, Hanming Fang, and Kenneth I. Spenner (2011), "Does Affirmative Action Lead to Mismatch? A New Test and Evidence," *Quantitative Economics,* 2(3), 303-333.

Ayres, Ian and Richard Brooks (2005), "Does Affirmative Action Reduce the Number of Black Lawyers?" *Stanford Law Review,* 57(6), 1807-1854.

Barnes, Katherine Y. (2007), "Is Affirmative Action Responsible for the Achievement Gap Between Black and White Law Students," *Northwestern University Law Review*, 101(4), 1759-1808.

Bertrand, Marianne & Hanna, Rema & Mullainathan, Sendhil (2010), "Affirmative Action in Education: Evidence from Engineering College Admissions in India," *Journal of Public Economics*, Vol. 94(1-2), 16-29.

Bowen, W.G. and D. Bok (2000), *The Shape of the River*. New Jersey: Princeton University Press.

Card, D. and Krueger, A. (2005). "Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence from California and Texas," *Industrial and Labor Relation Review*, 58(3), 416-434.

Chambers, David L., Timothy T. Clydesdale, William C. Kidder, and Richard O. Lempert (2005), "The Real Impact of Eliminating Affirmative Action in American Law Schools: An Empirical Critique of Richard Sanders' Study," *Stanford Law Review,* 57(6), 1855-1898.

Coate, S. and Loury, G. (1993), "Will Affirmative Action Policies Eliminate Negative Stereotypes?" *American Economic Review*, 83, 1220–1240.

Darity, William, Ashwini Deshpande, and Thomas Weisskopf (2011), "Who Is Eligible? Should Affirmative Action be Group- or Class-Based?" *American Journal of Economics and Sociology*, 70(1), 238-268.

Deshpande, Ashwini (2001), "Caste at Birth? Redefining disparity in India," *Review of Development Economics*, 5(1), 130-144.

Deshpande, A. (2005), "Affirmative action in India and the United States," Background Paper, *World Development Report*.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2011), "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101, 1739-1774.

Epple, D., Romano, R., and Sieg, H. (2002), "On the Demographic Composition of Colleges and Universities in Market Equilibrium," *American Economic Review: Papers and Proceedings*, 92 (2), 310–314.

_________ (2006), "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education," *Econometrica*, 74(4), 885-928.

_________ (2008), "Diversity and Affirmative Action in Higher Education," *Journal of Public Economic Theory*, 10(4), 475-501.

Fang, Hanming and Andrea Moro (2011), "Theories of Statistical Discrimination and Affirmative Action: A Survey," in Jess Benhabib, Matthew Jackson, and Alberto Bisin, eds., *Handbook of Social Economics*, vol. 1A, Netherlands: Elsevier.

Fischer, M., and D. Massey (2007), "The Effects of Affirmative Action in Higher Education," *Social Science Research*, 36, 531-549.

Frisancho Robles, Veronica, and Kala Krishna (2012), "Affirmative Action in Higher Education in India: Targeting, Catch Up, and Mismatch," *NBER Working Paper* w17727.

Fryer, R. and Loury, G. (2005), "Affirmative Action and its Mythology," *Journal of Economic Perspectives*, 19(3), 147-162.

Fryer, R., G. Loury, and T. Yuret (2007), "An Economic Analysis of Color-Blind Affirmative Action", *Journal of Law, Economics, and Organization*, 24(2), 319-355.

Hafalir, Isa E., M. Bumin Yenmez, and Muhammed A. Yilidirim (2013), "Effective Affirmative Action in School Choice," *Theoretical Economics*, 8, 325-363.

Heckman, J. (1979), "Sample Selection Bias As a Specification Error," *Econometrica*, 47(1), 153-161.

Hickman, Brent, "Effort, Race Gaps, and Affirmative Action: A Game-Theoretic Analysis of College Admissions, University of Chicago Working Paper, October 2009.

Hinrichs, Peter (2012), "The Effects of Affirmative Action Bans on College Enrollment, Educational Attainment, and the Demographic Composition of Universities," *Review of Economics and Statistics*, 94(3), 712-722.

Hnatkovska, Viktoria, Amartya Lahiri, Sourabh B. Paul, "Castes and Labor Mobility," *American Economic Journal: Applied Economics*, 4(2), 274-307.

Ho, Daniel (2005), "Why Affirmative Action Does Not Cause Black Students to Fail the Bar," *Yale Law Review*, 114(8), 1997-2004.

Howell, J. (2010), "Assessing the Impact of Eliminating Affirmative Action in Higher Education," *Journal of Labor Economics*, 28(1), 113-166.

Indian Express, "Scheduled Caste students discriminated against in IITs, IIMs, AIIMS: NCSC," September 6, 2012. http://www.indianexpress.com/news/scheduled-caste-students-discriminated-against-in-iits-iims-aiims-ncsc/998826/

Kochar, Anjini (2010), "Affirmative Action through Quotas: The Effect on Learning in India," working paper, Stanford Center for International Development.

Long, M. (2004), "Race and College Admissions: An Alternative to Affirmative Action," *The Review of Economics and Statistics*, 86(4), 1020-1033.

Lundberg, S. (1991), "The Enforcement of Equal Opportunity Laws Under Imperfect Information: Affirmative Action and Alternatives," *Quarterly Journal of Economics*, 106(1), 309–26.

Lundberg, S. and R. Startz (1983), "Private Discrimination and Social Intervention in Competitive Markets," *American Economic Review*, 73(3), 340-347.

Moro, A. and Norman, P. (2003), "Affirmative Action in a Competitive Economy," *Journal of Public Economics*, 87, 567–94.

__________________ (2004), "A General Equilibrium Model of Statistical Discrimination," *Journal of Economic Theory*, 114, 1–30.

Munshi, Kaivan and Mark Rosenzweig (2006), "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy," *American Economic Review*, 96(4), 1225-1252.

Muralidharan, Karthik and Venkatesh Sundararaman (2011), "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 119(1), 39-77.

Pop-Eleches, Cristian, and Miguel Urquiola (2013). "Going to a Better School: Effects and Behavioral Responses," *American Economic Review* 103(4), 1289-1324.

Rothstein, Jesse and Albert Yoon (2008), "Affirmative Action in Law School Admissions: What Do Racial Preferences Do?" *University of Chicago Law Review*, 75(2), 649-714.

Sander, Richard H. (2004), "A Systemic Analysis of Affirmative Action in American Law Schools," *Stanford Law Rev*iew, 367.

Shah, G., H. Mander, S. Thorat, S. Deshpande, and A. Baviskar (2006), "Untouchability in Rural India," *Sage Publications*, New Delhi, India.

Srinivas, M.N. (2009), "Social Change in Modern India," *Orient Black Swan*, New Delhi.

