Analysis Plan for Fall 2021 Follow-up to UpWork Experiment

The plan is to closely parallel the analysis presented for the UpWork experiment currently reported in the paper. One key definition is what it means for a participant to be **qualified** for the expert-level job. We focus on qualified in terms of the ASVAB score, as that is the desired qualification our treatments focus on. For our analysis, we will consider a participant to be qualified for the expert-level job if they have an ASVAB score strictly above the threshold used in the Positive and Normative advertisements - 5**.5 problems solved correctly**. Throughout the analysis, our main outcome variable will be the decision to apply to the expert-level job, which takes either 1 (applied to expert-level job) or 0 (did not apply to expert-level job).

We are running the sample on Prolific with a restriction to workers who have completed 100+ Prolific studies, with at least a 95% approval rate. Note that our plan was to recruit participants starting December 1st, 2021. For every day, following December 1 until the collection of 2400 responses has been completed, a coin toss will determine whether to recruit on that day or not. We will pause recruitment between December 18 and January 2nd. Recruitment will be concluded once 2400 participants have completed the study.

However, due to a technical issue on Prolific’s end, our launch was delayed until Dec. 2nd. On Dec. 2nd, we launched with a soft launch of 100 responses to check for possible technical issues before opening up the study to the full 2,400 participants. The technical components worked as expected but performance on the ASVAB test was substantially worse than we had planned in our calibration. The cutoff score is a crucial parameter in our design, as only participants above the cutoff are deemed qualified. Thus, to ensure adequate sample size of qualified participants, we decided to change the design to lower the cutoff from 6.5 to 5.5.

We have made this change and will exclude from analysis all observations from this soft launch on Dec. 2nd. We will collect 2,400 new observations using the new cutoff of 5.5 beginning on Dec. 5th.

We will exclude from analysis any participant who did not answer the attention check question correctly. Furthermore, we will exclude any worker who chose to apply to neither job (as indicated in their application decision), **and** indicated this was because of a reason other than perceived qualifications (as indicated in the follow-up question that asks why they did not apply). Workers who did not apply to either job and indicated in the follow-up question that they did not apply specifically because they did not feel qualified for either job will be included in the analysis and coded as not having applied to the expert-level job. We will exclude workers from analysis who indicated that their gender was neither “Woman” nor “Man”.

Below, we outline the analysis in detail.

1. Tests of characteristics across men and women

First, we will report the difference in average characteristics that appear on the “resume” across the men and women in the Prolific sample in terms of (i) average performance on the ASVAB test across men and women (two-tailed t-test) and (ii) education as reported on Prolific profile (test of proportions for each category of education). Because of our inclusion criterion, every worker will have at least 100+ Prolific studies completed with a 95% approval rating.

1. Descriptive summaries of application rates by gender, “qualified” indicator, and treatment

Our main question of interest is how application rates to the expert job vary by gender, treatment, and whether or not the applicant is qualified (according to definition above). We expect that reduced ambiguity in desired qualifications should (weakly) increase the likelihood of qualified candidates applying to the expert job. Being better informed about where the bar is, if you are above it, should increase your willingness to apply. On the other hand, reduced ambiguity should (weakly) decrease the likelihood of unqualified candidates applying.

We will report the rates of application to the expert-level job in the following ways:

* 1. Proportion of **qualified** women compared to the proportion of **qualified** men that apply to the expert-level job in the Control treatment (test of proportions)
	2. Proportion of **qualified** women compared to the proportion of **qualified** men that apply to the expert-level job in the Positive treatment (test of proportions)
	3. Proportion of **qualified** women compared to the proportion of **qualified** men that apply to the expert-level job in the Normative treatment (test of proportions)
	4. Proportion of **unqualified** women compared to the proportion of **unqualified** men that apply to the expert-level job in the Control treatment (test of proportions)
	5. Proportion of **unqualified** women compared to the proportion of **unqualified** men that apply to the expert-level job in the Positive treatment (test of proportions)
	6. Proportion of **unqualified** women compared to the proportion of **unqualified** men that apply to the expert-level job in the Normative treatment (test of proportions)
1. Regression Analysis focused on main hypotheses

