

Gandil, Mikkel; Leuven, Edwin

Working Paper

College Admission as a Screening and Sorting Device

IZA Discussion Papers, No. 15557

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Gandil, Mikkel; Leuven, Edwin (2022) : College Admission as a Screening and Sorting Device, IZA Discussion Papers, No. 15557, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/265778>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

DISCUSSION PAPER SERIES

IZA DP No. 15557

**College Admission as a Screening and
Sorting Device**

Mikkel Gandil
Edwin Leuven

SEPTEMBER 2022

DISCUSSION PAPER SERIES

IZA DP No. 15557

College Admission as a Screening and Sorting Device

Mikkel Gandil

University of Oslo

Edwin Leuven

University of Oslo and IZA

SEPTEMBER 2022

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

College Admission as a Screening and Sorting Device*

How can colleges find successful applicants? Criteria such as GPA, interviews, essays, and tests provide information about candidates, but which work and why? We shed light on these questions using unique data on the universe of objective and subjective rankings of all college applicants in Denmark, their applications, admissions and college outcomes. We implement a regression discontinuity design across multiple admission quotas to estimate how admission affects program and college completion, and investigate how this depends on the evaluative criteria used. We find that admission based on alternative criteria outperforms standard admission based on GPA. Alternative criteria are more effective in identifying good program matches, which ultimately leads to higher college completion rates because alternative evaluation is more likely to admit students that tend to struggle elsewhere. Most of the impact of alternative evaluation is found to be due to their impact on the applicant pool (sorting), and not because they are better at identifying successful students keeping the applicant pool fixed (screening). Our analysis of the evaluation technology shows that the use of individual grades leads to the admission of applicants that are less likely to succeed, while essays is the only criterion that is intrinsically better at screening out applicants that will do well in the program or in college more broadly. The use of tests, interviews and CVs do not outperform GPA in screening once we keep the applicant pool fixed. There is no evidence that interviews are an effective admission tool.

JEL Classification: D04, H43, I23, I28, J24

Keywords: college admission, higher education

Corresponding author:

Edwin Leuven
Department of Economics
University of Oslo
P.O. Box 1095 Blindern
0317 Oslo
Norway
E-mail: edwin.leuven@econ.uio.no

* We thank Peter Arcidiacono, Hessel Oosterbeek, Sarah Turner, and seminar participants for valuable feedback and suggestions. The Norwegian Research Council supported this research under project no. 275906.

1 Introduction

Persistent increases in skill demands put education central in discussions surrounding economic welfare, inequality, and social mobility (Autor, 2014). Education is now more important than ever in explaining individual economic and social success, and it is therefore not surprising that securing equitable and efficient access to education is a major policy preoccupation of governments around the world (OECD, 2012). This paper zooms in on a salient barrier to education participation: college admissions.

College admissions vary in the type of information used in the assessment of applicants. Quantitative measures of academic preparedness such as high-school GPA and tests like the American SAT play a key role in college admissions. On the other side of the spectrum lie more holistic admissions that may also use GPA and tests, but which typically involve the evaluation of recommendations, extracurricular activities, or other signals of talent or fit. Their use is exemplified by admission practices at highly selective institutions in the United States (f.e. Arcidiacono et al., 2020).

The primary motivation behind the use of different admission criteria rests on their ability to predict collegiate success.¹ Predicting academic achievement conditional on admission for all applicants is however challenging as outcomes are only observed for the selected sample of admitted students. Moreover, to evaluate the effects of changes in admission policies one not only must estimate potential outcomes for applicants under different admission regimes, but also recover counterfactual admissions. This paper tackles these issues and studies how different admission criteria perform in selecting more academically successful applicants.

We take advantage of two key characteristics of the Danish higher education system that allow us to make progress on this question. The first one is that applicants are not only admitted in a primary quota based on their high-school GPA, but that applicants also have the option to be admitted in a secondary quota where they are ranked on other criteria such as specific high-school grades, admission tests, writing assignments (essays), or more holistic assessments of applicants through considering their experience (CVs) and interviews. After programs have ranked all applicants within quotas the rankings must be submitted to a central governmental agency. Admission is based on central-

¹A related issue, that we study in a companion paper (Gandil and Leuven, 2022) relates to fairness. Critiques of the SAT for example point to its correlation with demographics and differential predictability across social background, and argue that this gives applicants from higher socio-economic groups an unfair admission advantage (e.g. Rothstein, 2004). Similarly, holistic admissions have also been criticized for favoring specific groups of applicants over others (Arcidiacono et al., 2020). In the Danish context, which we explore, Thomsen (2018) investigates the socioeconomic gradient in GPA for admitted students, but cannot observe the ordered latent eligibility scores which we are able to exploit in our companion paper to investigate differences in social gradients across admission criteria for admitted, non-admitted and marginal applicants.

ized deferred acceptance with multiple tie-breaking based on these rankings. The second feature of the admissions that we build on is that capacity in these two quotas is fixed which, together with the deferred-acceptance, gives rise to a regression discontinuity design that we use to address selection at the margin of college admission (f.e. [Öckert, 2001, 2010](#); [Kirkeboen et al., 2016](#)).

College application and admission in Denmark is at the program level, where a program is defined as a specific field-of-study at a given institution; for example economics at the University of Copenhagen. The criteria used in the alternative evaluation of applicants vary across programs, and programs are also responsible for evaluating and ranking the applicants to their secondary quota. Financing is ultimately also at the program level and is based on the number of graduates a program delivers. The analysis therefore takes a dual supply-side perspective. The first one is that of the program admission officer and we consider program completion as their primary outcome of interest. The second perspective is that of a social planner who cares about program completion rates more broadly. We therefore also investigate how the admission modality affects college completion. This allows us not only to investigate the extent to which programs are able to target program matches in their selection of applicants as opposed to more general college readiness when admitted, but also to assess how applicants fare when they are not admitted to the program of their choice.

We build on unique Danish registry data containing information on programs' objective and subjective rankings of all applicants, programs' admission decisions, as well as the subsequent enrollment and program and college completion outcomes for the applicants. Combining these data with the RD design generated by the deferred acceptance not only allows us to open the black box of college admissions, but also to estimate how the use of different evaluative criteria affects the college outcomes of the pool of admitted applicants, recover the causal impact of admission on college outcomes when using different admission criteria, and shed light on the underlying mechanisms.

We find that under alternative evaluations the program completion rate for the marginal applicant is 2 percentage points higher than that of marginal applicants admitted based on high-school GPA (who complete at a 0.50 rate). Using alternative criteria leads therefore to the admission of more successful applicants to the program. We do not observe the same admission advantage for college completion more broadly (i.e. completing any program). Alternative evaluation outperforms GPA-based admission at the college completion margin by only 0.1 percentage points (compared to a base of 0.81) and the difference is not statistically significant. This suggests that while alternative criteria are more effective in identifying program matches, they do not necessarily perform much better when it comes to identifying good college matches. This is con-

firmed by the effects for college completion value added – college completion relative to counterfactual college completion rates when applicants are not admitted – which are higher at 4 percentage points and comparable to the impact on program completion rates. Alternative evaluation thus admits applicants that do well in the program but are more likely to struggle when not admitted.

Alternative evaluation can outperform GPA-based admission because of two channels. The first one is screening; who gets admitted from a given pool of applicants. Changing criteria used in admission can however also affect sorting; who applies. That alternative evaluation affects the applicant pool is illustrated by the fact that on average only 40 percent of applicants to a program also apply to the secondary admissions of that same program. Knowing whether screening or sorting dominates is important for programs as this determines whether the information collected in secondary admissions is predictive of achievement (the screening channel) or whether imposing costs on applicants is more important (the sorting channel). We investigate the relative importance of screening and sorting as follows. To identify sorting we first keep screening fixed by comparing secondary quota applicants to the other applicants in primary admissions. We then consider screening by keeping sorting fixed and comparing secondary quota applicants who are admitted using high-school GPA to those who are admitted based on alternative criteria. The results from this exercise show that most of the impact of alternative admission is due to sorting.

We finally unpack the evaluation technology and investigate the extent to which the differential performance of secondary admission is explained by the type of criteria used in the evaluation of applicants. Programs select evaluative criteria themselves and typically use multiple criteria to select applicants. To estimate the marginal contribution of each criteria to the effectiveness of college admission we exploit the combinations of criteria to construct a research design similar to in spirit to difference-in-differences where Q1-applicants serve as “control”-units. We find that compared to GPA the use of individual grades, such as a single math grade, lowers the program completion rates of admitted applicants. The use of tests, CVs and essays has robust positive effects which are explained by their impact on sorting and not by an ability to select more successful students from a given pool of applicants. Given that the latter criteria are likely more costly for the applicant, these findings supports our hypothesis that application costs is the main driver of the better performance of alternative criteria relative to GPA. However, we do not find that interview is an effective tool in admission neither through screening nor sorting.

The current paper relates and contributes to different strands in the higher education literature. First there is a large body of work that studies the reliability and validity of

tests like the American SAT and the ACT (see f.e. [Burton and Ramist \(2001\)](#); [Zwick \(2007\)](#) for summaries of this literature). There is less work on how to optimally combine information in the admission portfolio. One example is [Bettinger et al. \(2013\)](#) who find that that the ACT would better predict outcomes in college if two out of four subtests would be dropped.² There is also large descriptive body of work on whether subjective admission criteria such as interviews, letters of recommendation, and essays predict educational outcomes (see f.e. [Kuncel et al., 2014](#); [Murphy et al., 2009](#); [Goho and Blackman, 2006](#), for reviews). These reviews of predominantly case studies conclude that subjective measures can provide some incremental validity over and above GPA or SAT scores. In contrast to these case studies we investigate an entire national admission system which allows us to unpack heterogeneity in admission outcomes across programs and evaluation criteria.

One important challenge that the validity literature faces, as pointed out by [Rothstein \(2004\)](#), is that most predictive studies are estimated on the selected sample of admitted or matriculated students as opposed to the applicants.³ We show that even when existing methods may be approximately predictive on average, they may fail to reliably estimate the predictiveness for the policy-relevant set of applicants who are at the margin of admission.⁴

A second limitation of the validity literature is that it is not informative about the effects of changes in admissions policies ([Rothstein, 2004](#)). A handful of papers directly study admission reforms. [Smith et al. \(2015\)](#) consider the effects of introducing admission essays and application fees on student applications, enrollment and first-year retention in American college markets. They find that requiring an essay and increasing application fees lead to a decrease in applications. Overall, requiring an essay decreases the number of matriculants at an institution, but the application fee has no discernible effect. These application criteria have no impact on freshman retention rates, suggesting that their use does not improve the quality of the match. [Grosz \(2021\)](#) finds that the replacement of lotteries and waiting lists with admissions that rely on grades lead to increases the average GPA of incoming cohorts at California's community college nursing programs. There is no evidence that this affected academic outcomes like completion and exam-passing rates, but the estimates are relatively imprecise. [Belasco et al.](#)

²In contrast [Silva \(2022\)](#) finds that adding the exam score in Portuguese to track-specific exam scores leads to the selection of more successful applicants.

³Using data from the University of California he estimates for the SAT that ignoring selection leads to a 20 percent bias in the predictiveness of the SAT (still assuming that individual campus admissions and student matriculation decisions are ignorable)

⁴We find that for the policy-relevant set of applicants, those who enroll only if crossing a cutoff, high-school GPA is more predictive than suggested by traditional approaches to test validity. In contrast, we find that rankings based on alternative criteria have little predictive power in terms of program completion.

(2015) estimate the impact of test-optional policies on applications but do not find conclusive evidence that they affected application numbers nor low-income and minority student enrollment. Rothstein (2022) finds that the inclusion of recommendation letters modestly improved application outcomes for the average underrepresented student at a highly selective US institution.

Our study also connects to the literature on college cost in a broader sense (see f.e. Hoxby and Turner, 2013; Dynarski et al., 2022b,a, for reviews) that also documents impacts on sorting, both at extensive and intensive margins. Hyman (2017); Hurwitz et al. (2015) for example find that mandatory college entrance examination policies for high-school students positively affects college enrollment rates. On the extensive application margin Pallais (2015) finds that students apply to more selective universities when the cost of sending ACT scores to an additional university is reduced. Top-percent programs in the US also affect application barriers (see f.e. Andrews et al., 2010). A recent example is Black et al. (2020) who study the introduction of a top percent plan in Texas that gave the 10 highest ranked students in a high-school access to the state’s flagship universities. They compare the students who gained access under this new admission scheme to the students that were crowded out, and find that the former experience increases in college enrollment and graduation while the latter attended less selective colleges but did not change their enrollment and completion rates.

Our analysis also speaks to the literature that studies the importance of so-called “match” effects in higher education. In our context, where students apply to the intersection of field-of-study and institutions, successful program admission must involve successful matching if admission increases overall college completion. The importance of matching has attracted substantial interest in the United States where affirmative action and the presence of highly selective institutions have the potential to increase the salience of match effects (Arcidiacono et al., 2015; Arcidiacono and Lovenheim, 2016). Dillon and Smith (2020) in their review of the college quality literature conclude that “most but not all of the literature finds little in the way of interaction effects, other than on intermediate outcomes such as transfer and major choice.” Our findings illustrates that matching can be salient along other dimensions than ability as reflected by GPA.

We start by providing the necessary institutional context, after which we document the registry data on which our analysis builds. A simple framework outlines the different mechanisms through which the use of different admission criteria can affect academic achievement, and we discuss the objective of admissions offices in our context. We then detail the empirical approach will allows us to estimate potential outcomes under admission, after which we turn to our pooled estimates and investigate the relative importance of sorting vs. screening, and heterogeneity by program type. In the final

part of the paper we investigate to what extent the differential performance of secondary admission depends on the criteria used there.

2 Institutional background

Danish higher education is essentially a public system that is accessible to everybody with a diploma from the academic high school track ("Gymnasiale uddannelser"). There are 8 universities and 7 university colleges which are spread out over 38 locations, as well as 8 so-called business academies also with multiple campuses. At entry, higher education offers 3-year bachelor programs as well as 2–4 year professional degrees, and more than 90% of those who complete a bachelor degree will later also complete a master. Higher education institutions do not charge tuition fees, and students can make use of a generous public grant and loan system to finance their studies.

Applications to higher education are submitted through a single online portal and managed centrally in a governmental agency. Applications are to programs which are defined as a specific field of study at a given university and applicants must rank the programs they apply to from most-preferred to least-preferred.

Applicants have the possibility to apply to up to eight programs and there are no application fees. While the maximum of eight may seem limited, 75 percent of the applicants do not apply to more than three programs, less than 10 percent apply to more than five programs, and only 3 percent of the applicants exhaust the list which puts an upper bound on the number of people that are constrained (probably in the neighborhood of 2 percent).

Programs typically have a limited amount of seats available, and their allocation is based on a deferred acceptance (DA) algorithm ([Abdulkadiroğlu and Sönmez, 2003](#)). Primary admissions, which we will refer to as Quota 1 or Q1, use applicants' average high-school GPA as the priority score in the DA. High-school GPA is based on a combination of central and externally graded exit exams and continuous assessment and is a number on a 7 point scale that is recorded up to 1 decimal place, and tie breaking in the DA is random. The Q1 cutoffs in terms of GPA are made publicly available on a governmental website and treated as front page news.⁵

Most programs also reserve a fixed number of seats for secondary admissions, referred to as Quota 2 or Q2, where students receive a priority ranking based on alternative criteria. These quotas were originally implemented when centralized admission

⁵If schools vary systematically in grading, students may sort into high-schools and thereby manipulate GPA. For our purpose this form of sorting is irrelevant, as the final GPA is the metric that the admission office can rank on. However, to investigate the importance of high schools we regress the difference between grades given by continuous assessment and grades given on centralized tests on school fixed effects and obtain an R^2 of 0.38, which is not negligible.

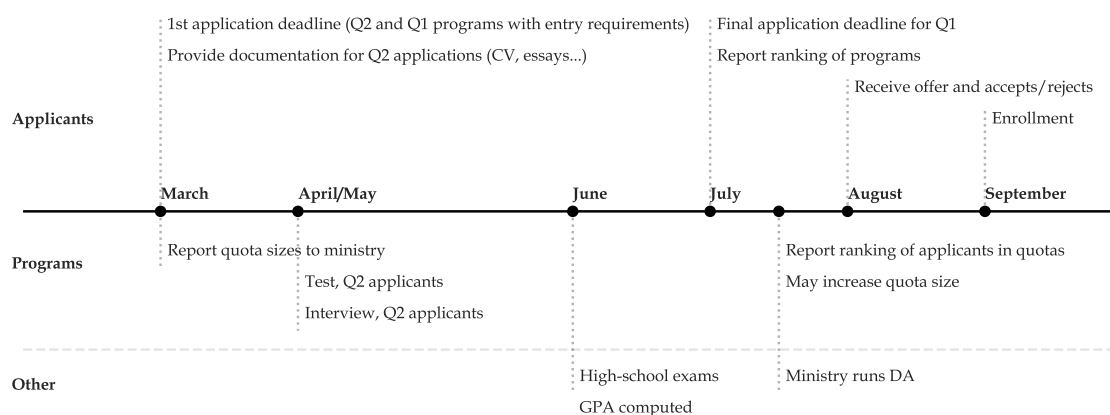


Figure 1. Timeline for the application process

was introduced in 1977 to ensure that applicants without a high-school qualification would also have access to higher education.⁶ The criteria that are used to evaluate applicants in Q2 vary across programs and can be “objective”, for example a subset of the high school grades (Grades), or college entry tests (Tests), but criteria can also be “subjective” and may for example involve the evaluation of relevant experience (CV), written assignments (Essay), or use interviews (Interview). Programs can partly choose their own criteria.