**Figure 1. Relationship between Entry Exam Rank and Effective Rank by Caste and Gender**

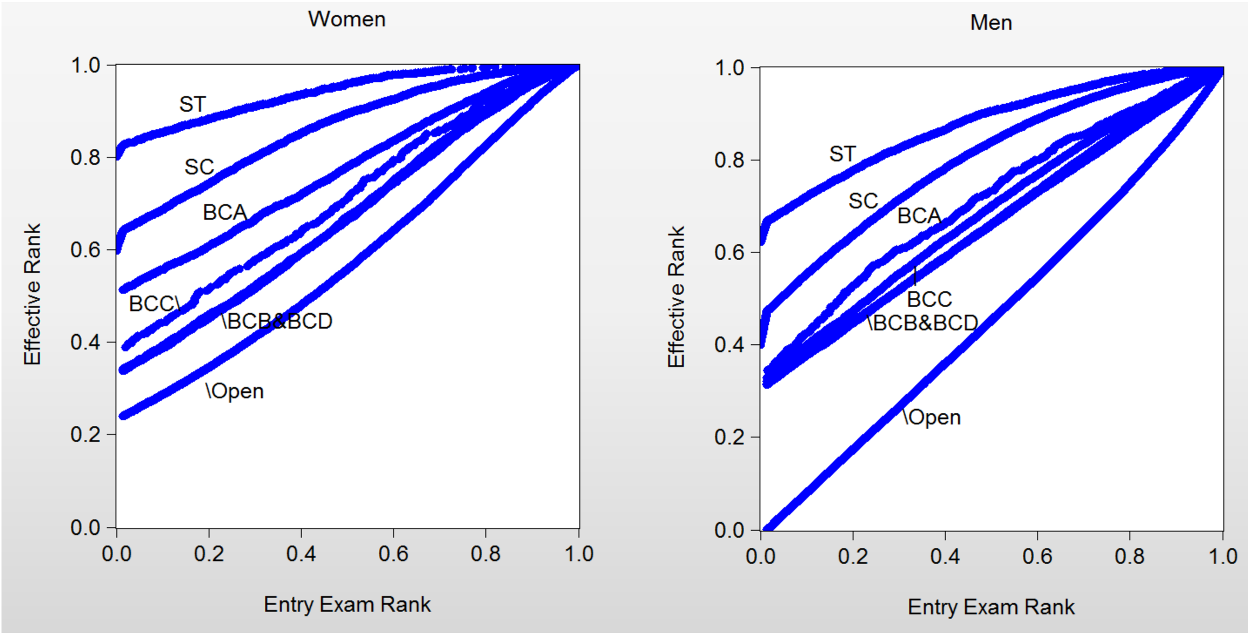

**Figure 2. Relationship between Effective Rank and College Attendance**

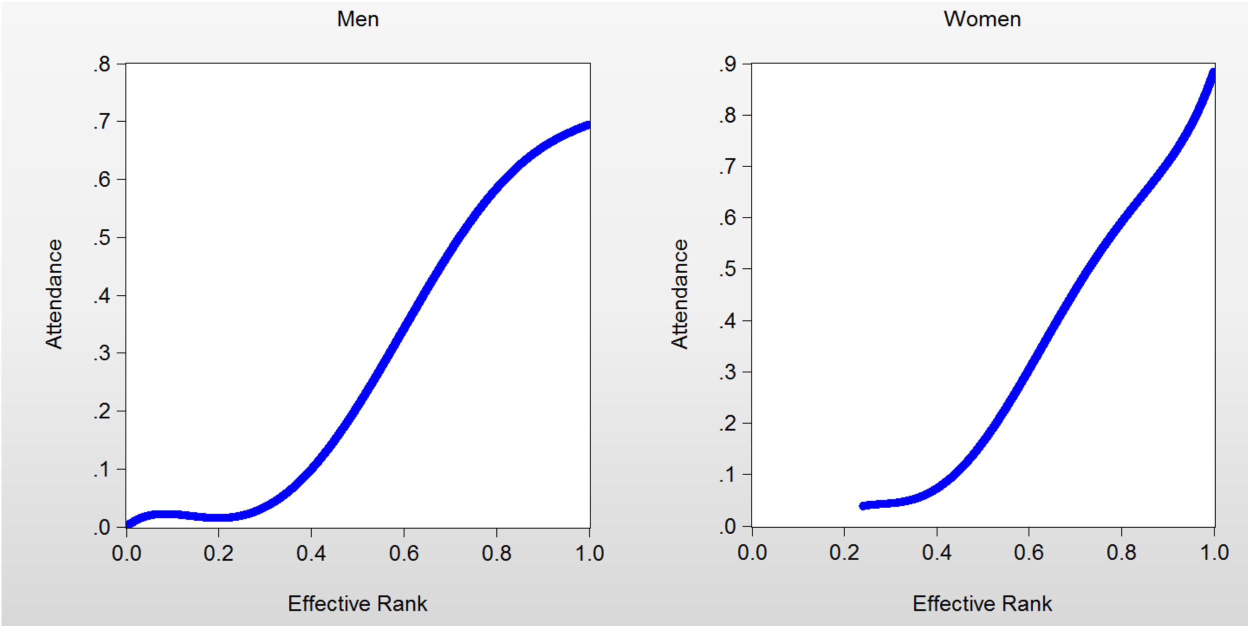

**Figure 3. Relationship between High School Score and Attendance**

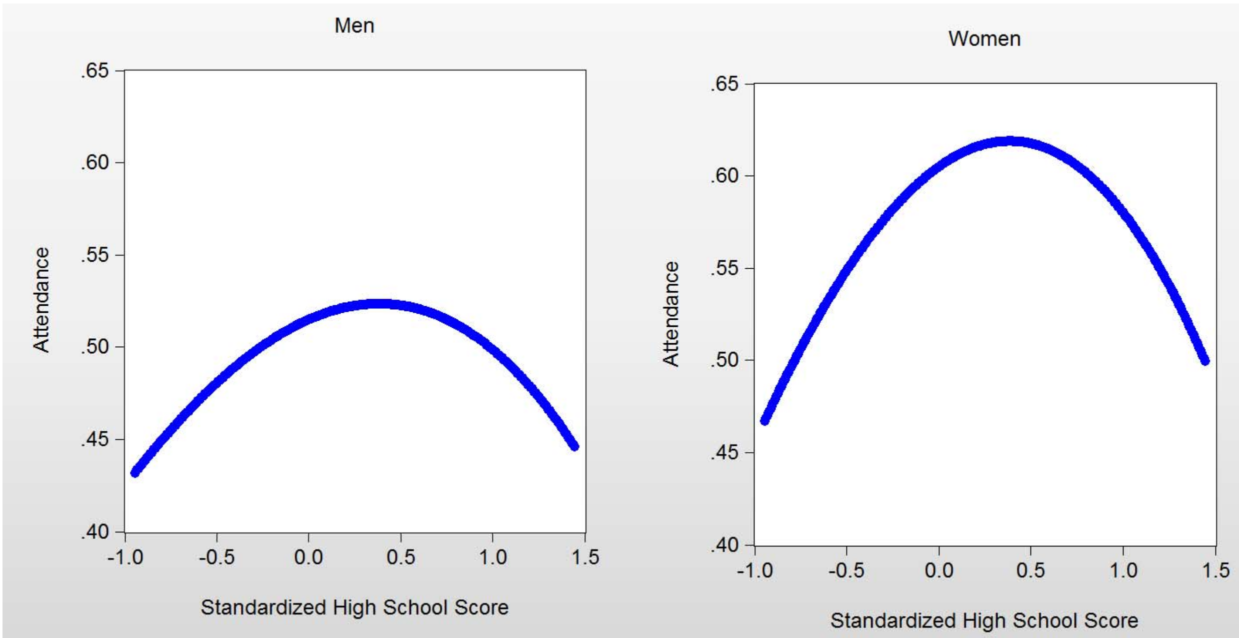

**Table 1. Summary Statistics of Variables Used to Study Matriculation**

| | Full Sample | ST | SC | BC-A | BC-B | BC-C | BC-D | Open |
|---|---|---|---|---|---|---|---|---|
| **A. Men** | | | | | | | | |
| Observations | 92,214 | 2486 | 9868 | 5619 | 13,613 | 719 | 10,571 | 49,338 |
| % of all Observations | 70.2% | 1.9% | 7.5% | 4.3% | 10.4% | 0.50% | 8.1% | 37.6% |
| Mean Entry Exam Score | 60.7 | 50.8 | 51.3 | 58.4 | 60.9 | 57.1 | 60.0 | 63.4 |
| (Standard Deviation) | (18.8) | (12.4) | (13.2) | (16.4) | (18.3) | (16.6) | (17.2) | (20.0) |
| Mean High School Score | 751 | 675 | 680 | 737 | 759 | 719 | 751 | 767 |
| (Standard Deviation) | (153) | (157) | (160) | (152) | (148) | (162) | (150) | (148) |
| Entry Exam Rank | 0.49 | 0.32 | 0.33 | 0.47 | 0.50 | 0.43 | 0.50 | 0.54 |
| (Standard Deviation) | (0.29) | (0.26) | (0.27) | (0.28) | (0.28) | (0.28) | (0.28) | (0.29) |
| Effective Rank | 0.60 | 0.82 | 0.72 | 0.68 | 0.67 | 0.69 | 0.67 | 0.52 |
| (Standard Deviation) | (0.26) | (0.11) | (0.17) | (0.19) | (0.19) | (0.19) | (0.20) | (0.37) |
| Proportion Attending | 0.44 | 0.51 | 0.55 | 0.55 | 0.54 | 0.54 | 0.53 | 0.29 |
| **B. Women** | | | | | | | | |
| Observations | 39,076 | 651 | 3,297 | 1,865 | 5,250 | 336 | 3,747 | 23,930 |
| % of all Observations | 29.8% | 0.5% | 2.5% | 1.4% | 4.0% | 0.3% | 2.9% | 18.2% |
| Mean Entry Exam Score | 60.6 | 49.2 | 51.3 | 57.5 | 59.5 | 57.6 | 58.9 | 62.9 |
| (Standard Deviation) | (16.9) | (10.5) | (12.3) | (14.3) | (15.4) | (14.3) | (15.1) | (17.8) |
| Mean High School Score | 744 | 654 | 685 | 712 | 738 | 700 | 727 | 761 |
| (Standard Deviation) | (156) | (136) | (151) | (158) | (155) | (159) | (156) | (153) |
| Entry Exam Rank | 0.51 | 0.30 | 0.34 | 0.47 | 0.50 | 0.47 | 0.49 | 0.55 |