Our main hypothesis is that reduced ambiguity about where the bar is (i.e. the Positive and Normative treatments) can draw in more qualified female candidates relative to the Control treatment, narrowing a gender gap in applications among qualified candidates. We can secondarily ask whether the more precise information about the bar deters unqualified female applicants more than unqualified male applicants (relative to the Control).

We plan to report the following table. Here the estimates are from the UpWork experiment but the coefficients would obviously be replaced with the estimates from the new experiment. We predict the decision to apply to the expert-level job from treatment assignment, using the control treatment as our baseline omitted category. We control for all resume characteristics we have (ASVAB score, education as reported on Prolific profile), as well as demographics reported in our survey excluding education. We will also control for their total number of jobs completed on Prolific and their approval rating.

We expect weakly positive signs on the Positive and Normative treatment indicators in Column VI. We expect a negative sign on the Female indicator in Column VII – indicating that qualified women applied at a lower rate than qualified women in the Control treatment. And, we expect a positive sign on the Positive x Female and Normative x Female interaction terms in Column VII – indicating that the decreased ambiguity about where the bar was helped to reduce the gender gap in application rates among qualified applicants.

We can also explore how more clearly stated qualifications impact behavior among unqualified applicants (Columns IV and V). We expect weakly negative signs on the Positive and Normative treatment indicators in Column IV. And, we can ask whether more clearly stated qualifications have a larger deterrence effect among women than men (negative sign on Positive x Female and Normative x Female in Column V). This is of secondary interest.

**Table 1. Application Rates to Expert-Level Job**

|  |  |
| --- | --- |
|  | **OLS Predicting Decision to Apply to Expert Job** |
|  | All Participants | All Unqualified | All Qualified |
|  | I | II | III | IV | V | VI | VII |
| Positive  | -0.026 | -0.043 | -0.039 | -0.046\* | -0.057 | 0.044 | 0.0067 |
| Treatment | (0.024) | (0.027) | (0.035) | (0.026) | (0.038) | (0.061) | (0.081) |
|  |  |  |  |  |  |  |  |
| Normative | -0.030 | -0.067\*\* | -0.033 | -0.070\*\*\* | -0.044 | 0.098 | -0.00076 |
| Treatment | (0.024) | (0.028) | (0.034) | (0.026) | (0.038) | (0.060) | (0.074) |
|  |  |  |  |  |  |  |  |
| Female | -0.029 | -0.026 | -0.039 | -0.0062 | 0.0023 | -0.075 | -0.20\*\* |
|  | (0.021) | (0.021) | (0.035) | (0.023) | (0.038) | (0.055) | (0.086) |
|  |  |  |  |  |  |  |  |
| Qualified |  | -0.057 |  |  |  |  |  |
|  |  | (0.046) |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Positive x  |  | 0.066 |  |  |  |  |  |
| Qualified |  | (0.057) |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Normative x |  | 0.15\*\*\* |  |  |  |  |  |
| Qualified |  | (0.056) |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Female x  |  |  | 0.024 |  | 0.020 |  | 0.10 |
| Positive |  |  | (0.049) |  | (0.052) |  | (0.12) |
|  |  |  |  |  |  |  |  |
| Female x Normative |  |  | 0.0050 |  | -0.047 |  | 0.28\*\* |
|  |  |  | (0.048) |  | (0.052) |  | (0.12) |
| Controls | Y | Y | Y | Y | Y | Y | Y |
| Observations | 1083 | 1083 | 1083 | 827 | 827 | 256 | 256 |
| Adj. R-squared | 0.035 | 0.039 | 0.034 | 0.037 | 0.037 | 0.012 | 0.026 |

Notes: Qualified candidates are those with a test score greater than or equal to the advertised threshold.