The application timeline is illustrated in Figure 1. For applicants who choose to be considered in Quota 2 the deadline is in March and applications are then sent to the programs for evaluation. The program only observes applications directed at them and the not position of the program on the applicant’s rank-ordered list. Programs evaluate and rank all applicants to Quota 2

Applicants for Quota 1 must submit their applications before mid-July which is after the final high-school GPA is known. In between the two deadlines, programs have the option of conducting interviews and tests in Quota 2. After the Quota 1 deadline in July the programs report their rankings of applicants within quotas to the ministry and the ministry applies the deferred acceptance algorithm, where each quota is represented as a separate “school”.

For the algorithm to run, the rank-ordered lists of applicants are expanded such that each quota is a separate priority on the expanded list. All applicants with a GPA will have Quota 1 as their first “within-program priority” followed by Quota 2 if the

⁶Denmark has a dual-track system, where following primary education applicants may continue on academic tracks (Gymnasiale uddannelser) or vocational tracks (Tekniske skoler) where the latter is a combination of school and apprenticeship. The vocational track does not provide a GPA which can be compared to the GPA from the academic track but may still grant access to higher education. For example a mechanic might apply to an engineering program through Quota 2.

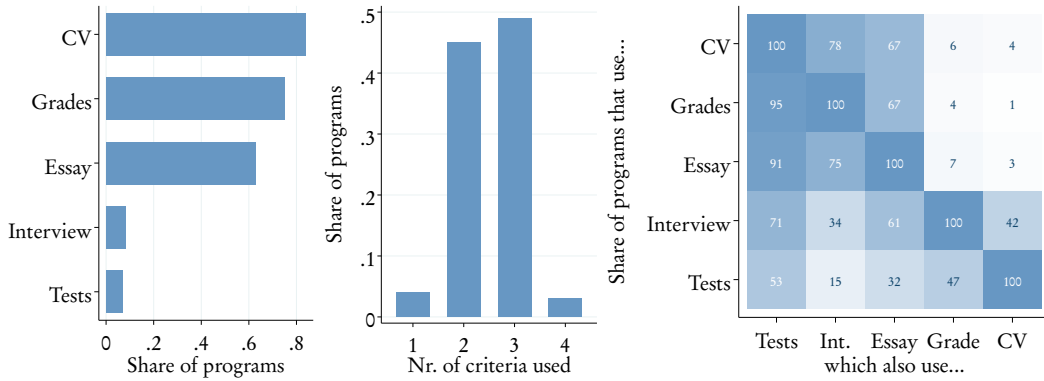


Figure 2. Use of admission criteria in secondary admissions

Note: The figure shows the use of criteria across programs as recorded by the Ministry of Higher Education and Science, see UFM (2020).

applicant filed a Quota 2 application. The presence of Quota 2 option does therefore not change the strategy-proof nature of the admission process which now becomes DA with multiple tie-breaking.⁷

About 30 percent of the applicants were admitted in Q2 during the period we study, and this share has been increasing somewhat over time. The extent to which alternative screening criteria are used varies across programs. The modal program admits about 50 percent of the applicants in Q2, but many programs admit between 10 and 30 percent of the applicants this way.

Figure 2 documents the use of different screening criteria in Q2 where the classification is based on UFM (2020). The left panel shows that screening of candidates on relevant work experience, grades or written assignments is very common, while interviews and tailored tests are used much less often, perhaps because this is more expensive to implement or develop. Most programs tend to combine two or three different criteria when ranking candidates in Q2. The right panel of Figure 2 shows that programs tend to combine relevant work experience, grades and written assignments, while programs that use tests and in particular interviews appear to screen more broadly.

3 Data

The starting point of our analysis is the application data for the years 2010-15 which provide us with information on applicants' preference lists, the ranking of the applicants in Q1 and Q2 (if applicable) and program offers. We link this information on individuals' applications to registry data from Statistics Denmark which provide us with individuals' educational trajectories and outcomes, as well as gender, age and family background; in

⁷Without Quota 2 all programs would use the same eligibility score and the mechanism would therefore be Serial Dictatorship.

particular parental income, education, and immigrant status. We restrict our analysis to people who apply to at least one restricted program, and do not consider programs where the criteria in Q2 could not be determined. We exclude Q2 applicants who do not have a GPA and therefore do not compete in Q1. Additionally, we exclude applicants over 30 years of age. We record the running variable prior to the sample exclusions, and the percentile in a quota therefore is relative to all applicants in the quota.⁸

Table 1 reports descriptive statistics for our main sample. The first column reports averages for all applicants (in Q1). We observe in total more than 245 thousand people applying to higher education over the period we study. More than 60 percent of the applicants are female, and 11 percent have an immigrant background. The applicants are from above average socio-economic backgrounds as their parents income rank is on average at the 74th percentile (SES). Applicants are also positively selected in terms of their high-school GPA and rank on average at the 54th percentile. About 40 percent of the applicants apply to the second quota. The second column shows that Q2 applicants are more likely to be female, slightly older, and to be from a non-immigrant background, are similar in SES but of substantially lower ability as measured by GPA.

It is worth pointing out that even though these applicants have significantly lower GPAs, their completion rates are similar to the overall population of applicants. We define “Completes on time” as a dummy which takes the value of 1 if the applicant completes her program within one year of formal completion time counting from the time of application. Around 17 percent of applicants, regardless of admission, complete the program in time, and we do not observe differences across Q1 and Q2 applicants.

Table 1 also breaks the sample statistics down along the criteria that are used by programs in their second admission quotas. In terms of gender and SES the applicants are relatively uniform, and only applicants for programs that use interviews and tests stand out in that they more often have immigrant backgrounds and higher GPAs.⁹

4 A stylized model of application

To illustrate the basic mechanisms of interest consider the stylized situation where individuals have to decide whether to apply for admission to a program in the primary quota (Q1) based on GPA, and also consider whether to apply for admission in the secondary quota (Q2) where other criteria are used. Assume that the application cost in the primary quota is negligible. Individuals apply to the program if their utility when admitted

⁸The restrictions decrease our sample from 416,562 applicants with 957,827 applications in 4,329 program-year combinations to 245,563 applicants with 536,327 applications in 1,333 programs.

⁹These applicants are also considerably more likely to complete in time. Note that criteria and applicants are non-randomly distributed across programs and differences in completion rates in Table 1 cannot be given a causal interpretation.

Table 1. Descriptive statistics – Program applicants

	Q1	Q2	Q2 criteria				
			Grades	Test	CV	Essay	Interview
Female	0.62	0.64	0.64	0.60	0.64	0.66	0.61
Age	21.7	22.2	22.2	22.0	22.2	22.3	22.3
Immigrant	0.13	0.10	0.10	0.04	0.10	0.10	0.08
SES	0.73	0.74	0.74	0.78	0.73	0.73	0.74
GPA rank	0.53	0.44	0.46	0.65	0.43	0.41	0.49
Apply in Q2	0.39						
Completes program	0.18	0.18	0.16	0.46	0.17	0.18	0.31
Completes college	0.78	0.78	0.78	0.84	0.78	0.78	0.81
Completion on time	0.17	0.16	0.15	0.40	0.16	0.17	0.27
# of applications	536,323	207,969	170,954	5,030	196,704	120,939	9,652
# of applicants	245,560	101,567	87,007	4,801	97,042	72,081	7,947

Note: Sample means for program applicants. Immigrant refers to having at least one non-native Danish parent. SES corresponds to the percentile rank of parents' average income in the year the applicants turn 15. GPA rank is the applicants' percentile ranking in their birth cohorts GPA distribution. College completion is measured in 2019. Applicants who apply in multiple years are included separately. Completes in time is defined as within one year after formal completion time.

to the program $U_1 \equiv Y^1 + \epsilon_1$ exceeds their utility when not admitted $U_0 \equiv Y^0 + \epsilon_0$:

$$\Delta U = Y^1 - Y^0 + (\epsilon_1 - \epsilon_0) > 0$$

and where utility depends on a potential outcome of interest Y^k – say program completion – which is indexed by admission, and a residual term ϵ_k . Applicants apply if the return to performance $Y^1 - Y^0$ exceeds the compensating differential $\epsilon_0 - \epsilon_1$.

Applicants who decide to apply to the secondary quota must provide additional information I at a cost $C(I) > 0$. Secondary screening depends on this new information I and we denote the probability of admission by $p_2 \equiv p_2(I)$. The probability of admission in the primary quota depends on applicants' GPA and is denoted by $p_1 \equiv p_1(GPA)$. Applicants apply to the secondary quota and provide information I if in expectation this is beneficial to them:

$$(1 - p_1)p_2(U_1 - U_0) - C(I) > 0$$

$$\Delta U > \frac{C(I)}{(1 - p_1)p_2}$$

This illustrates that the secondary quota can affect the outcomes of the pool of admitted

applicants through screening and sorting. Screening is directly affected by use of the additional criteria I in secondary admissions through p_2 as long as in expectation these provide differential information about applicants' performance in the program when admitted Y^1 .

Sorting arises because of the application cost $C(I)$ which not only leads to fewer people applying, $\Pr(Q2) \leq \Pr(Q1)$, but may also affect their composition. To illustrate this, assume that potential performance on admission Y^1 and net benefits $\Delta U \equiv U_1 - U_0$ follow a joint-normal distribution. In that case the average expected performance of students who applied to the secondary quota is higher (lower) than that in the primary quota if the correlation between Y^1 and ΔU is positive (negative). Introducing an application cost thus changes the pool of applicants depending on the nature of the selection.¹⁰

The above implies that given an application to primary admission the probability to apply for evaluation in secondary admission is a function of benefits of enrolling in the program, the probability of being admitted both primary and secondary admission, and application cost. We start out by modeling applicants latent propensity to apply to Q2 as follows

$$\Pr(Q2_i = 1 \mid x_i, \text{criteria}_j) = \Phi(x_i' \beta + \text{criteria}_j' \gamma) \quad (1)$$

where the expected net-benefits of enrolling, captured by the first term on the right-hand side of (1), may vary across background, field-of-study, institution, ability (GPA) and application cohort and are controlled for in the vector x_i . Q2 application cost are allowed to vary as a function of the application criteria by the second term. Though the criteria are used in combination we include them in a separable specification.

The results are displayed in Table 2. From model 1 we observe that males and immigrants are less likely to apply for admission in the secondary quota, while applicants from high SES background are more likely to apply. Applicants with higher GPA are less likely to apply, and a higher GPA-cutoff in the previous year increases the probability of observing a Quota 2 application. The role of GPA and cutoffs is in line with the presence of application costs in the model outlined above.

The decision to apply for admission in the secondary quota vary with criteria used to evaluate the applicant. The first criteria is grades, which are easily provided and thus should be associated with small application costs. In line with this, we find that additional grades as a criteria has negligible impact on application. We expect, on the other hand, that tests are associated with larger costs, as they require preparation and sitting the test. This is supported by the empirical evidence. Taking the model at face value,

¹⁰The assumption of joint normality is not necessary but makes the correlation between Y^1 and ΔU a sufficient statistic to determine the sign of the sorting.

Table 2. Probability of Q2 application

	(1)		(2)	
Female	0.011	(0.0022)	0.007	(0.0022)
Immigrant	-0.172	(0.0039)	-0.155	(0.0038)
SES	0.043	(0.0049)	0.051	(0.0049)
GPA	-0.616	(0.0039)	-0.625	(0.0039)
Lagged cutoff	0.052	(0.0008)	0.026	(0.0014)
Q2 Criteria				
- Grades	0.024	(0.0027)	-0.012	(0.0050)
- Test	-0.210	(0.0065)	-0.038	(0.0179)
- CV	0.067	(0.0038)	-0.001	(0.0056)
- Essay	-0.066	(0.0018)	0.004	(0.0042)
- Interview	-0.078	(0.0049)	-0.021	(0.0080)
Program FE			✓	
Observations	520,595		519,680	

Note: The table displays marginal effects from Probit models with standard errors in parentheses. Models are estimated on application level. The second model incorporates program fixed effects.

the introduction of tests is associated with a drop in applications to Q2 of 21 percentage points, which is the largest effect among the criteria used. CVs are associated with a positive effect, while essay has a negative effect of the same magnitude.¹¹ Interviews are also associated with a decrease in applications for Q2, which is in line with the presence of application costs.

Model 2 includes program fixed effects, thus discarding cross-program variation and solely exploiting variation across time. The effects of socio-demographic variables are invariant to the inclusion of program effects. The partial effect of the lagged cutoff halves, which is expected, as the cutoff varies less over time than across programs. Only exploiting time-variation in criteria changes the partial effects of criteria. Tests continue to be associated with the largest decrease in the probability of filing a Q2 application but falls by an order of magnitude compared to the specification without program fixed effects. The decrease in magnitudes of the partial effects of the lagged cutoff and criteria from including fixed effects indicate that cutoffs and criteria may be associated with unobserved program characteristics which may also influence the decision to apply for secondary admission. Thus we caution against attributing a causal interpretation to

¹¹We note that the linearly separable specification of criteria might hide complementarities in criteria. This is may especially be an issue with grades, CV and essays, as these are often used in combination as shown in Figure 2. We return to this issue in section 9.

the models presented in Table 2. This motivates the Regression Discontinuity Design which we present in the following section.

5 Program admissions

Assume that the admissions office aim is to identify applicants who (in expectation) are most likely to be academically successful on admission:

$$E[Y_i^1 | I_i] \tag{2}$$

where $Y_i^{A_i}$ is applicant i 's potential achievement as function of program admission $A_i \in \{0, 1\}$, and information I_i on applicant i . While programs can have admission objectives that relate to the grades of students, their success later in the labor market, or the demographic composition of the student body, we focus on program completion rates as the measure of academic performance that programs aim to maximize on admission as programs in Denmark are financed on the basis of the number of graduates. Admission decisions will be based on I_i , the available information on applicant i , through (3):

$$A_i(I_i) \tag{3}$$

There is a potential externality to admission when programs do not internalize the counterfactual outcomes of applicants who are not admitted to the program. Everything else equal it is optimal for a social planner to admit people who do not do well in the counterfactual, and a social planner that optimizes college completion will, therefore, consider the expected return or value added (VA) of an admission offer

$$E[Y_{\text{college},i}^1 - Y_{\text{college},i}^0 | I_i], \tag{4}$$

To investigate the extent to which program admissions are aligned with the aim of maximizing overall graduation rates we therefore also report estimates of value added for college completion. Another measure of interest is whether the applicant completes on time if admitted. We report estimates for this outcome in the appendix.

Estimating (2) (and (4)) without bias is typically challenging because of selection issues (Rothstein, 2004). Selection bias arises when information I_i is only observed conditional on matriculation, or when outcomes are only observed conditional on admission.

Another key challenge is when admissions are not random conditional on the infor-

mation used in the admission:

$$E[Y_i | A_i(I_i) = 1, I_i] \neq E[Y_i^1 | I_i]$$

in which case estimates of (2) are also biased. Similarly, bias also arises when admission is conditionally random but when not all information $I_i = (\tilde{I}_i, \tilde{I}_i^c)$ used in the admission is observed, but instead only a subset \tilde{I}_i is available:

$$E[Y_i | A_i(I_i) = 1, \tilde{I}_i] = E[Y_i^1 | A_i(I_i) = 1, \tilde{I}_i] \neq E[Y_i^1 | \tilde{I}_i] \neq E[Y_i^1 | I_i]$$

These challenges are compounded for 4 because then also the counterfactuals outcomes in absence of admission need to be estimated.

In addition to addressing selection bias, there is the question for what population we need to estimate (2). We are interested in the effects of changes in admissions policies, and more specifically the impact of using different information (I_{2i} vs. I_{1i}) in the admission process:

$$E[Y_i^1 | I_{2i}] - E[Y_i^1 | I_{1i}]$$

We start by investigating what type of information leads to more successful admissions *at the margin*. This not only tells us what criteria are more predictive of potential outcomes for marginal applicants, but is also of first-order policy interest as it indicates whether to expand or contract the different admission quotas.

The Danish registry data allow us to overcome the selection challenges that are due to the data limitations outlined above because we observe all applicants, their relevant (to admission) characteristics and subsequent admission and outcomes. We can estimate the potential outcomes for marginal applicants thanks to the regression discontinuity design implicit in the admission system which we turn to now.

