

Algorithmic Writing Assistance on Jobseekers’ Resumes Increases Hires

Emma van Inwegen
MIT

Zanele Munyikwa
MIT

John J. Horton
MIT & NBER

March 7, 2023

Abstract

There is a strong association between the quality of the writing in a resume for new labor market entrants and whether those entrants are ultimately hired. We show that this relationship is, at least partially, causal: a field experiment in an online labor market was conducted with nearly half a million jobseekers in which a treated group received algorithmic writing assistance. Treated jobseekers experienced an 8% increase in the probability of getting hired. Contrary to concerns that the assistance is taking away a valuable signal, we find no evidence that employers were less satisfied. We present a model in which better writing is not a signal of ability but helps employers ascertain ability, which rationalizes our findings.

1 Introduction

For most employers, the first exposure to a job candidate is typically a written resume. The resume contains information about the applicant—education, skills, past employment, and so on—that the employer uses to draw inferences about the applicant’s suitability for the job. Conveying this information is the most important function of the resume. A better-written resume—without any change in the underlying facts—might make it easier for the employer to draw the correct inferences, which could lead to a greater chance of an interview or job offer. We call this the “clarity view” of the role of resume writing quality. However, a resume might not merely be a conduit for match-relevant information; the resume’s writing itself could signal ability. In particular, the quality of the writing might be informative about the jobseeker’s communication skills, attention to detail, or overall quality. This is another reason better writing could lead to a greater chance of an interview or a job offer. We call this the “signaling view” of the role of resume writing quality.

In this paper, we explore how resume writing quality affects the hiring process using both observational data and a field experiment. First, using observational data from a large

online labor market, we document a strong positive relationship between writing quality and hiring (and not simply callbacks). This relationship persists even after controlling for other factors that might otherwise explain the relationship. In terms of magnitude, one additional error on a job seeker’s resume is associated with 1.4% fewer hires. However, this is only an association and there are other potential reasons writing quality could be correlated with hiring even with our controls. Second, we report the results of a field experiment in which we vary writing quality in the same market. We are primarily interested in understanding if resume writing quality has a causal effect on job market outcomes, and in distinguishing between the “clarity view” and “signaling view.”

In order to address the question of causality, we intercept new jobseekers at the resume-writing stage of registering for the platform and randomly offer some of them—the treatment group—algorithmic writing assistance. Others—the control group—had the status quo experience of no assistance. This writing assistance creates random variation in writing quality. The algorithmic writing assistance was a service provided by a company to the platform and embedded in the text box where jobseekers input their resumes. We will discuss in depth what this assistance, which we refer to as Algorithmic Writing Service provides, but generally, it makes writing better by identifying common errors and offering the writer suggestions on how to address those errors.

In the experimental data, there is a very strong “first stage,” in that those treated had better-written resumes on several quantifiable dimensions. For example, we find fewer spelling and grammar errors in the resumes of the treated group of jobseekers. Positive effects on resume quality were concentrated among the low-end of the distribution in writing quality, as jobseekers with already excellent resumes can benefit little from writing assistance.

After creating a resume, jobseekers engage in search, which may or may not lead to a job. We observe job search behavior and outcomes for both treated and control jobseekers. Treated workers did not send out more applications than workers in the control group, nor did they propose higher wages. This is a convenient result because our interest is in employers’ decision-making, even though randomization was at the level of the jobseeker. If jobseekers had altered their application behavior—perhaps sending more applications because they know they have a stronger case to make—we might wrongly attribute greater job-market success to the resume rather than this endogenous effort.

Our primary outcome of interest is the effect of writing assistance on hiring. We find that treated jobseekers had a 8% increase in their probability of being hired at all relative to the control group. The 95% confidence interval on the percentage increase in hiring is (3%, 13%). They also had 7.8% more job offers over the experimental period than those in the control

group. In terms of the matches themselves, treated workers' hourly wages were 10% higher than the hourly wages of workers in the control group. However, it is important to remember this is a conditional result and could simply be due to composition changes in which workers are hired.

In the “signaling view” the treatment would remove or at least weaken a credible signal of jobseeker ability. If this is the case, this should leave employers disappointed. Unique to our setting, we have a measure of employer disappointment, as both sides privately rate each other at the conclusion of the contract. Although these ratings have been shown to become inflated over time (Filippas et al., Forthcoming) and can be distorted when they are public and reciprocal (Bolton, Greiner and Ockenfels, 2013), they are still a useful signal of worker performance (Fradkin et al., 2021; Cai et al., 2014). If employers are disappointed with the performance of the worker, this would likely manifest in lower employer ratings at the conclusion of the contract. We find no evidence that this is the case.

We look at the impact of the treatment to the public and private numerical ratings the employers give to the workers, as well as the “sentiment” of the written text of reviews (which are less prone to inflation (Filippas et al., Forthcoming)). We find no significant treatment effects for any of these ratings, with both positive and negative point estimates. For example, the average rating of the “quality of work” completed by workers in the control group is 4.768, and 4.772 in the treatment group. The average private rating given to workers in the control group was 8.63 on a ten-point scale, with average ratings in the treatment group of 8.56. The sentiment of the review text in the treatment was slightly more positive, but the effect was close to zero and not significant. Given the 10% higher average wages in the treatment group, if employers were simply tricked into hiring worse workers generally, these higher wages should have made it even more likely to find a negative effect on ratings (Luca and Reshef, 2021).

One possible explanation for our results is that employers are simply wrong in regarding resume writing quality as informative about ability. However, the “clarity view” can also rationalize our results without making this assumption. It is helpful to formalize this notion to contrast it with the more typical signaling framing of costly effort and hiring. To that end, we present a simple model where jobseekers have heterogeneous private information about their productivity but can reveal their type via writing a “good” resume. This is not a signaling model where more productive workers face lower resume-writing costs—any worker, by writing a good resume, will reveal their information, and this cost is assumed to be independent of actual productivity. Our model has heterogeneous “good” resume writing costs. We show that writing assistance shifting the cost distribution can generate our findings of more hires, higher wages, and equally satisfied employers.

Our main contribution is to compare the “clarity view” and “signaling view” for the positive relationship between writing and hiring. Our main substantive finding is evidence for the “clarity view.” We can do this because we can trace the whole matching process from resume creation all the way to a measure of post-employment satisfaction. Helping jobseekers have better-looking resumes helped them get hired (consistent with both explanations), but we find no evidence that employers were later disappointed (which is what the “signaling view” explanation would predict). We also contribute more broadly by showing the importance of text in understanding matching (Marinescu and Wolthoff, 2020). The notion that better writing can help a reader make a better purchase decision is well-supported in the product reviews literature (Ghose and Ipeirotis, 2010) but is a relatively novel finding in labor markets. In one related example, (Sajjadi et al., 2019) analyze resumes of applicants to public school teaching jobs and find that spelling accuracy is associated with a higher probability of being hired. And Hong, Peng, Burtch and Huang (2021) show that workers who directly message prospective employers (politely) are more likely to get hired, but the politeness effect is muted when the workers’ messages contain typographic errors.

In addition to the general theoretical interest in understanding hiring decisions, there are practical implications to differentiating between these two views of the resume. If the “clarity view” is more important, then any intervention that encourages better writing is likely to be beneficial. There will likely be little loss in efficiency if parties are better informed. Even better, as we show, the kind of assistance that improves clarity can be delivered algorithmically. These interventions are of particular interest because they have zero marginal cost (Belot et al., 2018; Briscese et al., 2022; Horton, 2017), making a positive return on investment more likely, a consideration often ignored in the literature (Card et al., 2010). On the other hand, if the “signaling view” is more important, then providing such writing assistance will mask important information and lead to poor hiring decisions.

Unlike general advice, algorithmic interventions are adaptive. In our study, the algorithm took what the jobseeker was trying to write as input and gave targeted, specific advice on improvement. This is likely more immediately useful than more vague recommendations, such as telling jobseekers to “omit needless words.” This advice comes in the form of recommendations that are predicted to improve the resume’s effectiveness. A limitation of our study is that we cannot speak to crowd-out effects (Crépon et al., 2013), which are relevant to discuss the welfare implications of any labor market intervention. However, this concern is somewhat secondary to our narrower purpose of understanding how employers make decisions with respect to resumes. Additionally, given that in our setting, new entrants compete with established jobseekers on the platform, we anticipate the crowd-out effect will be small, and perhaps even welcome if at the expense of more established workers, given the obstacles

new entrants face (Pallais, 2013).

In addition to exploring an AI technology in a real labor market, we contribute to a large literature on how experimentally varying applicant attributes affects callback rates (Moss-Racusin et al., 2012; Bertrand and Mullainathan, 2003; Kang et al., 2016; Farber et al., 2016). While we are not the first to show that writing matters in receiving callbacks from employers (Sterkens et al., 2021; Martin-Lacroux and Lacroux, 2017), we are the first to do so on such a massive scale and with natural variation in writing quality¹. Our experiment involves 480,950 jobseekers which is an order of magnitude larger than the next largest experiments. Another benefit is that we do not need to guess how workers might make mistakes on their resumes, as it is workers and not researchers writing their resumes. Additionally, unique in this literature, we can follow the induced changes all the way through hiring and even post-employment assessment which allows us to answer our “clarity view” vs. “signaling view” questions.

The rest of the paper proceeds as follows. Section 2 describes the online labor market which serves as the focal market for this experiment. Section 3 reports the experimental results of the treatment effects on writing quality and subsequent labor market outcomes. In Section 4 we present a simple model that can rationalize our findings. Section 5 concludes.

2 Empirical context and experimental design

The setting for this experiment is a large online labor market. Although these markets are online, with a global audience, and with lower search costs (Goldfarb and Tucker, 2019), they are broadly similar to more conventional markets (Agrawal et al., 2015). Employers post job descriptions, jobseekers apply, and there are interviews followed by hiring and managing. One distinctive feature of online labor markets is that both the employer and the worker provide ratings for each other at the end of a contract.

Because of the many similarities between on and offline labor markets, a growing body of research uses online labor markets as a setting, often through randomized experiments. These studies contribute to the theory in longstanding questions about labor markets, such as deepening our understanding of the mechanisms and processes by which employers and workers find jobs. Online labor markets also allow researchers to broaden the range of questions in which it is possible to make causal estimates (Horton, 2010; Barach and Horton, 2021) because platforms store detailed data on things like applications, text, length of time spent working on an application, speed of hire, and much more.

¹While the reason this preference exists is not known, recruiters report, anecdotally, caring about a resume’s writing quality (Oreopoulos, 2011).

Many studies on online labor markets identify and measure phenomena that are relevant to labor markets both online and offline. Like the offline labor market, online labor markets have been shown to have hiring biases (Chan and Wang, 2018). But, Agrawal et al. (2016) shows that these biases tend to be ameliorated with experience and that general, employers are able to learn as they hire (Kokkodis and Ransbotham, 2022). And Stanton and Thomas (2016) shows that in an online labor market, agencies (which act as quasi-firms) help workers find jobs and break into the marketplace.

