CALIFORNIA POLICY LAB

The Impact of Letters of Recommendation on UC Berkeley Admissions in the 2016-17 Cycle

JESSE ROTHSTEIN

JUNE 2017

The California Policy Lab builds better lives through data-driven policy. We are a project of the University of California, with sites at the Berkeley and Los Angeles campuses.

http://www.capolicylab.org

Executive Summary

In the admissions cycle that began in November 2016, UC Berkeley carried out the second year of a pilot experiment with letters of recommendation. Unlike other highly selective universities, Berkeley has never previously asked applicants to submit letters from teachers and guidance counselors. This may limit the information available for use in holistic review, and some at Berkeley think that as the university gets more selective it is getting harder to make informed decisions with the evidence available. Others, however, are concerned that students from disadvantaged backgrounds may not have access to adults who can write strong letters, and that the use of letters will further disadvantage these students.

In the pilot experiment, a subset of applicants was invited to submit letters of recommendation if they wished. Any submitted letters were incorporated into the "second read" evaluations of their applications. I evaluate the impact of this on the outcomes of applicants from four groups underrepresented among successful applicants to Berkeley: students from families with low incomes, students whose parents did not attend college, students from low-scoring high schools, and students from underrepresented racial and ethnic groups. I use a variety of methods, including a within-subject design that compares application scores when readers had access to letters with scores from a parallel process that suppressed the letters and a regression discontinuity design that exploits sharp distinctions between otherwise similar students in the selection of students to be invited to submit letters.

My study yields three primary results:

1) Many applicants from underrepresented groups did not submit letters when invited to do so. This was concentrated among applicants whose chances of admission were very low in any case, and was much more common in 2016 than in 2015. I have not been able to uncover the reason for the change over the previous year. The impact might have been to reduce the share of admitted students from underrepresented groups by no more than two

Contents

percentage points (or five percent), and likely by much less.

2) Readers able to review letters of recommendation gave higher scores, on average, to applications than did other readers scoring the same applications without access to letters. The impact was larger for applicants from underrepresented groups than for others, and was concentrated among those with moderate likelihoods of admission for whom a small boost might have been quite important. This implies that consideration of letters raised the share of underrepresented students among admissions from those who submitted letters. The effect was to raise the underrepresented share of all admitted students by as much as four percentage points. This is about twice as large as the negative potential effect of some students' failure to submit letters, implying a positive net effect of the use of letters on the underrepresented share of admitted students.

3) An evaluation of the overall impact of an invitation to submit letters that exploits a comparison between students who just crossed the threshold for an invitation and those who just missed it is inconclusive but also suggests a positive effect of the invitation on the relative enrollment of students from underrepresented groups.

There is support here both for the concerns of those who worry that disadvantaged students will not be able to obtain letters and for the hopes of those who believe that additional information will allow for fairer decisions. Although the evidence is not fully conclusive, I interpret the weight of the evidence to indicate that the use of letters did not reduce diversity and likely increased it. Nevertheless, any effect is small, and would be easily swamped by other policy decisions. One path forward might be to take advantage of the information that can be provided by letters while using other tools (e.g., expanded outreach to high schools) to ensure that there is no net negative effect on diversity.

This report was prepared quickly to support policy decisions being made in Summer 2017, and focuses exclusively on the impact of letters of recommendation on the demographic composition of the admitted class. A full evaluation of the use of letters depends as well on whether they provide useful information to support better admissions decisions within demographic groups. In future work, I plan to investigate this question by examining the performance at Berkeley of students admitted under the letters of recommendation pilot, and by testing whether there are particular profiles of students or of letters that are associated with greater impacts on admissions decisions or on the quality of these decisions.

I. Introduction

In the 2016-17 admissions cycle, as in the prior cycle, many applicants for freshman admission to UC Berkeley were invited to submit letters of recommendation (LORs) as part of their applications.

On its face, the use of LORs in admissions seems consistent with the spirit of holistic review. The idea of holistic review is to look beyond reductive summaries like SAT scores and examine the whole applicant, taking account of contextual factors and obstacles overcome. LORs have the potential to offer insight into aspects of the applicant not captured by the available quantitative information or by the essays that applicants submit.

This view presumes, however, that all applicants have access to letter writers who are willing and able to provide well-written letters that provide insight into the student's background and suitability for Berkeley. This may not be the case. In particular, there is a legitimate concern that applicants from under-resourced high schools may have trouble identifying potential writers, and that when they do identify writers the letters that are provided may be less informative or prejudicial against the applicant (due, e.g., to poor writing or grammar, or to lower status of the letter writer).

The impact of the new opportunity to submit LORs on the types of students who are admitted is thus an empirical question. This report represents a preliminary, independent evaluation of this question. It is conducted under the auspices of the California Policy Lab (CPL, www.capolicylab.org). CPL is a new project, launched by faculty members at UC Berkeley and UCLA, that fosters research partnerships between UC faculty and state and local government in California, aimed at using data the government already collects to generate evidence to inform better policy.

Section II describes the origin of the study and more carefully defines the scope of inquiry. Section III reviews the admissions process at Berkeley in 2016-17, and in particular the role of letters of recommendation within it. Section IV analyzes the evidence regarding disparities in applicants' ability to obtain letters. Section V turns to a distinct question: When applicants submit letters, does their inclusion in holistic review create advantages or disadvantages for students from disadvantaged backgrounds? To assess this I rely on an experiment conducted in the UCB admissions office in Spring 2017, in which applicants who had submitted letters were re-scored without their letters to identify the independent effect of letters on readers' assessments. Section VI examines the overall impact of the letter invitation combining both the likelihood of submitting letters and the impact that the letters might have had on the evaluation of the application.

II. Origin and scope of the study

I am a Berkeley faculty member with experience doing academic research on admissions at UC and elsewhere. I have never been directly involved with Berkeley undergraduate admissions processes, nor did I have any role in the decision to implement the letter of recommendation pilot.

In Summer 2016 I was asked by Vice Chancellor Catherine Koshland and the UC Berkeley Academic Senate Committee on Admissions, Enrollment, & Preparatory Education (AEPE) to conduct an

independent review of the impact of LORs in the then-just-completed 2015-6 admissions cycle. Given the timing, the resulting study[1] was purely retrospective. In Fall 2016, VC Koshland, Interim Executive Vice Chancellor and Provost Carol Christ, and Berkeley Faculty Senate Chair Robert Powell asked me to design a new prospective study, to be implemented in the 2016-17 cycle, that would shed further light on the impact of LORs on Berkeley admissions.

Applications submitted in November 2016 were reviewed as normal in December 2016 and early 2017. As discussed below, some applicants were invited to submit letters, and when they did those letters were made available to the second readers and incorporated, as the readers saw fit, into their application scores. Following initial admissions decisions in early February, a number of the more experienced external application readers were retained to continue reading applications in February and March. This additional wave of application evaluations did not count toward admissions decisions, and was conducted solely for the purpose of the current study. Campus funding was made available to pay the application readers for their work, and to retain a contract employee to coordinate the process.

Beyond this funding, the current study is independent of the campus administration. I am of course a member of the Berkeley Academic Senate, but I act here as an independent evaluator with no stake in the outcome. I have received no financial support from the campus for this study (beyond the payments to application readers) nor any instructions intended to influence my results. My personal interest is in designing an effective admissions process for UC Berkeley that treats all applicants fairly and gives each a chance to demonstrate his or her strengths. My impression is that all of the campus actors with whom I have interacted are, like me, motivated primarily by a desire to understand the true effects of LORs on Berkeley admissions.

Scope

This preliminary report investigates a single question:

> *Did the inclusion of LORs in the admissions process in the 2016 admissions cycle affect the relative admissions chances of Berkeley freshman applicants from any of four underrepresented groups (low-income students, students from low performing high schools, first-generation college students, and students from underrepresented racial and ethnic groups)?*

I break this question into two parts. First, do those invited to submit LORs actually submit them? Second, do those submitted letters influence readers' evaluations of the students' applications? In each case, I focus exclusively on California resident, non-athlete applicants. The net impact of the addition of letters of recommendation to the admissions process combines the two components. It is quite plausible that both could operate, even in opposite directions – that some students from disadvantaged backgrounds would be further disadvantaged because they are unable to obtain letters, but that others would be advantaged because they are able to obtain letters and this leads to improvements in readers' evaluations of their applications. In this case, letters might change *who* is admitted, within demographic categories, more than they change the distribution of admissions *across* categories.

My final analysis examines the net effect of letters, taking advantage of arbitrariness in the selection of applicants who were invited to submit letters and comparing those who were invited with otherwise very similar applicants who were not. This shows the net impact of the LOR request, combining the two channels. I present estimates of the impact on the share of disadvantaged applicants who are admitted, but in this analysis I cannot identify changes in who was admitted within the disadvantaged group.

In future research, I plan to dig deeper into the two separate channels, with the goal of understanding changes in the composition of admitted students within demographic groups. I will examine two additional questions:

1. What characteristics of applicants and of the LORs themselves determine whether an applicant benefitted or was harmed by the availability of LORs to readers?

2. Does the consideration of LORs lead to the admission of students who are more or less successful, once at Berkeley, than the students who they displace?