5.1 Regression discontinuity design

The deferred acceptance admission mechanism described in section 2 results in admission for applicants who are ranked high enough while applicants with a marginally lower rankings are rejected. This setup gives rise to a fairly standard regression discontinuity design where the application threshold is defined by the application score of the last admitted student, and which has been used extensively to study the the impact college on individual outcomes (f.e. [Öckert, 2001, 2010](#); [Kirkeboen et al., 2016](#)).¹²

Programs rank applicants. For primary admissions (Q1) applicants are ranked based

¹²In the Danish case where the running variable is rounded on the first decimal and tie-breaking is random the last admitted student is selected through a lottery.

on their GPA (i.e. $I_i = I_{1i} = GPA_i$), and applicants that also applied to secondary admissions (Q2) are ranked based on alternative criteria I_{2i} . The applicants that rank highest based on their GPA will be admitted first. The number of applicants that receive an offer in Q1 is limited by programs' capacity and, if over-subscribed, the rank of the last admitted applicant defines the admission threshold c_1 . Applicants whose ranking in Q1 is below c_1 but who applied for alternative evaluation are now considered in the same manner, but this time based on their program ranking using alternative criteria. This again implicitly defines an application threshold c_2 , and applicants that reach this threshold are admitted in Q2. ¹³

In the analyses we define the assignment variable r_i as the percentile distance of applicant i 's application score to the threshold. Threshold crossing is defined by $z_i \equiv [r_i \geq 0]$ which equals 1 if an applicant is above the application threshold and is zero otherwise. For each quota we start with estimating non-parametric threshold crossing effects on the sample of applications with the following specification

$$E[W_i | z_i, r_i] = \delta z_i + f(r_i) + FE_i \quad (5)$$

where W_i can be i) program admission A_i , ii) program completion Y_i , and iii) college completion $Y_{\text{college},i}$. We control for a continuous function of the running variable $f(r_i)$ which is estimated non-parametrically on both sides of the cutoff using local linear regression. Most applicants apply to multiple programs and we estimate equation (5) at the application level using the implementation and default bandwidths of [Calonico et al. \(2017\)](#). Standard errors are clustered at the applicant level. Quota 1 is frequently much larger than Quota 2. In order to be able to compare movement along the running variables in the two quotas, we rescale the running variable in Quota 2 by the ratio of Quota 2 applicants to Quota 1 applicants in a given program-year combination.

6 Program applications

6.1 Validity

Before discussing the threshold crossing effect estimates we first perform standard checks to confirm the validity of our regression-discontinuity design. A first check consists of investigating whether applicants are able to strategically sort themselves around the application cutoff. If so, this would imply that the density of GPA is not continuous

¹³Applicants who did not apply to Q2 do not have an alternative ranking and are not shown in the Figure but are of course part of the admission in Q1. The exposition ignores the presence of other programs in the Deferred Acceptance mechanism. As explained in section 2 the admission system mechanically expands the rank-ordered list such that a Q2 application is listed immediately after the Q1 application,

Table 3. Bunching and balancing around the admission thresholds

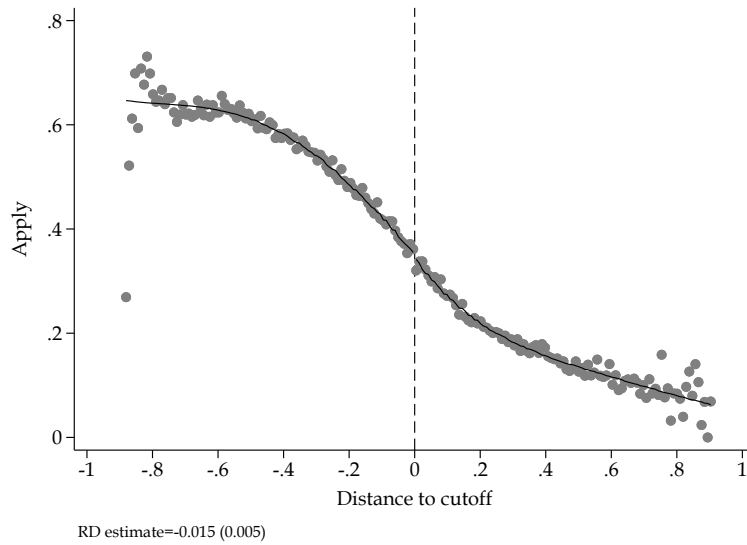
	Density Disc.	Imbalance (s.e.)				
	(%) [p-value]	Female	Immigrant	SES	Apply to Q2	GPA
Admission:						
- GPA-based (Q1)	0.017 [0.417]	0.006 (0.006)	-0.001 (0.004)	0.027 (0.011)	-0.010 (0.006)	
- Alternative (Q2)		-0.006 (0.008)	-0.001 (0.005)	-0.017 (0.017)		-0.017 (0.019)

Note: The table contains validity check on the regression discontinuity design in Quota 1 and Quota 2. Density check is performed using `rddensity`. Balance checks are performed using `rdplot` and `rdrobust` packages in Stata. Distance in GPA can only be computed within one decimal and applicants with zero distance are excluded from the density check as they are subject to a lottery.

around the threshold. The first column of Table 3 and Figure A1 in the Appendix shows that the estimates of the jump in the density are very small and highly insignificant, and there are therefore no signs of sorting in the primary admission quota.¹⁴

The second validity test is a balancing check where we investigate whether the applicants around the application threshold are comparable in terms of observed characteristics that are potential confounders. Balance would be the consequence of local quasi-randomization or continuity, and is required by the RD which relies on the continuity of potential outcomes around the application thresholds. We implement the balancing test by estimating threshold crossing “effects” on the background characteristics of the applicants, which are reported in the remaining columns of Table 3 (Figures A2 and A3 in the Appendix show the corresponding RD graphs). In the main admission quota we consider sex, immigrant status and SES, while in the second admission quota we also look at GPA. We do not find evidence of imbalance as the applicants are similar on both sides of the cutoff.

Figure 3 reports the fraction of applicants that also apply for admission on alternative criteria in Q2 as a function of the percentile rank in Q1. First, note that here too we do not observe a discontinuous change at the threshold. Second, there is a strong negative relationship between applicants ranking in Q1 and the likelihood of them applying to Q2. The negative slope is consistent with the model above where where the net-benefits of applying are decreasing in GPA since the likelihood of admission in Q1 increases with GPA.



Note: The figure plots the probability of applying to Quota 2 conditional on applying for Quota 1 as a function of percentile rank distance to the cutoff in Quota 1. Point estimate along with standard errors are presented below the graph. Graph and estimate are constructed using `rdplot` and `rdrobust` packages in Stata. Only bins with more than 20 individuals are shown.

Figure 3. Fraction of applicants that apply for alternative evaluation (Q2)

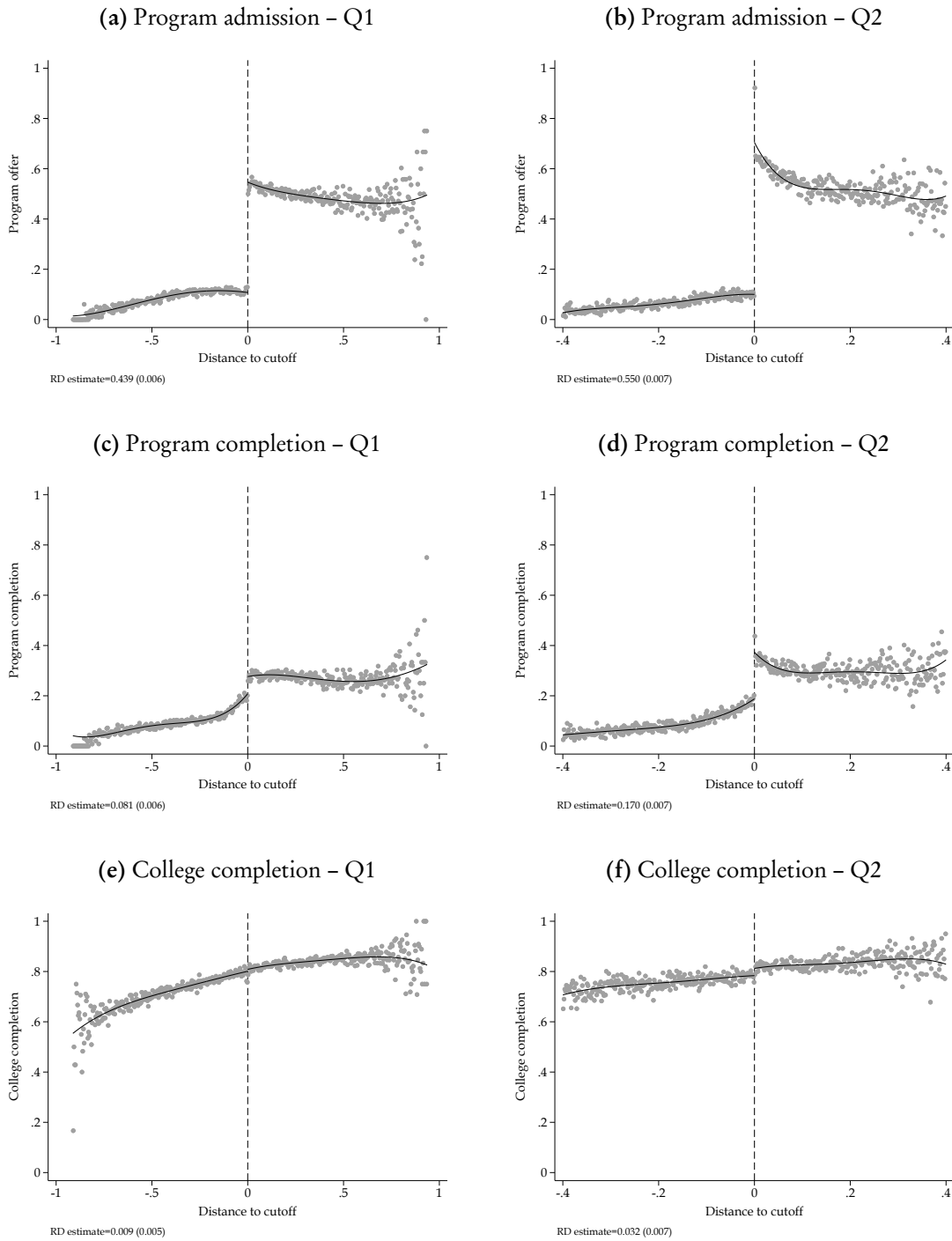
6.2 Admission thresholds

Figure 4 documents what happens to admission, program and college completion when applicants cross the threshold. The left column shows the findings for all applicants in the main admission based on high-school GPA regardless of whether they applied for Q2. In the top-left panel we see that about 50 percent of the applicants receive an offer if they find themselves above the application threshold, compared to a 10 percent offer rate for those below the threshold. The offer rate on threshold crossing does not equal 1 because applicants will only receive an offer if they were not admitted in a higher ranked program on their application list. The offer rate below the threshold is higher than 0 because applicants can receive an offer through the other quota.¹⁵ The effect of threshold crossing, the first-stage in the Q1 analyses below, is 0.44 and highly significant.

Slopes must be interpreted with caution. Applicants with a higher GPA face a higher likelihood to be admitted in a higher ranked program which would lead to a negative slope. A higher GPA may also increase the likelihood that people applied to more selective programs, which would lead to a positive slope. The observed slope is the net

¹⁴Applicants are ranked in the secondary admission quota. Since (percentile) rankings are always uniformly distributed, a bunching test for Q2 is not informative of selection on the margin of application.

¹⁵Programs may have additional quotas besides standard Q1 and Q2. For example, applicants from Greenland can be evaluated in an additional quota. Such additional quotas, while small, also contribute to a positive offer rate below the threshold in Quota 2.



Note: Figures contain first stages and reduced forms in Quota 1 and 2. Point estimates of value added, $Y^1 - Y^0$, with standard errors are presented below the graphs. Graphs and estimates are constructed using `rdplot` and `rdrobust` packages in Stata. The reduced forms for “completing in time” are show in Appendix figure A4.

Figure 4. Threshold crossing in Q1 and Q2

effect of these two mechanisms, and they appear to cancel each other out on average.

The left-middle panel shows that threshold crossing increases the likelihood of completing the program by 8 percentage points from about 0.20 below the threshold to 0.30 above, while the left-bottom panel shows that the probability of completing college increases with 1 percentage point by threshold crossing in the main admission quota.

The second column of Figure 4 reports the findings for admissions based on alternative criteria in Q2. Threshold crossing has a large impact on the likelihood of receiving an admission offer which increases by 55 percentage points. Note here that the slope above the threshold is negative because higher ranked applicants are more likely to receive an offer in a dominating program. The large increase in the program offer rate is also reflected in a large 17 percentage point increase in the program completion rate. Threshold crossing in Q2 increases the likelihood of completing college by 3 percentage points.

Threshold crossing gives therefore rise to sharp discontinuous increases in program offer and completion rates both in the main admission quota and the second one where alternative criteria are used to assess applicants. Moreover, admission in the main quota does not impact college completion while admission in the secondary quota does.

7 Evaluating college admissions

We start out by estimating whether, at the margin of admission, GPA-based admission or alternative evaluation is more successful in selecting applicants by estimating potential completion rates. We then investigate heterogeneity by program characteristics before turning to the mechanisms and quantify the importance of sorting relative to screening. Finally we investigate the importance of different evaluative criteria in explaining differences in admission outcomes.

7.1 Fuzzy RD

The regression-discontinuity in the admission process documented above is fuzzy because applicants only receive an offer above the threshold if they were not admitted to a higher ranked program on their application list. To identify the expected potential outcomes given admission under the different admission regimes we implement the fuzzy RD design generated by the admission process using 2SLS. The baseline model used to estimate the completion rate on admission for compliers (c) at the admission threshold, $E[Y_i^1 | c, r_i = 0]$, is as follows

$$E[A_i Y_i | A_i, r_i, F E_i] = \delta A_i + B_{3,0}(r_i) + F E_i \quad (6)$$

where A_i is a binary variable indicating whether applicant i received an admission offer in the year of application, and where the dependent variable $A_i Y_i$ ensures that we estimate counterfactual outcome *levels* Y_i^1 of compliers rather than effects of admission on outcomes (cf. [Abadie, 2003](#)). To estimate the value added of admission, $Y_i^1 - Y_i^0$, the dependent variable in (6) is changed to Y_i .

Since application thresholds are defined at the program-year level, we control for program \times year fixed effects (FE_i). This will also at least partially account for the non-linearities in the running variable that we observed in the non-parametric first-stage and reduced-forms threshold-crossing estimates above to the extent that these were explained by changes in composition. To account for any remaining non-linearities the 2SLS specification models $f(r_i)$ as a cubic spline with a knot at zero: $B_{3,0}(r_i)$.¹⁶

To take the mechanical fuzziness of the admission into account we instrument for A_i with threshold crossing $Z_i \equiv [r_i \geq 0]$:

$$E[A_i | Z_i, r_i, FE_i] = \pi Z_i + B_{3,0}(r_i) + FE_i$$

As before, estimation is at the application level and standard errors are clustered at the applicant level.¹⁷ We estimate (6) separately for GPA-based admissions and alternative evaluations and for program and college completion.

7.2 GPA-based admissions vs. alternative evaluations

Table 4 reports the OLS and 2SLS estimates of equation (6). To verify that the parametric 2SLS specification is sufficiently flexible the table also reports non-parametric fuzzy-RDD estimates. Comparing these to the parametric 2SLS estimates we see that the estimated potential completion rates of marginal applicants are very closely aligned across estimators. We therefore conclude that a specification with a cubic spline and waiting-list fixed effects is sufficiently flexible to capture the non-linearities in the data. Visual inspection of the fit of the reduced form in our 2SLS model with the data (reported in Appendix Figure A6) also confirms that the 2SLS regressions are correctly specified away from the application thresholds.

Is it better to admit the last student based on GPA (Q1) or alternative evaluations (Q2)? Table 4 shows that the 2SLS estimate of the average potential program completion rate for marginal applicants when admission is based on GPA is 0.5. The next row shows that the potential program completion rate for marginal applicants when admission is

¹⁶In order not to exhaust the Greek alphabet we use the generic notation $B_{3,0}(r_i)$ and FE_i throughout with the understanding that these are allowed to vary across specifications.

¹⁷For applicants with the same GPA, the ranking is determined by a lottery. We can therefore define the instrument as $Z_i = A_i$ for Quota 1 applicants exactly at the cutoff.

based on alternative criteria equals 0.52. This 2 percentage points relative to GPA-based admissions is substantial and statistically significant, and the average program that maximize completion rates should therefore increase take-up under alternative evaluation at the expense of regular GPA-based admission. However, the gain of 2 percentage points should be weighted against the costs of reviewing and ranking all Q2 applications regardless of whether they are from marginal applicants, clearly ineligible applicants or applicants who would get admitted anyway in Quota 1 on GPA. The cost of the Quota 2 admission process might therefore be considerable for the program whereas ranking on GPA in Quota 1 is essentially free.

