# Gender Differences in Reference Letters: Evidence from the Economics Job Market*

Markus Eberhardt[1], Giovanni Facchini[2], and Valeria Rueda[1]

[1]University of Nottingham and CEPR
[2]University of Nottingham, CEPR and IZA

This draft: January 23, 2023

**Abstract**

Academia, and economics in particular, faces increased scrutiny because of gender imbalance. This paper studies the job market for entry-level faculty positions. We employ machine learning methods to analyze gendered patterns in the text of 12,000 reference letters written in support of over 3,700 candidates. Using both supervised and unsupervised techniques, we document widespread differences in the attributes emphasized. Women are systematically more likely to be described using 'grindstone' terms and at times less likely to be praised for their ability. Using information on initial placement we highlight the implications of these gendered descriptors for the quality of academic placement.

*JEL Codes:* J16, A11.
*Keywords:* Gender, Natural Language Processing, Diversity.

*We gratefully acknowledge financial support from STEMM-CHANGE, the School of Economics, and the Faculty of Social Sciences (all at Nottingham University) and thank *Econ Job Market* for authorizing the use of the data. Malena Arcidiácono, Edoardo Cefalà, Cristina Griffa, Yuliet Verbel-Bustamante, Thea Zoellner, and Diego Marino-Fages have provided excellent research assistance. The University of Nottingham School of Economics Research Ethics Committee gave clearance to use the EJM data on March 17, 2020 and to conduct a survey of academics on January 18, 2021. A data use agreement for the authors was signed between EJM and the School of Economics at the University of Nottingham on April 13, 2020. The views expressed in this paper are those of the authors and do not necessarily represent the views of the University of Nottingham. We thank seminar participants at Bocconi University, Lund, St Andrews, SOFI, Surrey, Warwick, and Trinity College Dublin, as well as the Monash-Zürich text-as-data conference for comments and suggestions. The usual disclaimers apply. Corresponding author: Valeria Rueda, School of Economics, Sir Clive Granger Building, University of Nottingham, University Park, Nottingham NG7 2RD, U.K. Email: valeria.rueda@nottingham.ac.uk

# 1 Introduction

Gender disparities in the workplace have received significant attention in public debate. Academia is facing increased scrutiny due to its low female representation, especially in the field of economics (Valian, 1999; Lundberg, 2020, Part I). Recent empirical work has documented that the economics career pipeline for women is 'leaky', meaning that women tend to drop out of the profession at critical transitions, such as the jump from earning a Ph.D. to an assistant professorship, or from assistant to associate professor (for a broad review, see Lundberg and Stearns, 2019). This paper studies the first step of the academic career of an economist, the junior 'job market' —the stage at which the leak has grown the most in the past decade (Lundberg and Stearns, 2019)— and which so far has not received much systematic attention (Lundberg, 2020).

The academic job market in economics is unique in that it is a highly structured institution. It starts every year in late Fall, with universities posting their job advertisements and potential applicants preparing a 'job market package'. The latter consists of one or more academic papers, a CV and a set of recommendation letters written by scholars familiar with the candidate. All the parties involved, i.e. the candidates, the letter writers and the hiring committees, interact via centralized platforms. Typically, the same package is used for the vast majority of jobs, making the marginal cost of an additional application low. Reference letters are not tailored to a particular institution and the same letter is usually used for *all* job applications (for more details see Coles et al., 2010).

In this paper, we investigate the presence of differences in the language used in reference letters, depending on the gender of the candidate being recommended. We use a unique dataset encompassing all applications for entry-level positions received by a research-intensive university in the U.K. over the 2017-2021 period. Deploying Natural Language Processing tools, we analyze the text of almost 12,000 reference letters written in support of 3,700 candidates. A standard letter covers a lengthy discussion of the candidate's job market paper, some reference to their additional research, and to their teaching and citizenship skills. Importantly, the final section of the letter provides a summary assessment of the candidate's academic abilities and recruitment prospects. Since we are primarily interested in the way candidates are described, we focus much of our attention on this final section.