Limitations and caveats

There are several limitations to the analysis here, even relative to the narrowly defined research question defined above.

First, I investigate *only* the impact of LORs, not that of other aspects of the UCB admissions process. There are many other aspects of the process that influence the admissions rates of students from underrepresented groups, but these are out of scope for my study.

Second, the outcome measures available to me are relatively coarse: Readers score applications on a three-point scale (No / Possible / Yes), and admissions decisions are of course binary. The study does not shed light on whether LORs affect readers' assessments of applications within these coarse categories.

Third, I was not able to incorporate into this study data on UCB applicants' admissions outcomes at other UC campuses, as I obtained these data only days before the study was due. This limits the statistical power of the study. It also limits my ability to examine voluntary withdrawals from the UCB admissions process. Roughly half of students admitted to UCB do not enroll, and similarly half of students offered spots on the campus waitlist do not take them up. In each case, it is likely that many of these decisions reflect students who have chosen to attend other universities (including other University of California campuses) instead. Presumably, many students invited to submit LORs are also admitted elsewhere, and some may choose not to submit LORs because they have already made another choice (for example, due to an Early Decision that arrived between the original application submission in November and the letter request in December).

Finally, the study took place within the context of the actual 2016 admissions cycle. This is only the second year that readers had access to LORs for any applicants. Both this year and the previous year, the LORs were only available for a subset of applicants; it was not known until relatively late in the

process whether they would be available at all; and readers may not have been adequately trained in how to interpret LORs. Moreover, readers incorporated LORs into decisions that relied on a particular set of criteria used to evaluate applications. The impact of LORs might be different in an admissions process that used different criteria (e.g., putting more or less weight on traditional academic qualifications), or if readers were differently trained.

A related concern is that the results might be specific to the way that LORs were incorporated into the admissions process during the pilot years. In particular, some students who were unable to obtain letters on short notice, over a month after they thought their applications were finalized, might have been better able to identify letter writers if they (and their teachers and guidance counselors) had been given advance notice that this would be part of the application process. This consideration leads me to expect that estimates of the share of students who are able to obtain letters in the pilot year likely understate the share who would be able to obtain them outside of the pilot, perhaps particularly so for disadvantaged students who may have less access to helpful guidance counselors, who might be the only students in their high schools applying to UCB, and who may be less likely to be applying as well to private colleges and universities that also require letters.

Another implication of the way that LORs were used during the pilot years is that only the second readers of each application had access to the letters; first readers, by design, evaluated applications before letter requests were made. Admissions at Berkeley depend on both readers' scores, so the exclusion of letters from the first evaluation limits the potential impact on admissions. Letters might have larger impacts if they were incorporated into the regular admissions process and submitted at the same time as the rest of the application.

III. UC Berkeley admissions in 2016-17

I discuss here the admissions cycle that selected students who will begin their studies in August 2017 from among those who submitted applications in November 2016. Hereafter, I refer to this as the *2016* cycle, using the same nomenclature even for the parts of the cycle that occurred during calendar year 2017. I refer to the previous November 2015-August 2016 cycle as the *2015* cycle.

The admissions process

Students apply to the University of California in November, using a common application for all campuses. They are asked to provide essays, but no letters of recommendation.

In the 2015 and 2016 admissions cycles, every Berkeley application was evaluated by two separate readers. Each reader scored the applications on a three-point scale – Yes, Possible, or No. The first round of evaluation was conducted by a large pool of readers that included both regular UCB staff and outsiders retained to help with the process. The second round was conducted by a more carefully chosen pool of experienced readers, more of whom were internal staff. In each round, the campus used a holistic review process, where readers considered the whole application package rather than any particular numerical components in isolation. In the second round, readers had access to the first readers' scores and comments in addition to the application itself.

After both rounds of reading were complete, the files went to the central admissions staff for final admissions decisions. These decisions incorporated capacity constraints for the various undergraduate colleges – for example, the College of Engineering is much more competitive than the College of Letters and Science and for some majors. Within each division or major, decisions respected the rankings established by the two reader scores, with admissions staff breaking ties only within score groups. That is, all applicants to a division with two Yes scores were admitted before any students with just a single Yes. With only nine possible combinations of reader scores, however, there were inevitably ties. Tie-breaking decisions took into account a number of factors, such as applicant disadvantage (though not race/ethnicity) and faculty evaluations of applications.

Two aspects of this final selection process are particularly relevant for the current study. First, the only way that subjective elements of an application, such as an assessment of the letters of recommendation, can influence the admissions decision is through the reader scores – they are not used for tiebreaking within cells defined by the reader scores.

Second, the cells within which tiebreaking takes place are defined by both the first and the second readers' scores. Letters of recommendation were never available to first readers in the pilot. This limits the potential impact that letters might have had on admissions.

As an illustration, Table 1 shows the admissions rates for two illustrative cases, applicants to the College of Letters and Science and to the Mechanical Engineering major within the College of Engineering, for each of the nine possible combinations of first and second reader scores. In Letters and Science, nearly all applicants who receive Yes scores from either reader are admitted, regardless of the other reader's score. This means that a very strong letter might have helped an otherwise marginal applicant by raising the second reader's score to Yes. In contrast, a weak letter is unlikely to have hurt an otherwise strong applicant, as the first reader's Yes will have nearly ensured admission before the letter was even received. The situation is quite different in Mechanical Engineering, where to a first approximation only applicants who receive Yes scores from *both* readers are admitted. In this pool, a very strong positive letter that tips a second reader's score over from Possible to Yes is unlikely to help, since the first reader score will likely have already doomed the application, but a weak letter may well hurt an application that is otherwise very strong.

Quantitative summary of applicant pool and admissions outcomes

UC Berkeley is highly selective – only 21% of in-state, non-athlete applicants were admitted in 2015, and 19% in 2016. The 25th percentile SAT composite score among admitted students in 2015 was 2000, and the 10th percentile was 1775. These are the 93rd and 80th percentiles, respectively, of the national distribution. The SAT distributions for applicants and for admitted students in 2016 are somewhat lower, but remain extremely high.

High selectivity means that the great majority of applicants are not admitted. In some cases, students who are very highly qualified and observably quite similar to admitted students miss the cut. A large share of rejections, however, go to applicants with qualifications well outside the range of typical Berkeley admits.

Table 1. Admissions rates by first and second reader scores, selected colleges/majors.

		Reader 2 score		
		No	Possible	Yes
College of Letters and Science				
Reader 1 score	No	1%	16%	94%
	Possible	7%	38%	100%
	Yes	95%	97%	100%
Mechanical Engineering major				
Reader 1 score	No	0%	0%	
	Possible	0%	0%	1%
	Yes		2%	99%

Note: Cells show the share of applicants to the indicated college/major offered admission, either initially or off the waitlist. Samples include in-state, non-athlete applicants who did not withdraw their applications before the end of March. Applicants who failed to accept offered positions on the waitlist are excluded. Admissions rates are suppressed for cells with fewer than 20 applicants.

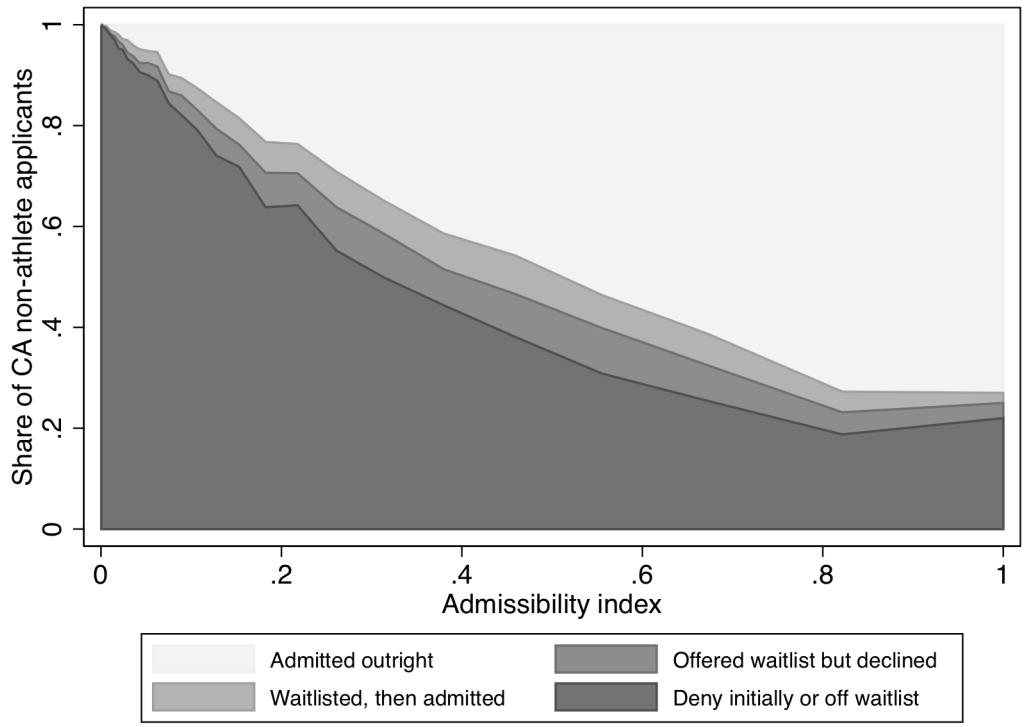
It is useful to have a simple summary measure of the strength of each application, considering the readily measurable factors that are quantified in the admissions office database. To do so, I estimated a logistic regression on 2015 applicant data. This is a statistical model that uses a set of explanatory variables to explain (in a statistical sense) a binary outcome, in this case an offer of admission. Explanatory variables included:

- the SAT score and high school GPA
- the average scores at the applicant's high school
- the applicant's course-taking choices relative to what was available at the high school
- the school's API rank (a measure of average scores)
- average parental income at the school
- an indicator for a school at which fewer than 5% of graduates apply to the UC
- the applicants' parents' education and income.