2.1 Search and matching on the platform

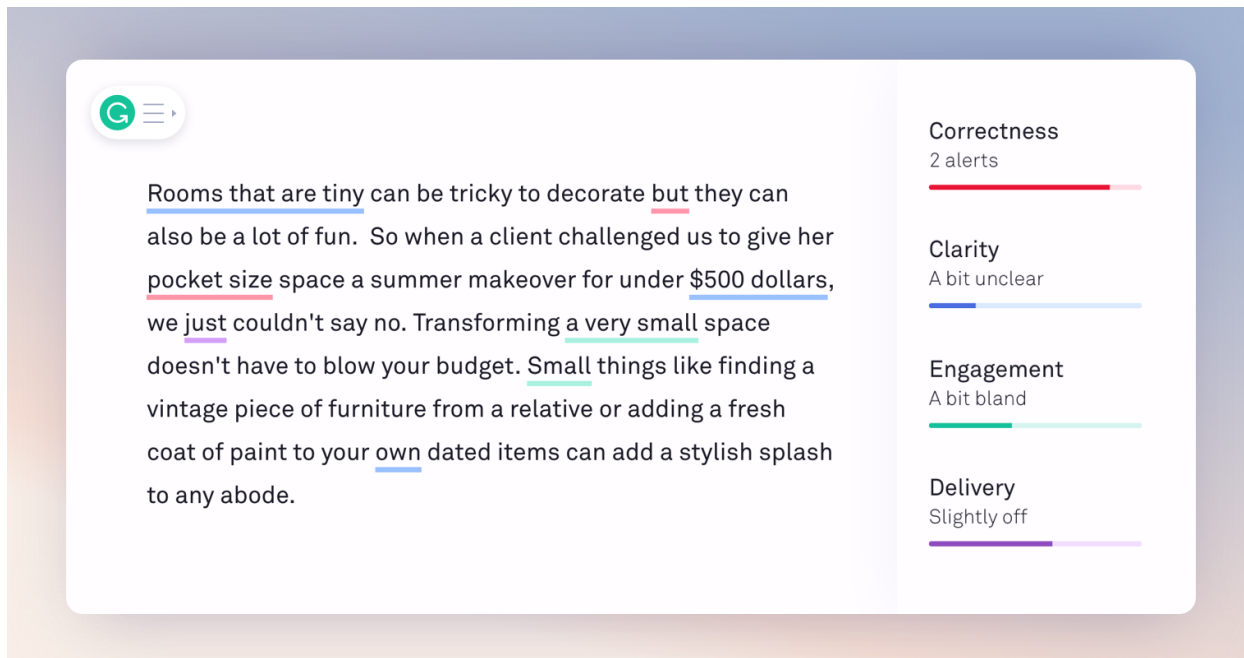
A would-be employer writes job descriptions, labels the job opening with a category (e.g., “Graphic Design”), lists required skills, and then posts the job opening to the platform website. Jobseekers generally learn about job openings via electronic searches. They submit applications, including a wage bid and a cover letter. In addition to jobseeker-initiated applications, employers can also use the interface to search worker profiles and invite workers to apply to particular jobs. The platform uses the jobseeker’s history and ratings on the platform to recommend jobseekers to would-be employers (Horton, 2017). Despite platforms making algorithmic recommendations, none are based on the writing quality of their resume. In terms of selection, Pallais (2013) shows that employers in an online labor market care about workers’ reputation and platform experience when hiring. After jobseekers submit applications, employers screen the applicants, decide whether to give interviews, and then whether to make an offer(s).

2.2 Experimental intervention at the resume-writing stage of profile creation

When new jobseekers sign up to work on the platform, their first step is to register and create their profile. This profile serves as the resume with which they apply for jobs. This profile includes a list of skills, education, and work experience outside of the platform, as well as a classification of their primary job category (e.g., “Graphic Design”), mirroring what employers select when posting a job. The interface consists of a text box for a profile title and a longer one for a profile description. Jobseekers either enter their profile information on the spot or they can copy and paste it from somewhere else.

During the experimental period, jobseekers registering for the platform were randomly assigned to an experimental cell. The experimental sample comprises jobseekers who joined the platform between June 8th and July 14th, 2021. For treated jobseekers, the text boxes for the profile description are checked by the Algorithmic Writing Service. Control jobseek-

Figure 1: Example of the Algorithmic Writing Service 's interface showing suggestions on how to improve writing



Notes: Example of the Algorithmic Writing Service applied to a paragraph of text. To receive the suggestions, users hover their mouse over the underlined word or phrase. For example, if you hover over the first clause “Rooms that are tiny” underlined in blue, “Tiny rooms” will pop up as a suggestion.

ers received the status quo experience. The experiment included 480,950 jobseekers, with 50% allocated to the treated cell. Table 1 shows that it was well-balanced and the balance of pre-treatment covariates was consistent with a random process.

2.3 The algorithmic writing assistance

Words and phrases which are spelled wrong or used incorrectly are underlined by the Algorithmic Writing Service. See Figure 1 for an example of the interface with an example of the text “marked up” by the Algorithmic Writing Service. By hovering a mouse cursor over the underlined word or phrase, the user sees suggestions for fixing spelling and grammar errors. The Algorithmic Writing Service also gives advice about punctuation, word usage, phrase over-use, and other attributes related to clarity, engagement, tone, and style.

2.4 Platform profile approval

When jobseekers finish setting up their profiles, they have to wait to be approved by the platform. The platform approves jobseekers who have filled out all the necessary information and uploaded an ID and bank details. The platform can also reject jobseekers at their

discretion. However, platform rejection is somewhat rare. About 10 percent of profiles are rejected, usually as a part of fraud detection or because the jobseekers leave a completely empty profile. 46% of workers who were allocated into the experiment upon registration complete and submit their profiles. About 41% of workers who begin registering get all the way through the approval process.

As approval is made following profile creation, this platform step creates a potential problem for interpreting any intervention that changes profile creation. For example, it could be that better writing just led to a greater probability of platform approval. Or, it could have caused jobseekers to be more likely to complete the registration process and submit their profile, both of which could effect hiring. While unlikely, this is possible, and we do several things to deal with this potential issue.

First, see whether there is any evidence of selection. We find no evidence that treated jobseekers were more likely to be approved. We show that treated jobseekers are no more likely to submit their profiles and that approval too is unaffected by the treatment².

Second, in our main analysis, we condition on profile approval in our regressions. We also do robustness checks where we report the same analysis not conditioned on profile approval and where we control for profile approval as a covariate. All our results are robust to these strategies and are described in Section 3.11.

Once a jobseeker is approved, they can begin applying for jobs posted on the platform. Their profile will include their resume and a “profile hourly wage” which is the wage offer to employers searching for workers. After they complete their first job on the platform, their profile also shows the worker’s actual wages and hours worked on jobs found through the platform.

2.5 Description of data used in the analysis

The dataset we use in the analysis consists of the text of jobseekers’ resumes as well as all of their behavior on the platform between the time they registered and August 14th, 2021, one month after allocations ended. We construct jobseeker level data including the title and text of their profile, the number of applications they send in their first month on the platform, the number of invitations to apply for jobs they receive, the number of interviews they give, and the number of contracts they form with employers. The most common categories worker’s list as their primary job categories are Design & Creative, Writing, Administrative Support, and Software Development, in order of frequency.

In Table 1 we present summary statistics about the jobseekers in the full experimental

²See Appendix Table 12.

Table 1: Comparison of jobseeker covariates, by treatment assignment

	Treatment mean: \bar{X}_{TRT}	Control mean: \bar{X}_{CTL}	Difference in means: $\bar{X}_{TRT} - \bar{X}_{CTL}$	p-value
<i>Full sample description: N = 480,948</i>				
Resume submitted	0.456 (0.001)	0.455 (0.001)	0.001 (0.001)	0.452
Platform approved	0.407 (0.001)	0.406 (0.001)	0.002 (0.001)	0.186
Resume length	32.911 (0.116)	32.860 (0.117)	0.051 (0.165)	0.755
Profile hourly rate	18.843 (0.126)	18.917 (0.126)	-0.074 (0.178)	0.676
<i>Flow from initial allocation into analysis sample</i>				
	<i>Treatment (N)</i>	<i>Control (N)</i>	<i>Total (N)</i>	
Total jobseekers allocated	240,232	240,718	480,950	
↪ who submitted their profiles	109,639	109,604	219,243	
↪ and were approved by the platform	97,860	97,610	195,470	
↪ with non-empty resumes	97,480	97,221	194,701	
<i>Pre-allocation attributes of the analysis sample: N = 194,700</i>				
From English-speaking country	0.182 (0.001)	0.183 (0.001)	-0.002 (0.002)	0.362
US-based	0.141 (0.001)	0.143 (0.001)	-0.002 (0.002)	0.222
Specializing in writing	0.166 (0.001)	0.168 (0.001)	-0.002 (0.002)	0.151
Specializing in software	0.115 (0.001)	0.115 (0.001)	0.000 (0.001)	0.770
Resume length	70.394 (0.222)	70.260 (0.222)	0.135 (0.314)	0.668

Notes: This table reports means and standard errors of various pre-treatment covariates for the treatment group and the control group. The first panel shows the post-allocation outcomes of the full experimental sample i) profile submission, ii) platform approval, iii) length of resume in the number of words, iv) profile hourly wage rate in USD. The means of profile hourly rate in treatment and control groups are only for those profiles which report one. The reported p-values are for two-sided t-tests of the null hypothesis of no difference in means across groups. The second panel describes the flow of the sample from the allocation to the sample we use for our experimental analysis. The complete allocated sample is described in the first line, with each following line defined cumulatively. The third panel looks at pre-allocation characteristics of the jobseekers in the sample we use for our analysis, allocated jobseekers with non-empty resumes approved by the platform. We report the fraction of jobseekers i) from the US, UK, Canada, or Australia, ii) from the US only, iii) specializing in writing jobs, iv) specializing in software jobs, and v) the mean length of their resumes in the number of words.

sample as well as the sample conditioned on platform approval. 17% of the jobseekers specify that writing jobs are their primary area of work. Only 14% of jobseekers are based in the US, and over 80% are based in a country where English is not the native language.

2.6 Constructing measures of writing quality

We do not observe the changes that the Algorithmic Writing Service suggested—we simply observe the resumes that result. As such, we need to construct our own measures of writing quality to determine if the treatment was delivered.

Algorithmic Writing Service gives suggestions to writers about how to improve text along several dimensions. Perhaps the most straightforward measure of writing quality is spelling. To see if the treatment impacted spelling errors, we take each worker’s profile and check if each word appears in an English language dictionary. We use the dictionary `hunspell`, which is based on MySpell dictionaries and is the basis for the spell checker for Google Chrome, Firefox, and Thunderbird.

As many of the resumes are for technical jobs, they often contain industry-specific terms such as “UX” or brand names like “Photoshop.” To prevent these from being labeled as errors, we augmented the list of words in the dictionary by checking the 1,000 most commonly “misspelled” words in our sample and adding non-errors manually.