| (Standard Deviation) | (0.28) | (0.23) | (0.26) | (0.26) | (0.26) | (0.26) | (0.26) | (0.27) |
| Effective Rank | 0.67 | 0.90 | 0.81 | 0.76 | 0.68 | 0.70 | 0.68 | 0.63 |
| (Standard Deviation) | (0.24) | (0.06) | (0.12) | (0.14) | (0.19) | (0.18) | (0.19) | (0.22) |
| Proportion Attending | 0.49 | 0.46 | 0.53 | 0.54 | 0.54 | 0.57 | 0.54 | 0.47 |

*Notes:* The high school score is available for 80,771 of the 92,214 men taking the entry exam, and 35,421 or the 39,076 women taking the entry exam, so statistics are calculated using these smaller samples.

**Table 2. Summary Statistics of Variables for Matriculants**

| | Full Sample | ST | SC | BC-A | BC-B | BC-C | BC-D | Open |
|---|---|---|---|---|---|---|---|---|
| **A. Men** | | | | | | | | |
| Mean of 1st year Score | -0.08 | -0.85 | -0.77 | -0.30 | -0.14 | -0.39 | -0.19 | 0.22 |
| (Standard Deviation) | (1.00) | (0.88) | (0.88) | (0.92) | (0.95) | (0.98) | (0.95) | (0.95) |
| On-Time Graduation | 0.75 | 0.60 | 0.54 | 0.70 | 0.76 | 0.61 | 0.75 | 0.82 |
| College Quality | 0.75 | 0.79 | 0.78 | 0.77 | 0.74 | 0.77 | 0.74 | 0.74 |
| Observations | 28,755 | 720 | 3117 | 2083 | 5270 | 232 | 4176 | 13,157 |
| % of Male Attendees | | 2.5% | 10.8% | 7.2% | 18.3% | 0.8% | 14.5% | 45.8% |
| **B. Women** | | | | | | | | |
| Mean of 1st year Score | 0.32 | -0.63 | -0.36 | 0.07 | 0.21 | 0.09 | 0.17 | 0.49 |
| (Standard Deviation) | (0.86) | (0.79) | (0.82) | (0.79) | (0.82) | (0.86) | (0.81) | (0.83) |
| On-Time Graduation | 0.90 | 0.73 | 0.77 | 0.87 | 0.90 | 0.92 | 0.91 | 0.93 |
| College Quality | 0.75 | 0.79 | 0.79 | 0.79 | 0.76 | 0.78 | 0.76 | 0.74 |
| Observations | 14,159 | 131 | 1,085 | 681 | 2,086 | 133 | 1,496 | 8,547 |
| % of Female Attendees | | 0.9% | 7.7% | 4.8% | 14.7% | 0.9% | 10.6% | 60.4% |

*Notes:* "College quality" is the mean effective rank among Open-caste men attending a given college, and is available for 210 of the 215 colleges (exceptions are all-female colleges). 1st year scores are normalized to have mean 0 and standard deviation 1. Graduation rates are available for 99.2% of men and 99.6% of women who completed the first year of college.