We do not observe the same difference at the college completion margin as average potential completion rates are similar under GPA-based and alternative evaluation. This suggests that while alternative evaluation is more effective than GPA-based admission at generating good program matches, and programs can increase completion rates by moving admission slots from Q1 to Q2, alternative evaluation does not outperform GPA-based admission when it comes to college completion.¹⁸ This is confirmed by the estimates for value added which show that the alternative evaluation outperforms GPA-based admission. Given that we do not find differences in expected Y^1 between Q1 and Q2 applicants, the difference in value added reflects lower completion rates of Q2 applicants if not admitted to the program. If college completion has a positive causal effect on labor market participation and earnings, then the difference in value added indicates that increased admissions via Quota 2 may substantially benefit government budgets in saved benefits and increased tax revenue.

7.3 Heterogeneity by program characteristics

Programs differ in content and applicant pool and this may have implications for the effectiveness of alternative evaluations compared to the use of GPA in admission. We therefore proceed to investigate heterogeneity in terms of field and selectivity of programs,

7.3.1 Field of study To investigate heterogeneity across fields we group programs by broad field-of-study and estimate our 2SLS-specification separately for fields. The results are presented in Figure 5. For program completion the largest difference between alternative evaluation and GPA-based admission is found for the Humanities, where completion levels are 5 percentage points larger for marginal applicants in Quota 2 relative to Quota 1. On the other extreme lie Social Science and Health where we do

¹⁸Results for completion in time, shown in Appendix Table A1, support the interpretation that Q2 outperforms Q1 in creating program matches.

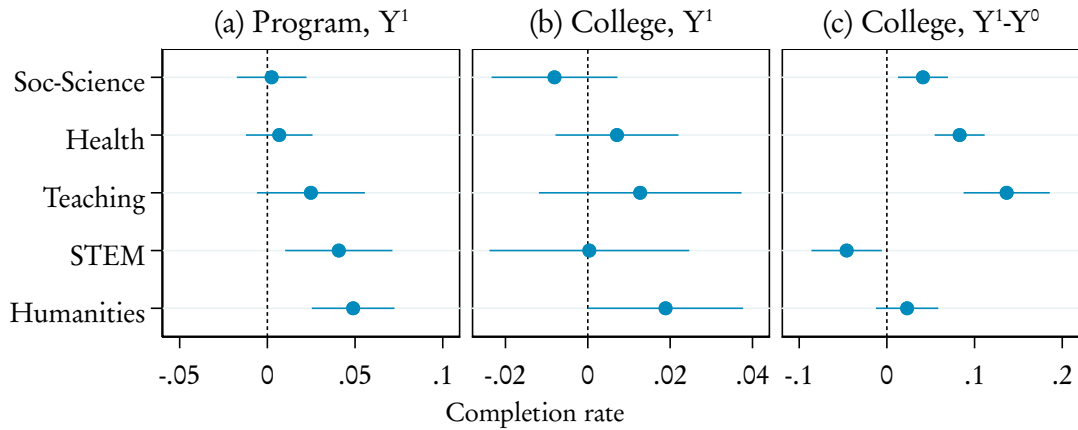
Table 4. Potential program and college completion rates (Y^1) and value added ($Y^1 - Y^0$) for marginal applications – GPA-based admission vs. Alternative Evaluation

	Program, Y^1			College, Y^1			College, $Y^1 - Y^0$		
Admission:									
- GPA-based (Q1)	0.517 (0.002)	0.498 (0.006)	0.490 (0.009)	0.820 (0.001)	0.805 (0.005)	0.801 (0.004)	0.027 (0.001)	0.025 (0.010)	0.021 (0.009)
- Alternative (Q2)	0.523 (0.002)	0.517 (0.006)	0.520 (0.007)	0.826 (0.002)	0.806 (0.005)	0.810 (0.005)	0.027 (0.002)	0.068 (0.008)	0.063 (0.012)
Q2 vs. Q1	0.006 (0.002)	0.019 (0.009)	0.030 (0.012)	0.006 (0.001)	0.001 (0.007)	0.009 (0.007)	-0.000 (0.002)	0.043 (0.013)	0.041 (0.017)
Estimator	OLS	2SLS	RD	OLS	2SLS	RD	OLS	2SLS	RD

Note: The table contains estimated average outcome of admittance (Y^1) with standard errors in parentheses. Standard errors are clustered on applicant level. Models are estimated for program and college completion using OLS, 2SLS and fuzzy RDD estimated in Stata using the rdrobust-package. Standard errors in the difference between Q2 and Q1 are bootstrapped with 50 replications. Results for program value added and completion in time are shown in Appendix Table A1.

not find evidence that alternative evaluation outperforms GPA-based admission. For college completion we once again fail to find any difference in completion levels. However, we find large differences in college completion value added ($Y^1 - Y^0$) for Health and Teaching of about 9 and 15 percentage points respectively. This indicates that if admission officers fail to take the outcome of non-admittance into account, programs will tend to admit too many applicants in the main quota relative to the quota using alternative criteria. A social planner may therefore prefer admission in Quota 2 within these fields, while individual programs would see little benefit.

7.3.2 Selectivity To investigate heterogeneity by program selectivity we rank program-year combinations into degrees of selectivity based on their GPA-cutoff in Quota 1 in previous years. Figure 6 reports the stratified estimation results. For very-low-selectivity programs we do not find evidence of a difference in expected program completion rates between Quota 2 and Quota 1, the point estimates is negative and very noisily estimated. The difference between Quota 2 and Quota 1 is however increasing in selectivity, and for moderate and highly-selective programs we estimate a difference of 4 percentage points. The positive average effect we found above is therefore driven by the more selective programs. One potential explanation can be framed in our model of admission in section 6. A high cutoff lowers the probability of admission, and a larger pool of applicants will find it attractive to apply for the alternative quota. If potential program completion is rising and concave in GPA, the marginal completion gain from GPA in highly selective



Note: These figures plot the estimated gap in outcomes (Y^1) and value added ($Y^1 - Y^0$) between GPA-based admissions and alternative evaluations by field of study. Standard 95-percent confidence intervals based on standard errors clustered at the applicant level. All estimates come from separate regression.

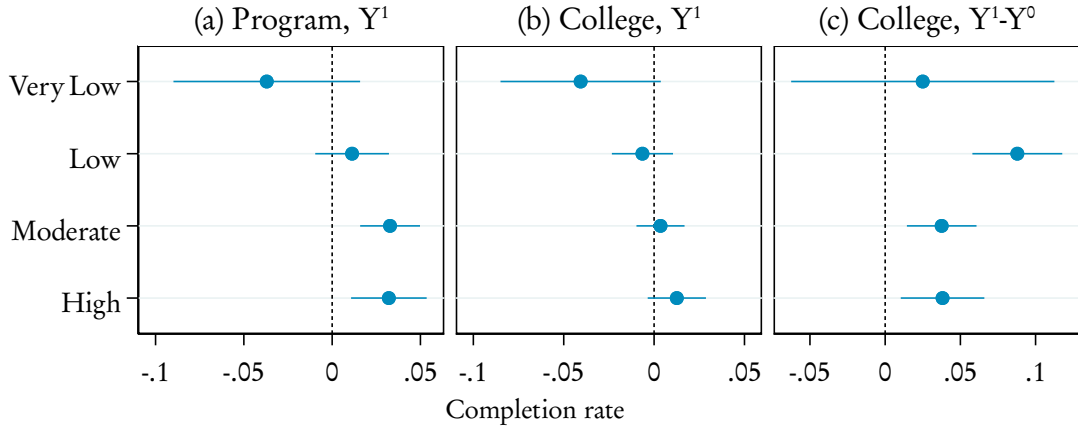
Figure 5. Potential program and college completion rates for marginal applications – GPA-based admission vs. Alternative Evaluation by field,

programs may be limited and the value of the signal of a Quota 2 application may be relatively more valuable.

For college completion the point estimates of the difference between alternative evaluation and GPA-based admission also increases in selectivity, but less so and differences are relatively small and statistically insignificant. Finally, we see a relatively uniform difference for college completion value added ($Y^1 - Y^0$). While programs with low selectivity appear to stand out somewhat we cannot reject that the effects are homogeneous. However, a large value-added for low selectivity programs reflect that for these programs, the outside option for Q2-applicants (with relatively low GPA) is worse than for the Q1-applicants.

8 Sorting vs. Screening

Our model of program applications in Section 4 highlighted that differences between admission regimes can arise because of differential selection into application, and because alternative evaluative criteria may admit different applicants *conditional on selection*. Knowing the channel is important for how the admission office should construct the Quota 2 criteria. If sorting dominates, then imposing costs on applicants is more important than using the additional information provided. In contrast, if screening dominates, then the criteria provide useful information on the potential completion rate of applicants. We now assess why alternative evaluation outperforms GPA-based admission by quantifying the relative importance of these sorting and screening channels.



Note: These figures plot the estimated gap in outcomes (Y^1) and value added ($Y^1 - Y^0$) between GPA-based admissions and alternative evaluations by program selectivity. Low selectivity programs had a cutoff below 4, Low between 4 and 7, Moderate between 7 and 10, and highly selective programs had cutoffs above 10. Standard 95-percent confidence intervals based on standard errors clustered at the applicant level. All estimates come from separate regression.

Figure 6. Potential program and college completion rates for marginal applications – GPA-based admission vs. Alternative Evaluation by selectivity.

To make these concepts precise let $E[Y_i^1 | Q_{2i}, I_i]$ refer to the average completion rate of marginal applicants if admitted based on information I_i . All respondents apply to Q1, so we can characterize the applicants with a dummy for a Quota 2 application, Q_{2i} . To quantify sorting note that the difference in potential outcomes between Q2 and Q1-only applicants who were admitted in Q1 keeps the evaluation constant and only reflects differences in composition induced by the alternative screening criteria. This comparison therefore quantifies sorting:

$$E[Y_i^1 | Q_{2i} = 1, I_i = GPA_i] - E[Y_i^1 | Q_{2i} = 0, I_i = GPA_i] = \text{Sorting}$$

Similarly, the difference in potential outcomes between Q2 applicants who were admitted under different criteria keeps the applicant pool constant and any remaining difference will be the result of screening:

$$E[Y_i^1 | Q_{2i} = 1, I_i = Alt_i] - E[Y_i^1 | Q_{2i} = 1, I_i = GPA_i] = \text{Screening}$$

and the total difference in potential outcomes between i) applicants who applied to *and* were admitted in Q2 and ii) applicants who only applied to Q1, therefore encapsulates both these sorting and screening effects:

$$E[Y_i^1 | Q_{2i} = 1, I_i = Alt_i] - E[Y_i^1 | Q_{2i} = 0, I_i = GPA_i] = \text{Total} = \text{Sorting} + \text{Screening}$$

To illustrate this decomposition, Figure 7 shows the first stages in our data for applicants with an application for the secondary quota.¹⁹ The top plot shows the first stage for GPA-admission and the right plot shows the first stage using alternative criteria. These plots are similar to the first stages presented in Figure 4. The scatter plot in the center shows the distribution of observations, colored according to admission probability, with lighter color corresponding to higher admission rate.

As the figure only contains Q2-applicants, $E[Y^1|Q_{2i} = 1, I = Alt_i]$ is estimated at the threshold at the vertical axis for the marginal applications in programs' Q2-admissions based on alternative evaluations. $E[Y^1|Q_{2i} = 1, I = GPA]$ is estimated at the threshold on the horizontal axis for programs' applications at the margin of GPA-based admissions in Q1. The difference between the two constitute the screening effect. The sorting effect is computed by comparing the Q2-applicants at the margin in the Q1 admissions to the marginal applicants without a Q2-application, $E[Y^1|Q_{2i} = 0, I = GPA]$. These latter Q1-only applicants are located around the same threshold on the horizontal axis but cannot be located on the vertical axis as they have no Q2-ranking.

Crossing a threshold is sufficient to get an offer, and the marginal applicants can therefore only be located on the dashed lines. An advantage of our quota-specific RDDs and the assignment mechanism is that we recover the potential outcomes of these marginal applicants without the need to model the joint distribution of running variables.^{20, 21}

We implement the decomposition into sorting and screening using the following specification where we pool all applications:

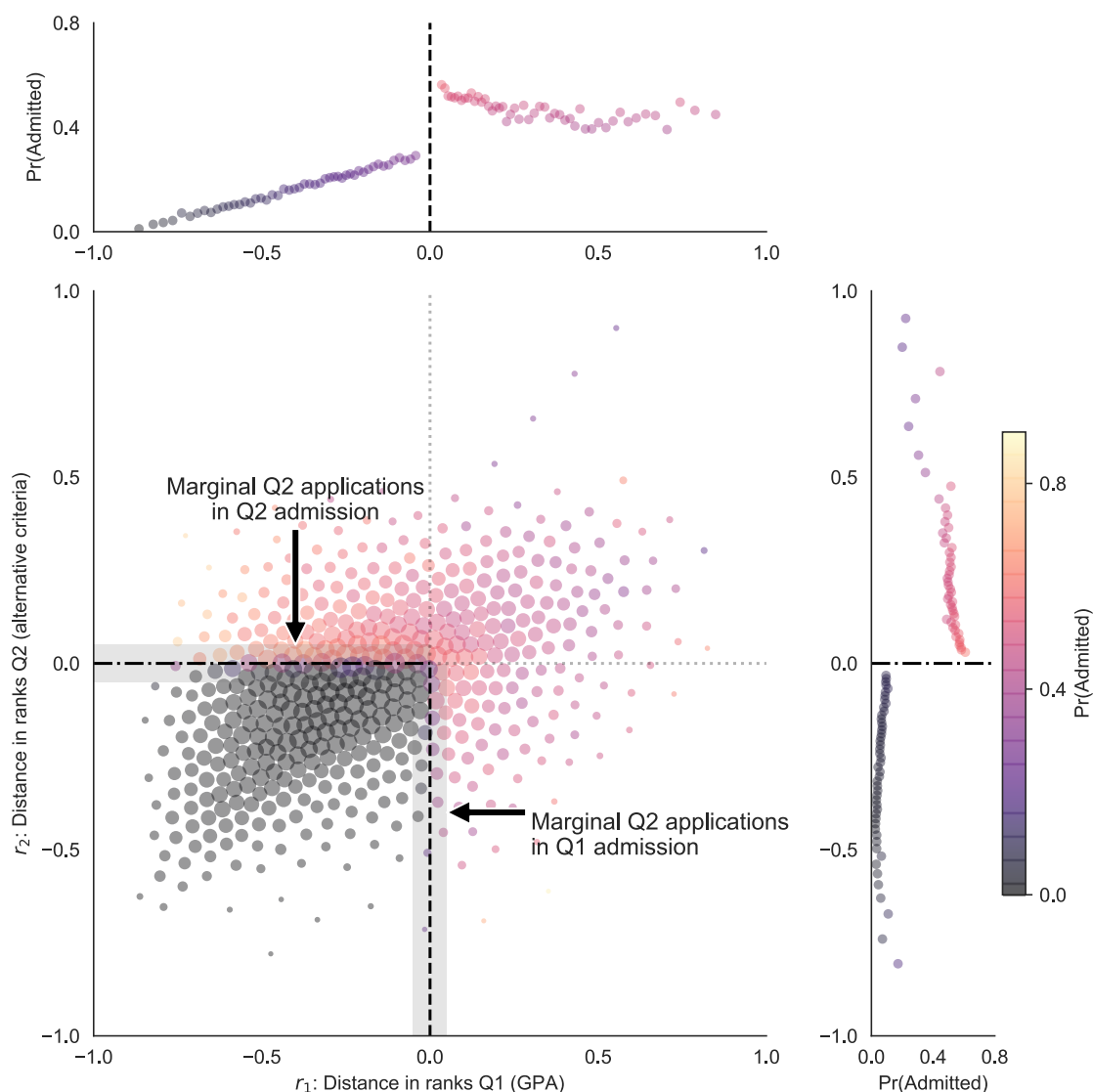
$$E[A_i Y_i | A_i, Z_{1i}, Q_{2i}, r_{1i}, r_{2i}, FE_i] = \delta_{q1} A_i + \delta_{\text{sorting}} Q_{2i} A_i + \delta_{\text{screening}} (1 - Z_{1i}) Q_{2i} A_i + \alpha Q_{2i} + \gamma Z_{1i} Z_{2i} + B_{3,0}(r_{1i}) + Q_{2i} B_{3,0}(r_{1i}) + Q_{2i} B_{3,0}(r_{2i}) + FE_i \quad (7)$$

The variable A_i indicates whether applicant i was admitted to the program, and Q_{2i} is a dummy indicating whether i applied for alternative evaluation in Q2. The variable Z_{1i} equals one if the applicant's GPA is above the application threshold and thus indicates whether i is predicted to receive an offer in the main GPA-based admission in Q1. Similarly Z_{2i} equals one if applicants' ranking in the alternative evaluation is higher than

¹⁹Applicants without an application for Q2 are not ranked in the alternative evaluation and therefore omitted.

