Job Credentials in the Labor Market: Pre-Analysis Plan

Lisa Abraham

August 14, 2018

Abstract

Women are underrepresented in certain jobs relative to men and that disparity may be due to the composition of the applicant pool. Anecdotal evidence suggests that women are less likely than men to apply for jobs when they do not meet all the posted requirements and that produces gender gaps in job applications and subsequent hiring. Together with a large IT firm (Uber), I run a randomized control trial (RCT) to test if posted requirements matter for gender differences in applications. I examine whether deleting optional credentials and reframing language in the company's job postings could encourage more women to apply and, thereby, progress through the application process. This document describes the pre-analysis plan for the RCT.

1 Introduction

An important component of gender differences in the labor market stems from inequities in representation. There is evidence that females are underrepresented in certain sectors of the U.S. economy and in higher level positions. While some of this may be attributable to different preferences for task content and culture, another potential explanation focuses on pipeline challenges, particularly at the point of job application.

Though there is research examining factors that influence job application decisions (e.g., advertised amenities), as well as research on gender differences in confidence, there is less work investigating how gender differences in beliefs about meeting all the listed credentials in a job application interacts with a female's decision to apply for the job, particularly a traditionally "masculine" or highly technical job. The best evidence is anecdotal in nature: Sheryl Sandberg's *Lean In* highlights an internal Hewlett-Packard report which found that males apply for open jobs when they meet 60% of the credentials, whereas females only apply if they meet 100% of the listed credentials. While this descriptive anecdote has been referenced in numerous media sources, there are no rigorous economic analyses confirming this fact or analyzing potential solutions. My RCT examines whether deleting optional credentials and reframing language in job postings could encourage more women to apply and progress through the application process.

2 Experimental Design

2.1 Sample

I have partnered with Uber to implement the RCT. As of August 2018, Uber advertised approximately 1,000 U.S. job postings on their career website,[1] comprising both technical and non-technical roles. Individuals can apply to Uber by going directly to the career website, or by linking to the career website from a job posting on a third party website (e.g., LinkedIn, Glassdoor, Indeed). The sample consists of individuals who *directly* view a job posting on Uber's career website. I exclude individuals that link to the career website from a third party site,[2] as well as referrals and external hires (given the different nature of the referral / external hiring process).[3] I consider the first three weeks of the RCT a pilot; I plan on including data from the pilot in the analyses, unless there is a problem with the execution of the experiment.

2.2 Design

The experiment randomizes individuals who directly[4] visit the Uber career website. Individuals are identified by their IP address and randomly sorted[5] into one of two groups: (1) a Control group where individuals see the original version of the job posting, or (2) a Treatment group, where individuals see a version of the job posting which deletes optional credentials, deletes adjectives describing skills, and reframes vague credentials. Treatment or Control status is maintained across all jobs an individual views. Through the use of cookies, the same treatment is served to an individual who returns to the website at a later date if she uses the same browser and same device.

[1] https://www.uber.com/careers

[2] I am unable to include individuals that link from a third party website in our RCT due to technical constraints.

[3] Individuals who have uBlock, an ad-blocking browser extension, are also excluded from the experiment since this software blocks Optimizely, the A/B testing platform that we are using to implement the experiment.

[4] It is possible to identify and eliminate individuals who link to the career website from a third party website.

[5] Randomization is at the IP address level, but for ease of exposition I refer to randomization as being at the "individual" level.

Upon seeing the job posting on the Uber career website, individuals decide whether or not to apply for the position. For those who decide to apply, the first step is to click on the "Apply Now" button, after which individuals are given several options for applying: (1) by signing in (if they already have an account on the career website); (2) by logging into their LinkedIn profile (so that data from the individual's LinkedIn profile is transmitted to the Company); (3) by submitting their resume; (4) by manually filling out the application form.

It is possible to track the treatment status of the applicants in the experiment at all stages of the hiring process (i.e., application, phone screen, interview, offer, and hire). If an individual applies to more than one job posting over the duration of the experiment, it is possible to examine the portfolio of her job application choices.

Importantly, recruiters and hiring managers will not be able to easily discern whether an applicant is coming from the Treatment or Control group.

2.3 Treatment

The RCT only alters language in the section of the job posting listing required credentials and desired credentials. In some job postings, desired credentials are interspersed in the required list (e.g., "Python experience is a plus"), while other job postings break out the desired credentials in a separate section. The required credentials are labeled "What You'll Need," (or a variant of this) and the desired credentials, when broken out, are labeled "Bonus Points" (or a variant of this).