**Table 3. OLS Regression: Impact of Effective Rank on First-Year Achievement**

| | Regression (1) | | Regression (2) | | Regression (3) | | Regression (4) | |
|---|---|---|---|---|---|---|---|---|
| | Coefficient (s.e.) | | Coefficient (s.e.) | | Coefficient (s.e) | | Coefficient (s.e) | |
| Intercept | -0.851** | (0.087) | -0.811** | (0.090) | -0.611** | (0.062) | -0.927** | (0.096) |
| Female (F) | 0.233** | (0.049) | 0.293** | (0.036) | 0.316** | (0.033) | 0.288** | (0.040) |
| ST | -0.674* | (0.330) | -0.371** | (0.079) | -0.425** | (0.081) | -0.269** | (0.078) |
| SC | -0.29 | (0.156) | -0.268** | (0.054) | -0.322** | (0.059) | -0.212** | (0.055) |
| BC-A | -0.136 | (0.112) | -0.135** | (0.035) | -0.166** | (0.039) | -0.108** | (0.034) |
| BC-B | -0.169 | (0.084) | -0.093** | (0.026) | -0.099** | (0.028) | -0.063* | (0.026) |
| BC-C | -0.124 | (0.236) | -0.143* | (0.059) | -0.194** | (0.064) | -0.110 | (0.062) |
| BC-D | -0.208* | (0.090) | -0.129** | (0.028) | -0.148** | (0.030) | -0.116** | (0.029) |
| ST × F | -1.269 | (0.830) | 0.020 | (0.067) | -0.080 | (0.073) | -0.047 | (0.070) |
| SC × F | 0.008 | 0.162) | 0.095** | (0.025) | 0.023 | (0.028) | 0.088** | (0.025) |
| BC-A × F | 0.163 | (0.187) | 0.070* | (0.025) | 0.032 | (0.030) | 0.058* | (0.028) |
| BC-B × F | 0.307** | (0.094) | 0.070** | (0.018) | 0.056** | (0.022) | 0.055** | (0.018) |
| BC-C × F | -0.065 | (0.261) | 0.122 | (0.067) | 0.076 | (0.082) | 0.130 | (0.067) |
| BC-D × F | 0.256** | (0.099) | 0.099** | (0.021) | 0.078** | (0.026) | 0.097** | (0.024) |
| Effective rank ($r$) | 0.620** | (0.155) | 0.542** | (0.157) | 0.561** | (0.134) | 0.615** | (0.166) |
| $r$ × F | -0.024 | (0.056) | -0.095* | (0.045) | 0.014 | (0.039) | -0.094* | (0.046) |
| $r$ × ST × M | 0.313 | (0.340) | | | | | | |
| $r$ × SC × M | -0.014 | (0.163) | | | | | | |
| $r$ × BC-A × M | -0.022 | (0.130) | | | | | | |
| $r$ × BC-B × M | 0.084 | (0.092) | | | | | | |
| $r$ × BC-C × M | -0.054 | (0.299) | | | | | | |
| $r$ × BC-D × M | 0.086 | (0.102) | | | | | | |
| $r$ × ST × F | 1.683 | (0.977) | | | | | | |
| $r$ × SC × F | 0.085 | (0.258) | | | | | | |
| $r$ × BC-A × F | -0.141 | (0.218) | | | | | | |
| $r$ × BC-B × F | -0.227* | (0.115) | | | | | | |
| $r$ × BC-C × F | 0.207 | (0.356) | | | | | | |
| $r$ × BC-D × F | -0.117 | (0.116) | | | | | | |
| HS Score (HSS) | 0.642** | (0.011) | 0.643** | (0.011) | | | 0.635** | (0.012) |
| Entry Score (ES) | 0.086 | (0.054) | 0.117* | (0.046) | 0.695** | (0.049) | 0.129** | (0.046) |
| HSS × ES | 0.142** | (0.018) | 0.141** | (0.018) | | | 0.151** | (0.020) |
| HSS Squared | 0.069** | (0.009) | 0.069** | (0.009) | | | 0.068** | (0.010) |
| ES Squared | 0.040 | (0.027) | 0.030 | (0.016) | 0.025 | (0.016) | 0.028 | (0.017) |
| HSS Squared × ES | 0.020 | (0.013) | 0.020 | (0.013) | | | 0.025 | (0.015) |
| HSS × ES Squared | 0.003 | (0.013) | 0.004 | (0.013) | | | -0.005 | (0.015) |
| HSS Cubed | -0.003 | (0.007) | -0.003 | (0.007) | | | -0.007 | (0.008) |
| ES Cubed | -0.027** | (0.008) | -0.026** | (0.007) | -0.042** | (0.006) | -0.023** | (0.008) |

*Notes:* $n$ = 41,528 in regressions (1) – (3); $n$ = 37,580 in regression (4). Dependent variable: in (1) – (3) the standardized first-year test score; in (4) the standardized first-year test score only for tests administered and graded externally. $R^2$ = 0.61 in regressions (1) and (2); $R^2$ = 0.49 in regression (3); and $R^2$ = 0.60 in regression (4). Standard errors clustered at the college level. *Significant at 0.05; **significant at 0.01.

**Table 4. 2SLS Regression: Impact of College Quality on First-Year Achievement**

| Panel A: Second Stage | Regression (1) | | Regression (2) | |
|---|---|---|---|---|
| | Coefficient | (s.e.) | Coefficient | (s.e.) |
| Intercept | -3.057* | (1.239) | -2.184** | (0.642) |
| Female | 0.174** | (0.044) | 0.204** | (0.025) |
| ST | -0.795* | (0.314) | -0.496** | (0.186) |
| SC | -0.707* | (0.283) | -0.472** | (0.168) |
| BC-A | -0.415* | (0.175) | -0.274** | (0.102) |
| BC-B | -0.225* | (0.096) | -0.126* | (0.055) |
| BC-C | -0.431* | (0.195) | -0.277* | (0.119) |
| BC-D | -0.240** | (0.089) | -0.163** | (0.052) |
| ST × Female | 0.073 | (0.089) | -0.031 | (0.075) |
| SC × Female | 0.150** | (0.038) | 0.115** | (0.026) |
| BC-A × Female | 0.061 | (0.032) | 0.038 | (0.031) |
| BC-B × Female | 0.034 | (0.021) | 0.000 | (0.021) |
| BC-C × Female | 0.119 | (0.082) | 0.089 | (0.070) |
| BC-D × Female | 0.035 | (0.028) | 0.031 | (0.024) |
| College quality | 3.859* | (1.866) | 2.513* | (0.994) |
| HS Score | 0.646** | (0.013) | 0.629** | (0.013) |
| Entry Score | -0.137 | (0.184) | 0.001 | (0.107) |
| HS Score × Entry Score | 0.102** | (0.026) | 0.126** | (0.022) |
| HS Score Squared | 0.110** | (0.024) | 0.090** | (0.014) |
| Entry Score Squared | -0.033 | (0.031) | -0.027 | (0.023) |
| HS Score Squared × Entry Score | 0.018 | (0.015) | 0.028 | (0.016) |
| HS Score × Entry Score Squared | 0.015 | (0.015) | 0.005 | (0.016) |
| HS Score Cubed | 0.011 | (0.011) | 0.000 | (0.009) |
| Entry Score Cubed | -0.008 | (0.013) | -0.010 | (0.010) |

| Panel B: First Stage | | | | |
|---|---|---|---|---|
| Effective rank | 0.140** | (0.049) | 0.242** | (0.054) |
| Marginal F statistic | F = 8.04 | | F = 19.79 | |

*Notes:* $n$ = 41,528 in regression (1); $n$ = 37,580 in regression (2). Dependent variable: in (1), the standardized first-year test score; in (2), the standardized first-year test score only for tests administered and graded externally. $R^2 = 0.47$ in regression (1), $R^2 = 0.54$ in regression (2). Standard errors clustered at the college level. *Significant at 0.05; **significant at 0.01. The first stage regression includes also all variables included in the second stage (except of course college quality).

**Table 5. 2SLS Regression: Impact of College Quality on On-Time Graduation**

Panel A: Second Stage

| | Coefficient | (s.e.) |
|---|---|---|
| Intercept | 1.132* | (0.536) |
| Female | 0.114** | (0.020) |
| ST | 0.048 | (0.142) |
| SC | -0.027 | (0.129) |
| BC-A | 0.012 | (0.079) |
| BC-B | 0.024 | (0.043) |
| BC-C | -0.046 | (0.090) |
| BC-D | 0.021 | (0.040) |
| ST × Female | 0.005 | (0.042) |
| SC × Female | 0.088** | (0.021) |
| BC-A × Female | 0.052* | (0.017) |
| BC-B × Female | 0.014 | (0.011) |
| BC-C × Female | 0.164* | (0.045) |
| BC-D × Female | 0.038* | (0.014) |
| College quality | -0.607 | (0.809) |
| High School Score | 0.205** | (0.009) |
| Entry Exam Score | 0.077 | (0.082) |
| High School Score × Entry Exam Score | 0.037** | (0.013) |
| High School Score Squared | -0.050** | (0.010) |
| Entry Exam Score Squared | -0.008 | (0.010) |
| High School Score Squared × Entry Exam Score | -0.024** | (0.005) |
| High School Score × Entry Exam Score Squared | -0.003 | (0.003) |
| High School Score Cubed | -0.019** | (0.008) |
| Entry Exam Score Cubed | 0.007 | (0.009) |

Panel B: First Stage

| | | |
|---|---|---|
| Effective rank | 0.140** | (0.049) |

*Notes:* $n = 41{,}268$. Dependent variable is the standardized score on the 1st year test. $R^2 = 0.47$. Standard errors clustered at the college level. *Significant at 0.05; **significant at 0.01. The first stage regression included also all variables included in the second stage (except of course college quality).