This corpus is then transformed into a term-frequency-inverse-document-frequency (tf-idf) representation. Borrowing from methods developed in cognitive psychology and linguistics, we quantify whether letters written in support of female candidates emphasize systematically different attributes. We use three complementary approaches. First, we employ an unsupervised method to ascertain the terms in the letters that are the best predictors of a candidate's gender. We adopt a LASSO technique that selects the strongest predictors. Among these, we frequently observe terms related to research interests, but also to personality ('nice', 'pleasant') and 'grindstone' attributes ('determined', 'hardworking', etc.). Second, we rely on a supervised method, building dictionaries of words for common attributes emphasized

in reference letters. These dictionaries are informed by existing research on the topic (Trix and Psenka, 2003; Schmader et al., 2007). We validate our dictionaries through an original comprehensive survey of academic economists based in U.K. research-intensive universities. Corroborating the exploratory results from the LASSO, we observe that descriptions of female candidates tend to emphasize significantly more 'grindstone' attributes. In further specifications, we also uncover a tendency to use fewer terms related to ability and research. Third, we also qualify the strength of support received by a candidate by analysing the type of placement recommendation received (e.g. "I recommend this person to any institution, including the very best"). We observe that women receive fewer positive signals, but no difference in negative ones. They are also more likely to be compared to other candidates.

This paper thus documents differences in the language chosen for female and male candidates. In line with previous research, we observe that women are described with more 'grindstone' attributes and at times fewer 'ability' ones (Bourdieu and Passeron, 1977; Trix and Psenka, 2003; Schmader et al., 2007). A natural question following on from these findings is whether the differences uncovered matter. Diligence and working hard are positive attributes (see Alan et al., 2019, on 'grit'). However, given the overwhelmingly positive tone of recommendation letters in the job market, it may be misleading to interpret our findings as suggesting that women receive 'better' recommendations. The opposite may well be true. In fact, as noted by Valian (1999, p. 170) "[a]lthough working hard is a virtue, labelling a woman a hard worker can be damning with faint praise. If someone is not considered able to begin with, working hard can be seen as confirmation of his or her inability." More generally, sociologists have pointed out that minorities are more often praised for their diligence than for their innate ability and that the signal of diligence is often interpreted as a lack of innate talent (Bourdieu and Passeron, 1977, p.201).

To illustrate the importance of language in reference letters, we study the correlation between the attributes emphasized in letters and job market placement. We manually collected information from personal websites, academic departments' placement records, and LinkedIn profiles, to establish whether the candidate placed in academia or elsewhere. For academic placements, we also link the hiring institution to its RePEc (Research Papers in Economics) rankings. To the best of our knowledge, we are the first to collect such information.[1] The results indicate that language matters differentially for women and men. In particular, male candidates tend to benefit from 'standout' terminology while for females the patterns vary. For academic placements, letters emphasizing 'grindstone' are associated with obtaining a job in a less prestigious institution, but only so for women.

In studying gendered patterns in the language of reference letters, we address several empirical challenges. More specifically, the attributes emphasized in reference letters may be influenced by many factors, such as the institution the candidate graduated from or their research

[1] In a related paper, Baltrunaite et al. (2022) study placement in the same year for all candidates, i.e. between one and up to ten years after their initial placement, and focus on the attainment of the associate professor rank.

field. Some of these determinants may differ systematically for male and female candidates. We tackle this problem in a variety of ways. In our baseline specifications, we control for the observable candidate and writer characteristics obtained from the application platform and from additional information we collected manually. On the writer's side, we control for their gender, the number of letters they provide in our sample, and the RePEc ranking of their institution. On the candidate side, we control for ethnicity, years since Ph.D. completion, broad field of specialization, publication record, and the ranking of their Ph.D.-awarding institution. The baseline results are not sensitive to these controls, nor to alternative definitions of the reference letter ends. We also check that the results are not driven by alternative explanations that could correlate with the gender of the letterwriter such as the location of the Ph.D.-granting institution, the gender or the cultural background of the letterwriter, the academic field of the candidate, or the extent of networking conducted before the market, among others.