These include many, but of course not all, of the characteristics that are considered under holistic review. I used the regression model to generate a single predicted probability for each 2015 and 2016 applicant, corresponding to the probability that an applicant with that particular set of characteristics would have been admitted in the 2015 cycle. I refer to this predicted probability as the "admissibility index," or "AI."

The AI is a useful, if incomplete, measure of an application's strength. Of course, some applicants with very low AIs had stellar essays or other strengths that were visible to application readers but not captured by the variables listed above, and so were admitted; similarly, some applicants with high AIs underwhelmed in other ways and were not admitted. But these are unusual outcomes.

Figure 1. Admissions outcomes by admissibility index (AI), 2016.



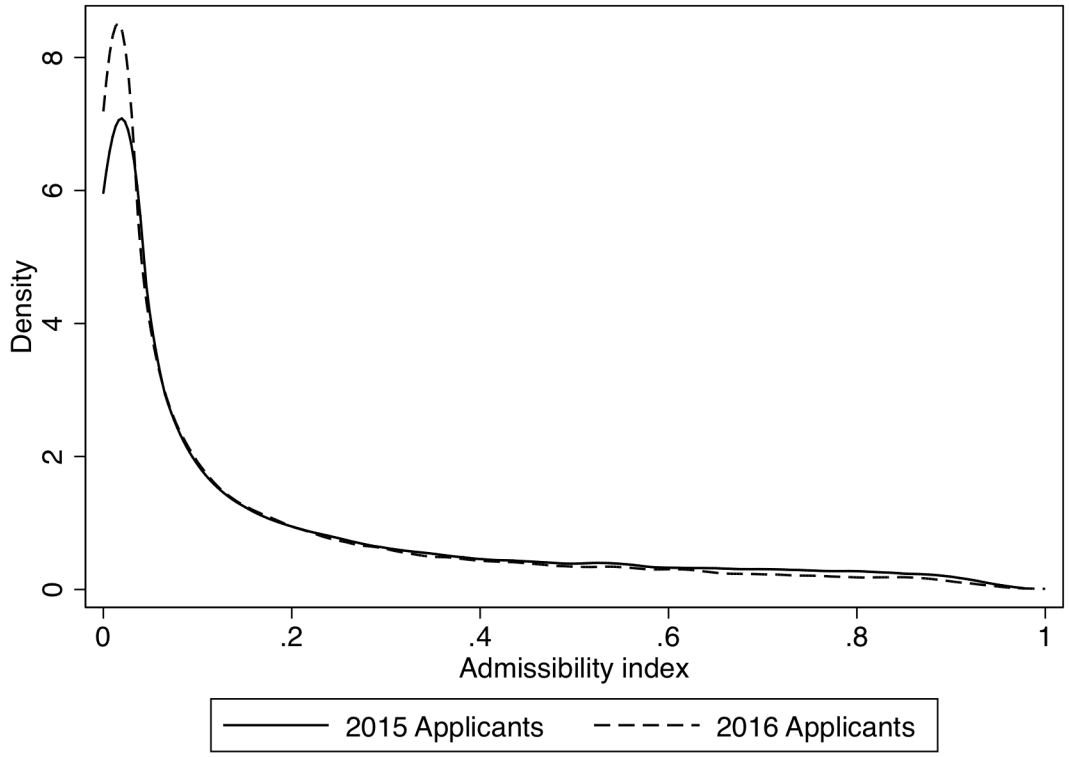
Note: Sample restricted to California resident, non-athlete applicants.

Figure 1 shows the distribution of admissions outcomes, following holistic review, for 2016 applicants at various points in the admissibility index distribution. In analyzing admissions outcomes, the waitlist presents complications: Some students are not admitted because they did not take an offered position on the waitlist, even though they would have been admitted eventually had they accepted it. I thus treat those admitted off the waitlist and those who decline positions on the waitlist as separate categories. (Those who agree to be on the waitlist but are not admitted are grouped with the initial denials.)

Figure 1 shows that applicants with very low admissibility scores were, not surprisingly, quite unlikely to be admitted, while both the probability of being admitted in the first round and the probability of being admitted at all rose nearly linearly with the admissibility index.

It is important to emphasize that this strong relationship does not indicate a failure of the holistic review process. Rather, the AI simply summarizes the outcomes of that process, and captures the quantifiable characteristics of students who typically did well or poorly in holistic review. The variation in outcomes among students with the same AIs reflects other factors that are captured by holistic review but not by the AI.

Figure 2. Distribution of admissibility index (AI), 2015 and 2016 applicants.



Note: Sample restricted to California resident, non-athlete applicants.

The distribution of the admissibility index among applicants is very highly skewed toward the lower end. The average 2016 applicant had an AI of 16%. This is similar to the overall admissions rate, not surprising since the selectivity of UCB admissions did not change dramatically between 2015 and 2016. Figure 2 shows the distribution of admissibility indices among all applicants in 2015 and 2016. It shows that there are a great many applicants with very low AIs in each year, with even more in 2016 than in 2015. Fully 49% of 2016 applicants had AIs below 5%. As Figure 1 shows, these applicants were quite unlikely to be admitted (though of course a very small share of them were admitted due to other strengths uncovered in the holistic review).

Because the very-low-AI applicants are such a large share of the applicant pool, overall summaries of the pool weight the experience of these applicants quite heavily. This can be misleading, as although this is a large group of applicants, it contributes a quite small share of admitted students – only 7% of students admitted in 2016 had AIs below 5%.

Table 2 summarizes the characteristics of the applicant pools in the two years. In addition to the 49% of 2016 applicants with admissibility indices below 5%, another quarter have indices between 6

Table 2. Applicant pool and outcomes (California resident, non-athletes only).

	2015	2016		
	All applicants	All applicants	Not underrepresented	Underrepresented
N	45,626	49,172	22,941	26,231
SAT score composite				
Mean	1898	1867	2019	1734
25th percentile	1690	1680	1890	1540
75th percentile	2140	2080	2170	1925
Admissibility index (AI)				
0-5%	46%	49%	41%	57%
6-10%	12%	12%	11%	13%
11-20%	12%	13%	13%	12%
21-50%	17%	16%	19%	13%
51-100%	13%	10%	15%	5%
Underrepresented groups				
Fee waiver	33%	33%	0%	62%
Low API	22%	22%	0%	42%
First generation	23%	24%	0%	44%
Race/ethnicity	34%	35%	0%	66%
Any	53%	53%	0%	100%
Letters of recommendation				
Invited	32%	36%	26%	45%
Requested	28%	27%	23%	31%
Submitted	27%	27%	23%	30%
Admissions outcomes				
Initial admit	18%	16%	20%	13%
Offered waitlist	9%	7%	8%	6%
Ever admitted	21%	19%	24%	15%
SIR	11%	9%	12%	7%

Notes: Sample restricted to California resident, non-athlete applicants. The admissibility index is the predicted probability of admission, based on 2015 admissions patterns and a long list of applicant characteristics including SAT, high school GPA, course-taking, demographics (excluding race/ethnicity), and effort relative to high school offerings.

and 20%. At the other end of the scale, only 10% have indices above 50%. The 2015 pool was roughly similar.

Table 2 also shows several measures of applicant disadvantage. In this study, I rely upon four proxies that are used by the Berkeley admissions office. An applicant is considered to come from an underrepresented group if he or she:

- Qualifies for a waiver of the application fee (based on low family income);
- Comes from a high school with an Academic Performance Index (API) decile rank of 5 or lower;
- Does not have a parent who attended college; or
- Is a member of an underrepresented racial or ethnic group.

I use the first three of these in constructing my admissibility index.[2]

53% of 2016 Berkeley applicants came from one of the four underrepresented groups (which of course overlap). Columns 3 and 4 of Table 2 show the characteristics of these applicants and of other applicants in 2016.