**Table 6. Changes in Attendance from Affirmative Action**

| | (1) | (2) | (3) |
|---|---|---|---|
| Attended with Affirmative Action? | Yes | Yes | No |
| Would Have Attended without Affirmative Action? | Yes | No | Yes |
| ST Men | 303 | 863 | 0 |
| SC Men | 2,089 | 2,516 | 0 |
| BC-A Men | 2,274 | 607 | 0 |
| BC-B Men | 6,139 | 796 | 0 |
| BC-C Men | 246 | 109 | 0 |
| BC-D Men | 4,626 | 663 | 0 |
| Open Caste Men | 16,487 | 0 | 5,702 |
| ST Women | 18 | 191 | 0 |
| SC Women | 450 | 1,108 | 0 |
| BC-A Women | 563 | 398 | 0 |
| BC-B Women | 2,263 | 377 | 0 |
| BC-C Women | 123 | 46 | 0 |
| BC-D Women | 1,601 | 278 | 0 |
| Open Caste Women | 10,505 | 0 | 1,053 |

**Table 7. Mean Difference (MD) in Predicted Attendance: Rates By Caste and Entry Exam Rank**

| Rank (R) on Entry Exam | ST MD | ST *n* | SC MD | SC *n* | BC-A MD | BC-A *n* | BC-B MD | BC-B *n* | BC-C MD | BC-C *n* | BC-D MD | BC-D *n* | Open MD | Open *n* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **A. Men** | | | | | | | | | | | | | | |
| 0 < R < .2 | 0.44 | 774 | 0.21 | 2755 | 0.07 | 932 | 0.05 | 2106 | 0.10 | 146 | 0.05 | 1685 | -0.02 | 6894 |
| .2 < R < .4 | 0.45 | 487 | 0.34 | 1806 | 0.13 | 1103 | 0.08 | 2355 | 0.19 | 116 | 0.08 | 1844 | -0.12 | 7830 |
| .4 < R < .6 | 0.32 | 388 | 0.27 | 1497 | 0.12 | 1032 | 0.07 | 2567 | 0.17 | 135 | 0.08 | 2008 | -0.19 | 8197 |
| .6 < R < .8 | 0.14 | 282 | 0.13 | 1063 | 0.07 | 985 | 0.03 | 2534 | 0.09 | 103 | 0.04 | 2023 | -0.14 | 9249 |
| .8 < R < 1 | 0.04 | 129 | 0.03 | 628 | 0.02 | 860 | 0.01 | 2591 | 0.02 | 101 | 0.01 | 1869 | -0.03 | 11697 |
| **B. Women** | | | | | | | | | | | | | | |
| 0 < R < .2 | 0.60 | 215 | 0.38 | 933 | 0.20 | 318 | 0.03 | 747 | 0.07 | 55 | 0.03 | 567 | 0.00 | 2813 |
| .2 < R < .4 | 0.61 | 148 | 0.48 | 778 | 0.29 | 380 | 0.08 | 985 | 0.16 | 55 | 0.08 | 662 | -0.03 | 3743 |
| .4 < R < .6 | 0.48 | 128 | 0.39 | 532 | 0.25 | 408 | 0.09 | 1128 | 0.16 | 74 | 0.10 | 823 | -0.07 | 4642 |
| .6 < R < .8 | 0.31 | 53 | 0.25 | 404 | 0.15 | 368 | 0.07 | 1117 | 0.12 | 63 | 0.08 | 819 | -0.04 | 5380 |
| .8 < R < 1 | 0.16 | 17 | 0.14 | 180 | 0.09 | 229 | 0.05 | 826 | 0.07 | 38 | 0.06 | 528 | 0.00 | 5265 |

*Notes:* MD is the mean difference in predicted attendance due to the affirmative action policy.

**Table 8. Estimated Mean Impact of Affirmative Action on Academic Achievement**

| | Men | | Women | |
|---|---|---|---|---|
| ST | 0.181*** | (0.062) | 0.195* | (0.102) |
| SC | 0.123*** | (0.039) | 0.136** | (0.055) |
| BCA | 0.042* | (0.022) | 0.069* | (0.041) |
| BCB | 0.022 | (0.013) | 0.025 | (0.021) |
| BCC | 0.059 | (0.059) | 0.045 | (0.076) |
| BCD | 0.025* | (0.015) | 0.028 | (0.024) |
| OC | -0.053*** | (0.017) | -0.010 | (0.010) |
| All | 0.005 | (0.005) | 0.016* | (0.009) |

*Notes:* Standard errors in parentheses. See Appendix D for calculation methods. Significance: *0.10, **0.05, ***0.01.

**Table 9. Discipline Choice by Caste**

| Discipline (Ave. HS Score) | All | ST | SC | BC-A | BC-B | BC-C | BC-D | Open |
|---|---|---|---|---|---|---|---|---|
| **A. Men** | | | | | | | | |
| Electronics & Comm. (0.55) | 30.2% | 34.3% | 36.0% | 31.7% | 31.9% | 32.3% | 28.7% | 28.1% |
| Computer Science (0.40) | 26.1% | 30.1% | 33.3% | 30.4% | 26.0% | 30.6% | 24.5% | 24.0% |
| Information Tech. (0.26) | 14.7% | 9.9% | 12.1% | 13.3% | 14.4% | 11.2% | 13.8% | 16.2% |
| Electrical Eng. (0.24) | 16.3% | 19.7% | 11.4% | 14.0% | 16.4% | 13.4% | 19.2% | 16.7% |
| All Other (0.02) | 4.9% | 2.2% | 2.5% | 3.7% | 3.7% | 2.6% | 5.4% | 6.2% |
| Mechanical Eng. (-0.08) | 7.8% | 3.8% | 4.7% | 6.8% | 7.5% | 9.9% | 8.4% | 8.7% |
| **B. Women** | | | | | | | | |
| Electronics & Comm. (0.65) | 30.1% | 40.5% | 36.2% | 34.9% | 34.5% | 33.8% | 32.0% | 27.4% |
| Computer Science (0.56) | 29.1% | 39.7% | 40.7% | 36.7% | 32.3% | 40.6% | 31.6% | 25.4% |
| Information Tech. (0.43) | 16.2% | 5.3% | 12.5% | 14.2% | 15.0% | 9.0% | 14.9% | 17.6% |
| Electrical Eng. (0.40) | 14.5% | 13.0% | 8.2% | 9.4% | 12.0% | 12.8% | 12.7% | 16.6% |
| All Other (0.30) | 6.5% | 0.8% | 1.7% | 3.4% | 4.7% | 2.3% | 5.7% | 8.1% |
| Mechanical Eng. (0.15) | 3.7% | 0.8% | 0.6% | 1.3% | 1.6% | 1.5% | 3.1% | 4.9% |

*Notes:* The sample is students who completed the first year of college, were enrolled in one of 184 colleges for which we have graduation data, and for whom we have High School Score (28,754 men and 14,159 women). The order of Discipline in the Table is determined by the aptitude of students enrolling in the major, as indicated by average scores on the High School Scores.