²⁰In Figure 7 our 2SLS approach corresponds to estimating RDD on the marginal distribution (for Q2-applicants). In line with the workings of the allocation mechanism, we observe sharp discontinuities in Figure 7 at the borders of the lower left quadrant, illustrated by the sudden shift in color. Above the two thresholds the applicants are either always- or never-takers on each threshold. In line with this reasoning, we do not observe discontinuities in admission probabilities along the dotted lines in the figure. This is essentially a placebo test and serves as further validation of our research design.

²¹Note that the results displayed in Figure 7 exclude Q1-applicants without a Q2-application and does not include fixed effects. The results are therefore not directly comparable to the 2SLS results below.



Note: This figure illustrates the first stages in our data for applicants with a Quota 2 application. To preserve anonymity the center plot is constructed using K-means where bins are colored according to the share admitted to the program. The top and right marginal plots show the first stages for Q1 and Q2 respectively. Observations in the top and right plots are clustered into 100 bins using K-means. As we exclude applicants without a Q2-application and due to difference in binning technique, the first stages presented here differ from those in Figure 4. Note that we do not control for program-year fixed effects. The possible location in the joint distribution of the marginal applicants on the marginal RDDs are illustrated by the dashed lines. The dotted lines serve as placebo tests for cutoffs, for those who should be admitted in the other quota.

Figure 7. Program admission under regular GPA-based admission (Q1) and alternative evaluation (Q2) for Q2 applicants

that of the last admitted applicant and is zero for everybody else.²²

The endogenous admission variables A_i , $Q_{2i}A_i$ and $(1 - Z_{1i})Q_{2i}A_i$ in (7) are instrumented with their corresponding predicted offers: Z_{1i} , $Q_{2i}Z_{1i}$ and $(1 - Z_{1i})Q_{2i}Z_{2i}$. Since Q2 applicants who are predicted to receive an offer in the GPA-based admission ($Z_{1i} = 1$) will never receive an offer in the alternative evaluation, we dummy out $Z_{1i}Z_{2i}$ in order to exploit the correct variation in Z_{2i} (i.e. threshold crossing from the bottom left to bottom right, and bottom left to top left for the Quota 2 applicants in Figure 7, and Q1 threshold crossing for the Q1-only applicants). Finally, to complete the 2SLS implementation of the RD design, the specification also includes quota-specific cubic splines in the running variables r_{1i} and r_{2i} in Q1 and Q2, as well as program \times year specific fixed effects FE_i .

The baseline outcome for GPA-based admission for applicants who did not apply to alternative evaluation is estimated by δ_{q1} . The key coefficients of interest in equation (7) are δ_{sorting} which quantifies sorting; the differential outcome of marginal applicants who applied to alternative evaluation and those who did not under GPA-based admission, and $\delta_{\text{screening}}$ which quantifies screening; the differential outcome of marginal applicants who applied to Q2 and who were admitted based on alternative criteria compared to those who were admitted based on GPA. The sum of δ_{sorting} and $\delta_{\text{screening}}$ is the total differential outcome between Q2 marginal applicants and Q1 marginal applicants without a Quota 2 application. This specification differs from the specification in Table 4, where the marginal applicant is a weighted mean of applicants with and without a Quota 2 application.²³

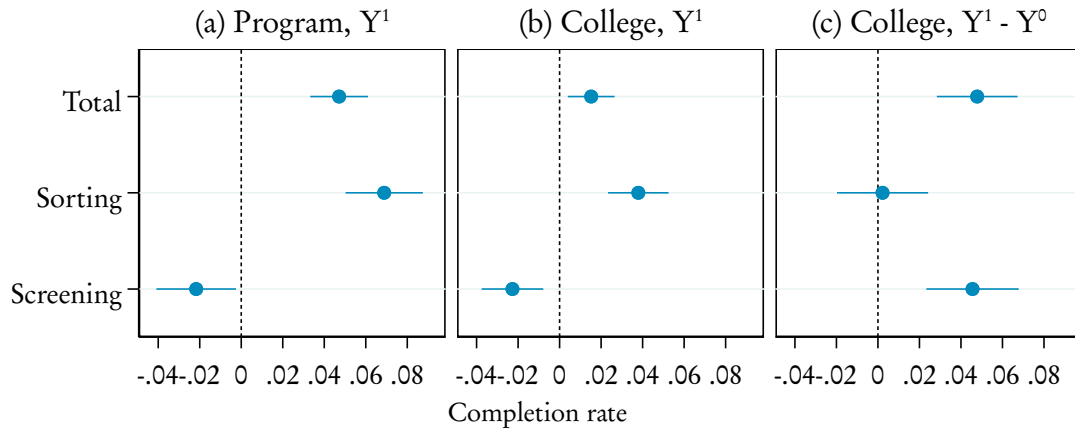
Figure 8 reports the expected performance when admitted for these different applicants and admission regimes. The left panel reports the estimates for program completion. The first estimate shows that the total effect of alternative evaluation compared to GPA-based admission is in the neighborhood of 5 percentage points.²⁴ This differs from the 2 percentage point completion gap between GPA-based admission and alternative evaluation in Table 4. The reason is that a non-negligible share of the marginal applicants in Q1 consists of Q2 applicants who will be admitted in either way.

The next estimates show that the higher program-completion rate under alternative evaluation is completely explained by sorting, and that the screening differential is even

²²For applicants who do not apply to Quota 2 we set $Z_{2i} = 0$ and $r_{2i} = 0$.

²³An additional advantage of our specification is that Q1 and Q2 applicants are compared within programs. We therefore avoid comparing Q2 and Q1 across programs with different levels of Q2 intake, which may be correlated with unobserved program and applicant characteristics. The specification is thus similar in spirit to a difference-in-difference design.

²⁴This is higher than the corresponding estimate in Table 4 which compared all Q1 applicants to all Q2 applicants which are overlapping groups. In the decomposition here we are comparing the Q2 applicants to those who only applied to Q1.



Note: Decomposition estimates of Equation (7) along with 95-percent confidence intervals. Value added is computed by replacing $A_i Y_i$ with Y_i as the dependent variable in (7). We cluster standard errors on the applicant level. Appendix figure A7 shows the corresponding figure for completion in time.

Figure 8. The impact of alternative evaluation of college applications on program and college completion. Total effect and mechanisms (sorting and screening).

negative. The middle panel of Figure 8 reports similar estimates for college completion. Here we see that on the margin college completion rates are about 2 percentage points higher for students when admitted based on alternative evaluations rather than GPA. This is partially explained by large sorting effects of about 4 percentage points, which are again partly offset by a negative 2 percentage point screening disadvantage.

These results show that alternative evaluation of college applications outperforms GPA-based admission in raising program-completion rates at the margin of admission. This advantage is completely explained by sorting. We interpret these results as showing that the increased application cost induced by tests, interview, or other requirements leads to a pool of more motivated/qualified applicants. On average there is however no evidence that for a given pool of applicants these alternative sources of information perform better in singling out successful students than GPA. Moreover, the results for college completion show that alternative evaluation is much better at generating good program matches than good college matches.

For college completion we also observe positive sorting effects, but alternative criteria actually perform worse than GPA in picking good college matches when keeping the application pool fixed. This is consistent with programs designing the alternative admissions to maximize their own outcomes while paying less attention to how well their admission process predicts collegiate performance elsewhere. The final right panel of Figure 8, which reports the estimates for college completion value added ($Y^1 - Y^0$), shows, however, that admitting the marginal applicant based on alternative evaluations rather than GPA increases their college completion rates by 6 percentage points. This is

because Q2 applicants not only perform better than Q1-only applicants when admitted, but are also less likely to ultimately complete college if they are not admitted.

8.1 Predicting performance

Screening is directly related to the predictiveness of the running variable, GPA or the subjective ranking of the applicants. The literature on the validity of measures of academic preparedness such as the SAT or ACT typically quantifies the validity of a predictive measure for a population of interest using an R^2 , where a large R^2 reflects that a measure is predictive of the outcome of interest Y^1 . With a single measure of preparedness, r , the R^2 has a simple formulation:

$$R^2 = \beta^2 \sigma_r^2 / \sigma_{Y^1}^2,$$

where β is the slope from a simple linear regression of Y^1 on r (including a constant) and $\sigma_{Y^1}^2$ is the variance of the outcome. Validity measures depend on unbiased estimates of slopes and the variance of the predictor r and outcome for the population of interest. However, as Rothstein (2004) notes, estimates are often based on samples of *admitted and enrolled* applicants which might differ from the applicants more generally which introduces scope for bias. Even in the absence of bias there is an unresolved question, which is: to what extent these measures are informative for the marginal applicant? This set of applicants is arguably the policy-relevant set for program admission officers. Our design not only allows us to address selection bias, but also to estimate the relevant parameters for the marginal applicants (compliers in our 2SLS estimation).

Since we are looking at marginal applicants the focus shifts to β because at the margin r is fixed and σ_r^2 therefore zero. We are thus interested in determining the partial effect $\beta = \partial E[Y^1 | c, r] / \partial r$ where c stands for compliers (marginal applicants). From Abadie (2003); Hahn et al. (2001) we know that at the threshold we can identify counterfactuals of interest for the local compliers:

$$E[Y^1 | c, r = 0] = \frac{\lim_{r \downarrow 0} E[A \cdot Y | Z = 1, r] - \lim_{r \uparrow 0} E[A \cdot Y | Z = 0, r]}{\lim_{r \downarrow 0} E[A | Z = 1, r] - \lim_{r \uparrow 0} E[A | Z = 0, r]} \quad (8)$$

While $E[Y^1 | c, r]$ and its derivative are identified at the threshold when $r = 0$, away from the threshold we observe an unknown mixture of always-takers and compliers. To estimate $E[Y^1 | c, r]$ for $r > 0$ we therefore need to extrapolate both $E[A \cdot Y | z = 0, r]$ to estimate Y^1 for compliers as well as $E[A | z = 0, r]$ to estimate the share of compliers.

For example, in a linear potential outcome model and (partial) linear first-stage:²⁵

$$\begin{aligned} E[A \cdot Y \mid A, r] &= \delta_1 A + \delta_2 A \cdot r \\ E[A \mid Z, r] &= \pi_0 + \pi_1 Z + \pi_2 Z \cdot r + \pi_3 r \end{aligned}$$

we get

$$E[Y^1 \mid c, r] = \delta_1 + \delta_2 r$$

and corresponding partial effect:

$$\frac{\partial E[Y^1 \mid c, r]}{\partial r} = \delta_2.$$

The relevant slope for the compliers is therefore consistently estimated with two-stage least squares, where we instrument the interaction of the running variable and admission above the threshold with the corresponding interaction between the instrument and the running variable. We implement this approach in our model (6) above as follows

$$E[A_i Y_i \mid A_i, r_i, F E_i] = \delta_1 A_i + \delta_2 A_i r_i + B_{3,0}(r_i) + F E_i \quad (9)$$

and estimate (9) using 2SLS. This allows us to compare the slope for marginal applicants to the conventional estimate from a OLS regression which recovers the average slope for marginal applicants (compliers) and above-marginal ones (always takers). Before assessing the importance of accounting for the presence of such always takers, we briefly investigate the correlation between the the running variables for Q2 applicants. Table 5 presents correlations between the two running variables for Q2 applicants in the quadrants displayed in Figure 7 along with their marginals. The aggregate correlation is 0.49, implying that a higher GPA is associated with a higher ranking in Quota 2. Conditional on crossing at least one cutoff, the correlation between the running variables is very small at -0.01.

As a baseline we estimate the linear slope of the running variables among *admitted* students using standard OLS where we control for waiting list fixed effects. This is in line with the standard approach taken in the literature described by Rothstein (2004). Table 6 presents the results. Moving a decile in the ranking in Quota 1 is on average associated with an increase in completion of 1 percentage point. In Quota 2 the slopes

²⁵These give the following reduced form

$$E[A \cdot Y \mid z, r] = \delta_1 \pi_0 + \delta_1 \pi_1 Z + (\delta_1 \pi_3 + \delta_2 \pi_0) r + (\delta_1 \pi_2 + \delta_2 \pi_1) Z \cdot r + \delta_2 \pi_2 Z \cdot r^2 + \delta_2 \pi_3 r^2$$

Table 5. Correlation between GPA rank and Alternative-Evaluation ranking

	$Z_1 = 0$	$Z_1 = 1$	Total
$Z_2 = 1$	0.10	0.28	0.33
$Z_2 = 0$	0.35	0.06	0.31
Total	0.40	0.28	0.49
Conditional on receiving offer ($Z_1 = 1$ or $Z_2 = 1$)			-0.01
Conditional on admission			-0.03

Note: The table contains linear correlation coefficients between running variables in Q1 and Q2 for Q2 applicants conditional on threshold crossings and unconditionally.

Table 6. Predicting program completion among admitted, OLS & 2SLS estimates

	All admitted applicants			Marginal applicants		
	(1)	(2)	(3)	(4)	(5)	(6)
Percentile rank, Q1	0.098 (0.005)		0.107 (0.008)	0.340 (0.068)		0.114 (0.050)
Percentile rank, Q2 (rescaled)		0.296 (0.019)	0.299 (0.020)		0.198 (0.131)	0.094 (0.109)
Estimator	OLS	OLS	OLS	2SLS	2SLS	2SLS
Admission Quota	Q1	Q2	Q2	Q1	Q2	Q2

Note: Column 1 to 3 present results from OLS regressions of program completion on percentile ranking for the two quota. Columns 1 and 2 present estimates for Quota 1. Column 3 presents estimates for the Quota 2 sample where percentile ranks for Quota 1 and Quota 2 are included jointly. Columns 4 and 5 present estimated slopes of the running variable for compliers using 2SLS in the corresponding samples.

are three times larger, though less precisely estimated. As the ranks in Q1 and Q2 are largely orthogonal given admission, we see little change in slopes when estimating jointly in Q1 and Q2.

We then turn to our 2SLS-estimates of the slopes for marginal applicants. The slope in Quota 1 for marginal applicants is over three times larger than the OLS estimates. This suggests that the lack of predictiveness in Quota 1 is explained by the presence of always-takers, which in Quota 1 largely consist of highly ranked Quota 2 applicants. In Quota 2 we no longer find evidence that the subjective evaluation is predictive of completion as the slope for marginal applicants is smaller and highly insignificant. Lastly, we instrument the slopes for both rankings for Q2 applicants in column 6. Both coefficients are attenuated reflecting that, for the marginal applicants in Q2, the rankings in the two quotas are not orthogonal. These results show that applicants at the margin of admission are meaningfully different from always-takers, which highlights a potential limitation to the SAT-validity literature over and above those pointed out by [Rothstein](#)

(2004). As the subjective ranking in Quota 2 does not statistically predict completion, these results are consistent with our finding that sorting rather than screening is the main driver of the wedge between alternative evaluating and GPA-based admission.

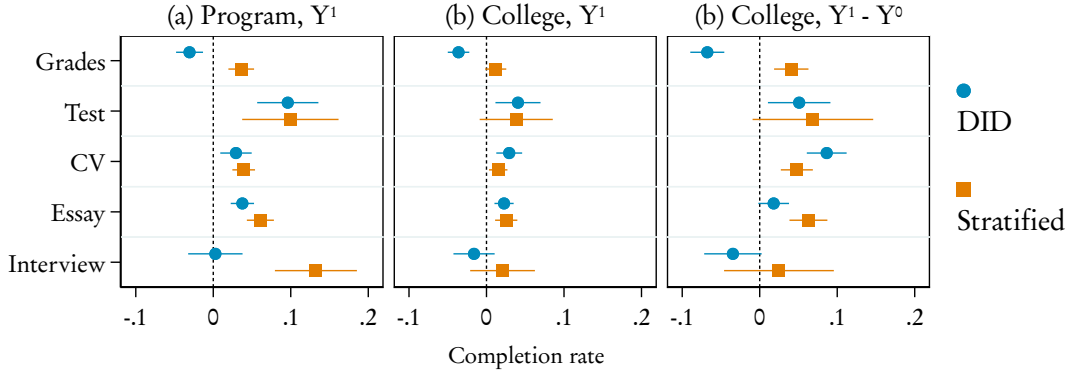
9 Evaluation technology

Alternative evaluation increases program and college completion compared to regular admission based on high-school GPA. Figure 2 above documented that most programs combine two or three criteria in their evaluations of applicants, and that programs differ in the extent to which they combine screening criteria. This raises the question whether the effectiveness of alternative evaluation depends on the type of information collected in the admission process. There are several reasons why one may expect such heterogeneity. First, different criteria could entail different application cost for prospective applicants and thus induce differential sorting. For example, the use of specific high-school grades would add little cost to applying, while a program-entry testing requirement will most likely demand a non-trivial time investment from applicants through preparation and sitting the test. Second, alternative criteria may vary in the extent to which they are able to identify potentially successful applicants and therefore have different screening effects.

Figure 2 documented that programs typically use multiple evaluation criteria. We observe in total 21 different combinations in our data.²⁶ This variation allows us to implement a difference-in-difference type approach to estimate the marginal contribution of a given criterion. Consider a simplified example where one program admits Q2 applicants based on Grades and CVs and a second program uses only Grades. Like above, a first within-program difference identifies the impact of alternative evaluation at the margin of admission. Assuming that the first-order effects of combining criteria is reasonably well approximated by a separable specification, we can then use a second difference between the two programs to estimate the marginal contribution of CVs.