1. Direct Analysis of Perceptions of the Bar

One important goal of the follow-up experiment is to investigate directly candidates’ perceptions of where the bar is. We include several questions following the application decision that we will analyze.

1. We ask, “what do you think is the probability that, based upon your resume, you would be considered qualified for the expert-level job?” Similar to application decisions, we will analyze how answers to this question vary based upon gender, treatment, and whether or not a participant was qualified according to our definition. We expect that our Positive and Normative treatments will increase the self-reported probability among qualified applicants, and decrease it among unqualified applicants (relative to the Control). We anticipate that there is a gender gap in reported probability in the Control among qualified applicants (men feel more qualified than women on average), and that the Positive and Normative treatments should reduce this gender gap among qualified applicants. In particular, we will present (Xs indicate coefficients to be reported):

**Table 2. Perceptions of Being Qualified**

|  |  |
| --- | --- |
|  | OLS Predicting Self-Reported Probability of Being Qualified for Expert-Level Job |
|  | All Unqualified | All Qualified |
|  | I | II | III | IV |
| Positive  | X | X | X | X |
| Treatment |  |  |  |  |
|  |  |  |  |  |
| Normative | X | X | X | X |
| Treatment |  |  |  |  |
|  |  |  |  |  |
| Female | X | X | X | X |
|  |  |  |  |  |
|  |  |  |  |  |
| Female x  |  | X |  | X |
| Positive |  |  |  |  |
|  |  |  |  |  |
| Female x Normative |  | X |  | X |
|  |  |  |  |  |
| Controls | Y | Y | Y | Y |

Notes: Qualified candidates are those with a test score greater than or equal to the advertised threshold.

1. We ask, “what do you expect the lowest ASVAB skills test score to be among all the hired applicants for the expert-level job?” We will report the distributions of answers to this question by treatment (Control, Positive, Normative) and gender. We expect that answers will be more variant in the Control treatment and concentrated more tightly around the threshold in the Positive and Normative treatments. We expect that there may be a gender gap in the Control treatment (women expecting a higher bar than men), and that this gender gap should be reduced in the Positive and Normative treatments.
2. We ask participants how sure they are about their guess to the answer above. We expect that participants will be quite unsure in the Control treatment and significantly more sure in the Positive and Normative treatments. We will report the average response to this question by treatment and use two-tailed t-tests to compare.
3. We ask participants how objective, specific, and clear the stated requirements for the expert-level job were. Each participant makes this assessment first for the treatment they were in, and then for the other two treatments (they see just the qualifications language from the other treatments). The key prediction is that participants will believe the language in the Positive and Normative treatments is significantly more objective, specific, and clear than the language in the Control. We are also interested in whether there are any differences between the Normative and the Positive treatments. In particular, if one of these two treatments has a larger impact on behavior (in terms of application rates), we will want to see whether that corresponds with a different perception of the ambiguity as measured here. For instance, suppose, as in the original study, that the Normative treatment has a directionally larger impact of qualified female applicants than the Positive treatment. Then, we will check whether it is the case that the qualifications in the Normative treatment are perceived as less ambiguous than the qualifications in the Positive treatment.

We will analyze average differences across treatment to this question, first using all the data (all participant answers for each treatment; within-subject comparisons included), and then using only data for each treatment from participants who were assigned that treatment; only across-subject comparisons). In the analysis using all data, we will use OLS to predict the answer to this question from a set of dummies for the treatment language that was being evaluated (Positive, Normative, and Control as omitted group), and cluster standard errors at the individual level. We will include the same controls as included above, including ASVAB score, education, demographics from the survey, and a dummy for whether they were qualified. In the analysis using only across-subject comparisons, we will use this same specification but no longer need to the clustered standard errors as there will be just one observation per person.