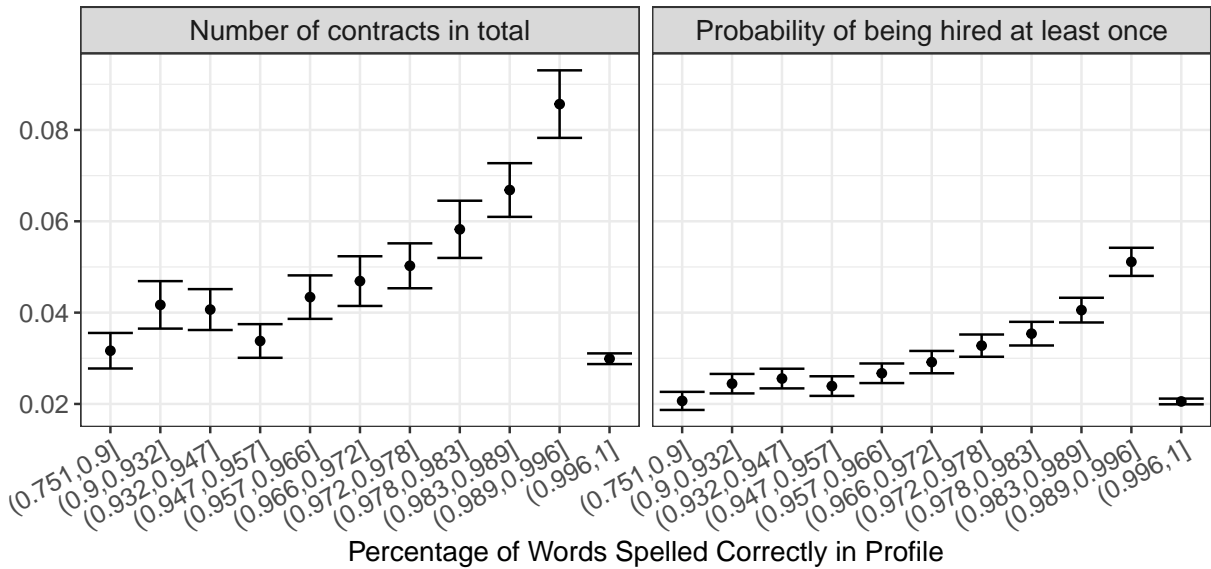
Spelling is not the only measure of writing quality. To broaden our measures, we use LanguageTool, an open-source software that finds many errors that a simple spell checker cannot detect, to understand employers care about measures of writing quality other than simply the number of spelling mistakes. LanguageTool is a rule-based dependency parser that identifies errors (rule violations) and categorizes them. Some example categories include “Nonstandard Phrases,” “Commonly Confused Words,” “Capitalization,” and “Typography.” For example, the nonstandard phrase “I never have been” would be flagged with a suggestion to replace it with “I have never been.”³

2.7 Spelling errors are associated with lower hiring probabilities in the observational data

Before presenting results of the experiment, we explore the relationship between resume writing quality and hiring in observational data from this market. We begin by studying the most unambiguous measure of writing quality: spelling. In Figure 2 we plot the relationship between hiring outcomes and the percentage of words spelled correctly on the resumes of all jobseekers who registered for the platform over the month of May 7th through June 7th, 2021, prior to the experiment. Because the distribution of percent correctly spelled is so left skewed, we truncate the sample to only those who spell at least 75% of the words in their resumes correctly. This window includes 97% of jobseekers. The x-axis is deciles between 75% and 100% of words spelled correctly.

³For a more detailed explanation of all of the rule categories, see Table 7 in Appendix A.

Figure 2: Association between spelling errors and hiring outcomes in the observational data



Notes: These data show the relationship between the percentage of correctly spelled words on a jobseekers' resume with various hiring outcomes. A 95% confidence interval is plotted around each estimate. The sample is of all new jobseekers who registered and were approved for the platform between June 1st and June 7th, 2021, and had resumes with more than 10 words. Plots are truncated at those who spelled at least 75% of the words in their resume correctly.

Job seekers with resumes with fewer spelling errors are more likely to be hired. In the left facet, the y-axis is the number of contracts a jobseeker forms in their first month on the platform. Jobseekers with over 99% of the words in their resume spelled correctly are hired nearly three times more in their first month on the platform than jobseekers with less than 90% spelled correctly. In the right facet, the y-axis is the probability that a jobseeker is ever hired in their first month on the platform. However, as is visible in both facets, resumes with 100% of words spelled correctly are much *less* likely to receive interest from employers. This is likely because those resumes tend to be much shorter than the others—the average length of a resume that has zero spelling errors is only 47 words long.

2.8 The association between various kinds of writing errors and hiring probabilities

Moving beyond spelling, in Table 2, we show the correlation between hiring outcomes on each type of language error in the resumes in the observational data.⁴ In the first specification we show the correlation between the error rate for the various types of language errors

⁴In Table 8 of Appendix A we summarize the occurrence of other types of errors within the observational data.

and the number of contracts formed over the jobseeker’s first month on the platform. In the second specification the outcome is simply whether or not the jobseeker was ever hired in their first month. In Columns (3) and (4), we control for the jobseekers’ profile hourly rate and their job category. Resumes with more per word grammar errors, typos, typography errors, and miscellaneous errors are all hired less. This linear model places some unreasonable assumptions like constant marginal effects on the relationship between various writing errors and hiring. There may be interactions between these error types. However, it is still useful to summarize the relationships. We can see generally negative relationships between a higher writing error rate and hiring.

Interestingly, more style errors *positively* predict hiring. While initially surprising, style errors are often caused by language being unnecessarily flowery. Some examples of style errors are “Moreover, the street is almost entirely residential” and “Doing it this way is more easy than the previous method.” This implies that despite employers’ dislike of most writing errors, they forgive or even prefer this kind of flowery language.

For robustness we repeat these analysis in levels in Appendix Table 17. And in Appendix Table 18 we collapse all error types into one measure of Total Errors and report these results in both levels and normalized by resume length. The negative relationship between writing errors and hiring persists in all specifications.

3 Effects of the treatment

We look at two main kinds of experimental results. First, we examine how the treatment affected the text of resumes. We are looking to see whether there is a “first stage.” Next, we look at market outcomes for those treated workers. For convenience, we present these treatment effects as percent changes, in Figures 3 and 5.

3.1 Algorithmic writing assistance improved writing quality

The first step is to measure the effect the Algorithmic Writing Service has on writing in the treatment group. We start with the fraction of words in the resume spelled incorrectly. In the control group, resumes are 70 words long on average. Even the worst spellers spell most of the words correctly, and an average resume has 96% of the words spelled correctly.

To understand the effects of the treatment on other types of writing errors we return to the more fine-grained LanguageTool definitions of writing errors. In Figure 3, we look at the effect of treatment on the number of each type of writing error, normalized by resume

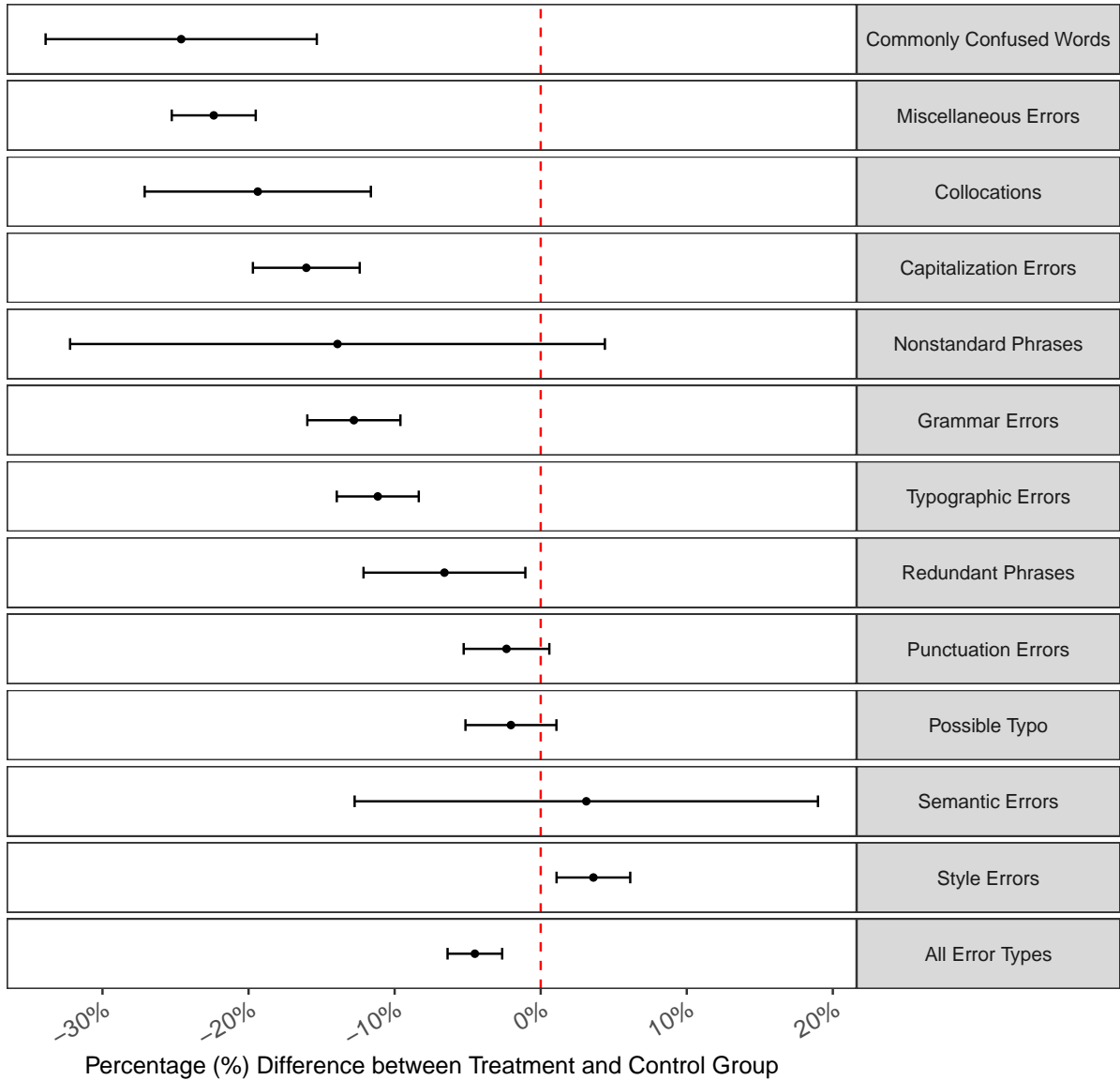
Table 2: Hiring outcomes predicted based on language errors (normalized by word count) in observational data

	<i>Dependent variable:</i>			
	Number of Contracts (1)	Hired (2)	Number of Contracts (3)	Hired (4)
Capitalization Error	-0.075 (0.048)	-0.038 (0.025)	-0.055 (0.045)	-0.026 (0.023)
Possible Typo	-0.030** (0.013)	-0.022*** (0.007)	-0.016 (0.012)	-0.013** (0.006)
Grammar Error	-0.534*** (0.097)	-0.314*** (0.051)	-0.360*** (0.092)	-0.210*** (0.047)
Punctuation Error	-0.0001 (0.006)	0.0002 (0.003)	-0.0002 (0.006)	0.0001 (0.003)
Typography Error	-0.098*** (0.026)	-0.069*** (0.014)	-0.066*** (0.025)	-0.050*** (0.013)
Style Error	0.261** (0.119)	0.130** (0.062)	0.234** (0.112)	0.115** (0.058)
Miscellaneous Error	-0.414*** (0.151)	-0.220*** (0.079)	-0.252* (0.143)	-0.121 (0.074)
Redundant Phrases	-0.433 (0.437)	-0.264 (0.229)	-0.240 (0.414)	-0.149 (0.213)
Nonstandard Phrases	0.804 (1.681)	-0.124 (0.882)	0.699 (1.591)	-0.193 (0.819)
Commonly Confused Words	-0.761 (0.618)	-0.331 (0.324)	-0.531 (0.584)	-0.190 (0.301)
Collocations	-0.637 (0.434)	-0.380* (0.228)	-0.438 (0.411)	-0.262 (0.211)
Semantic Error	-0.340 (1.112)	-0.532 (0.583)	-0.191 (1.052)	-0.445 (0.541)
Constant	0.053*** (0.002)	0.036*** (0.001)	0.036*** (0.002)	0.026*** (0.001)
Controls			X	X
Observations	65,114	65,114	65,114	65,114
R ²	0.001	0.002	0.106	0.140

Notes: This table analyzes correlation between various writing errors on jobseekers' resumes and their hiring outcomes. The independent variables, writing errors, are divided by the number of words in the jobseekers' resume. Number of Contracts is defined as the number of unique jobs they work over the month after they register for the platform. Hired is defined as 1 if the jobseeker was ever hired over that month, and 0 if else. Columns (3) and (4) include controls for profile hourly rate and job category. Writing errors are defined by LanguageToolR. The sample is made up of all jobseekers who registered for the platform in the week before the experiment who submitted non-empty resumes.