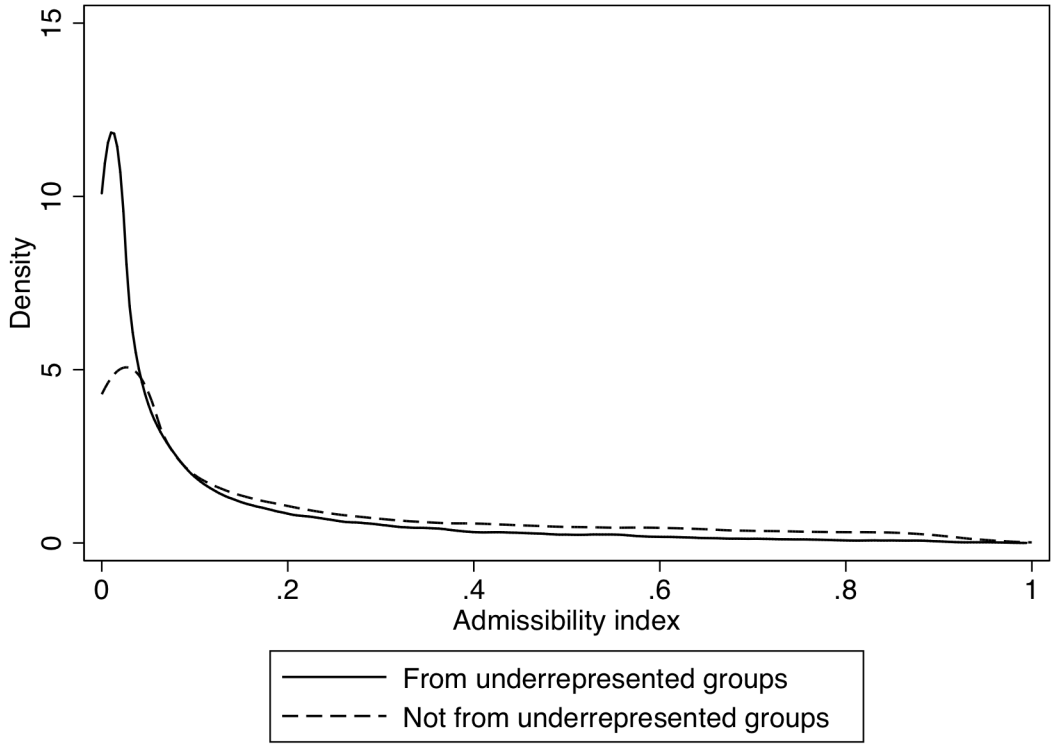
Figure 3 shows the distribution of the admissibility index for underrepresented and other applicants. Not surprisingly, applicants from underrepresented groups have lower admissibility indices, on average, than do other applicants. This is not because the index directly penalizes applicants for being members of these groups. To the contrary, all things equal a low-income or first generation applicant, or one from a low-API high school, is more likely to be admitted than does a student without these characteristics.[3] But all things are rarely equal; underrepresented applicants typically score lower on other dimensions (SAT scores, GPAs, course taking, etc.), despite the inclusion in the AI construction of a number of measures of performance relative to one's local context.

The letters of recommendation pilot

UCB piloted the use of letters of recommendation in the 2015 cycle and continued the pilot in 2016. Applicants submitted applications to UCB as to the other UCs in November, without letters. Applications went through an initial screen, and a subset of applications judged to be intermediate – neither clear admits nor clear denials – were identified. Those intermediate applicants were invited in mid-December to submit letters of recommendation.

The intent was to identify the intermediate subset via the first application reads – applications scored by the first readers as Possible would be invited to submit letters. However, timing was very tight: Invitations needed to go out in mid-December, not long after the last applications came in. It was not feasible to read every application by this point. Thus, two methods were used. Those applications that could be read by the deadline for requesting LORs, and that were scored as Possibles, were invited. In addition, a computer model was used to assess *every* application and identify some as likely to receive Possible scores, based on quantifiable characteristics (roughly, the same ones that I used above to construct the admissibility index) of the application. Any applicant identified by the

Figure 3. Distribution of admissibility index in 2016, by group.



Note: Underrepresented groups are first-generation college students, those from low-income families or low-API high schools, and underrepresented racial and ethnic groups.

computer model as a likely Possible was also invited to submit LORs, whether or not his or her application had been read by a human reader at that point.

This two-pronged approach was used in both 2015 and 2016. The computer models used in 2015 and 2016 were quite different, however, and led to quite different sets of students being selected for LOR invitations. I discuss this further below.

In each year, applicants identified for LOR invitations were sent e-mails directing them to a website where they were asked to provide contact information for recommenders. These recommenders were then contacted directly and asked to provide letters through a secure web interface. This process makes it possible to distinguish applicants who did not provide names of recommenders from those who provided names but whose named recommenders failed to provide letters.

Any letters that were submitted by the deadline, at the end of January, were made available to the second readers. Readers were given instructions that letters could help an applicant but not harm him or her, and that an applicant's failure to submit letters when invited should not be interpreted negatively. It is not clear whether readers followed these instructions or indeed how often readers

actually considered the letters at all; several readers told me that they did not put great weight on them.

I examined the 2015 pilot in my July 2016 study[4], and I describe it more fully there. There were several important differences in the implementation of the pilot in 2016. As Table 2 shows, a somewhat larger share of applicants was invited to submit letters in 2016 – 36% of in-state, non-athlete applicants in 2016, vs. 32% in 2015. The LOR requests were more closely aligned with actual first read scores in 2016: 61% of applicants invited to submit LORs in 2016 received Possibles on their first read, as compared to 52% of LOR invitees in 2015.

This in part reflects differences in reader scoring in the two years – 23% of applications were scored Possible by the first readers in 2016, as compared to 20% in 2015. A more important change was the use of a different computer model in 2016. In each year, the model constructed a single index out of a long list of applicant characteristics and assigned LOR invitations based on the value of this index. However, the index used to generate LOR invitations weighted the various characteristics quite differently in 2015 and 2016.

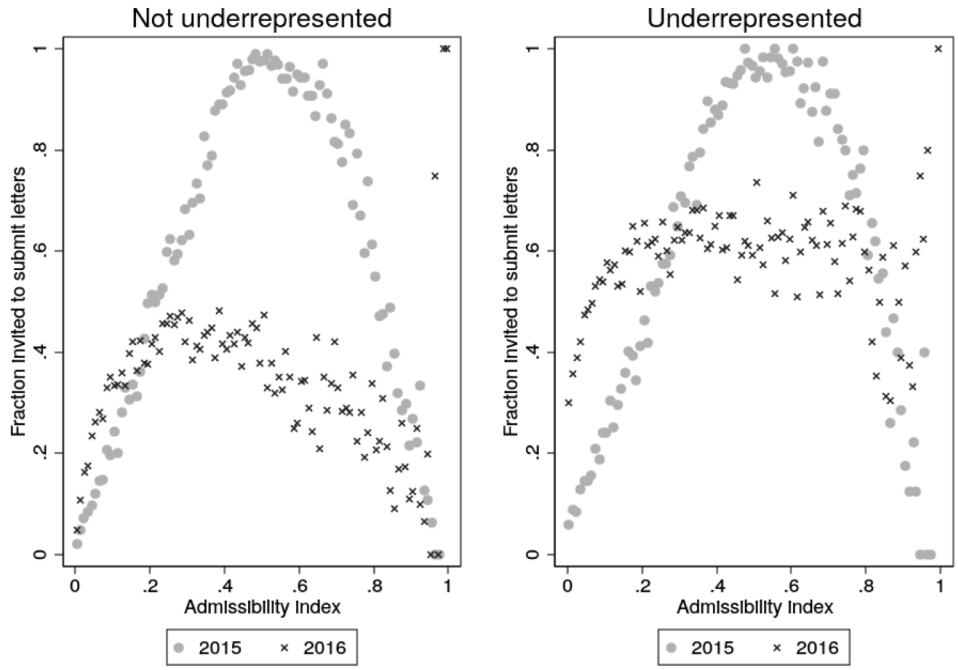
In 2015, the weights were based on an analysis of data from the 2014 admissions cycle. A linear regression was fit to predict the 2014 first reader's score, which in that year was on a 1-5 scale. The regression equation was used to generate a predicted first read score for each 2015 applicant, and those with predicted scores in an intermediate range, between 2.38 and 3.26, were selected as likely Possibles and invited to submit letters.

In 2016, the weights were reconstructed based on an analysis of data from the 2015 admissions cycle. In 2015, however, the first readers assigned only categorical ratings of Yes, Possible, and No, rather than the 1-5 rating used previously. Rather than estimate a predicted score, as in 2015, the admissions office instead used a logistic regression to estimate the probability of a Possible rating for applicants with various characteristics, with the alternative outcome encompassing both Yes and No ratings. The regression model was then used to assign each 2016 applicant a predicted probability of getting a Possible rating, and those with predicted probabilities above 50% were invited to submit letters.

The changing models led to quite different groups of students being selected for LOR invitations in the two years.[5] Nearly all (97%) of the students that the 2016 model identified as likely Possibles came from the four underrepresented groups. However, these students were more diverse in terms of their admissibility indices than the group selected in 2015.

Statistics for LOR invitations are shown in the lower portion of Table 1. In 2016, applicants from the four underrepresented groups were substantially more likely to be invited to submit letters than were other applicants. This reflects the role of the statistical model, and represents a dramatic change from 2015. Applicants from underrepresented groups were somewhat less likely than other applicants to receive a Possible score on the first read, in both 2015 and 2016. In 2015, however, underrepresented applicants were less likely to be selected by the model than their non-underrepresented peers, while in 2016 they were overwhelmingly more likely to be selected by the model.