Our final specification is therefore similar in spirit to a difference-in-difference approach (DID) but where the endogenous admission is instrumented with predicted admission based on threshold crossing. The corresponding second stage that performs the

²⁶Figure A5 shows that with the exception of Teaching which, in our sample, does not use interviews and tests in admissions, there is good support in the use of admission criteria across field-of-study.



Note: The figure shows estimated average outcome of admittance (Y^1) for marginal applicants along with 95-percent confidence intervals using 2SLS. Standard errors are clustered on applicant level. Models are estimated for program and college completion. Stratified models are estimated separately for programs using a given criteria. A program with more than one criteria can therefore enter into more than one sample. The separable model (10) is estimated on the full sample. Values are reported in appendix table A2.

Figure 9. Stratified vs. Marginal contribution of individual admission criteria – program & college completion

total decomposition reads as follows

$$\begin{aligned}
 E[A_i Y_i \mid D_i^c, A_i, Q_{2i}, Z_{1i}, r_{1i}, r_{2i}, F E_i] = & \\
 \sum_c [D_i^c = 1] \cdot (\delta_{c11} A_i + \delta_{c12} A_i Q_{2i} + \delta_{c22} (1 - Z_{1i}) A_i Q_{2i}) & \\
 + \alpha Q_{2i} + \gamma Z_{1i} Z_{2i} + B_{3,0}(r_{1i}) + Q_{2i} B_{3,0}(r_{1i}) + Q_{2i} B_{3,0}(r_{2i}) + F E_i & \quad (10)
 \end{aligned}$$

where the admission variables of interest A_i , $A_i Q_{2i}$ and $(1 - Z_{1i}) A_i Q_{2i}$ are now interacted with D_i^c , an indicator variable for the admission criteria, c , used in the secondary admissions of the program that i applied to, and where $c \in \{\text{Grades, Test, CV, Essay, Interview}\}$. The specification has quota-specific cubic splines in the running variables r_{1i} and r_{2i} in the two quotas. The fixed effects $F E_i$ are program times application year-specific fixed effects as before and standard errors are clustered at the applicant level. Programs chose criteria themselves the choice of criteria may be correlated with unobserved program characteristics. Our specification uses Quota 1 applicants within program-years to control for unobserved program-specific components shared across Q1 and Q2 applicants. For comparison we also present results of stratified regressions, where we select all programs using a given criteria.²⁷

Figure 9 reports the resulting estimates for program and college completion. First consider the estimates for program completion. On the margin the use of tests, CVs

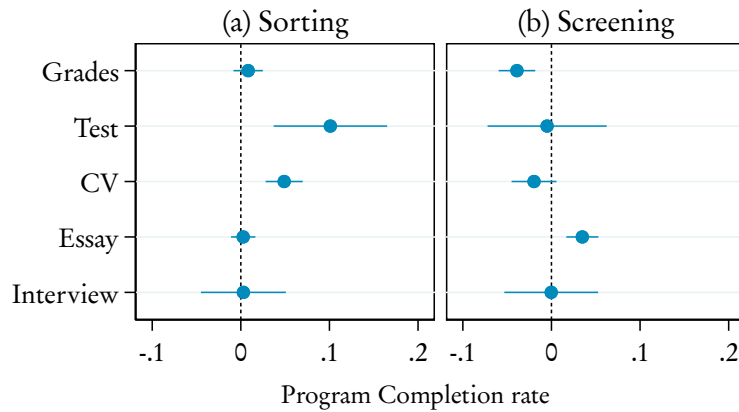
²⁷These stratified models disregard that criteria are used in combinations. Also note that the same program may enter into multiple sub-samples.

and Essays leads to higher program completion rates compared to admission based on high-school GPA. Estimates range from 12 percentage points for Tests to 3 percentage points for CVs. In contrast, the use of interviews on the other hand does not appear to lead to increased program completion and the use individual grades even appears to lower program completion rates compared to average high-school GPA. It is important to note here that basically none of the criteria are used in isolation, and these effects must therefore be interpreted as marginal effects in addition to the use of at least one other criteria. In the stratified specifications, interviews and grades appear to have positive effects, but this reflects that these criteria are used in combination with other criteria. Once we take these combinations into account in our DID-specification, interviews appear to add little, while the additional use of individual grades appears to have adverse effects. This latter negative impact can potentially be explained by the fact that there will be substantially more noise in a single grades than in GPA as the latter one is a weighted average of about 30 different grades. The difference in noise might therefore drown out the subject specific signal contained in a single grade. While individual grades do worse than GPA, the additional use of other criteria leads to overall positive effects for program completion.

The DID model gives effect estimates for college completion that tend to be qualitatively similar to those at the program completion margin, but of a smaller magnitude. As a consequence, the use of additional criteria does not compensate for the negative use of individual grades, explaining why alternative evaluation does not do better than GPA-based admission for college completion. The negative impact of using single grades in admission is amplified for college completion value added, while the use of tests and CVs stand out positively.

To understand how the use of different evaluative criteria affect the outcomes of the admitted applicant pool we can decompose their impact on sorting and screening. Figure 10 reports estimates of the decomposition of (7) that quantifies the sorting and screening channels for the DID model (10) as well as the stratified models. On the margin we do not find positive sorting effects for the use of grades, essays and interviews, while the use of CVs and tests in admission lead to stronger applicant pools. The use of individual grades is arguably low cost for applicants and one would not expect this to induce much sorting, the absence of sorting for the use of essays and interviews is harder to understand. We find a negative screening estimate for grades which is consistent with the explanation offered above that grades tend to be noisier than GPA. Only essays appear to outperform GPA with respect to screening, although the point estimate for tests is rather imprecisely estimated.

The estimates of the marginal contributions of individual criteria rely on the sep-



Note: The figure shows estimated average program completion rate when admitted (Y^1) for marginal applicants following the decomposition of Equation (7) along with 95-percent confidence intervals following equation (10). Standard errors are clustered on applicant level. Values are reported in appendix table A2

Figure 10. Marginal contribution to screening + sorting of individual admission criteria – Program completion

arable specification in (10) which rules out interaction effects and also assumes homogeneity. To assess the robustness of these estimates to violations of the homogeneity assumption we re-estimated the DID model on subsets of programs that do not make use of a given criteria. We for example drop all programs that use individual grades in the alternative evaluation of applicants and re-estimate (10). While we then no longer can estimate the marginal contribution of grades, the support in our data allows us to recover the marginal contribution of the other criteria. We repeat this for all criteria.

The results from this exercise for the total effect estimates in Figure 9 are reported in the Appendix Figure A8 for program completion and A9 for college completion, show that the estimates are broadly robust to these sample changes. If anything, some of the estimates that are already statistically different from zero tend to become somewhat larger. These results are therefore consistent with homogeneity being a reasonable approximation in our data.

This does however not rule out the possibility of interaction effects, namely that the impact of the combined use of different criteria is different from the sum of its parts. When two criteria are complements the interaction effect is positive, and when they are substitutes the interaction effect is negative. To investigate this possibility we estimated an extended version of equation (10) that not only includes the five main criteria effects, but also their interactions and report the estimates in appendix tables A4 and A5.

With a few exceptions the main effects in the interacted model are broadly consistent with the estimates from the simple separable specification (10). Overall we do not find evidence that the use of individual grades contributes positively to the admission of

successful students to the program, while the other point estimates tend to be larger but more noisily estimated. We continue to find that the use of grades does not induce sorting and has a negative but now statistically insignificant point estimate for screening. More generally we do not find evidence for screening for the other criteria either, and the main difference between the separable and interacted model is that we now find that the use of essays had a positive sorting effect.

The estimates from the interacted model generally do not provide systematic evidence for interaction effects (see appendix tables [A4](#) and [A5](#)). For the overall effects only the interactions between tests and grades and tests and interviews are statistically significant, but this finding is not repeated when decomposing these effects into sorting and screening. In fact, for screening none of the interactions are statistically significant, while for sorting two interactions are significant but they differ for program completion and college completion.

In summary, compared to GPA the use of individual grades lowers the program completion rates of admitted applicants. We find however robust effects for the use of tests, CVs and essays which are explained by their impact on sorting and not by an ability to select more successful students from a given pool of applicants. We do not find that interviews are an effective tool in admission.

10 Conclusion

This paper asks the question how different admission criteria perform in selecting more academically successful applicants. Building on Danish administrative data and features of the admission system we find that using alternative criteria such as essays, CVs, interviews, tests or single grades in admission identifies more successful program applicants than standard GPA-based admission. We do not find evidence that alternative evaluation outperforms GPA-based admission at the college completion margin. Alternative criteria are therefore more effective in identifying program matches but not necessarily college matches. One open issue is the relative importance of sorting as opposed to screening when changing the modality of admission. We find that most of the impact of alternative admission is due to sorting, and that the use of other information is not better at singling out successful students. Unpacking the evaluation technology shows that the use of individual grades leads to the admission of applicants that are less likely to succeed. The use of tests, CVs and essays does however have robust effects which are explained by their impact on sorting and not because they allow programs to select more successful students from a given pool of applicants. There is no evidence that interviews are an effective admission tool.

In our context these findings imply that completion rates can be increased by shifting admission capacity from primary admissions based on GPA to secondary admissions based on alternative criteria but that the information gained through secondary admission is of little use in selecting applicant who are likely to complete. Rather, our results indicate that applicants self-select due to application costs. Our results also point the way to additional areas of investigation which we leave for further research. This paper studies the supply-side of admission, and even though alternative evaluations are found to increase value added on average at the margin of admission, there may still be winners and losers. A first question therefore, concerns how different admission criteria affect the composition of the admitted students in terms of gender, race or socio-economic background. In other words, how does it affects *who* gets admitted?

A second question is the importance of selection criteria for outcomes beyond program and college completion, such as labor market performance.

A third question relates to extrapolation. While our results are at the margin of admission and therefore directly policy relevant, one may not only wonder how changes in quotas sizes affect overall outcomes as one moves away from the application thresholds, but also what happens when people start to change application behavior. This can both have repercussions for program level outcomes as well as distributional implications at the individual level.

A final question concerns the design of alternative evaluations. Our results show that sorting is the main mechanism why alternative evaluations outperform regular GPA-based selection of applicants, which raises the question how to optimally design the evaluation portfolio and whether there are more cost-effective ways to achieve similar outcomes.

References

- Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113(2):231–263.
- Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American economic review*, 93(3):729–747.
- Andrews, R. J., Ranchhod, V., and Sathy, V. (2010). Estimating the responsiveness of college applications to the likelihood of acceptance and financial assistance: Evidence from Texas. *Economics of Education Review*, 29(1):104–115.
- Arcidiacono, P., Kinsler, J., and Ransom, T. (2020). Asian American discrimination in Harvard admissions. NBER Working Paper No. 27068, National Bureau of Economic Research.

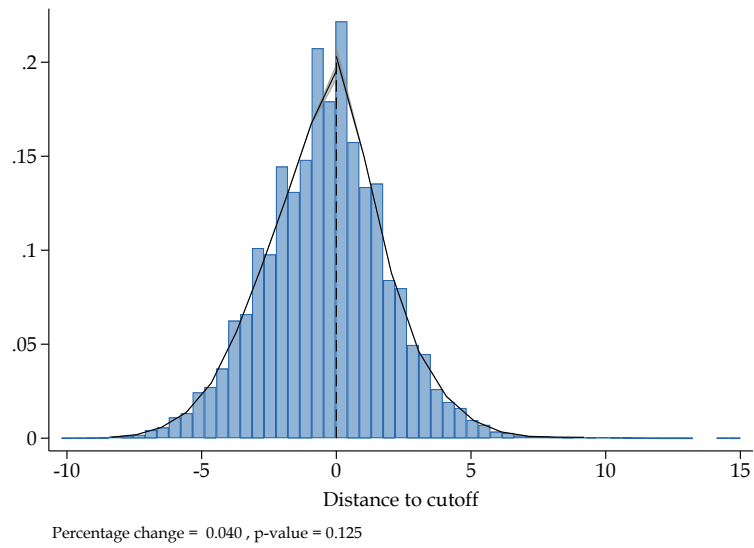
- Arcidiacono, P. and Lovenheim, M. (2016). Affirmative action and the quality-fit trade-off. *Journal of Economic Literature*, 54(1):3–51.
- Arcidiacono, P., Lovenheim, M., and Zhu, M. (2015). Affirmative action in undergraduate education. *Annual Review of Economics*, 7(1):487–518.
- Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, 344(6186):843–851.
- Belasco, A. S., Rosinger, K. O., and Hearn, J. C. (2015). The test-optional movement at Americas selective liberal arts colleges: A boon for equity or something else? *Educational Evaluation and Policy Analysis*, 37(2):206–223.
- Bettinger, E. P., Evans, B. J., and Pope, D. G. (2013). Improving college performance and retention the easy way: Unpacking the ACT exam. *American Economic Journal: Economic Policy*, 5(2):26–52.
- Black, S. E., Denning, J. T., and Rothstein, J. (2020). Winners and losers? The effect of gaining and losing access to selective colleges on education and labor market outcomes. NBER Working Paper No. 26821, National Bureau of Economic Research.
- Burton, N. W. and Ramist, L. (2001). Predicting success in college: SAT studies of classes graduating since 1980. Research report no. 2001-2, College Entrance Examination Board.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2):372–404.
- Dillon, E. W. and Smith, J. A. (2020). The consequences of academic match between students and colleges. *Journal of Human Resources*, 55(3):767–808.
- Dynarski, S., Nurshatayeva, A., Page, L. C., and Scott-Clayton, J. (2022a). Addressing non-financial barriers to college access and success: Evidence and policy implications. NBER Working Paper No. 30054, National Bureau of Economic Research.
- Dynarski, S., Page, L. C., and Scott-Clayton, J. (2022b). College costs, financial aid, and student decisions. NBER Working Paper No. 30275, National Bureau of Economic Research.
- Gandil, M. and Leuven, E. (2022). College admission criteria and inequality in access to education. Mimeo, University of Oslo.
- Goho, J. and Blackman, A. (2006). The effectiveness of academic admission interviews: an exploratory meta-analysis. *Medical Teacher*, 28(4):335–340.
- Grosz, M. (2021). Admissions policies, cohort composition, and academic success: Evidence from California. *Journal of Human Resources*, pages 0820–11131R3.
- Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

- Hoxby, C. and Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. Research discussion paper 12-014, Stanford Institute for Economic Policy.
- Hurwitz, M., Smith, J., Niu, S., and Howell, J. (2015). The maine question: How is 4-year college enrollment affected by mandatory college entrance exams? *Educational Evaluation and Policy Analysis*, 37(1):138–159.
- Hyman, J. (2017). ACT for all: The effect of mandatory college entrance exams on postsecondary attainment and choice. *Education Finance and Policy*, 12(3):281–311.
- Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111.
- Kuncel, N. R., Kochevar, R. J., and Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment*, 22(1):101–107.
- Murphy, S. C., Klieger, D. M., Borneman, M. J., and Kuncel, N. R. (2009). The predictive power of personal statements in admissions: A meta-analysis and cautionary tale. *College and University*, 84(4):83.
- Öckert, B. (2001). *Effects of higher education and the role of admission selection*. PhD thesis, Stockholm University, Faculty of Social Sciences, The Swedish Institute for Social Research (SOFI).
- Öckert, B. (2010). What's the value of an acceptance letter? Using admissions data to estimate the return to college. *Economics of Education Review*, 29(4):504–516.
- OECD (2012). *Equity and quality in education: Supporting disadvantaged students and schools*. Organisation for Economic Co-operation and Development, Paris.
- Pallais, A. (2015). Small differences that matter: Mistakes in applying to college. *Journal of Labor Economics*, 33(2):493–520.
- Rothstein, J. (2022). Qualitative information in undergraduate admissions: A pilot study of letters of recommendation. *Economics of Education Review*, 89:102285.
- Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics*, 121(1-2):297–317.
- Silva, P. L. (2022). Specialists or all-rounders: How best to select university students? Iza discussion paper no. 15271, IZA, Bonn.
- Smith, J., Hurwitz, M., and Howell, J. (2015). Screening mechanisms and student responses in the college market. *Economics of Education Review*, 44:17–28.
- Thomsen, J.-P. (2018). Test-based admission to selective universities: A lever for first-generation students or a safety net for the professional classes? *Sociology*, 52(2):333–350.
- UFM (2020). Evaluering af optagelsessystemet til de videregående uddannelser. Tech-

nical report, Ministry of Higher Education and Science.