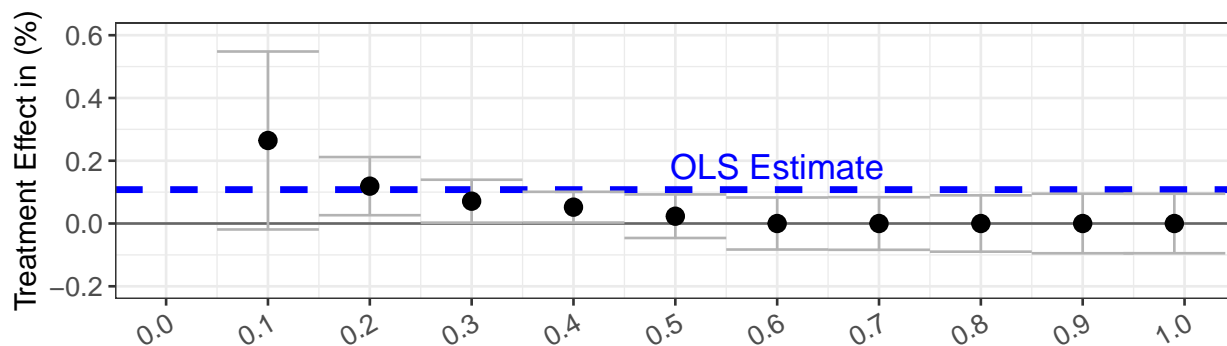
Significance indicators: $p \leq 0.10$: *, $p \leq 0.05$: ** and $p \leq .01$: ***.

Figure 3: Effect of the algorithmic writing assistance on writing quality measures



Notes: This plot shows the effect of the treatment on various writing errors in jobseekers' resumes. Point estimates are the percentage change in the dependent variable versus the control group for the treatment groups. A 95% confidence interval based on standard errors calculated using the delta method is plotted around each estimate. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes, with $N = 194,701$. Regression details can be found in Tables 10 and 11 of the Appendix.

Figure 4: Effect of treatment on percentage of words spelled correctly, by deciles



Notes: This plot shows the effect of the treatment on the percentage of words spelled correctly in jobseekers’ resumes, by deciles. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes, with $N = 194,701$.

length.⁵ Our outcomes of interest are the error rate for each type, so we normalize each type of error to the number of words in the resume. For treatment effects measured in percentage terms we calculate the standard errors using the delta method.

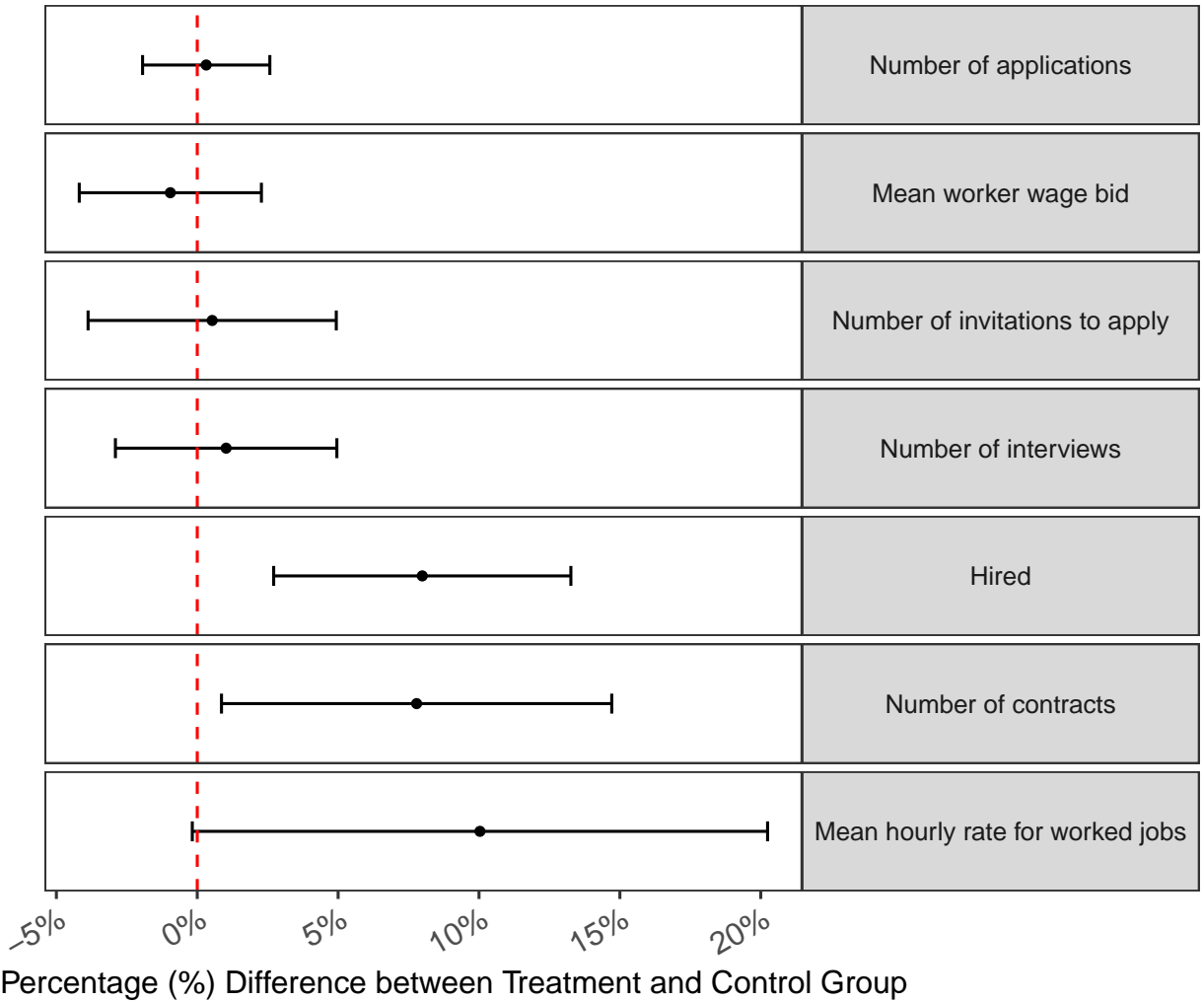
We find that jobseekers in the control group had significantly higher rate of errors of the following types: capitalization, collocations, commonly confused words, grammar, spelling, possible typos, miscellaneous, and typography. We find larger treatment effects for errors associated with writing clarity than for many others. For example, two of the largest magnitudes of differences in error rate were commonly confused words and collocations, where two English words are put together that are not normally found together. Interestingly, the treatment group had more “style” errors, paralleling our results from the observational data, Table 2.

3.2 Algorithmic assistance helped the worst writers more

The treatment was predominantly effective for jobseekers at the bottom of the spelling distribution. In Figure 4 we report results from a quantile regression on the effect of the treatment on the percentage of words they spelled correctly. The effect is concentrated in jobseekers in the bottom half of the spelling distribution. The treatment effect is largest for jobseekers below the 30% decile, with effects decreasing at each decile until the median at which point the treatment did not affect spelling.

⁵The treatment had no effect on the length of resumes—see Table 9 in Appendix A.

Figure 5: Effect of algorithmic writing assistance on hiring outcomes



Notes: This analysis looks at the effect of treatment on hiring outcomes on jobseekers in the experimental sample. The x-axis is the difference in the mean outcome between jobseekers in the treated group and the control group. A 95% confidence interval based on standard errors calculated using the delta method is plotted around each estimate. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021, and had non-empty resumes, with $N = 194,701$. Regression details on the number of applications and wage bid can be found in Table 4. Regression details on invitations, interviews, hires, and the number of contracts can be found in Appendix Table 13. Regression details on hourly wages can be found in Table 6.

3.3 Heterogeneous treatment effects to spelling

A natural question is whether effects differed by jobseeker background. In Table 3 we interact pre-randomization jobseeker attributes with the treatment. We can see that jobseekers from the US, from English-speaking countries,⁶ and who are writers all do better in “levels.” We find that jobseekers from countries that are not native English speaking experience significantly larger treatment effects to the fraction of words they spell correctly than their anglophone counterparts.

Table 3: Effects of writing assistance on error rate

	<i>Dependent variable:</i>			
	Total Error Rate x 100			
	(1)	(2)	(3)	(4)
Algo Writing Treatment (Trt)	-0.578*** (0.073)	-0.697*** (0.081)	-0.667*** (0.079)	-0.581*** (0.080)
Anglophone Country		-4.767*** (0.133)		
Trt × Anglo		0.609*** (0.188)		
US			-4.683*** (0.147)	-4.388*** (0.104)
Trt × US			0.561*** (0.208)	
Writer				-0.729*** (0.138)
Trt × Writer				-0.046 (0.195)
Constant	7.978*** (0.052)	8.854*** (0.057)	8.651*** (0.056)	8.732*** (0.058)
Observations	187,858	187,858	187,858	187,858
R ²	0.0003	0.012	0.010	0.010

Notes: In Column (1) we show the overall effect of the treatment to the number of errors on a jobseekers’ resume divided by the number of words. In Column (2) we interact the treatment with a dummy variable for if the jobseeker is from the US, UK, Canada, or Australia. In Column (3) we interact the treatment with a dummy for if the jobseeker is in the US. In Column (4) we interact the treatment with a dummy for if the jobseeker lists Writing as their primary category of desired work. The experimental sample is of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10$: *, $p \leq 0.05$: ** and $p \leq .01$: ***.

⁶We define whether a jobseeker is from a native English-speaking country, by whether they login to the platform from USA, UK, Canada, or Australia.

3.4 Treated workers did not change their job search strategy or behavior

One potential complication in our desire to focus on employer decision-making is that the treatment could have impacted jobseekers search behavior or intensity. Suppose treated jobseekers changed their behavior, knowing they had higher quality resumes. In that case, we could not interpret our treatment effect as being driven by employers' having improved perceptions of treated jobseekers. However, we find no evidence that jobseekers changed their search behavior due to the treatment. In the first facet of Figure 5, the outcome is the number of applications a jobseeker sends out over their first 28 days after registering. We find no effect of the treatment on the number of applications sent.

In the second facet, the outcome is the mean wage bid proposed by the jobseekers on their applications in their first 28 days on the platform. Average wage bids in both the treatment and control group were \$24 per hour. The lack of effects on jobseekers' behaviors makes sense because they were unaware of the treatment.

Table 4 show the effects of the treatment on jobseekers application behavior. In Column (1) we see whether treated jobseekers applied for more jobs than those in the control group over the experimental period and find they did not. In Column (2) we find that treated jobseekers do not apply to more hourly jobs than those in the control group. They also could have bid for higher wages knowing they had better-looking resumes. In Column (3) we see no evidence of this, where we narrow the sample to only applications to hourly jobs and look at the effect of the treatment on hourly wage bids.

3.5 The treatment did not affect employer recruiting

Employers were able to seek out workers using the platform's search feature to invite jobseekers to apply to their job openings. In Figure 5's third facet from the top, the outcome is the number of invitations to apply for a job that the jobseeker receives in their first month. We find the effect of the treatment on employer invitations is a precise zero. In the fourth facet from the top, the outcome is the number of interviews a jobseeker gives over their first month on the platform. We find that this is also zero. Table 13 Column (4) provides the details of this regression.