Treatment varies significantly depending on the job posting. While some job postings have a lot of optional credentials (e.g., "PhD preferred"), adjectives describing skills (e.g., "Excellent" before "coding skills"), and vague credentials (e.g., "think like your enemy"), others have less scope to be edited. Given the heterogeneity in treatment, for every job posting included in our experiment, I document the original version of the credentials and the treated version of the credentials. I also indicate how many adjectives were deleted, how many vague credentials were reframed, how many optional credentials were deleted, and whether the optional credentials were interspersed in the required credentials or broken out in a separate section. There are different ways that these edits can be classified,[6] and the strength of the treatment across jobs could also

[6] In my initial classification, *Adjectives Deleted* represent the number of deleted adjectives (at the word level) which are not included in a vague credential or optional credential, *Vague Credentials Reframed* represent the number of vague credentials that are reframed (at the phrase level), and *Optional Credentials Deleted* represent

be a function of the rest of the text in the job posting, so my classification represents an initial attempt at quantifying the treatment. I plan on having workers on Mechanical Turk rate the job postings on various dimensions after the RCT concludes as this might provide a more accurate measure than the quantitative metrics. I also preserve the text of the Treatment and Control versions for alternative classification and finer textual analysis.[7]

2.4 Slippage

I anticipate that there might be some slippage in assignment to treatment, given the fact that individuals can view an Uber job posting on a third party website, and later find that same job posting directly on the career website; this would result in slippage if the version the individual sees on the third party website does not match the version the individual is randomly assigned on the career website. This could be partly mitigated if the credentials only become salient to individuals when they are deciding whether or not to apply.

Slippage might also occur if individuals access the career website from multiple, different browsers (e.g., Mozilla Firefox and Google Chrome), or multiple, different devices (e.g., ipad and computer). I believe this is somewhat mitigated by the relative ease of the applying in one sitting (detailed above), which makes it seem unlikely that individuals are repeatedly viewing the same posting from different browsers or devices over the brief span of time that they are deciding whether or not to apply.

There is also the potential for slippage due to the presence of "faulty" job links on Uber's career website. These faulty job links are those which have more than five digits in the job requisition ID, representing approximately 10% of all U.S. job postings. When browsing the career website, these faulty links sometimes present a cached version of the last job the individual viewed on the career website, as opposed to the actual job associated with the digits. As a result of these links, there may be slippage for individuals assigned to Treatment (i.e., they will see the Control version of the cached job on the faulty link even though they were assigned to Treatment). Users who are assigned to the Control group will not experience slippage. However, applications from these faulty job links from both Treatment and Control groups will

the number of optional credentials that are deleted (at the bullet point level).

[7] Due to technical constraints, there are some jobs for which the swapped in Treated version of the qualifications section appears in a different font, size, or bullet indentation relative to other sections of the text. However, given that several of Uber's job postings display inconsistent font, size, and indenting, this does not seem like a significant departure from what an individual might expect to see.

not be included in the experiment, so while the slippage might make individuals aware of the experiment, it is less likely to induce bias.

Finally, slippage could occur as a result of a split second delay that sometimes occurs when the webpage content is loading on a particular browser, which could allow an individual to see the Control version of a job for a split second before seeing the Treatment version. However, given that individuals are used to the dynamic nature of websites, and the fact that the updates to the "What You'll Need" section occur below the fold (i.e., to the lower half of the webpage which individuals must scroll down to view), it seems likely that slippage resulting from this source is minimal.

3 Empirical Strategy

3.1 Individual-level Regressions

The main specification is the regression at the individual level. This regression answers the question: is there a difference between the overall fraction of female applicants out of all applicants in the Treatment and Control conditions? The regression is given by:

$$I(Female)_i = \alpha + \beta Treated_i + \epsilon_i$$

where $I(Female)_i$ is an indicator for whether person i is female and $Treated_i$ is a dummy variable for treatment status of the individual. β is the percentage point change in the fraction of female applicants between the Treatment and Control groups (before controlling for any other variables). This analysis can be repeated at all other stages of the interview process (i.e., I can examine phone screens, on-site interviews, and offers instead of applicants).