**Appendix A. Ranking within Caste and Effective Rank**

Our empirical approach requires a constructed variable, *effective rank*, which indicates the priority a student has in admission, as determined by performance on the entry exam and affirmative action policy.

We proceed by first constructing an *entry exam rank* variable based on exam scores and following the tie-breaking procedures outlined in the text (e.g., footnote 11). We rank students 1 through *N*, and let $R_i$ denote the rank for student *i*. Then we define entry rank, $E_i = 1 - \frac{R_i}{N}$, thereby normalizing *entry exam rank* to lie between 0 and 1. Importantly, gender and caste play *no* role in the construction of this variable.

Now affirmative action rules mandate that a specified share of seats in each college be reserved for members of disadvantaged castes, and the rules mandate furthermore that within each caste (including the Open category) one third of seats be reserved for women. Panel A of Appendix Table A shows seat share assignments under the applicable rules. The law thus mandates that college priority be determined by *exam rank within caste-gender groups*, rather than by overall rank on the entry exam.

To illustrate, consider the following hypothetical case with two groups: 1000 Open caste students (say caste 0) and 300 disadvantaged caste students (caste 1) who generally score less well on entry exams than Open students. In our example, affirmative action policy reserves one third of seats for caste-1 students at each college (one caste-1 seat for every two caste-0 seats). College choice proceeds as follows: students are ranked by exam *within* caste, and then the top-ranked student chooses first, the second-ranked student choose next, and so forth. Of course, the top-ranked caste-1 students have the same choice as the two top-ranked caste-0 students. Then as the choice process proceeds, if caste-1 and caste-0 students hold similar views about the desirability of available college seats, the $10^{th}$ ranked caste-1 student will have a similar choice set as the $20^{th}$ ranked caste-0 student, the $100^{th}$ ranked caste-1 student will have a similar choice set to the $200^{th}$ ranked caste-0 student, and so forth.

*Effective rank* is a construct that reflects the process we have just described. Let effective rank for student $i$ in caste $j$ be

(A1) $$r_{ij} = 1 - \frac{s_0 R_{ij}}{N_0 s_j},$$

where $s_0$ is the share of seats reserved for open students, $N_0$ is the number of open students, $R_{ij}$ is the rank of student $i$ within caste $j$, and $s_j$ is the share seats reserved for caste $j$. In our example, with $s_1 = \frac{1}{3}$ and $s_0 = \frac{2}{3}$, effective rank for a caste-1 student is $r_{i1} = 1 - \frac{s_0 R_{i1}}{N_0 s_1} = 1 - \frac{2R_{i1}}{1000}$, while effective rank in the open caste is $r_{i0} = 1 - \frac{R_{i0}}{1000}$. Notice that a 10[th] ranked caste-1 student ($R_{i1} = 10$) indeed has the same effective rank as a 20[th] ranked caste-0 student ($R_{i0} = 20$). Also notice that the lowest rank caste-1 student ($R_{i1} = 300$) has an effective rank of 0.40, the same as the caste-0 student with rank $R_{i0} = 600$. The lowest rank caste-0 student has effective rank 0. Figure 1 shows patterns similar to our example.

We construct effective rank as defined in (A1) for the 14 caste/gender groups in our analysis. Notice that effective rank is scaled to be essentially 1 for the top-scoring student in each demographic group, and 0 for the lowest ranked open male student. Effective rank values for men and women of all castes are shown in Figure 1, and are used in all regressions reported in the paper.[27]

For our counterfactual analysis, we compare outcomes (attendance, achievement, graduation) if choice priority were determined without affirmative action to outcomes when choice is determined with affirmative action. Without affirmative action, choice priority is determined by entry rank, $E_i$, which has mean of 0.5 when averaged across all individuals. Hence, to assure that effects attributed to affirmative action are not an artifact of scaling, we require a normalization of effective rank such that its mean is also 0.5. We define *normalized effective rank* as $r_{ij}^N = 1 - \frac{R_{ij}}{k s_j}$, with $k$ chosen so that the mean of this construct equals 0.5. Variables $r_{ij}$ and $r_{ij}^N$ are linearly related, and thus either of these measures of effective rank yields the same fit and statistical significance in the equations we estimate. For ease of interpretation, we report in the text regression results using $r_{ij}$. Then to avoid over-stating the effects of affirmative action in our counterfactual analysis, we use normalized effect rank to obtain the results reported in Tables 6 – 8.

[27] If *effective rank* ($r_{ij}$) works as intended, it should be a substantially better predictor of a student's college quality than a student's *entry exam rank* ($E_i$). In our analysis of the impact of college quality on student performance (Section III.C) we estimate a "first-stage regression" that has college quality as a function of effective rank, caste/gender fixed effects, and a latent ability construct, finding a coefficient on $r_{ij}$ of 0.140 (s.e., 0.049). Remarkably, when we include also $E_i$, the coefficient is on $r_{ij}$ is still 0.140 (s.e., 0.051) and the coefficient on $E_i$ is -0.002 (s.e., 0.041).

A modification was made in 2001 to the seat selection process by *Government Order 550 of the Department of Higher Education*. The following detail regarding the seat selection process is needed to explain this change: In the seat selection process, Open seats for men are filled first. At this stage, seats are provisionally filled based on overall rank on the entry exam. All individuals, regardless of caste and gender provisionally take a seat at this stage based on their entry exam rank until all seats are filled. Then Open-caste women seats are filled next, with all women regardless of caste provisionally taking seats based on entry exam rank among women until all seats are filled. Any woman who qualifies for a more-preferred seat at this stage may take it in lieu of the seat provisionally chosen on the previous round. Next, within each disadvantaged caste, seats for males are filled first. Caste members, regardless of gender, provisionally take seats at this stage based on rank within caste until all seats are filled, and any caste member who can obtain a preferred seat to that taken on the Open round may do so. Next, within each disadvantaged caste, seats for women are filled in order of rank among women within caste, and any woman who can obtain a more preferred seat than provisionally occupied on a previous round may take the more-preferred seat.

Prior to 2001, seats provisionally taken and subsequently vacated reverted to the caste/gender group to which the seats were originally allocated. Virtually all members of disadvantaged castes are able to improve their seat selection by using their caste/gender quota. Hence, prior to 2001, almost all seats provisionally taken by disadvantaged castes in the Open round reverted to members of Open castes. This was changed by *Government Order 550*. The new implementation specified that, rather than reverting to Open caste members, all vacated Open seats first be offered to students of the same caste/gender as the student who vacated the seat. For students from the highly-disadvantaged ST and SC groups, this has little impact on the allocation of seats. Very few of these students are ranked highly enough in the entry exam to obtain attractive seats during the Open round, and, hence, almost all seats provisionally chosen by ST's and SC's on the Open round revert to Open castes. However, many members of the less-disadvantaged BC-B and BC-D castes have sufficiently high entry exam rank to obtain relatively attractive seats on the Open round. When they subsequently exercise their caste priority to obtain a still better seat, the seats they vacate are taken by lower-ranked members of their castes. The impact, then, is essentially to expand the allocation of seats to BC-B and BC-D students, while reducing the number of seats actually available to Open caste members, particularly Open caste men.

The net impact for the cohort we study is given in Panel B of the Appendix Table A. In calculation of effective rank for our empirical analysis, we use the effective seat shares shown in Panel B.[28] For students in the most disadvantaged groups—ST, SC, BC-A, and BC-C—there was negligible effect on the proportion of allocated seats. However, the implementation of the law effectively increased the allocation of seats to the BC-B and BC-D men and women, while lowering the remaining seats available to Open students.