Still, we may worry that unobservable determinants could affect our findings. Therefore, we run more restrictive models that allow us to account for unobserved, time-invariant institutional and letterwriter characteristics. A first set of models, which include fixed effects for the Ph.D.-granting institution, confirm the gendered patterns observed even for candidates of the same cohort at the same institution. In further analysis, we restrict the sample to referees who have written letters for both male and female candidates and employ writer fixed effects. These more demanding specifications confirm that differences in describing male and female candidates are detectable even when we focus on individual writers. Further probing indicates that more experience in writing for female candidates attenuates some of these differences.

This article is related to the literature on gender representation in academia. Several papers have shown that women are under-represented in math-intensive fields (for a detailed review of the literature see Ceci and Williams, 2009, p.3-16; Kahn and Ginther, 2017). Investigations of different aspects of academic life have uncovered significant barriers. For example, Nittrouer et al. (2018) and Hospido and Sanz (2021), among others, observe that female academics are less likely to be accepted to present their work at academic conferences. Many researchers have emphasized systematic gender biases in student evaluations of teachers, which are frequently-used indicators of performance in promotion and tenure packages (MacNell et al., 2015; Boring, 2017; Fan et al., 2019; Mengel et al., 2019; Boring and Philippe, 2021). These patterns are persistent, despite evidence of a demand for diversity (Funk et al., 2019).

While other math-intensive fields have shown some improvement, economics has been in the spotlight for its persistently low representation of women (Bayer and Rouse, 2016; Lundberg and Stearns, 2019). Not only is there low female representation at the earliest stages of the profession, but the career pipeline is also 'leaky'. In trying to understand barriers to women's advancement in economics, researchers have looked at different stages of an academic career.

Focusing on the first one, Boustan and Langan (2019) document the wide variation of gender representation across Ph.D. programs, and that this representation tends to be a persistent attribute of a department. Turning to the next steps as academic professionals, other limitations to the advancement of women have been observed. In particular, there is evidence that females face barriers to promotion (Ginther and Kahn, 2004; Sarsons, 2017; Bosquet et al., 2019), higher standards to judge the quality of their research (Card et al., 2020; Dupas et al., 2021; Grossbard et al., 2021; Hengel, 2022), and that their work gets cited less (Koffi, 2021). Taken together, all these factors are likely to hamper the progression of women in their academic careers. We contribute to this burgeoning literature by focusing on a major and to date unexplored stepping stone: the junior job market. At this stage, beyond institutional credentials, little information about the candidate's research or teaching is observed. Therefore, reference letters play a crucial role in supporting the applicant.

The professional culture in economics may also be problematic for women's advancement. Wu (2018) reports evidence of gender biases in posts about women in a well-known and widely used anonymous forum in the profession. Similarly, Dupas et al. (2021) study the seminar culture and present evidence that female speakers face more hostile audiences. By analyzing recommendation letters, we are investigating a different aspect of the professional culture, namely mentorship. As opposed to these previous studies, our focus is on a setting in which economists fulfil a supportive and nurturing role.

Existing literature has uncovered gendered patterns in academic reference letters in other disciplines. For example, Trix and Psenka (2003) show that letters written in support of female applicants to medical faculty positions are shorter and emphasize more 'grindstone' and 'teaching characteristics'. Looking at job applicants in chemistry and biochemistry, Schmader et al. (2007) observe similar patterns. Madera et al. (2009) documents that letters for female applicants in psychology emphasize their 'communal' attributes ('nice', 'collegial', etc.). This line of research has also uncovered systematic differences in the presence of doubt raisers in geosciences (Dutt et al., 2016), and psychology (Hebl et al., 2018; Madera et al., 2019).

We contribute to this literature in three main ways. First, we validate the 'sentiment' classification previously used by carrying out an original survey of academic economists.

Second, by focusing on economics, we can leverage a substantially larger sample of letters that are broadly representative of a highly structured and globalized academic job market. This allows us to rely on unsupervised techniques to describe gendered patterns in the language used when writing references. More specifically, fitting a LASSO, we show that many of the words that best predict letters written for women relate to 'grindstone' or 'teaching and citizenship' traits, whereas many 'ability' terms are more predictive of letters written for men. In other words, gendered differences in language used are already salient when describing the data with an unsupervised approach. This suggests in turn that the patterns uncovered in this literature with supervised techniques are unlikely to be driven by biases in the selection of the relevant terms.