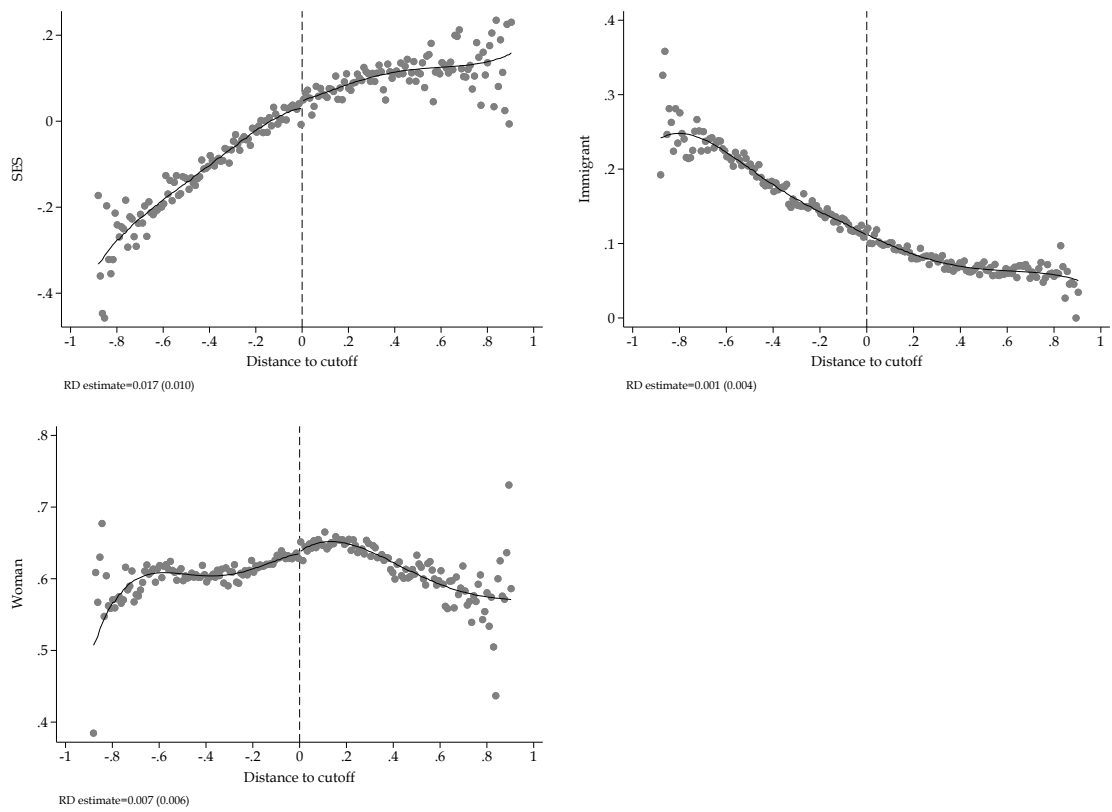
Zwick, R. (2007). College admission testing. Report, National Association for College Admission Counseling.

A Appendix Tables and Figures



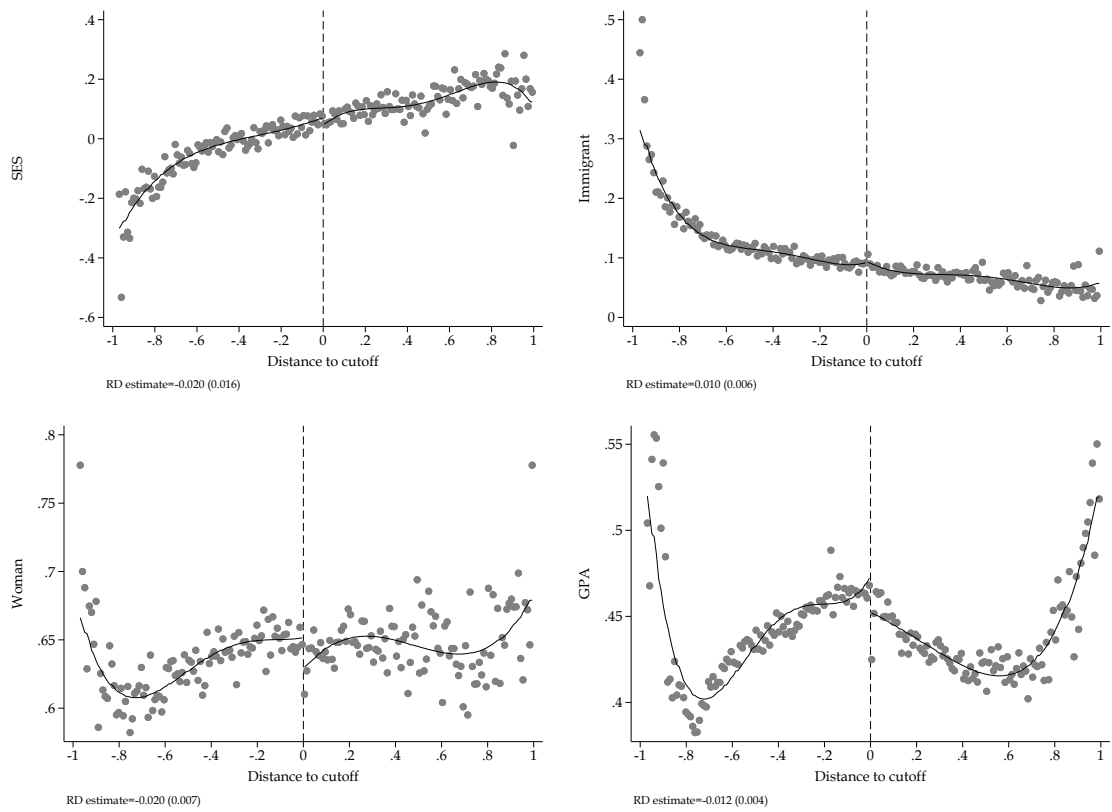
Note: Density is estimated using `-rddensity-` in Stata. A corresponding plot for Quota 2 is not provided as we do not observe the latent Q2 score but solely the ranking which is uniformly distributed by nature.

Figure A1. Bunching Q1: Density of centered-GPA around the Q1 cutoff



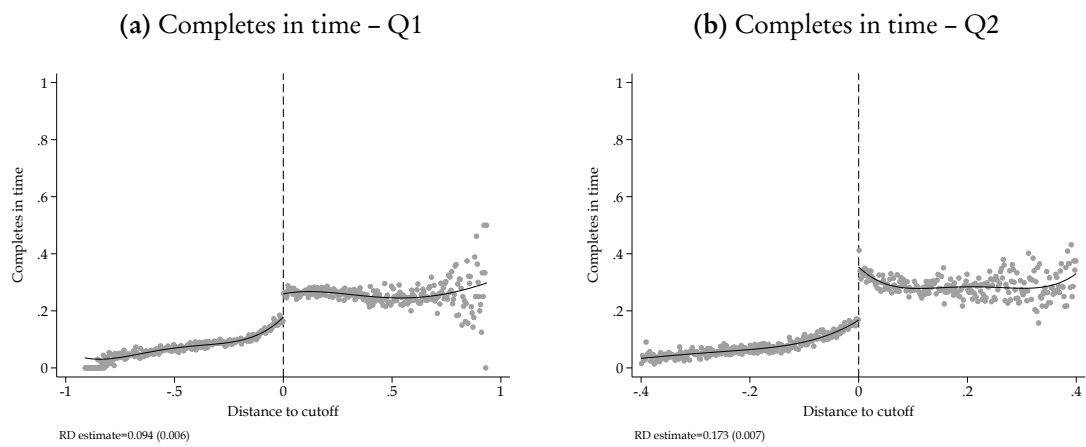
Note: Figures contain balancing checks in Quota 1. Point estimates along with standard errors are presented below the graphs. Graphs and estimates are constructed using rdplot and rdrobust packages in Stata. Only bins with more than 20 individuals are shown.

Figure A2. Balancing - Q1



Note: Figures contain balancing checks in Quota 2. Point estimates along with standard errors are presented below the graphs. Graphs and estimates are constructed using `rdplot` and `rdrobust` packages in Stata. Only bins with more than 20 individuals are shown.

Figure A3. Balancing - Q2



Note: Figures contain first stages and reduced forms in Quota 1 and 2. Point estimates of value added, $Y^1 - Y^0$, with standard errors are presented below the graphs. Graphs and estimates are constructed using `-rdplot-` and `-rdrobust-` packages in Stata.

Figure A4. Threshold crossing in Q1 and Q2

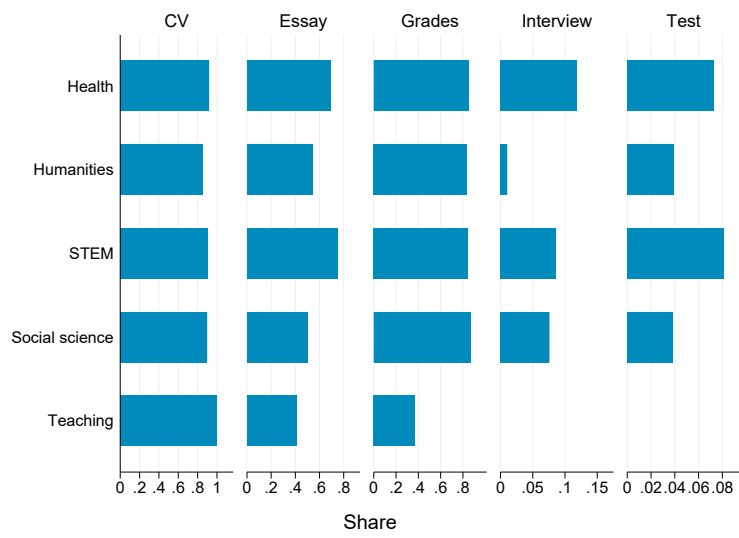
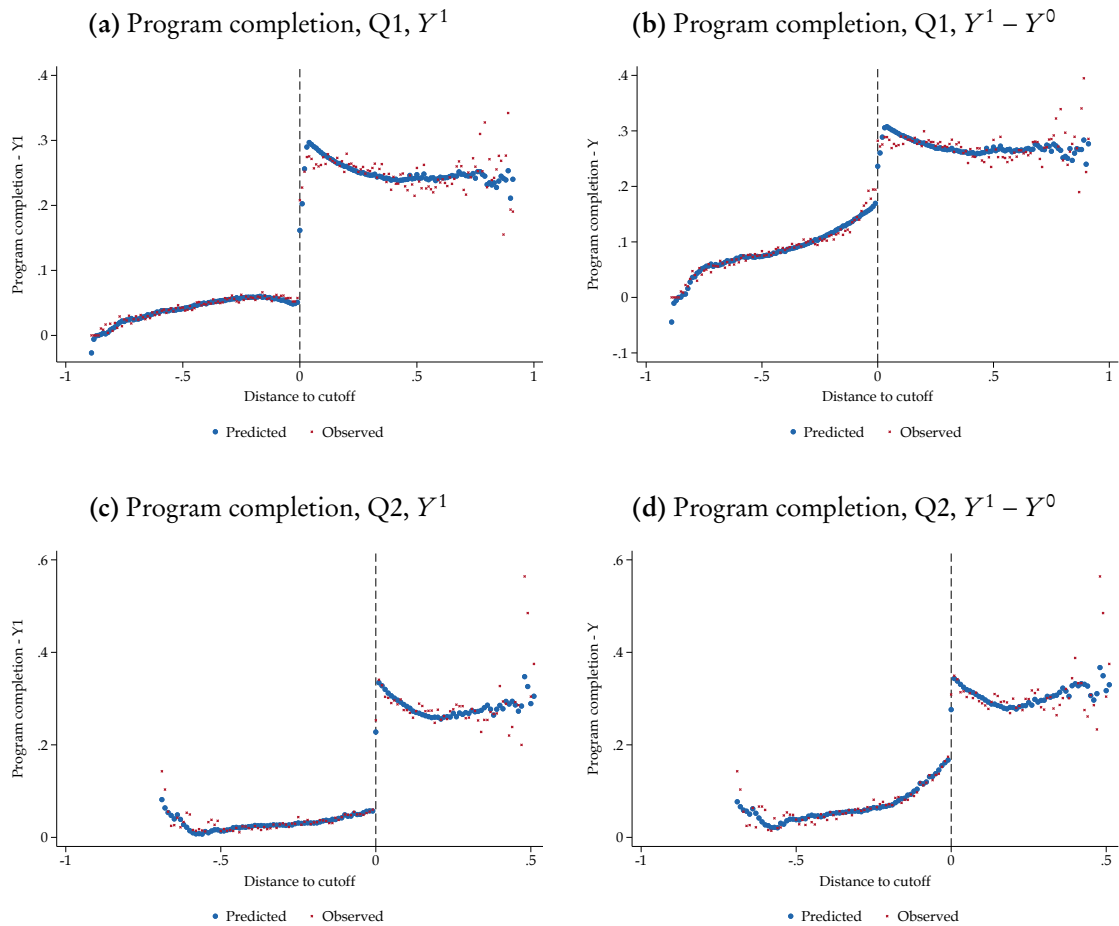


Figure A5. Evaluation criteria of Q2 applications by program field-of-study



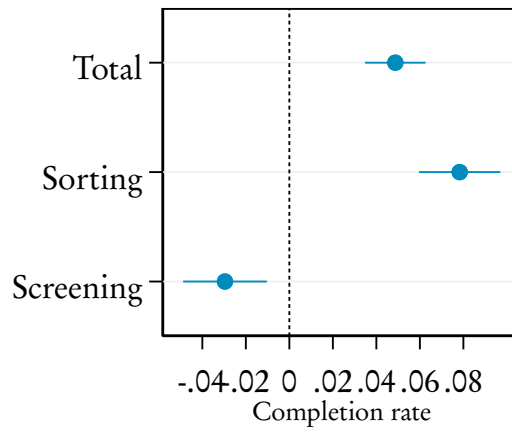
Note: The figures plot predicted values using the reduced forms of our 2SLS models and compares them to means within corresponding bins or realized outcomes. Model fit is close indicating that the inclusion of cubic splines and program-year fixed effects are sufficient to model the non-linearities of outcomes in the running variables.

Figure A6. Fit of reduced form compared to non-parametric RDD

Table A1. Potential program, college completion rates, and “completion in time” (Y^1) and value added ($Y^1 - Y^0$) for marginal applications – GPA-based admission vs. Alternative Evaluation

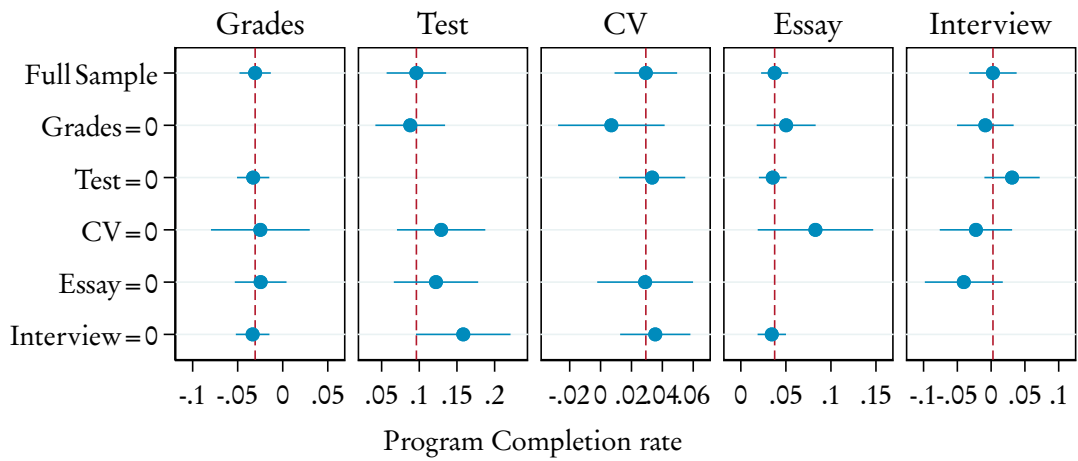
Admission	Program, Y^1			Program, $Y^1 - Y^0$		
- GPA-based (Q1)	0.517 (0.002)	0.498 (0.006)	0.490 (0.009)	0.462 (0.002)	0.193 (0.009)	0.186 (0.015)
- Alternative (Q2)	0.523 (0.002)	0.517 (0.006)	0.520 (0.007)	0.464 (0.003)	0.331 (0.008)	0.310 (0.011)
Q2 vs. Q1	0.006 (0.002)	0.019 (0.009)	0.030 (0.012)	0.002 (0.002)	0.138 (0.013)	0.124 (0.021)
	College, Y^1			College, $Y^1 - Y^0$		
- GPA-based (Q1)	0.820 (0.001)	0.805 (0.005)	0.801 (0.004)	0.027 (0.001)	0.025 (0.010)	0.021 (0.009)
- Alternative (Q2)	0.826 (0.002)	0.806 (0.005)	0.810 (0.005)	0.027 (0.002)	0.068 (0.008)	0.063 (0.012)
Q2 vs. Q1	0.006 (0.001)	0.001 (0.007)	0.009 (0.007)	-0.000 (0.002)	0.043 (0.013)	0.041 (0.017)
	Completes on time, Y^1			Completes on time, $Y^1 - Y^0$		
- GPA-based (Q1)	0.492 (0.002)	0.470 (0.006)	0.464 (0.008)	0.447 (0.002)	0.218 (0.009)	0.217 (0.014)
- Alternative (Q2)	0.502 (0.002)	0.491 (0.006)	0.494 (0.007)	0.452 (0.003)	0.334 (0.008)	0.315 (0.009)
Q2 vs. Q1	0.010 (0.002)	0.021 (0.009)	0.030 (0.011)	0.005 (0.002)	0.115 (0.012)	0.098 (0.017)
Estimator	OLS	2SLS	RD	OLS	2SLS	RD

Note: The table contains estimated average outcome of admittance (Y^1) with standard errors in parentheses. Standard errors are clustered on applicant level. Models are estimated for program and college completion, and “completion in time” using OLS, 2SLS and fuzzy RDD estimated in Stata using the rdrobust-package.