Although it may seem surprising given the results on hires and contracts, it makes sense given that our experimental sample consists of only new jobseekers to the platform. New entrants almost never appear in the search results when employers search for jobseekers, given that their rank is determined by their platform history.

In the fourth facet of Figure 5, we show no effect of the treatment on number of inter-

Table 4: Effects of writing assistance on jobseekers’ application behavior

	<i>Dependent variable:</i>		
	Num Applications (1)	Num Hourly Applications (2)	Mean Hourly Wage Bid (3)
Algo Writing Treatment	0.008 (0.027)	0.008 (0.017)	-0.232 (0.402)
Constant	2.337*** (0.019)	1.235*** (0.012)	24.230*** (0.284)
Observations	194,701	194,701	65,411
R ²	0.00000	0.00000	0.00001

Notes: This table analyzes the effect of the treatment on jobseekers’ application behavior. The experimental sample is made up of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes. The outcome in Column (1) is the number of total applications a jobseeker sent out between the time the experiment began and one month after it ended. The outcome in Column (2) is the number of specifically hourly applications sent out in that same time period. The outcome in Column (3) is the mean hourly wage bid they proposed for those hourly jobs, and the sample narrows to only jobseeker who submitted at least one application to an hourly job.

Significance indicators: $p \leq 0.10$: *, $p \leq 0.05$: ** and $p \leq .01$: ***.

views. Interviews, while technically feasible, are very uncommon on this platform. In the control group the average jobseeker gives 0.2 interviews over the course of their first month after registering. Table 13 Column (3) provides the details of this regression.

3.6 Treated jobseekers were more likely to be hired

The treatment raised jobseekers’ hiring probability and the number of contracts they formed on the platform. In the fifth facet of Figure 5, the outcome is a binary indicator for whether or not a jobseeker is ever hired in their first 28 days on the platform. During the experiment, 3% of jobseekers in the control group worked at least one job on the platform. Treated jobseekers see an 8% increase in their likelihood of being hired in their first month on the platform. In Table 5 Column (1) we report these results in levels.

Jobseekers in the treated group formed 7.8% more contracts overall. In the sixth facet of Figure 5, the outcome is the number of contracts a jobseeker worked on over their first month.

3.7 Hourly wages in formed matches were higher

Treated workers had 10% higher hourly wages than workers in the control group.

Table 5: Effects of writing assistance on hiring, by sub-groups

	<i>Dependent variable:</i>			
	Hired x 100			
	(1)	(2)	(3)	(4)
Algo Writing Treatment (Trt)	0.247*** (0.080)	0.223** (0.088)	0.242*** (0.086)	0.237*** (0.088)
Anglophone Country		2.508*** (0.146)		
Trt × Anglo		0.155 (0.207)		
US			2.602*** (0.161)	
Trt × US			0.072 (0.228)	
Writer				-0.293* (0.151)
Trt × Writer				0.060 (0.214)
Constant	3.093*** (0.057)	2.632*** (0.063)	2.719*** (0.061)	3.142*** (0.062)
Observations	194,703	194,703	194,703	194,703
R ²	0.00005	0.003	0.003	0.0001

Notes: This table analyzes the effect of the treatment on whether or not a jobseeker was ever hired on the platform in the month after they joined, times 100. In Column (1) we show the overall effect of the treatment to hiring. In Column (2) we interact the treatment with a dummy variable for if the jobseeker is from the US, UK, Canada, or Australia. In Column (3) we interact the treatment with a dummy for if the jobseeker is in the US. In Column (4) we interact the treatment with a dummy for if the jobseeker lists Writing as their primary category of desired work. The experimental sample is of all new jobseekers who registered and were approved by the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10$: *, $p \leq 0.05$: ** and $p \leq .01$: ***.

In the seventh facet, the outcome is the mean hourly rate workers earned in jobs they worked over their first month on the platform.⁷

In the control group, workers on average made \$16.80 per hour. In the treatment group, workers made \$18.48 per hour, a significant difference at the 0.042 level. Since workers did not bid any higher, this result suggests that employers are hiring more productive workers, or that they thought the treated workers were more productive. If it is the latter, the “signaling view” would predict that employers would then be disappointed with the workers they hired, which we should be able to observe in worker ratings.

Table 6: Effect of algorithmic writing assistance on average wages and ratings of worked jobs

	<i>Dependent variable:</i>				
	Hourly wage rate	Private rating	Positive text review	Recieved rating	Recieved text review
	(1)	(2)	(3)	(4)	(5)
Algo Writing Treatment	1.685** (0.830)	-0.077 (0.082)	-0.004 (0.009)	-0.002 (0.012)	0.006 (0.008)
Constant	16.796*** (0.605)	8.633*** (0.059)	0.874*** (0.006)	0.624*** (0.008)	0.138*** (0.006)
Observations	2,816	4,250	4,529	6,263	6,263
R ²	0.001	0.0002	0.00005	0.00001	0.0001

Notes: This analysis looks at the effect of treatment on outcomes of worked jobs for jobseekers in the experimental sample. Column (1) defines hourly wage rate as the mean hourly wage rate paid for all hourly jobs worked. Column (2) defines private rating as the mean private rating on all jobs given by employers to the workers after the job ended. In Column (3) we take the text of the reviews left by employers on each job and use sentiment analysis (model: distilbert-base-uncased-finetuned-sst-2-english) to impute whether the review is positive or negative, labeled one or zero. The outcome is the mean of these ratings over all worked jobs in the sample. Column (4) is the percentage of contracts worked where the freelancer recieved any private rating. And Column (5) is the percentage of contracts worked where the freelancer recieved any text based review. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10$: *, $p \leq 0.05$: ** and $p \leq .01$: ***.

⁷Hourly wage rates for new entrants are not representative of rates on the platform. If a new entrant gets hired for their first job, they tend to experience rapid wage growth.

3.8 Employers satisfaction was unaffected by the treatment

At the end of every contract, employers rate and review the workers by reporting both public and private rating to the platform. Private ratings are not shared with the worker. In the control group, workers had an average private rating of 8.63. In Table 6 we show that treated workers who formed any contracts over the experimental period did not have statistically different private ratings than workers in the control group. In Column (2) we show that workers in the treated group have an average private rating of 8.56 with a standard error of 0.08.

When the employers give these ratings they are also able to leave text reviews. While numerical ratings have become inflated in recent years, [Filippas et al. \(Forthcoming\)](#) show that the sentiments associated with the text of reviews has increased significantly less over time. This means that text reviews are likely more informative about the workers' quality than the numerical ratings. We use a BERT text classification model ([HF Canonical Model Maintainers, 2022](#)) to label each review as having positive or negative sentiment. These classifications are significantly correlated with the private ratings, with a Pearson correlation coefficient of 0.54. In Column (3) of Table 6 we show that the treated workers' average text reviews are not statistically different from the average sentiment of the reviews for control workers. We may also worry that if employers are less happy with the workers quality or productivity, that they may be less likely to leave a review at all. In Column's (4) and (5) we show that workers in the treatment group are not more or less likely to receive any rating or text reviews than workers in the control group.

Lastly, in Appendix Table 16 we report the results of the effect of the treatment on the employers' public ratings of the workers. Each outcome is a public rating the employers give to the workers at the end of a contract. Employers rate the workers communication, skills, quality of work, availability, cooperation, and ability to make deadlines. Each rating is given on a five point scale. There is less variation in the public ratings than in the private ones, and the average rating for each attribute is over 4.75 stars. Like the private ratings, there are no significant effects of the treatment to any of the ratings, including to workers' communication skills. And the point estimate of the treatment effect to the quality of the work done is even positive.

3.9 How much power do we have to detect worse contractual outcomes?

Given the null effect of the treatment to ratings, a natural question is how much power is available to detect effects. While we do find a substantial increase in hiring—8%—these

marginal hires are mixed in with a much larger pool of “inframarginal” hires that would likely be hired anyway, but for our intervention. How much worse could those marginal applicants have been and still get our results to private ratings in the treatment?

Let I indicate “inframarginal” jobseekers who would have been hired in the treatment or control. Let M indicate “marginal” jobseekers who are only hired in the treatment. For workers in the control group, the average private rating will be $\bar{r}_C = \bar{r}_I$. But for the treatment, the mean rating is a mixture of the ratings for the inframarginal and the ratings for the induced, marginal applicants, and so

$$\bar{r}_T = \frac{\bar{r}_I + \tau \bar{r}_M}{1 + \tau} \quad (1)$$

where τ is the treatment effect. We assume no substitution, making our estimates conservative. The sampling distribution of the mean rating for the marginal group is

$$\bar{r}_M = \frac{\bar{r}_T(1 + \tau) - \bar{r}_C}{\tau} \quad (2)$$

Our course, \bar{r}_T , τ and \bar{r}_C are all themselves random variables. Furthermore, they are not necessarily independent. To compute the sampling distribution of \bar{r}_M , we bootstrap sample both the hiring regressions and the private feedback regressions on the experimental sample.⁸ Because we do not have feedback on workers who are never hired, we use the estimates values to calculate \bar{r}_M . Figure 6 shows the sampling distribution of \bar{r}_M .

The treatment actual rating is plotted as a dotted line and the and control actual rating is plotted as a solid vertical line. The distribution is centered at these mean values.

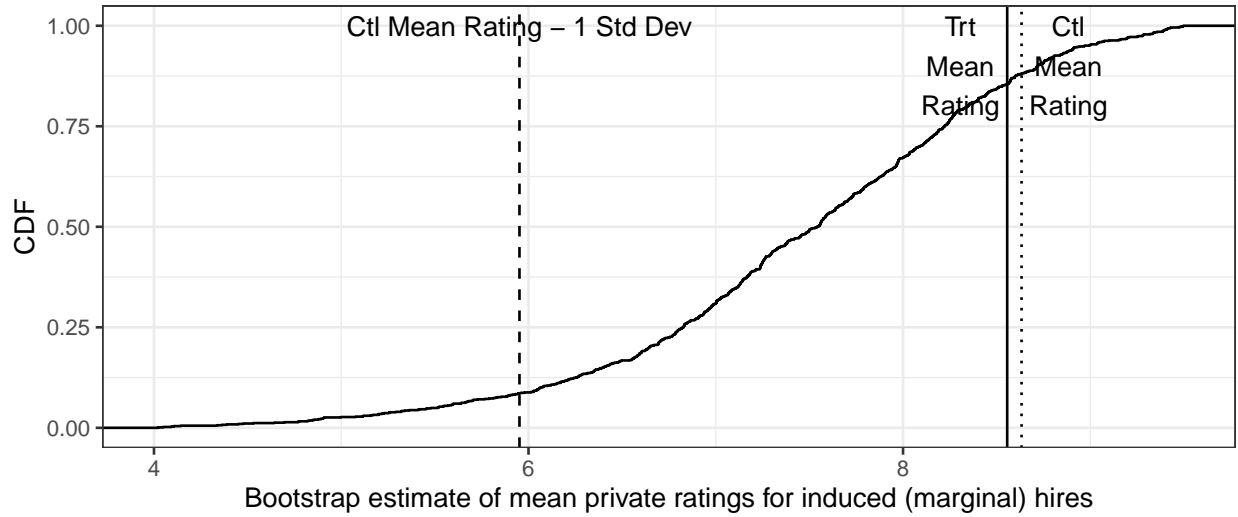
The dashed line indicates the control mean rating minus one standard deviation in the private ratings (where the standard deviation is 2.4). Comparing this value to the distribution of \bar{r}_M , this value (at the dashed line) lies at only about 0.025 of the density. In short, it would be quite surprising for us to get the results we have—an 8% increase in hires and slightly higher (but not significant ratings) if the actual marginal hires were a standard deviation worse.