Given the heterogeneity in treatment strength, I also consider a different specification that interacts the "intensity" of the treatment with treatment status. This regression is given by:

$$I(Female)_i = \alpha + \beta Treated_i + Intensity_i + \sigma(Treated * Intensity)_i + \epsilon_i$$

where $Intensity_i$ is a categorical variable denoting the intensity of the treatment.

3.2 Job-level Regressions

I can also run a regression at the job-treatment level. This regression answers the question: does the overall fraction of females increase between the Treatment and Control conditions after collapsing to the job-treatment level and accounting for the number of applicants in each job-treatment cell? This regression examines treatment effect heterogeneity and is given by:

$$\%\ Female\ Applicants_{j,t} = \alpha + \beta Treated_{j,t} + \lambda_j + \epsilon_{j,t}\ [aweight = \#applicants_{j,t}]$$

where $\%\ Female\ Applicants_{j,t}$ indicates the fraction of female applicants out of all applicants in job j and treatment condition t, $Treated_{j,t}$ is a dummy variable for treatment status (note that each job has both a treatment and control condition), and λ_j represents job fixed-effects. β is the percentage point change in the fraction of female applicants between the Treatment and Control groups after collapsing to the job-treatment level and weighting by the number of applicants in the job-treatment cell. I also plan to run this reqression without weighting by by the number of applicants in the job-treatment cell; doing so weights all jobs equally. This analysis can also be repeated at all other stages of the interview process (i.e., I can examine phone screens, on-site interviews, and offers instead of applicants).

As with the individual-level regression, I also consider a different specification that interacts the "intensity" of the treatment with treatment status. This regression is given by:

$$\%\ Female\ Applicants_{j,t} = \alpha + \beta Treated_{j,t} + \lambda_j + Intensity_{j,t} + \sigma(Treated * Intensity)_{j,t}$$
$$+\epsilon_{j,t}\ [aweight = \#applicants_{j,t}]$$

where $Intensity_{j,t}$ is a categorical variable denoting the intensity of the treatment.

3.3 Additional Sample Details

There are several details that I plan to control for in the analysis. Over the duration of the experiment, recruiters and hiring managers (who are blind to applicant assignment to Treatment) may edit a job posting after it is live on the website. Since the Treatment version of the job might not match the Control version of the job after editing, I plan to control for this in the analysis by excluding all applications from jobs that were edited if the application came in after the recruiter edited it.

I also plan to include specifications that control for delayed refreshing of the career website and imperfect information about when an individual views a job. For example, supposing a job is entered into the A/B test at 1:00pm PST. Then an applicant who has been assigned to the Treatment group could view the Control version of a job posting at 12:58pm PST that same day and apply at 1:05pm PST, but the website might not have refreshed the job posting to the appropriate Treatment version of the job. In this case, the applicant would be included in the experimental data, since she would have applied after the A/B test started, even though she never viewed the Treatment version of the job. As such, I plan to conduct sensitivity analyses that exclude applications which come in at varying intervals after the A/B test is initiated.

I also plan to control for an institutional feature of Uber's hiring process, whereby recruiters and hiring managers can move an applicant to a different job (i.e., not the one that the applicant originally applied for) if another job seems like a better fit. For these "moved" applicants, while I cannot examine progression through the hiring process in the job they originally applied to, I can examine progression in the job which they were moved to.

3.4 Other Analyses

3.4.1 Total Number of Applicants

While the primary focus on the experiment is on gender differences, I can also examine whether more individuals apply and reach subsequent stages of the application process via the Treatment condition versus the Control condition; differences between the Treatment and Control condition (irrespective of gender composition) could be indicative of potential inefficiencies in screening.

3.4.2 Conversion Rates

I can examine conversion rates from application to subsequent stages of the interview process (e.g., phone screen, on-site interview, and offer). This analysis addresses how the quality of applicants in the Treatment and Control conditions differs (at the aggregate level and by gender).

3.4.3 Portfolio of Job Applications

If an individual applies to more than one job over the duration of the RCT, I can examine the portfolio of jobs she applies to. This analysis has the potential to inform our understanding of how listed job credentials interact with job search.

4 Variables of Interest

4.1 Outcome Variables

To execute our analyses, I examine the following outcomes:

1. Gender of applicants
2. Gender of phone screens
3. Gender of on-site interviews
4. Gender of offers
5. Gender of hires
6. Number of applicants
7. Number of phone screens
8. Number of on-site interviews
9. Number of offers
10. Number of hires
11. Data on applicant quality (parsed from candidate resumes / application information)
12. Portfolio of jobs an individual applies to