Figure 4. Fraction of applicants invited to submit letters by group, admissibility index, and year.



Note: Underrepresented groups are first-generation college students, those from low-income families or low-API high schools, and underrepresented racial and ethnic groups.

Figure 4 shows the share of applicants at various points in the admissibility index distribution who were invited to submit letters of recommendation, separately for 2015 and 2016 applicants who were and were not from the four underrepresented groups. Applicants with AIs between about 0.2 and 0.9, whether from underrepresented groups or not, were much less likely to be invited to submit letters in 2016 than in 2015. By contrast, the more numerous group of applicants with AIs below 0.2 – which, recall, account for nearly three-quarters of all applicants – were much more likely to be invited to submit LORs in 2016 than in 2015. This reflects both the changes in the model, discussed above, and changes in the profile of students scored as Possible by the first readers.

Based upon my conversations with admissions office staff, I do not believe that they intended to make such a dramatic change in 2016. The model was developed very quickly, in the heat of the admission review season, without time for adequate study, and generated unanticipated results.

Nevertheless, the change has important implications for my study. The group of students invited to submit LORs has a wider range of admissions qualifications – some corresponding to very low chances of admission and some to quite high chances – in 2016 than in 2015. This enables me to generalize more broadly to applicants with a range of qualifications, where the results of my 2015 study pertained more specifically to students with intermediate qualifications.

IV. Outcomes of LOR invitations

A central concern in the debate about LORs is that students from disadvantaged backgrounds may be further disadvantaged by the consideration of LORs, as they may not have access to adults who know them well and are willing and able to write effective letters. This could manifest in two ways: Disadvantaged students might not submit letters in the first place, or these letters might be poorly written in comparison to those submitted by advantaged students, leading to worse reader scores. Of course, the intent of the LOR policy is the opposite of this: LORs might allow students from disadvantaged backgrounds to demonstrate strengths that are not apparent either in the quantifiable aspects of the application or in the personal essays.

In this section, I examine the objectively measurable outcomes of LOR invitations: Did students who were invited to submit LORs identify potential letter writers? And did those identified individuals actually submit letters? In the next section, I focus on those who did submit letters, and ask whether the letters led to better or worse reader scores, on average. In each case, I emphasize differences between underrepresented and other applicants.

Table 3 summarizes the objective LOR outcomes for 2015 and 2016 applicants, separately for those who are and are not members of the underrepresented groups. In 2015, 87% of applicants invited to submit letters produced at least one letter. This fell to 74% in 2016, a dramatic decline. The decline was driven by applicants from underrepresented groups, for whom the submission rate fell from 84% in 2015 to 67% in 2016. By contrast, the submission rate for applicants not from underrepresented groups was 89% in 2015 and 88% in 2016.

Table 3. Outcomes of invitations to submit letters of recommendation.

	N	Letters received	Request sent	Received given request
2015				
All	14,406	87%	88%	99%
Not underrepresented	8,069	89%	90%	99%
Underrepresented	6,337	84%	85%	98%
2016				
All	17,747	74%	76%	97%
Not underrepresented	5,951	88%	89%	99%
Underrepresented	11,796	67%	70%	96%

Note: Samples consist of applicants invited to submit letters. In column 4, samples are restricted to those who requested letters from potential recommenders.

One would not expect 100% of applicants to return letters, even if all were able to. It is relatively easy for a student already applying to other UC campuses to add an application to Berkeley as well – it simply involves checking an additional box on the application and paying an additional fee (if this is not waived). For many students who do so, Berkeley is not the first choice. Recall that only half of students admitted to Berkeley accept their admission, and only half of students offered spots on the waitlist accept them (even though this is free).

Submitting a letter of recommendation is at least somewhat costly – a student must request that an adult in his or her life write and send one. For a student who already knows that he or she will not attend Berkeley in any case, it may be simpler to just abandon the application. While this decision must happen relatively early in the cycle, before admissions decisions have been received from other UC campuses, it is likely that some applicants had received admissions (perhaps under Early Action or Early Decision programs) to private colleges and universities by that time. There is no way for me to know just how many students would have decided not to submit letters for this reason, but 10% seems a reasonable guess. In other words, it appears that nearly all applicants not from underrepresented groups who wanted to submit letters were able to do so.

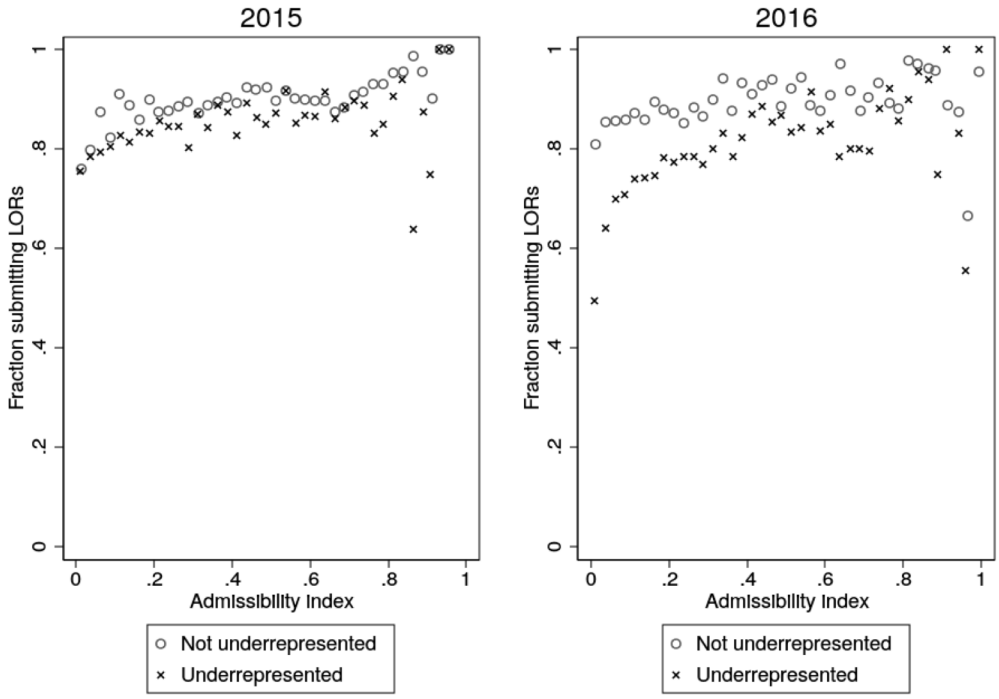
The same is not true for applicants from underrepresented groups in 2016 – it is not plausible that one-third of these applicants had already decided to go elsewhere by January. To further examine the drop-off in their submission rate, columns 3 and 4 of Table 3 show the share of invitees who provided contact information for one or more potential recommenders (the *request rate*), and the share of those who provided contact information for whom at least one letter was eventually returned (the *receipt rate*).

The receipt rate was uniformly high – across each demographic group and each year, Berkeley received letters for 96% or more of applicants who nominated letter writers. The low response rate for underrepresented students in 2016 was driven, rather, by requests: While 85% of these students submitted contact information for potential recommenders in 2015, only 70% did so in 2016.

To shed further light on this, Figure 5 shows how the request rate varied with the applicant's admissibility index, separately for underrepresented and other applicants in 2015 and 2016. It indicates that there are two explanations for the decline in the underrepresented student request rate in 2016. First, in each year, request rates were much lower for applicants with very low AIs than for those with higher AIs. (Figure 5 shows low response rates for underrepresented applicants at very high AIs in 2016 as well. But there are very few such applicants, and this likely represents statistical noise.) Because a much larger share of 2016 invitations went to students with low AIs, this depressed the 2016 invitation rate. This accounts for about one-third of the dropoff. Second, among applicants from underrepresented groups with the same AIs, the request rate was notably lower in 2016 than in 2015, particularly at the lowest AI scores. This accounts for about two-thirds of the dropoff.

I have not uncovered an explanation for the decline in request rates among low-AI students from underrepresented groups in 2016. On its face, it validates concerns that many applicants from underrepresented groups may have trouble obtaining letters of recommendation. On the other hand, one would expect these applicants to have had similar trouble in 2015, so this kind of difficulty

Figure 5. Fraction requesting letters among those invited, by year, admissibility index, and group.



Notes: Underrepresented groups are first-generation college students, those from low-income families or low-API high schools, and underrepresented racial and ethnic groups.

cannot explain the change in 2016. Moreover, it is not clear why applicants with weaker applications would have had more difficulty than demographically similar applicants with stronger applications.

One possible explanation for the change is that by 2016 applicants and their counselors understood the LOR pilot, which was never explained clearly to the public or even to applicants, better than they had in 2015. If applicants who believed themselves to be unlikely to be admitted to Berkeley interpreted an LOR request as a positive signal in 2015 but not in 2016, this could explain the observed pattern. But this is entirely speculative.

It is worth noting that low request rates for invitees with very low AIs are unlikely to have large effects on the composition of the admitted class, regardless of the explanation. This is because these invitees are very unlikely to be admitted in any case. Figure 5 shows that request rates are quite high for applicants with AIs of 0.5 or above, for both underrepresented students and others.