3.10 Heterogeneous treatment effects to hiring

We might have expected the treatment to have differential hiring effects on the subgroups of interest, particularly since the treatment disproportionately impacted the fraction of words

⁸We define this sample as the workers allocated into the experiment who were approved by the platform and had non-empty resumes. From this we bootstrap sample with replacement. We run the hiring regressions on this sample and the ratings regressions on the same samples, narrowed to only those workers who were ever hired.

Figure 6: Sampling distribution of the private ratings of marginal hired jobseekers



spelled correctly in non-native English speakers’ resumes. In hiring outcomes, we might expect, for example, that native English or US-based jobseekers would benefit less, while writers might benefit more—though as we saw earlier, writers already make few errors. However, for these same jobseekers, the treatment might do less.

We have already shown above in Table 3 that the treatment disproportionately impacted the fraction of words spelled correctly in non-native English speakers’ resumes. If we look downstream to hiring outcomes, in Table 5, we interact the same groups with the treatment and look at their effect on the probability they were hired. The point estimates are generally quite imprecise and we lack the power to conclude much. While non-native English speakers’ writing might benefit more from the treatment, it does not translate into more hires relative to native English speakers.

3.11 Robustness checks

In our main analysis we narrow the sample to only those jobseekers whose profiles were approved by the platform. In Appendix Table 14 we run a similar regression on the full experimental sample, but we include profile approval as a control to see if it affects the estimates. In this analysis, we find that the treatment effect on the number of hires is slightly smaller than in the analysis conditional on platform approval—conditioning the sample on only jobseekers whose profiles were approved has an estimate of 7.8% while it is 10% in the full sample. The effect on the probability of any hire is 8% in the sample of only approved jobseekers and 8% in the unconditional sample. This approach and narrowing

the sample to only approved jobseekers would “block” the approval channel. In Appendix Table 15 we report the same analysis not conditioned on profile approval. None of these robustness checks change the direction or significance of any of the hiring estimates, and the slightly larger estimates in the unconditional sample are unsurprising because platform approval is a necessary condition for a jobseeker to be hired.

4 A simple model of the “clarity view” of resume writing

In this section, we formalize a rational model of how the writing intervention could (a) increase hiring but (b) not lead to worse matches. We formalize the argument that better writing allowed employers to better ascertain who was a potential match with a simple model, and show how this kind of interplay between resume quality and hiring could exist in equilibrium.

4.1 A mass of jobseekers with heterogeneous productivity

There is a unit mass of jobseekers. If hired, their productivity is θ_i . Workers are either high-type ($\theta = \theta_H$) or low-type ($\theta = \theta_L$), with $\theta_H > \theta_L$. Workers know their own type. It is common knowledge that the fraction of high types in the market is γ . All workers, if hired, are paid their expected productivity, from the employer’s point of view. Hires only last one unit of time.

4.2 Jobseekers decide whether to put into resume-writing

Before being hired, jobseekers write resumes. Jobseekers must decide whether to put effort $e \in \{0, 1\}$ into writing that resume. Effort itself is not observable. The cost of this effort is jobseekers-specific and there is a distribution of individual resume effort costs. The support of the cost distribution is $[0, \bar{c}]$. The distribution has mass everywhere and the CDF is F and PDF is f . Jobseekers who put in no effort have resume costs of 0, while those that put in effort have a cost of c_i . Critically, this cost is independent of a jobseeker’s type i.e., there is no Spence-like assumption that better workers find it cheaper to create better resumes (Spence, 1978).

Before making an offer, firms observe a signal of jobseekers’ type on their resume, $R \in \{0, 1\}$. With effort, a high-type jobseeker generates an $R = 1$ signal; without effort, $R = 0$. A low-type jobseeker generates $R = 0$ no matter what.

Clearly, low-types will never put in effort. The question is whether a high type will put in effort. The decision hinges on whether the cost of resume effort is worth the wage premium it creates. Let $w_{R=0}$ be the wage paid in equilibrium to jobseekers with $R = 0$. Note that $w_{R=1} = \theta_H$, as there is no uncertainty about a jobseeker’s type if $R = 1$.

A jobseeker i who is a high-type will choose $e = 1$ if $\theta_H - w_{R=0}(c_i) > c_i$. The marginal high-type jobseeker is indifferent between putting in effort or not, and has a resume-writing cost

of \hat{c} , where

$$\hat{c} = \theta_H - w_{R=0}(\hat{c}). \quad (3)$$

This implies that there are $F(c)\gamma$ jobseekers that choose $e = 1$. These are the high-type jobseekers with relatively low resume writing costs. The remaining $[1 - F(c)]\gamma$ high-type jobseekers choose $e = 0$. They are pooled together with the $1 - \gamma$ jobseekers that choose $e = 0$ because they are low-types.

From the employer's perspective, if they believe that the resume effort cost of the marginal high-type jobseekers is \hat{c} , the probability an $R = 0$ jobseekers is high-type is

$$p_H^{R=0}(\hat{c}) = \frac{1 - F(\hat{c})}{1/\gamma - F(\hat{c})}. \quad (4)$$

The wage received by an $R = 0$ worker is

$$w_{R=0}(\hat{c}) = \theta_L + (\theta_H - \theta_L)p_H^{R=0}(\hat{c}) \quad (5)$$

When the cost of the marginal jobseeker is higher, more jobseekers find it worth choosing $e = 1$, as $F'(\hat{c}) > 0$. This leaves fewer high-types in the $R = 0$ pool, and so

$$\frac{dp_H^{R=0}}{d\hat{c}} < 0. \quad (6)$$

4.3 The equilibrium fraction of high-type workers putting effort into resume-writing

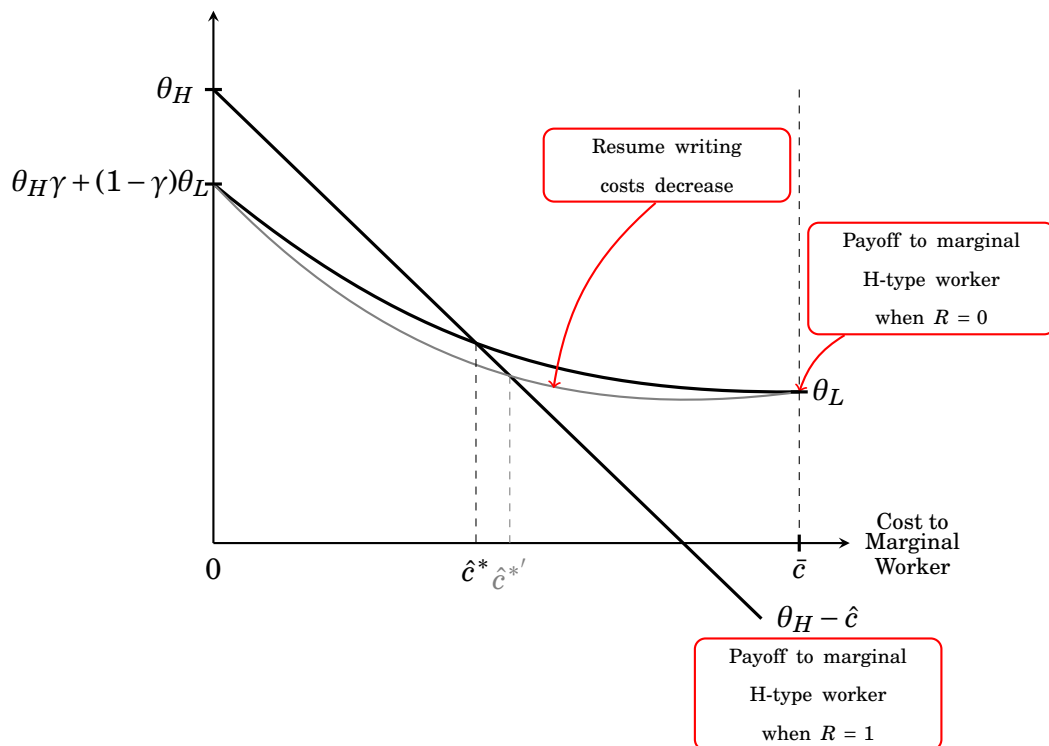
In equilibrium, there is some marginal high-type jobseeker indifferent between $e = 0$ and $e = 1$, and so

$$(\theta_H - \theta_L)(1 - p_H^{R=0}(\hat{c}^*)) = \hat{c}^*.$$

Figure 7 illustrates the equilibrium i.e., the cost where the marginal jobseeker is indifferent between $e = 0$ and $e = 1$. The two downward-sloping lines are the pay-offs to the marginal jobseeker for each \hat{c} . The pay-off to $R = 1$ is declining, as the wage is constant (at θ_H) but the cost is growing linearly. The pay-off to $R = 0$ is also declining, from Equation 6. Both curves are continuous.

Note that when the marginal jobseeker has $\hat{c} = 0$, there is just a point-mass of high-types that have a cost that low, i.e., $f(\hat{c})$. Because the marginal jobseeker is indifferent between

Figure 7: Equilibrium determination of the marginal high-type jobseeker indifferent between putting effort into a resume



putting in effort and not putting in effort, jobseekers with costs of even ε will not put in effort. Since no one finds it worthwhile to put in effort the $R = 0$ pool is just the expected value of all jobseekers. And the wage is $w_{R=0}(\hat{c}) = \gamma\theta_H + (1 - \gamma)\theta_L$. The marginal jobseeker pays nothing, so the pay-off is θ_H .

At the other extreme, $\hat{c} = \bar{c}$, all but a point mass of jobseekers have a cost less than this. Since the marginal jobseeker is indifferent between putting in effort at a cost of \bar{c} , any jobseeker with cost $\bar{c} - \varepsilon$ or below will put in effort. Then the $R = 0$ pool is purely low-types and the wage is θ_L . For the $R = 1$ market, the marginal jobseeker has a cost of \hat{c} so the pay-off is $\theta_H - \hat{c}$. We know $\theta_H > \gamma\theta_H + (1 - \gamma)\theta_L$. And by assumption, $\theta_L > \theta_H - \hat{c}$, and so by the intermediate value theorem, an equilibrium \hat{c}^* exists on $(0, \bar{c})$.

4.4 A shift in the resume writing cost distribution leads to more high-type workers choosing to exert effort

Now suppose a technology comes along that lowers—or at least keeps the same—resume writing costs for all jobseekers. This would shift F higher for all points except the endpoints of the support, creating a new distribution of costs that first-order stochastically dominates

the other.

Before determining the new equilibrium, note that no matter the marginal \hat{c} , when F increases, the probability that an $R = 0$ worker is a high-type declines, as

$$\frac{dp_H}{dF} = -\frac{1}{(F-2)^2} < 0. \quad (7)$$

This shifts the $w_{R=0}$ curve down everywhere, without changing the endpoints.

Because $w_{R=1} - \hat{c}$ is downward sloping, it intersects $w_{R=0}(\hat{c})$ at a higher value of \hat{c} . At the new equilibrium, the marginal jobseeker has resumes costs of $\hat{c}^{*'} > \hat{c}^*$. At this new equilibrium, more jobseekers choose $e = 1$, causing more $R = 1$ signals. This lowers wages for the $R = 0$ group.