There is one final twist to the admission process. There were too few applicants among ST and SC students to fill allocated seats. This doesn't change the way we calculate effective rank for these students; indeed, it is precisely why even very low-performing ST and SC students have such a high effective ranks. However, under Government Rule 550 these seats revert to Open students. So, in the end, 0.300 of seats effectively went to Open men and 0.192 to Open women.

**Appendix Table A. Effect of Seat Allocation Due to Affirmative Action**

| | ST | SC | BC-A | BC-B | BC-C | BC-D | Open |
|---|---|---|---|---|---|---|---|
| *A. Allocation of Seats by Initial Quota* | | | | | | | |
| Men | 0.040 | 0.100 | 0.047 | 0.067 | 0.007 | 0.047 | 0.337 |
| Women | 0.020 | 0.050 | 0.023 | 0.033 | 0.003 | 0.023 | 0.168 |
| All | 0.060 | 0.150 | 0.070 | 0.100 | 0.010 | 0.070 | 0.505 |
| *B. Effective Allocation of Seats Under Government Order 550* | | | | | | | |
| Men | 0.040 | 0.100 | 0.051 | 0.122 | 0.007 | 0.097 | 0.250 |
| Women | 0.020 | 0.050 | 0.023 | 0.049 | 0.003 | 0.035 | 0.153 |
| All | 0.060 | 0.150 | 0.074 | 0.171 | 0.010 | 0.132 | 0.403 |

[28] Notice that the sum of the "effective allocations" is somewhat greater than 1. This is because some ST and SC seats are essentially allocated twice—first they are provisionally allocated to ST/SC students, and, when subsequently vacated, they are made available to Open caste students. We adopt a straightforward extension of the normalization described above so that mean effective rank equals mean entry rank.

**Appendix B. Modelling College Attendance**

An applicant to an engineering college in the State we study potentially has three choices: attend an engineering college in the State, attend some other academic institution, or choose the no-college option. We observe only whether the applicant attends an engineering college in the State. To frame our choice model, let $U_{ij}^{e}(a_{ij}, r_{ij}, c_{ij}^{e})$ denote utility in the engineering college to which student *i* in caste/gender group *j* is admitted. Here, $a_{ij}$ denotes the student's academic aptitude, $r_{ij}$ denotes effective rank of the student in his or her caste/gender group, and $c_{ij}^{e}$ is the cost of attending an engineering college. We then write utility of the engineering college option as the sum of a deterministic component, denoted with lower-case $u$, and an idiosyncratic term $\varepsilon_{ij}^{e}$:

(B1) $$U_{ij}^{e}\left(a_{ij}, r_{ij}, c_{ij}^{e}\right) = u_{ij}^{e}\left(a_{ij}, r_{ij}, c_{ij}^{e}\right) + \varepsilon_{ij}^{e} .$$

For most students in our data, an engineering college will be the best available academic option. However, exceptionally able students can be expected to gain admission to a more prestigious institution, e.g., one of the Indian Institutes of Technology (IIT). Let $\delta_{ij} = 1$ for a candidate who is admitted to such an institution, and who also prefers that institution to the no-college option, and let $\delta_{ij} = 0$ otherwise. Also, let $c_{ij}^{a}$ be the cost of attending this alternative academic institution. Then if $\varepsilon_{ij}^{a}$ is the idiosyncratic utility shock of the non-engineering college option, we can write

(B2) $$U_{ij}^{n}\left(a_{ij}, c_{ij}^{a}\right) = \delta_{ij}\left[ u_{ij}^{a}\left(a_{ij}, c_{ij}^{a}\right) - u_{j}^{0}(a_{ij}) \right] + u_{j}^{0}(a_{ij}) + \varepsilon_{ij}^{a} ,$$

where $u_{ij}^{a}\left(a_{ij}, c_{ij}^{a}\right)$ and $u_{j}^{0}(a_{ij})$ denote, respectively, the deterministic component of utility in an alternative academic institution and in the no-college option. Note that effective rank is not included in these two utility expressions because effective rank affects priority only for admission to an engineering college in the State we study. Also note that we have a subscript *j* on each utility option, permitting the possibility of systematic differences across castes and gender in the valuations of benefits of college attendance or non-attendance.

Of course the probability of admission to an alternative competitive institution, such as an IIT, is itself a function of ability and also plausibly of caste and gender. Letting $\rho_j(a_{ij}) = E(\delta_{ij})$ be the probability, conditional on aptitude, that applicant *i* in group *j* obtains admission to a more elite institution, we have

(B3) $$U_{ij}^{n}(a_{ij}, c_{ij}^{a}) = \rho_j(a_{ij})[\, u_{ij}^{a}(a_{ij}, c_{ij}^{a}) - u_j^{0}(a_{ij}) \,] + u_j^{0}(a_{ij}) + \varepsilon_{ij}^{n} \, ,$$

where

(B4) $$\varepsilon_{ij}^{n} = \varepsilon_{ij}^{a} + [\delta_{ij} - \rho_j(a_{ij})] \, [\, u_{ij}^{a}(a_{ij}, c_{ij}^{a}) - u_j^{0}(a_{ij}) \,].$$

Thus $\varepsilon_{ij}^{n}$ impounds both an idiosyncratic preference shock from (B2), as well as factors that influence whether a member of group *j* with aptitude $a_{ij}$ is admitted to a superior alternative academic institution.

Now an applicant who gains admission to an engineering college in our State will matriculate if (B1) is greater than (B3), that is, when

(B5) $$[\, u_{ij}^{e}(a_{ij}, c_{ij}^{e}, r_{ij}) - u_j^{0}(a_{ij}) \,] - \rho_j(a_{ij}) \, [\, u_{ij}^{a}(a_{ij}, c_{ij}^{a}) - u_j^{0}(a_{ij}) \,] > \varepsilon_{ij}^{n} - \varepsilon_{ij}^{e} \, .$$

The first term in the left-hand side is the difference in the deterministic components of utility between and engineering college and the no-college option. The second term is the probability of admission to an academic institution preferred to an engineering college multiplied by difference in the deterministic components of utility between the alternative academic institution and the no-college option. Below, we assume that the idiosyncratic shocks are i.i.d. normal, implying a probit specification for the binary choice of attending or not attending an engineering college.

For most students who take the entry examination for engineering colleges, an engineering college will be their best academic option. The probability of admission to an IIT or comparable institution will be an increasing and convex function that is near zero throughout much of its domain and increases sharply for aptitudes in the right tail of the distribution. Thus, for most applicants, the second term of the left-hand side of equation (B5) will be approximately zero, implying that the deterministic portion of the choice between an engineering college and the no-college option is

(B6) $$[\, u_{ij}^{e}(a_{ij}, c_{ij}^{e}, r_{ij}) - u_j^{0}(a_{ij}) \,] \, .$$

Clearly (B6) must be increasing in $r_{ij}$ because those of higher rank within their caste/gender group have higher priority in engineering college choice. If the academic gain from attending an engineering college is greater for more able students, then this latter expression is also increasing in $a_{ij}$; for most ability levels, we expect the probability of matriculation in our engineering colleges to be increasing in $a_{ij}$. At a relatively high levels of aptitude, though, the middle term in (B5) comes to dominate. For very-high aptitude students, an increase in $a_{ij}$ improves the probability of admissions to an IIT or other high-prestige institution and thus *reduces* the probability of matriculation to one of our engineering colleges.