To illustrate this point, suppose that the admissibility index accurately represents each applicant's chances of being admitted, but that applicants who are invited to submit letters and fail to do so are never admitted. This is in direct contravention of Berkeley's actual policy, which instructed readers not to draw negative inferences from failures to submit letters. It is also counterfactual: In fact, 9% of

applicants who were invited to submit letters but did not do so were eventually admitted. Nevertheless, the extreme assumption is useful for bounding the possible impact of non-submissions. Under this assumption, if all LOR invitees had submitted letters, 39% of 2016 admissions offers would have gone to students from underrepresented groups, and the failure of some students to submit letters would have pushed this down to 37%. An impact of this magnitude would be important, but is not enormous. Moreover, it is very much an upper bound; the true impact must have been much smaller than this.

V. Within-subject study of the impact of LORs on reader evaluations

The results in Section IV indicate that a sizeable number of applicants from underrepresented groups, particularly those with low AIs, failed to request letters from potential recommenders when invited in the 2016 cycle. These applicants may have been disadvantaged by the use of LORs. But even with the decline in request rates, fully two-thirds of underrepresented applicants in 2016 who were invited to provide letters managed to do so, and those who didn't were disproportionately applicants who were very unlikely to be admitted in any case.

In this section, I consider whether the availability of letters to second readers led to better or worse scores, and whether this effect was different for underrepresented students. Did the letter writers effectively explain the hurdles that students had overcome, providing useful context not otherwise available in the application that might have helped to bolster applications that otherwise would have been assessed poorly? Or were the letter writers available to advantaged students so much better at the task that the use of LORs had the effect of magnifying these students' advantages?

The supplemental reading process incorporated into the 2016 admissions cycle was designed to answer these questions. The application, LOR invitation and submission, review, and decision process was carried out as described above. After admissions decisions were made, a random sample of 10,000 applicants who had submitted letters was selected. The admissions office then recruited 35 experienced application readers to stay on and conduct additional evaluations of applications. These readers scored the 10,000 applications in a supplemental evaluation cycle.

The readers conducting the supplemental evaluations, which was deliberately segregated from the actual admissions decision, were not given access to the LORs. Otherwise, the supplemental evaluations were designed to be as similar as possible to the second reads that were part of the regular admissions cycle. Readers were given access to the first readers' scores (as in the second read cycle) and all application materials except the letters; they were not given access to the second readers' scores, as these might have been influenced by the letters. Wherever possible, readers were assigned applications from the same region (e.g., the Inland Empire, or the Northern Sierras) to which they were assigned during the regular application review process, to take advantage of any familiarity they might have had with high schools in that region. They were not asked to review applications that they had already seen in the first or second read cycles; where this rule was inadvertently violated, scores were discarded and the applications were reassigned (even if this meant going to a reader from outside the applicant's region).

At the end of data collection, there were two alternative "second read" scores for each of 9,993 applicants – the actual second read, which incorporated the LORs, and the supplemental read, generated under similar conditions but not influenced by the LORs.[6] The comparison between these reflects, in part, the impact of LORs on readers' assessments of applicants. Of course, this is not the only determinant of differences – different readers may simply assess the same applicant differently, even when presented with the same information – but in aggregate, across large groups of applicants, other determinants should cancel out. Thus, insofar as LORs help (harm) the prospects of applicants from underrepresented groups, one would expect these applicants to have higher (lower) scores on the actual second read with LORs than on the supplemental read without them.

The comparison between second read scores and supplemental read scores is what is known as a "within subjects" research design for assessing the impact of letters on readers' evaluations. It offers as much internal validity – if not more – as a randomized experiment, as it in effect uses each applicant as his or her own control. The only threat to the validity of this research design is that the second and supplemental readers might have applied different standards.

There is some reason to be concerned about this here. In the file reading process, the admissions office devotes substantial effort to calibration, aimed at keeping the various readers using the same scale and thresholds for Yes and Possible ratings. Readers are given a target for the share of applications that should get each score, and are brought together regularly to discuss individual applications as a way of ensuring a consensus around the grading standards. In the supplemental reading process, all of the readers were experienced, so collective scoring sessions were deemed unnecessary. Moreover, providing readers with targets was complex: The applications included in the supplemental process were not a random sample of the full pool, so the overall pool targets were inappropriate. Instead, readers were told to aim for 19% Yes, 44% Possible, and 37% No scores, the distribution that was seen for the same 10,000 applications in the second reads. If this were enforced, it would ensure that there was no difference in score distributions between the second and supplemental reading waves, regardless of what information letters provide. In fact, it was not enforced – supplemental readers were told that the figures were approximations that would not be checked or enforced the same way that they are in the regular process.

Nevertheless, the comparison of *average* scores between the actual and alternative second reads combines the effects of LORs on average scores and possible drift in readers' calibration during the supplementary read process. It therefore does not provide strong evidence about whether the availability of LORs helps applicants on average. Fortunately, this is not a particularly interesting question – the number of seats available in Berkeley's entering class does not depend on whether LORs are used in admissions, so any average effect of LORs on scores would need to be adjusted away by changing the calibration in any case.

The more important question is whether the availability of LORs helps or hurts the *relative* chances of applicants from underrepresented groups. This question can be answered without potential confounding from mis-calibration in the supplementary scoring process, by examining whether these applicants do better or worse, *relative to non-underrepresented applicants*, when LORs are available than when they are not.

Table 4. Impact of availability of letters on reader scores, from within-subject comparison.

	Score better with letters	Same	Score worse with letters	Better - Worse	Yes with letters	Yes without letters	Net impact of letters on Yes
Overall							
All	26%	59%	15%	10 p.p.	19%	13%	+6 p.p.
Not underrepresented	28%	53%	19%	9 p.p.	22%	16%	+5 p.p.
Underrepresented	24%	64%	13%	11 p.p.	18%	11%	+7 p.p.
By admissibility index							
All applicants							
Admissibility index < 5%	17%	73%	10%	7 p.p.	4%	2%	+3 p.p.
Admissibility index 6-10%	28%	58%	14%	14 p.p.	10%	5%	+6 p.p.
Admissibility index 11-20%	30%	54%	16%	13 p.p.	17%	9%	+8 p.p.
Admissibility index 21-50%	30%	51%	19%	11 p.p.	29%	20%	+9 p.p.
Admissibility index 51%+	26%	56%	18%	8 p.p.	49%	42%	+7 p.p.
Not underrepresented							
Admissibility index < 5%	28%	52%	20%	8 p.p.	6%	4%	+2 p.p.
Admissibility index 6-10%	32%	53%	15%	17 p.p.	11%	4%	+8 p.p.
Admissibility index 11-20%	27%	54%	19%	8 p.p.	16%	12%	+4 p.p.
Admissibility index 21-50%	28%	52%	20%	9 p.p.	24%	18%	+6 p.p.
Admissibility index 51%+	27%	54%	19%	8 p.p.	46%	39%	+7 p.p.
Underrepresented							
Admissibility index < 5%	14%	80%	7%	7 p.p.	4%	1%	+3 p.p.
Admissibility index 6-10%	26%	60%	14%	12 p.p.	10%	5%	+5 p.p.
Admissibility index 11-20%	32%	54%	14%	17 p.p.	17%	7%	+11 p.p.
Admissibility index 21-50%	32%	50%	18%	13 p.p.	33%	21%	+11 p.p.
Admissibility index 51%+	24%	59%	17%	8 p.p.	52%	45%	+7 p.p.

Notes: Samples restricted to the 9,993 students in the supplemental read sample, selected at random from among applicants who returned letters. Read scores with letters are from the second read; read scores without letters are from the supplemental evaluations.

Table 4 shows the overall comparison between the actual second reads and the supplemental reads. The first row shows that 26% of applicants included in the study sample earned higher scores in the actual second reads, when LORs were available to readers, than in the supplemental reads where they were not, while 15% earned lower scores. Thus, calibration issues aside, it appears that LORs led to higher average scores. The second and third rows show estimates for applicants from the four underrepresented groups and for other applicants. Both groups did better when LORs were available than when they were not, but the difference is somewhat larger for applicants from underrepresented groups. This points to a positive effect of LORs on readers' ability to assess the challenges that underrepresented students have overcome. In further analyses not reported in Table

4, I find that this reflects both more Yeses and fewer Nos when letters were available to readers, particularly for applicants from underrepresented groups.

The lower portion of Table 4 separates out applicants by their AIs. The availability of LORs seems to have led to higher scores throughout the AI distribution, but there are particularly large positive effects for underrepresented students with AIs between 10 and 50%. This is exactly where one might expect LORs to be particularly valuable to readers, and where a boost to an applicant's evaluation is likely to be important to his or her eventual admission.

It is difficult to precisely quantify the effect of the availability of LORs on admissions decisions, as reader scores do not translate directly into decisions. However, it can be approximated if we assume that a reader score of Yes on the second read equates to admission. This is a crude approximation – as discussed above, admissions decisions depend on both first and second reader scores. Nevertheless, it is not too far off. Pooling together applicants from all colleges and majors, over 90% of second read Yeses are eventually admitted, though one-third of students admitted get Possibles (or, rarely, Nos) from the second reader.