4.5 The effects of lower costs are theoretically ambiguous

Note that this shift in costs is not Pareto improving—low-types are made worse off as they find themselves in a pool with fewer high-types. Furthermore, because workers are all paid their expected product, the surplus maximizing outcome would be for everyone to choose $R = 0$. Resume effort purely changes around the allocation of the wage bill, not the total amount. Total surplus is

$$\theta_H \gamma + (1 - \gamma) \theta_L - \int_0^{\hat{c}} cf(c)dc, \quad (8)$$

which is maximized at $\hat{c} = 0$, i.e., when no one finds it worthwhile to choose effort. However, with a *shift* in cost distribution (raising F), what matters is whether the marginal decrease in costs for all inframarginal workers i.e, those with $c < \hat{c}$ outweighs the costs borne by the (newly) marginal jobseekers who choose to put in effort.

In our model, all job offers are accepted. However, if we think of jobseekers as having idiosyncratic reservation values that determine whether they accept an offer, the shift in costs makes it more likely that high-types will accept an offer, while making it less likely that low-types will accept an offer. This is consistent with results where there is a greater chance an employer hires at all in the treatment. It is also consistent with our result of higher wages. Finally, if we think of employer ratings being a function of surplus, our finding of no change in employer satisfaction is also consistent, as employers are, in all cases, just paying for expected productivity.

5 Conclusion

Employers are more likely to hire new labor market entrants with better-written resumes. We argue that better writing makes it easier for employers to decide to hire a particular worker. We show results from a field experiment in an online labor market where treated workers were given algorithmic writing assistance. These jobseekers were 8% more likely to get hired and formed 7.8% more contracts over the month-long experiment. While one might have expected writing quality to be a valuable indicator of worker quality, the treatment did not affect employers’ ratings of hired workers. We provide a model of the hiring process where the cost of exerting effort on a resume is lowered by the algorithmic writing assistance, which helps employers to distinguish between high and low-type workers.

One possibility is that the benefits to treated workers came at the expense of other workers, as both treated- and control-assigned workers compete in the same market. Crowd-out concerns have been shown to be important with labor market assistance (Crépon et al., 2013). However, even if additional hires came from experienced workers, this is likely still a positive result. New labor market entrants are uniquely disadvantaged (Pallais, 2013) in online labor markets. To the extent that the gains to new workers come partially at the expense of experienced workers, this is likely a good trade-off.

Conceptualizing AI/ML innovation and proliferation as a fall in the cost of prediction technology fits our setting (Agrawal et al., 2018b,a). Writing a resume is, in part, an applied prediction task—what combination of words and phrases, arranged in what order, are likely to maximize my pay-off from a job search? The Algorithmic Writing Service reduces the effort or cost required for making these decisions. When revising their resumes, rather than identifying errors in their own predictions themselves, jobseekers with access to Algorithmic Writing Service specify their target audience and writing goals are given suggestions for error correction and cleaned up writing. Furthermore, the treatment, by lowering the costs of error-free writing for at least some jobseekers, causes them to do better at writing their resumes.

These kinds of algorithmic writing assistance will likely “ruin” writing as a signal of ability. With the proliferation of writing technologies with capabilities far beyond what is explored here (Brown et al., 2020), even if the “signaling view” was at one time true, technological changes are likely to make it not true in the future.

References

- Agrawal, Ajay, John Horton, Nicola Lacetera, and Elizabeth Lyons**, “Digitization and the contract labor market,” *Economic analysis of the digital economy*, 2015, 219.
- , **Joshua Gans, and Avi Goldfarb**, “Prediction, judgment, and complexity: a theory of decision-making and artificial intelligence,” in “The economics of artificial intelligence: An agenda,” University of Chicago Press, 2018, pp. 89–110.
- , —, and —, *Prediction machines: the simple economics of artificial intelligence*, Harvard Business Press, 2018.
- , **Nicola Lacetera, and Elizabeth Lyons**, “Does standardized information in online markets disproportionately benefit job applicants from less developed countries?,” *Journal of international Economics*, 2016, 103, 1–12.
- Barach, Moshe A and John J Horton**, “How do employers use compensation history? Evidence from a field experiment,” *Journal of Labor Economics*, 2021, 39 (1), 193–218.
- Belot, Michèle, Philipp Kircher, and Paul Muller**, “Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice,” *The Review of Economic Studies*, 10 2018, 86 (4), 1411–1447.
- Bertrand, Marianne and Sendhil Mullainathan**, “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.(2003),” *Amer. Econ. Rev.*, 2003, 94, 991.
- Bolton, Gary, Ben Greiner, and Axel Ockenfels**, “Engineering trust: reciprocity in the production of reputation information,” *Management science*, 2013, 59 (2), 265–285.
- Briscese, Guglielmo, Giulio Zanella, and Veronica Quinn**, “Providing Government Assistance Online: A Field Experiment with the Unemployed,” *Journal of Policy Analysis and Management*, 2022, 41 (2), 579–602.
- Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell et al.**, “Language models are few-shot learners,” *Advances in neural information processing systems*, 2020, 33, 1877–1901.
- Cai, Hongbin, Ginger Zhe Jin, Chong Liu, and Li an Zhou**, “Seller reputation: From word-of-mouth to centralized feedback,” *International Journal of Industrial Organization*, 2014, 34, 51–65.

- Card, David, Jochen Kluge, and Andrea Weber**, “Active labour market policy evaluations: A meta-analysis,” *The Economic Journal*, 2010, 120 (548), F452–F477.
- Chan, Jason and Jing Wang**, “Hiring preferences in online labor markets: Evidence of a female hiring bias,” *Management Science*, 2018, 64 (7), 2973–2994.
- Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora**, “Do labor market policies have displacement effects? Evidence from a clustered randomized experiment,” *The Quarterly Journal of Economics*, 2013, 128 (2), 531–580.
- Farber, Henry S, Dan Silverman, and Till Von Wachter**, “Determinants of callbacks to job applications: An audit study,” *American Economic Review*, 2016, 106 (5), 314–18.
- Filippas, Apostolos, John Joseph Horton, and Joseph Golden**, “Reputation Inflation,” *Marketing Science*, Forthcoming.
- Fradkin, Andrey, Elena Grewal, and David Holtz**, “Reciprocity and unveiling in two-sided reputation systems: Evidence from an experiment on Airbnb,” *Marketing Science*, 2021, 40 (6), 1013–1029.
- Ghose, Anindya and Panagiotis G Ipeirotis**, “Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics,” *IEEE transactions on knowledge and data engineering*, 2010, 23 (10), 1498–1512.
- Goldfarb, Avi and Catherine Tucker**, “Digital economics,” *Journal of Economic Literature*, 2019, 57 (1), 3–43.
- HF Canonical Model Maintainers**, “distilbert-base-uncased-finetuned-sst-2-english (Revision bfdd146),” 2022.
- Hong, Yili, Jing Peng, Gordon Burtch, and Ni Huang**, “Just DM Me (Politely): Direct Messaging, Politeness, and Hiring Outcomes in Online Labor Markets,” *Information Systems Research*, 2021, 32 (3), 786–800.
- Horton, John J.**, “Online labor markets,” *Internet and Network Economics: 6th International Workshop, WINE 2010, Stanford, CA, USA, December 13-17, 2010. Proceedings*, 2010.
- , “The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment,” *Journal of Labor Economics*, 2017, 35 (2), 345–385.

- Kang, Sonia K, Katherine A DeCelles, András Tilcsik, and Sora Jun**, “Whitened résumés: Race and self-presentation in the labor market,” *Administrative Science Quarterly*, 2016, 61 (3), 469–502.
- Kokkodis, Marios and Sam Ransbotham**, “Learning to successfully hire in online labor markets,” *Management Science*, 2022.
- Luca, Michael and Oren Reshef**, “The effect of price on firm reputation,” *Management Science*, 2021, 67 (7), 4408–4419.
- Marinescu, Ioana and Ronald Wolthoff**, “Opening the black box of the matching function: The power of words,” *Journal of Labor Economics*, 2020, 38 (2), 535–568.
- Martin-Lacroux, Christelle and Alain Lacroux**, “Do Employers Forgive Applicants’ Bad Spelling in Résumés?,” *Business and Professional Communication Quarterly*, sep 2017, 80 (3), 321–335.
- Moss-Racusin, Corinne A, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman**, “Science faculty’s subtle gender biases favor male students,” *Proceedings of the national academy of sciences*, 2012, 109 (41), 16474–16479.
- Oreopoulos, Philip**, “Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes,” *American Economic Journal: Economic Policy*, 2011, 3 (4), 148–71.
- Pallais, Amanda**, “Inefficient Hiring in Entry-level Labor Markets,” *American Economic Review*, 2013.
- Sajjadiani, Sima, Aaron J Sojourner, John D Kammeyer-Mueller, and Elton Mykerezzi**, “Using machine learning to translate applicant work history into predictors of performance and turnover,” *Journal of Applied Psychology*, 2019, 104 (10), 1207.
- Spence, Michael**, “Job market signaling,” in “Uncertainty in economics,” Elsevier, 1978, pp. 281–306.
- Stanton, Christopher T and Catherine Thomas**, “Landing the first job: The value of intermediaries in online hiring,” *The Review of Economic Studies*, 2016, 83 (2), 810–854.
- Sterkens, Philippe, Ralf Caers, Marijke De Couck, Michael Geamanu, Victor Van Driessche, and Stijn Baert**, “Costly Mistakes: Why and When Spelling Errors in Resumes Jeopardise Interview Chances,” *Working paper*, 2021.

A Appendix

Table 7: Description of Error Rule Categories with Examples

Category	Description	Examples
American English Phrases	Sentence favors the American English spelling of words.	<i>apologize, catalog, civilization, defense</i>
British English, Oxford Spelling	Sentence favors the British English spelling of words.	<i>apologise, catalogue, civilisation, defence</i>
Capitalization	Rules about detecting uppercase words where lowercase is required and vice versa.	<i>This house is old. it was built in 1950. I really like Harry potter.</i>
Collocations	A collocation is made up of two or more words that are commonly used together in English. This refers to an error in this type of phrase.	<i>Undoubtedly, this is the result of an extremely dynamic development of Lublin in the recent years. I will take it in to account. It's batter to be save then sorry.</i>
Commonly Confused Words	Words that are easily confused, like 'there' and 'their' in English.	<i>I have my won bed. Their elicit behavior got the students kicked out of school.It's the worse possible outcome.</i>
Grammar	Violations related to system of rules that allow us to structure sentences.	<i>Tom make his life worse. A study like this one rely on historical and present data.This is best way of dealing with errors.</i>
Miscellaneous	Miscellaneous rules that don't fit elsewhere.	<i>This is best way of dealing with errors. The train arrived a hour ago. It's nice, but it doesn't work. (inconsistent apostrophes)</i>
Nonstandard Phrases		<i>I never have been to London. List the names in an alphabetical order. Why would a man all of the sudden send flowers?</i>
Possible Typo	Spelling issues.	<i>It'a extremely helpful when it comes to homework. We haven't earned anything.This is not a HIPPA violation.</i>
Punctuation	Error in the marks, such as period, comma, and parentheses, used in writing to separate sentences and their elements and to clarify meaning.	<i>"I'm over here, she said. Huh I thought it was done already. The U.S.A is one of the largest countries.</i>
Redundant Phrases	Redundant phrases contain words that say the same thing twice. When one of the words is removed, the sentence still makes sense. Sometimes the sentence has to be slightly restructured, but the message remains the same.	<i>We have more than 100+ customers. He did it in a terrible way. The money is sufficient enough to buy the sweater.</i>
Semantics	Logic, content, and consistency problems.	<i>It allows us to both grow, focus, and flourish. On October 7, 2025 , we visited the client.This was my 11nd try.</i>
Style	General style issues not covered by other categories, like overly verbose wording.	<i>Moreover, the street is almost entirely residential. Moreover, it was named after a poet. Doing it this way is more easy than the previous method. I'm not very experienced too. Anyways, I don't like it.</i>
Typography	Problems like incorrectly used dash or quote characters.	<i>This is a sentence with two consecutive spaces. I have 3dogs.The price rose by \$12,50. I'll buy a new T—shirt.</i>

Table 8: Summary statistics on error counts and rates in the control group

	Total Errors	Error Rate
All Error Types	4.390 (10.257)	0.080 (0.252)
Capitalization Errors	0.140 (0.543)	0.004 (0.029)
Possible Typo	2.312 (8.768)	0.041 (0.106)
Grammar Errors	0.219 (0.589)	0.004 (0.014)
Punctuation Errors	0.425 (1.629)	0.008 (0.213)
Typographic Errors	0.821 (2.985)	0.016 (0.051)
Style Errors	0.317 (0.890)	0.004 (0.011)
Miscellaneous Errors	0.103 (0.391)	0.002 (0.009)
Redundant Phrases	0.025 (0.162)	0.000 (0.003)
Nonstandard Phrases	0.002 (0.050)	0.000 (0.001)
Commonly Confused Words	0.010 (0.117)	0.000 (0.002)
Collocations	0.012 (0.121)	0.000 (0.003)
Semantic Errors	0.003 (0.061)	0.000 (0.001)

Notes: This table reports means and standard errors of the writing errors in the resumes of the control group. The first column displays the average total error count and the second column displays the average error rate (total errors normalized by word count). Writing errors are defined by LanguageToolR. The sample is made up of all jobseekers in the control group of the experimental sample who submitted non-empty resumes and were approved by the platform.