To summarize, our model has two clear predictions: First, the probability of matriculation at an engineering college is increasing in $r_{ij}$, and second, it is an inverted U-shaped function in aptitude $a_{ij}$. This motivates an empirical specification in which we estimate the attendance probability using a polynomial in effective rank and a polynomial in aptitude—allowing flexibility in the effects of these constructs on matriculation. We could proceed with a probit model of college attendance, and indeed did so in an earlier version of the paper. In the current version, we use a linear probability model, which yields qualitatively similar results.

**Appendix C. Regression Results for Attendance**

Estimates of our attendance equation (1) are not easily interpretable because of the presence of higher-order terms. Thus in the main paper we show key results using graphs (Figures 2 and 3). Below are coefficient estimates:

| | (1) Men | | (2) Women | |
|---|---|---|---|---|
| Independent Variables | Coefficient | (s.e.) | Coefficient | (s.e.) |
| Constant | 0.121** | (0.029) | -0.552 | (0.501) |
| ST | -0.123** | (0.036) | -0.316** | (0.068) |
| SC | 0.027 | (0.026) | -0.126** | (0.045) |
| BC-A | 0.063** | (0.016) | -0.078** | (0.028) |
| BC-B | 0.050** | (0.013) | 0.028* | (0.013) |
| BC-C | 0.041 | (0.024) | 0.032 | (0.031) |
| BC-D | 0.050** | (0.013) | 0.032* | (0.015) |
| (Effective rank) | 0.646** | (0.199) | 9.087* | (4.630) |
| $(\text{Effective rank})^2$ | -6.040** | (1.170) | -41.916* | (17.02) |
| $(\text{Effective rank})^3$ | 19.877** | (3.095) | 88.409** | (30.15) |
| $(\text{Effective rank})^4$ | -21.067** | (3.528) | -82.074** | (25.62) |
| $(\text{Effective rank})^5$ | 7.279** | (1.459) | 28.078** | (8.418) |
| High School Score | -0.017** | (0.004) | 0.009 | (0.007) |
| Entry Exam Score | 0.073** | (0.022) | 0.095* | (0.037) |
| High School Score × Entry Exam Score | 0.074** | (0.005) | 0.079** | (0.010) |
| High School Score Squared | -0.065** | (0.003) | -0.077** | (0.006) |
| Entry Exam Score Squared | -0.034** | (0.005) | -0.059** | (0.010) |
| High School Score Squared × Entry Exam Score | 0.011** | (0.004) | -0.012 | (0.010) |
| High School Score × Entry Exam Score Squared | -0.017** | (0.004) | -0.006 | (0.011) |
| High School Score Cubed | -0.007** | (0.001) | -0.009** | (0.003) |
| Entry Exam Score Cubed | -0.003 | (0.001) | 0.002 | (0.003) |

*Notes:* $n$ = 80,771 for men and $n$ = 35,421 for women. This is a linear probability regression (dependent variable is 1 for matriculation). $R^2 = 0.31$ for regression (1); $R^2 = 0.35$ for regression (2).

## Appendix D. Calculating Standard Errors in Table 8

In this Appendix we provide details for our calculation of standard errors on the estimated mean impacts of affirmative action reported in Table 8.

As noted in Appendix A, all analysis can be done either with our *effective rank* variable or a *normalized effective rank* variable that this linearly related to the former; fit and statistical significance is the same either way. Our counterfactuals ask what would happen if admission policy was based on *entry exam rank*, $E_i$ (with no preference by caste/gender) and it is appropriate to use the *normalized effective rank*, $r_i^N$, for this exercise because it has been normed so that the two constructs have the same mean.[29]

Let $\beta_m$ and $b_m$ be, respectively, the population coefficient and estimated coefficient on normalized effective rank in the first-year achievement regression. For student $i$ define $y_{i1}$ to be achievement with affirmative action and define $y_{i2}$ be the corresponding achievement in the absence of affirmative action. Let $\varepsilon_{i1}$ be the error term in the achievement equation with affirmative action and let $\varepsilon_{i2}$ be the error term that would have appeared in the achievement equation if there were no affirmative action. Then the difference in achievement for student $i$, with and without affirmative action, is

(D1) $$y_{i1} - y_{i2} = \beta_m\left(r_i^N - E_i\right) + \varepsilon_{i1} - \varepsilon_{i2}\ .$$

The corresponding difference in *predicted* outcomes is

(D2) $$\hat{y}_{i1} - \hat{y}_{i2} = b_m\left(r_i^N - E_i\right).$$

The error in the estimated impact of affirmative action is thus

(D3) $$e_i = (\hat{y}_{i1} - \hat{y}_{i2}) - (y_{i1} - y_{i2}) = (\beta_m - b_m)\left(r_i^N - E_i\right) + \varepsilon_{i1} - \varepsilon_{i2}.$$

Under our assumptions, $E(e_i) = 0$.

With his in mind, consider a caste/gender group $j$ of size $n_j$ individuals. The error in the mean estimated effect of affirmative action group $j$ is

(D4) $$\bar{e}_j = (\beta_m - b_m)\left(\bar{r}_j^N - \bar{E}_j\right) + \textstyle\sum_i(\varepsilon_{i1} - \varepsilon_{i2})\,/n_j.$$

[29] Effective rank has a *j* subscript, indicating caste/gender, which we suppress here to make notation cleaner.

If we let $s_m$ be the standard error on $b_m$, the variance of the object in (D4) is

(D5) $$Var(\bar{e}_j) = s_m^2\left(\bar{r}_j^N - \bar{E}_j\right)^2 + \frac{\sum_i Var(\varepsilon_{i1} - \varepsilon_{i2})}{n_j^2},$$

which can be written,

(D6) $$Var(\bar{e}_j) = s_m^2\left(\bar{r}_j^N - \bar{E}_j\right)^2 + \frac{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)}{n_j}.$$

Consider the numerator in last term of (D6). The variance under the counterfactual ($\sigma_2^2$) is unknown, but it is natural to assume that it will be approximately the same as under the existing college choice regime, which suggests taking $\sigma_2^2 = \sigma_1^2$. The correlation of the error terms $\rho$ is also unknown. The natural interpretation of the error in the achievement equation is that it is a combination of several factors. One factor is the student-specific fit for higher education in engineering that is revealed when the student attends a college. Other factors might be termed idiosyncratic luck (the student had the good fortune to connect with a motivating teacher, the student was ill during the exam week, etc.). The student-specific component would tend to impart a positive value to $\rho$, while the luck component would favor a zero value of $\rho$. There is no natural interpretation that would yield a negative value of $\rho$. The most conservative reasonable approach (yielding the largest variance estimate) thus sets $\rho = 0$, in which case (D6) becomes

(D7) $$Var(\bar{e}_j) = s_m^2\left(\bar{r}_j^N - \bar{E}_j\right)^2 + \frac{(\sigma_1^2 + \sigma_2^2)}{n_j}.$$

To give an example, in the estimated achievement equation the standard error of the regression is 0.61, so $\hat{\sigma}_1^2 = 0.61^2$. The standard error of the estimated coefficient on normalized effective rank is $s_m = 0.124$. Consider the smallest male caste group, ST. For this group $n_j = 720$ and $\left(\bar{r}_j^N - \bar{E}_j\right) = 0.424$. Substituting into (D7), and taking the square root, we have the estimated standard error (0.062) reported in the first row of Panel A in Table 8. If we had taken the less-conservative route of ignoring the second term in (D7), our estimated standard error would have been 0.053.