With this assumption, I can compare the pool of students who receive Yeses from the actual second reader to the pool who would have received Yeses had the supplemental reads been used instead. This indicates that consideration of LORs raises the share of underrepresented students among admitted students by fully four percentage points, from 33% to 37%.

This does not take account of any negative effect of the use of LORs on students who do not submit them, as I do not have counterfactual scores for them. But the analysis discussed above indicates that that effect was no larger than two percentage points, so the positive effect of letters on reader scores appears to be much larger than any negative effect on students unable to obtain letters.

VI. Regression discontinuity analysis of the net impact of LOR invitations on admissions decisions

The results thus far are mixed. Students from underrepresented groups, when invited to submit letters of recommendation, were less likely than other applicants in 2016 to submit requests for letters from adults in their lives. This plausibly reflects, at least in part, limited access to adults with the time and ability to provide letters. However, the failure to submit letters is concentrated among applicants who were unlikely to be admitted in any case. Among those students who do submit letters, the evidence indicates that their availability to readers led to higher scores, particularly for applicants from underrepresented groups on the cusp of admission.

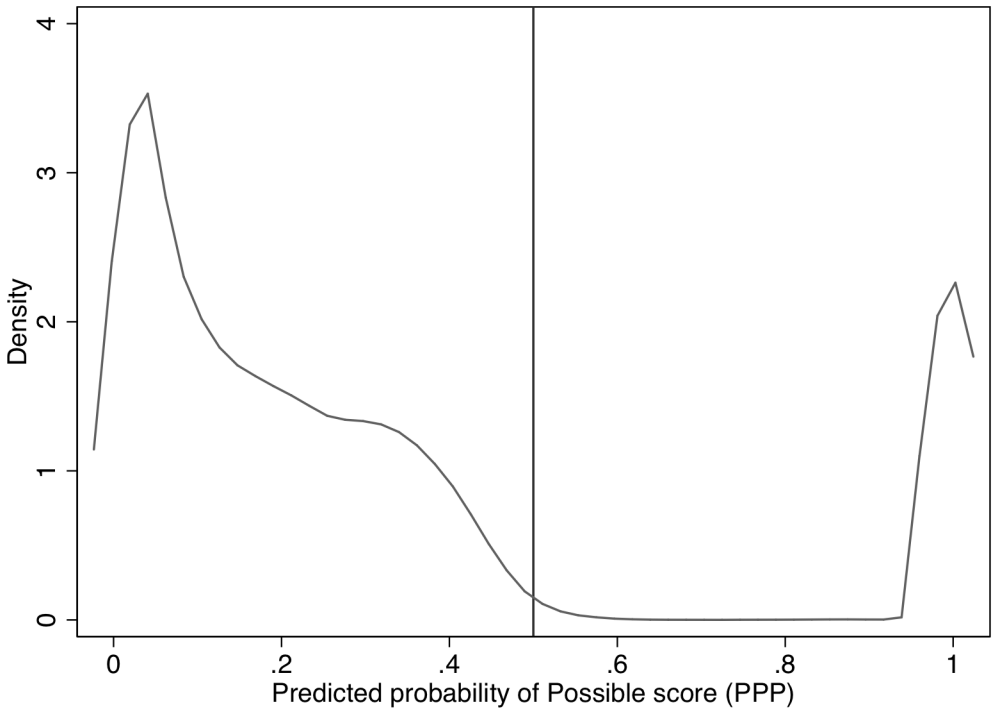
The latter effect appears to be at least twice as large as the former, indicating that letters increase the share of underrepresented students, on average. However, this estimate is predicated on the assumption that second read scores translate directly to admissions decisions, where in fact they do not. Actual decisions are in principle heavily dependent on the second read scores but in practice may wind up placing even more weight on "tiebreak" factors.

In this Section, I examine the net effect of an LOR invitation, whether or not it was taken up and regardless of the impact of any resulting letters on readers' scores, on the eventual admissions decision. To do this, I adopt a strategy from my initial 2016 report, taking advantage of sharp breaks in the allocation of LOR invitations introduced by the use of a computer model to select applicants for invitations.

Recall that in 2016, all applicants were assigned a predicted probability of receiving a Possible score on the first read, based on a computer model that took into account a long list of quantifiable aspects of the application. Applicants with predicted probabilities in excess of 50% were automatically invited to submit letters, regardless of the actual first read score, while those with predicted probabilities below that threshold were invited to submit letters only if the first read took place early enough and yielded an actual Possible score.

The predicted probabilities were continuous, and there is no reason to think that an applicant with a predicted probability of 51% was meaningfully stronger (or weaker) than an applicant with a predicted probability of 49%. Thus, the causal effect of the LOR invitation can be identified by comparing the admissions outcomes of applicants just above the 50% threshold to those of applicants just below the threshold, as the only meaningful difference between them is that all of the former were invited to submit letters and many of the latter were not. This is known as a *regression discontinuity* design. Under the proper conditions – satisfied here – it offers a very high degree of

Figure 6. Density of predicted probability of a Possible score (PPP), 2016 applicants.



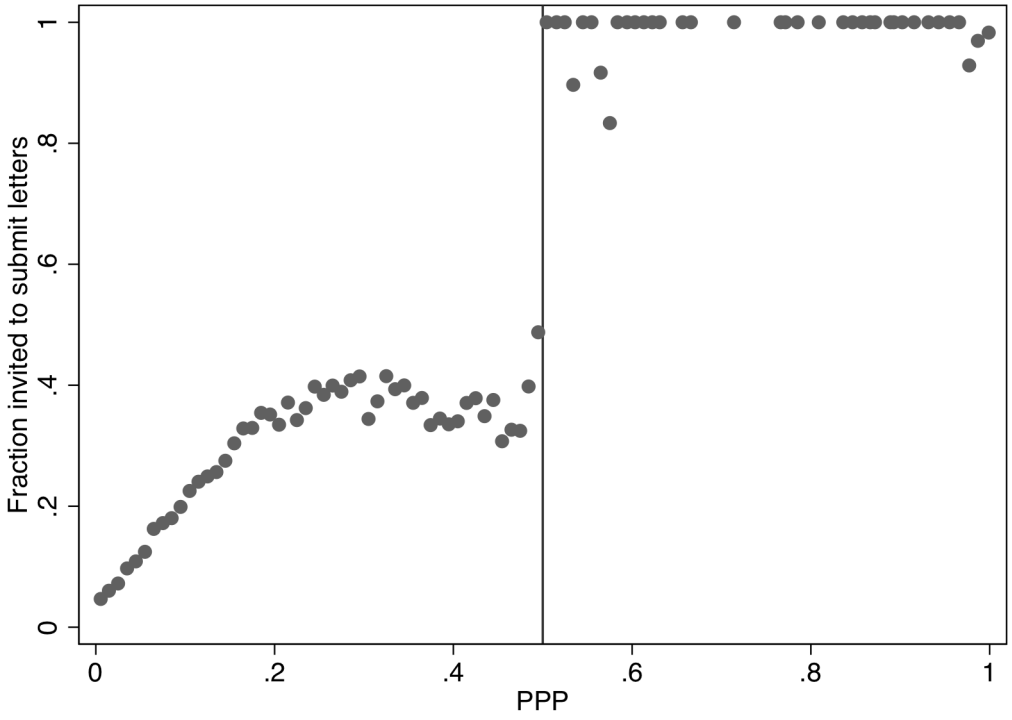
Notes: Sample restricted to California resident, non-athlete applicants.

credibility to the estimated causal effects, albeit not as high as in a true randomized experiment or within-subjects design.

Figure 6 shows the distribution of the 2016 predicted probabilities of receiving a Possible score (hereafter, *PPP*s). The most common PPP values are close to zero. This largely reflects the many applicants with very low admissibility indices; very few of these applicants receive Possible scores, so their PPPs are very low. There is another group of applicants with PPPs very close to one – these are applicants with combinations of characteristics that nearly always get Possible scores from the first readers. Nearly all of these applicants are from underrepresented groups; while many non-underrepresented applicants get Possible scores, there is no set of characteristics that seems to nearly assure a Possible score for these applicants.

The vertical line in Figure 6 shows the 0.5 threshold for an automatic LOR request. There are relatively few applicants around this threshold. This limits the precision of the regression discontinuity (RD) estimates, as they are identified solely from the small subset of applicants with PPPs close to 0.5. Although this group is small, it is reasonably diverse. It contains substantial numbers of applicants both from underrepresented groups and not, and admissibility indices are roughly uniformly distributed between about 0.3 and 0.8. Thus, while the RD estimates of the impact of an LOR invitation will be less precise than might be hoped, they are generally reflective of the experiences of a broad range of applicants (though not of those with the lowest AIs, who are rarely admitted in any case).

Figure 7. Fraction invited to submit letters in 2016, by predicted probability of Possible score (PPP).



Notes: Sample restricted to California resident, non-athlete applicants.