Table 9: Effects of writing assistance on length of resume

	<i>Dependent variable:</i>
	Number of words in resume
Algo Writing Treatment	0.128 (0.314)
Constant	70.541*** (0.223)
Observations	194,701
R ²	0.00000

Notes: This table analyzes the effect of the treatment on the number of words in a jobseeker’s resume. The sample is made up of all jobseekers in the experimental sample who submitted non-empty profiles and were approved by the platform. Significance indicators: $p \leq 0.10$: *, $p \leq 0.05$: ** and $p \leq .01$: ***.

Table 10: Effect of Treatment to Writing Errors, Page 1

	<i>Dependent variable:</i>						
	Spelling (1)	Capitalization (2)	Possible Typo (3)	Grammar (4)	Punctuation (5)	Typography (6)	Style (7)
Algo Writing Treatment	-0.001* (0.0005)	-0.0005*** (0.0001)	-0.002*** (0.0005)	-0.0005*** (0.0001)	-0.0004 (0.0004)	-0.002*** (0.0003)	0.0003*** (0.0001)
Constant	0.036*** (0.0003)	0.003*** (0.00005)	0.041*** (0.0003)	0.004*** (0.00004)	0.010*** (0.0003)	0.015*** (0.0002)	0.004*** (0.00004)
Observations	187,858	187,858	187,858	187,858	187,858	187,858	187,858
R ²	0.00002	0.0003	0.0001	0.0004	0.00000	0.0004	0.0001

Notes: This table analyzes the effect of the treatment on all types of writing errors, normalized by resume length. Writing errors are defined by LanguageToolR, and divided by the number of words in a jobseekers' resume to calculate their error rate. The sample is made up of all jobseekers in the experimental sample who completed the platform registration page and submitted non-empty resume. Significance indicators: $p \leq 0.10$:

*, $p \leq 0.05$; **, $p \leq .01$; ***, $p \leq .001$.

Table 11: Effect of Treatment to Writing Errors, Page 2

	<i>Dependent variable:</i>					
	Miscellaneous (1)	Redundant Phrases (2)	Nonstandard Phrases (3)	Commonly Confused Words (4)	Collocations (5)	Semantics (6)
Algo Writing Treatment	-0.0004*** (0.00003)	-0.00003** (0.00001)	-0.00001 (0.00000)	-0.00005*** (0.00001)	-0.0001*** (0.00001)	0.00001 (0.00001)
Constant	0.002*** (0.00002)	0.0004*** (0.00001)	0.00003*** (0.00000)	0.0001*** (0.00001)	0.0003*** (0.00001)	0.0001*** (0.00000)
Observations	187,858	187,858	187,858	187,858	187,858	187,858
R ²	0.001	0.00002	0.00001	0.0002	0.0001	0.00000

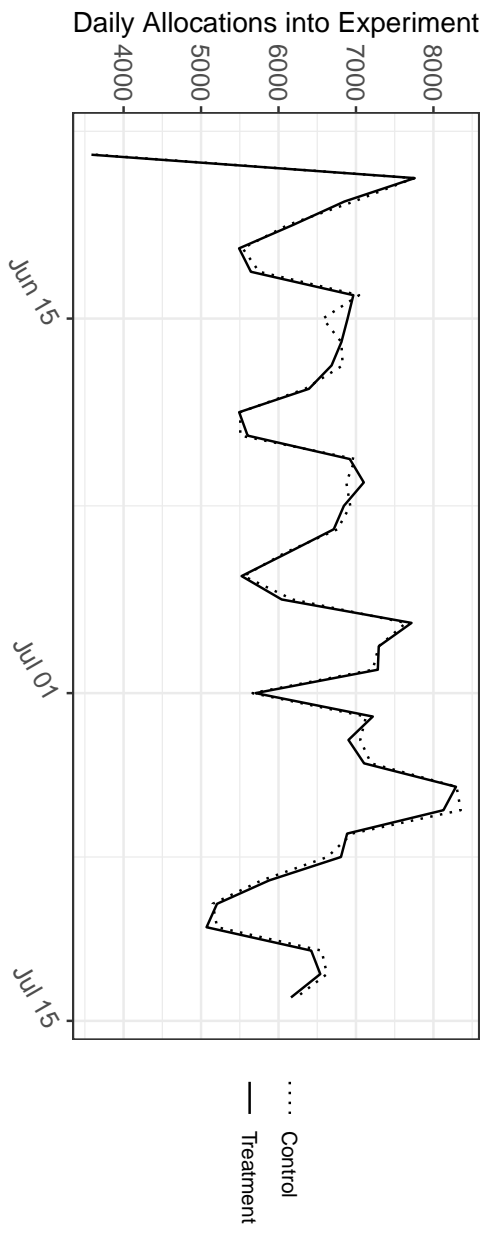
Notes: This table analyzes the effect of the treatment on all types of writing errors, normalized by resume length. Writing errors are defined by LanguageToolR, and divided by the number of words in a jobseekers' resume to calculate their error rate. The sample is made up of all jobseekers in the experimental sample who completed the platform registration page and submitted non-empty resume. Significance: $p \leq 0.10$: †, $p \leq 0.05$: *, $p \leq 0.01$: ** and $p \leq .001$: ***.

Table 12: Effects of writing assistance on profile submission and platform approval

	<i>Dependent variable:</i>		
	Profile submitted x 100 (1)	Approved x 100 (2)	Approved x 100 (3)
Algo Writing Treatment	0.107 (0.144)	0.200 (0.133)	0.186 (0.142)
Constant	45.532*** (0.102)	89.057*** (0.094)	40.550*** (0.100)
Observations	480,950	219,243	480,950
R ²	0.00000	0.00001	0.00000

Notes: This table analyzes the effect of the treatment on whether or not a jobseeker's profile was submitted and approved. In Column (1) the outcome is 100 times a binary indicator for whether or not the jobseeker completed platform registration and submitted their resume, on the full experimental sample. In Column (2) the outcome is 100 times a binary indicator for whether or not the platform approved the resume, on the sample of only those jobseekers who submitted their resumes. In Column (3) the outcome is 100 times a binary indicator for whether or not the platform approved the resume, on the full experimental sample. Significance indicators: $p \leq 0.10$: *, $p \leq 0.05$: ** and $p \leq 0.01$: ***.

Figure 8: Daily allocations of jobseekers into experimental cells



Notes: This plot shows the daily allocations into the treatment and control cells for the experimental sample of 480,950 new jobseekers to the platform.

Table 13: Effect of algorithmic writing assistance on hiring outcomes

	<i>Dependent variable:</i>			
	Num Contracts (1)	Hired x 100 (2)	Num Hourly Interviews (3)	Num Invitations (4)
Algo Writing Treatment	0.004** (0.002)	0.247*** (0.080)	0.002 (0.004)	0.001 (0.003)
Constant	0.047*** (0.001)	3.093*** (0.057)	0.210*** (0.003)	0.142*** (0.002)
Observations	194,703	194,703	194,703	194,703
R ²	0.00003	0.00005	0.00000	0.00000

Notes: This analysis looks at the effect of treatment on hiring outcomes on jobseekers in the experimental sample. Column (1) defines Number of Contracts as the number of unique jobs they work over the month after they register for the platform. Column (2) defines Hired x 100 as one hundred times the probability the jobseeker was hired over that month. Column (3) is the number of interviews they gave over that month. And the Column (4) outcome Invitations is the number of times they were recruited to a job over their first month. The experimental sample is of all new jobseekers who registered and were approved for the platform between June 8th and July 14th, 2021 and had non-empty resumes. Significance indicators: $p \leq 0.10$: *, $p \leq 0.05$: ** and $p \leq .01$: ***.

Table 14: Effect of algorithmic writing assistance on hiring outcomes, controlling for platform approval

	<i>Dependent variable:</i>			
	Num Contracts (1)	Hired x 100 (2)	Num Hourly Interviews (3)	Num Invitations (4)
Algo Writing Treatment	0.001** (0.001)	0.100*** (0.032)	0.001 (0.002)	0.0004 (0.001)
Approved by Platform	0.049*** (0.001)	3.204*** (0.033)	0.210*** (0.002)	0.142*** (0.001)
Constant	-0.001 (0.001)	-0.050* (0.027)	-0.0003 (0.001)	0.00001 (0.001)
Observations	480,952	480,952	480,952	480,952
R ²	0.011	0.019	0.030	0.023

Notes: This analysis looks at the effect of treatment on hiring outcomes on jobseekers in the experimental sample. Column (1) defines Number of Contracts as the number of unique jobs they work over the month after they register for the platform. Column (2) defines Hired x 100 as one hundred times the probability the jobseeker was hired over that month. Column (3) is the number of interviews they gave over that month. And the Column (4) outcome Invitations is the number of times they were recruited to a job over their first month. The sample used in this analysis is the entire experimental sample. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq 0.01 : ***$.

Table 15: Effect of algorithmic writing assistance on hiring outcomes, unconditional on platform approval

	<i>Dependent variable:</i>			
	Num Contracts (1)	Hired x 100 (2)	Num Hourly Interviews (3)	Num Invitations (4)
Algo Writing Treatment	0.002** (0.001)	0.106*** (0.033)	0.001 (0.002)	0.001 (0.001)
Constant	0.019*** (0.0005)	1.249*** (0.023)	0.085*** (0.001)	0.058*** (0.001)
Observations	480,952	480,952	480,952	480,952
R ²	0.00001	0.00002	0.00000	0.00000

Notes: This analysis looks at the effect of treatment on hiring outcomes on jobseekers in the experimental sample. Column (1) defines Number of Contracts as the number of unique jobs they work over the month after they register for the platform. Column (2) defines Hired x 100 as one hundred times the probability the jobseeker was hired over that month. Column (3) is the number of interviews they gave over that month. And the Column (4) outcome Invitations is the number of times they were recruited to a job over their first month. The sample used in this analysis is the entire experimental sample. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$ and $p \leq 0.01 : ***$.