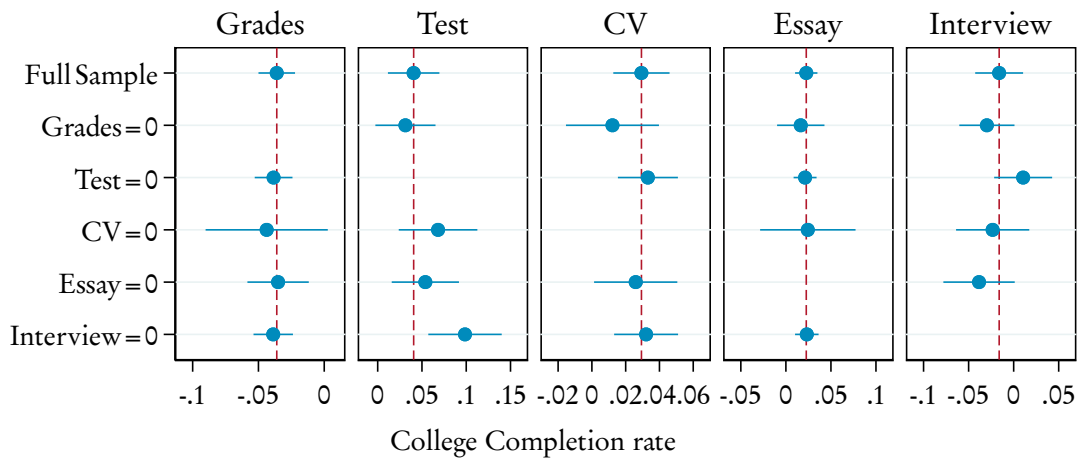
Note: Decomposition estimates of Equation (7) along with 95-percent confidence intervals. Value added is estimated by replacing $A_i Y_i$ with Y_i as the dependent variable in (7). We cluster standard errors at the applicant level.

Figure A7. The impact of alternative evaluation of college applications on the probability of completion in time. Total effect and mechanisms (sorting and screening).



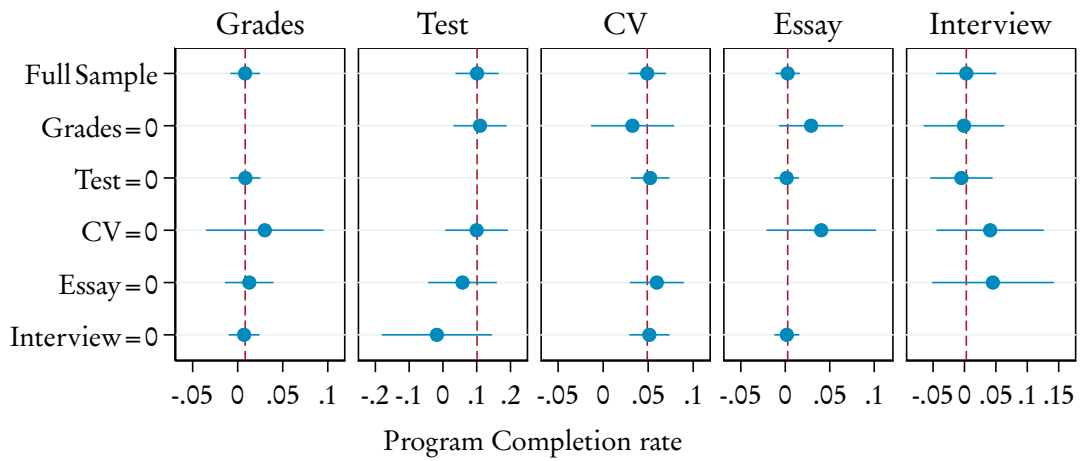
Note: The figure present estimates the marginal contributions of individual criteria to program completion as modeled in equation 10 estimated on subsets of programs that do not make use of a given criteria. 95-percent confidence intervals are computed using standard errors clustered at the applicant level.

Figure A8. Sample robustness checks marginal contribution of individual admission criteria - Program completion



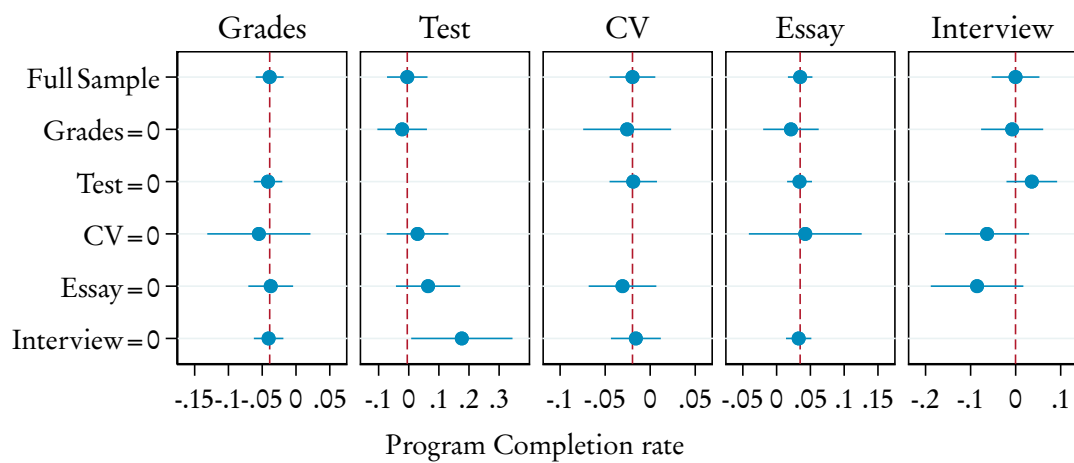
Note: The figure present estimates the marginal contributions of individual criteria to college completion as modeled in equation 10 estimated on subsets of programs that do not make use of a given criteria. 95-percent confidence intervals are computed using standard errors clustered at the applicant level.

Figure A9. Sample robustness checks marginal contribution of individual admission criteria - College completion



Note: The figure present estimates the marginal contributions of individual criteria to sorting for program completion as modeled in equation 10 estimated on subsets of programs that do not make use of a given criteria. 95-percent confidence intervals are computed using standard errors clustered at the applicant level.

Figure A10. Sample robustness checks marginal contribution of individual admission criteria - Sorting, Program completion



Note: The figure present estimates the marginal contributions of individual criteria to screening for program completion as modeled in equation 10 estimated on subsets of programs that do not make use of a given criteria. 95-percent confidence intervals are computed using standard errors clustered at the applicant level.

Figure A11. Sample robustness checks marginal contribution of individual admission criteria - Screening, Program completion

Table A2. Screening technology, Program completion Y^1

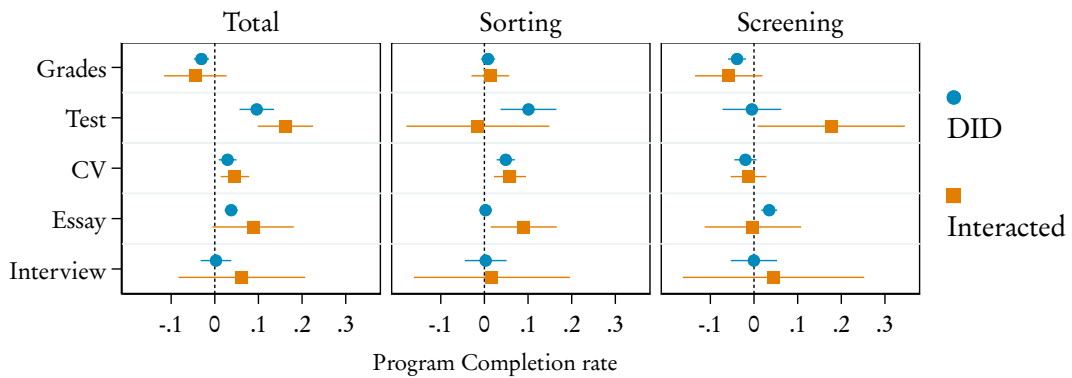
	Sorting			Screening			Total		
	Strat.	DID	DID+Int.	Strat.	DID	DID+Int.	Strat.	DID	DID+Int.
Grades	0.067 (0.011)	0.008 (0.008)	0.013 (0.022)	-0.031 (0.011)	-0.039 (0.011)	-0.058 (0.039)	0.036 (0.008)	-0.031 (0.009)	-0.045 (0.036)
Test	0.136 (0.060)	0.101 (0.033)	-0.015 (0.084)	-0.037 (0.062)	-0.005 (0.034)	0.177 (0.086)	0.099 (0.032)	0.096 (0.020)	0.162 (0.032)
CV	0.064 (0.010)	0.049 (0.011)	0.059 (0.019)	-0.025 (0.010)	-0.020 (0.013)	-0.013 (0.021)	0.039 (0.007)	0.029 (0.010)	0.046 (0.017)
Essay	0.068 (0.013)	0.003 (0.007)	0.090 (0.039)	-0.008 (0.013)	0.035 (0.009)	-0.003 (0.056)	0.061 (0.009)	0.038 (0.008)	0.088 (0.048)
Interview	0.213 (0.052)	0.003 (0.024)	0.017 (0.091)	-0.081 (0.054)	-0.000 (0.027)	0.045 (0.106)	0.132 (0.027)	0.003 (0.018)	0.062 (0.074)

Note: The table present estimates the marginal contributions of individual criteria to program completion as illustrated in Figures 9 and 10. See figure texts for additional details.

Table A3. Screening technology, College completion Y^1

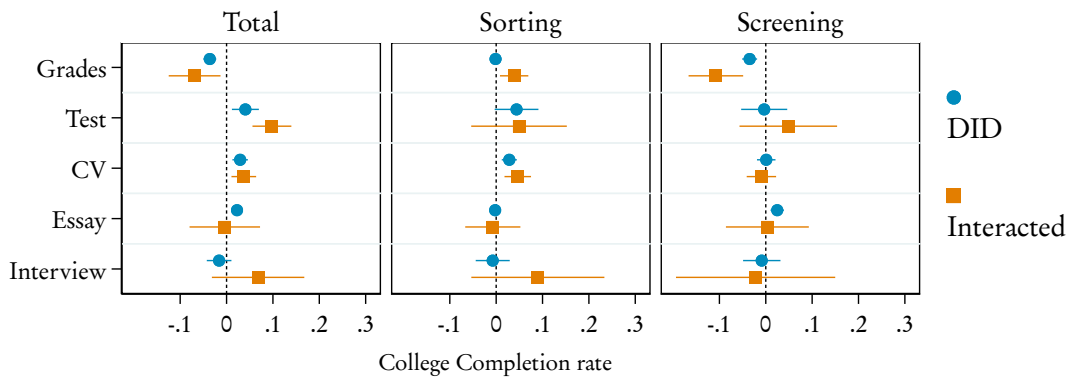
	Sorting			Screening			Total		
	Strat.	DID	DID+Int.	Strat.	DID	DID+Int.	Strat.	DID	DID+Int.
Grades	0.035 (0.008)	-0.001 (0.006)	0.039 (0.016)	-0.023 (0.009)	-0.035 (0.008)	-0.108 (0.030)	0.012 (0.007)	-0.036 (0.007)	-0.069 (0.028)
Test	0.019 (0.046)	0.044 (0.024)	0.049 (0.053)	0.019 (0.046)	-0.003 (0.025)	0.049 (0.054)	0.038 (0.024)	0.041 (0.015)	0.098 (0.021)
CV	0.036 (0.008)	0.028 (0.008)	0.047 (0.015)	-0.021 (0.008)	0.001 (0.010)	-0.009 (0.016)	0.015 (0.006)	0.029 (0.008)	0.037 (0.014)
Essay	0.030 (0.010)	-0.002 (0.005)	-0.007 (0.030)	-0.005 (0.010)	0.025 (0.007)	0.003 (0.046)	0.025 (0.007)	0.023 (0.006)	-0.004 (0.039)
Interview	0.008 (0.040)	-0.007 (0.019)	0.090 (0.073)	0.013 (0.041)	-0.009 (0.021)	-0.022 (0.088)	0.021 (0.021)	-0.016 (0.014)	0.068 (0.051)

Note: The table present estimates the marginal contributions of individual criteria to college completion as illustrated in Figures 9 and 10. Standard errors clustered at the applicant level in parentheses. See figure texts for additional details.



Note: The figure shows estimated average program completion rate when admitted (Y^1) for compliers along with 95-percent confidence intervals for the separable specification of equation (10) and an extended specification that also includes criteria interactions. For the latter model only main effects are shown in the figure and appendix tables A4 and A5 also report the interaction effects. Standard errors are clustered at the applicant level.

Figure A12. Marginal contribution of individual admission criteria to program completion rates (Y^1) – Separable and fully interacted model



Note: The figure shows estimated average college completion rate when admitted (Y^1) for compliers along with 95-percent confidence intervals for the separable specification of equation (10) and an extended specification that also includes criteria interactions. For the latter model only main effects are shown in the figure and appendix tables A4 and A5 also report the interaction effects. Standard errors are clustered at the applicant level.

Figure A13. Marginal contribution of individual admission criteria to college completion rates (Y^1) – Separable and fully interacted model

Table A4. Full interaction - Program completion Y^1

		Grades	Test	CV	Essay	Interview
Grades	Sorting	0.013 (0.022)	0.010 (0.086)	0.002 (0.027)	-0.026 (0.021)	-0.007 (0.056)
	Screening	-0.058 (0.039)	0.125 (0.092)	0.008 (0.042)	0.022 (0.025)	-0.026 (0.066)
	Total	-0.045 (0.036)	0.135 (0.064)	0.010 (0.037)	-0.003 (0.021)	-0.033 (0.051)
Test	Sorting		-0.015 (0.084)	-0.076 (0.083)	0.038 (0.135)	0.134 (0.124)
	Screening		0.177 (0.086)	0.053 (0.095)	-0.052 (0.159)	-0.273 (0.137)
	Total		0.162 (0.032)	-0.023 (0.070)	-0.014 (0.100)	-0.139 (0.081)
CV	Sorting			0.059 (0.019)	-0.074 (0.034)	0.011 (0.061)
	Screening			-0.013 (0.021)	0.019 (0.054)	-0.030 (0.074)
	Total			0.046 (0.017)	-0.055 (0.046)	-0.019 (0.055)
Essay	Sorting				0.090 (0.039)	-0.077 (0.070)
	Screening				-0.003 (0.056)	0.041 (0.078)
	Total				0.088 (0.048)	-0.036 (0.057)
Interview	Sorting					0.017 (0.091)
	Screening					0.045 (0.106)
	Total					0.062 (0.074)

Note: The table present estimates the marginal contributions of individual criteria to program completion as illustrated in Figures A12 from a fully interacted model. Standard errors clustered at the applicant level in parentheses. See figure texts for additional details.

Table A5. Full interaction - College completion Y^1

		Grades	Test	CV	Essay	Interview
Grades	Sorting	0.039 (0.016)	0.137 (0.068)	-0.048 (0.020)	-0.009 (0.016)	-0.024 (0.047)
	Screening	-0.108 (0.030)	-0.012 (0.072)	0.068 (0.032)	0.029 (0.019)	0.039 (0.055)
	Total	-0.069 (0.028)	0.125 (0.051)	0.020 (0.029)	0.019 (0.017)	0.015 (0.040)
Test	Sorting		0.049 (0.053)	-0.018 (0.067)	0.100 (0.107)	-0.107 (0.091)
	Screening		0.049 (0.054)	-0.004 (0.078)	-0.072 (0.126)	-0.058 (0.103)
	Total		0.098 (0.021)	-0.022 (0.053)	0.029 (0.077)	-0.165 (0.056)
CV	Sorting			0.047 (0.015)	0.009 (0.027)	-0.052 (0.050)
	Screening			-0.009 (0.016)	0.005 (0.044)	0.039 (0.062)
	Total			0.037 (0.014)	0.014 (0.038)	-0.013 (0.042)
Essay	Sorting				-0.007 (0.030)	-0.071 (0.051)
	Screening				0.003 (0.046)	0.022 (0.058)
	Total				-0.004 (0.039)	-0.049 (0.040)
Interview	Sorting					0.090 (0.073)
	Screening					-0.022 (0.088)
	Total					0.068 (0.051)

Note: The table present estimates the marginal contributions of individual criteria to program completion as illustrated in Figures A13 from a fully interacted model. Standard errors clustered at the applicant level in parentheses. See figure texts for additional details.