Figure 7 shows the share of applicants at each PPP who are in fact invited to submit letters. As expected, essentially every applicant with a PPP above 0.5 was invited. There are a few exceptions, which apparently reflect errors in the admissions office's processes – students who should have been invited but for some reason were not. But these constitute fewer than 2% of the applications with PPPs above 0.5. Below the 0.5 threshold, where applicants were invited to submit letters only if the first reader scored them as Possibles, the share of students invited is around 40% for PPPs between 0.2 and 0.5, and lower below that. There is thus a sharp contrast in the invitation rate between students with PPPs just below and just above 0.5.

Table 5 shows regression discontinuity estimates of the effect of an LOR invitation on admissions, with standard errors in parentheses. I examine three admissions outcomes: Whether a student was admitted on the initial round; whether he or she was ever admitted, either initially or off the waitlist; and whether he or she filed a Statement of Intent to Register (SIR). For each outcome, I show estimates for all applicants, for applicants from underrepresented groups, and for other applicants. For example, the upper right entry indicates that the effect of being invited to submit letters is to reduce the likelihood of being admitted in the initial round by 0.06 percentage point. The standard error of 0.08, so this is not remotely statistically significant.

The results are imprecise, and only one (for SIRs among underrepresented applicants) is statistically significantly different from zero. Nevertheless, the pattern of estimated effects is interesting. They indicate negative effects of LOR requests on the admissions rates of non-underrepresented applicants and zero effects on those of applicants from underrepresented groups. Impacts on SIRs are more positive than on admissions decisions. This is consistent with the view that one

Table 5. Regression discontinuity estimates of the effects of LOR invitations on 2016 admissions.

	All	Not underrepresented	Underrepresented
Initial admit	-0.06	-0.19	0.01
	(0.08)	(0.12)	(0.12)
Ever admit	-0.04	-0.16	0.04
	(0.08)	(0.12)	(0.11)
Statement of Intent to Register (SIR)	0.12	0.00	0.23
	(0.07)	(0.11)	(0.09)

Notes: Each entry represents a separate fuzzy regression discontinuity analysis. Samples consist of all 2016 applicants with PPPs between 0.4 and 0.6. Models control for separate linear terms in the PPP on either side of the discontinuity, and are weighted using a triangular kernel centered at 0.5. Standard errors in parentheses.

consequence of an LOR invitation is to induce some applicants to reveal that they don't want to come to Berkeley after all, leading them not to be admitted but not having an effect on their enrollment since they would not have enrolled in any case. (Another potential explanation is that LOR invitations begin to build psychological bonds between the student and the university, making them more likely to respond positively to eventual admissions offers.) For underrepresented applicants, the LOR effect on SIRs is positive and statistically significant.

There is certainly no indication here that LOR invitations reduce the relative admissions chances of applicants from underrepresented applicants, and some (albeit imprecise) evidence to the contrary.

VII. Conclusion

This analysis of the second year of UC Berkeley's LOR pilot is not fully conclusive. Those who were predisposed to think that the use of letters of recommendation in Berkeley admissions would further disadvantage students who are already disadvantaged can find some evidence to support that view here, in the high share of underrepresented students who failed to respond to LOR invitations in 2016. But so can those who think that LORs can provide useful context to inform holistic admissions decisions, enabling readers to better see the obstacles that students have overcome.

There is evidence of both negative and positive effects on applicants from underrepresented groups: Many of these applicants did not submit letters when requested, but those who did submit them seem to have benefitted from being given the additional opportunity to provide evidence in support of their admission that was not otherwise included in the application. The overall net impact on admissions is not very precisely estimated, but there is no indication that LOR invitations hurt the relative chances of underrepresented applicants, and some evidence that invitations increased these students' enrollment.

Ultimately, any impact of the use of letters of recommendation on the diversity of the Berkeley student body seems likely to be small. It would be easily overwhelmed by other aspects of the admissions process, including such decisions about how much effort to put into outreach, how much weight to put on extracurricular activities, and the importance accorded to SAT scores relative to high school grades or to evidence of applicants' effort relative to the opportunities available to them.

In my understanding, AEPE's goal in including letters of recommendation in admissions was not to increase diversity (and certainly not to reduce it), but rather to support better decisions in an increasingly competitive applicant pool. It is too soon to tell whether this goal was met. Are the students who are admitted under an LOR admissions process but would not be otherwise students who do well once at Berkeley? Or do LORs lead to the admission of students who might have impressed their guidance counselors but are not strong enough academically to succeed? There would be great value in a follow-up study that used evidence on grades and persistence rates, both for students matriculating at Berkeley and for those who enroll at other UC campuses, to assess this question. I hope to pursue this in the future.

Pending further evidence, I cannot make a recommendation about whether letters should be used in Berkeley admissions going forward. My study of admissions processes during the pilot years, however, does support recommendations about how letters should be included, if they are at all:

- First, there should be much more transparency about whether and under what circumstances letters will be requested. Applicants and counselors should not be left to guess whether an invitation, or its absence, represents a signal of the eventual outcome.
- Second, if possible, letters should be obtained early enough to be considered by both the first and second application readers; if this is not possible, selection procedures should be reformed to place much less weight (if any) on the first reader's score, conditional on the second. There is no sense in requesting letters when the first reader's score has already determined the fate of the application.
- Third, there are a number of changes that the campus might adopt to offset any negative effect of letters of recommendation on diversity. These might include greater outreach to underrepresented potential applicants before applications are due; improved communication to applicants from whom letters are requested (and, perhaps, communication directly with those applicants' schools or counselors); and/or modifying the weight put on student background, as compared with traditional academic variables, in the admissions decision. If the information contained in letters is truly valuable, any of these small changes would be worth it in order to support the collection of this information.

Endnotes

[1] Rothstein, Jesse, "UC Berkeley Admissions in 2015 and 2016: An Analysis of the Role of Letters of Recommendation and Augmented Review." Report to the UC Berkeley faculty committee on Admissions, Enrollment, and Preparatory Education, July 2016. http://eml.berkeley.edu/~jrothst/otherwriting/rothstein_report_july122016.pdf

[2] The fourth underrepresented category is not used in the academic index or in any of the other prediction models considered below (e.g., those used for assigning LOR invitations) because race and ethnicity are not considered in Berkeley admissions under California law.

[3] That is, in the regression model that is used to generate the admissibility index, these three proxies for applicant disadvantage have positive coefficients: Holding constant other characteristics, applicants from these groups are more likely to be admitted than applicants not from these groups. Note that the fourth proxy, race/ethnicity, is not used either in the construction of the admissibility index or in admissions decisions.

[4] Rothstein 2016, *op. cit.*

[5] For example, consider an applicant profile for which one-third of applicants are scored Yes, one-third Possible, and one-third No. The expected, average score for applicants with this profile is a Possible, but there is only a one-third chance that these applicants will actually receive Possibles. The 2015 model would have selected these applicants for LOR invitations on the basis of the expected score, but the 2016 model would not have because the predicted probability of a Possible was not high enough. By contrast, a group of applicants with a 51% chance of being scored Possible and a 49% chance of being scored Yes would have been selected in 2016 but not 2015.

[6] The missing seven applications were found after selection to be ineligible for the study (due, for example, to being from out of state), or were inadvertently assigned to readers who had seen them previously.

About the Author

Jesse Rothstein is a Professor of Public Policy and Economics at the University of California, Berkeley. He serves as Director of the UC Berkeley Institute for Research on Labor and Employment and of the California Policy Lab-Berkeley. He previously served as Chief Economist at the U.S. Department of Labor and as Senior Economist with the Council of Economic Advisers, Executive Office of the President, both in the Obama Administration. Rothstein's research focuses on education policy and on the labor market. His recent work includes studies of teacher quality, of school finance, of intergenerational economic mobility, and of the labor market during the Great Recession. His work has been published in leading journals in economics, public policy, education, and law. He was named the John T. Dunlop Outstanding Scholar by the Labor and Employment Relations Association in 2011. He is a member of the editorial boards of the *American Economic Review*, *Industrial Relations*, and the National Education Policy Center. He is a research associate of the National Bureau of Economic Research and a fellow of the National Education Policy Center, the CESifo Research Network, the IZA, and the Learning Policy Institute. Rothstein received a Ph.D. in economics and a Master's in Public Policy, both from the University of California, Berkeley, and an A.B. from Harvard.

Acknowledgments

I thank Audrey Tiew for excellent research assistance and Greg Dubrow, Amy Jarich, Bob Powell, Catherine Koshland, Carol Christ, and the members of the UC Berkeley AEPE and UCOP BOARS and UCAADE committees for helpful conversations. I am especially grateful to Colleen Lim for her help in overseeing the supplemental application reading process, and to several dozen UC Berkeley application readers for carrying out that process. I thank the UC Berkeley Office of Undergraduate Education for funding to support the supplemental reading process. This research was conducted under the auspices of the California Policy Lab, a joint Berkeley-UCLA project that aims to support better evidence-based policy through the analysis of administrative data in California.

Errata

The following is a description of the changes made to the report since it was originally published.

July 24, 2017

Page 17: The original document included an incomplete version of **Figure 4**. This has been corrected.

Page 20: The original document included an incomplete version of **Figure 5**. This has been corrected.