Third, we also further advance the literature by analysing gendered differences in the quality of placement recommendations (e.g. whether people are explicitly recommended to a top institution). More importantly, we also analyse the implications of gendered language on the initial placement *outcomes* of candidates. Ongoing work on reference letters by Baltrunaite et al. (2022), which relies on word embedding representation of words rather than tf-idf, confirms these patterns for two Italian institutions and focusing on longer-term career outcomes.

The remainder of the paper is organized as follows. In Section 2 we discuss our sample as well as the general approach of our main textual analysis. Section 3 explains the process of data cleaning and preparation, followed by the exploratory analysis using unsupervised methods in Section 4. Section 5 outlines the supervised approach and presents the baseline results, with extensions and additional robustness checks. Section 6 discusses the analysis of job market placement, followed by concluding remarks.

## 2 Data

We collected and cleaned the text of almost 12,000 reference letters written in support of over 3,700 candidates who applied for entry-level positions between 2017 and 2021 at a research-intensive economics department in the U.K.[2] In each year in our sample the department advertised multiple positions open to all fields.

The department is one of the largest in the U.K., with over 55 regular faculty members and was ranked in the top–5 in the most recent public evaluation of scientific research carried out in U.K. universities (Research Excellence Framework, 2021). It has been consistently ranked in the top-75 worldwide according to the RePEc rankings. The majority of the faculty has an international background, with 53% having earned a Ph.D. outside the U.K. (half of them in the U.S., the other half in other European countries). The department has a large Ph.D. program, with over 50 students in residence in a given year. 23% of staff is female.[3]

The applications were collected from the *EconJobMarket* (EJM) platform. Access to and handling of these confidential data were in accordance with the data processing agreement signed between the researchers and EJM, which obtained appropriate ethical approval.

For each letter, we know a number of characteristics of the candidate and the letter writer. For candidates, we know characteristics they entered on *EconJobMarket*, such as gender, ethnicity, and the institution granting their Ph.D.[4] We also manually collect data from the candidates' CVs: we add information on their publication record at the time of application and their graduation date. The institutional ranking of both letter writers and candidates are taken from

[2] All applications were filed exclusively through EJM, without any additional paperwork required.

[3] A figure slightly below the average for U.K. research-intensive institutions in the so-called Russell Group. For more details see De Fraja et al. (2019).

[4] If the gender was withheld in the EJM application, it was determined using a manual internet search.

RePEc.[5] Information on the main advisor is also collected. Finally, we manually searched online for each individual candidate to establish their first professional placement in the year following their first appearance in our sample. Combining information from personal websites, academic departments' placement records, and LinkedIn profiles, we establish whether the candidate placed in academia or elsewhere. For academic placements, we also collect the name of the institution, which we link to RePEc rankings.

For each letterwriter, we have information on the institution where they were based at the time the letter was written. Using the R library 'GenderizeR', we also infer their gender from their first names. For this procedure, we adopt a conservative approach and manually search for cases in which the gender probability reported by the algorithm is below 0.75.[6] We also manually collected information on their academic rank, their seniority (year of Ph.D. completion), and their country of birth.[7]

Summary statistics of these characteristics are presented in Tables 1 and 2. The majority of applicants and reference letter writers are based in the top-100 ranked institutions, with slightly more letter writers concentrated at the very top, as also shown in Figure 1. We have 5,655 writers (female share 17.4%) in our sample, and on average each writer has written slightly more than two letters. Overall, approximately 30% of the candidates in our sample are women. This statistic is consistent with the figures reported by Lundberg and Stearns (2019) and has remained stable over time as shown in Figure 2. Table 3 shows the share of applicants by country. Approximately 50% of the candidates are based at U.S. institutions and 14% in the U.K. (see also Appendix Table 3 for a detailed breakdown).

Reference letters for the economics job market have a mean length of 1,089 words, which corresponds to around three pages A4, with a standard deviation of 554 words (around 1.5 pages). A standard letter covers a lengthy discussion of the candidate's job market paper, some reference to their additional research, and to their teaching and citizenship attributes. Importantly, the final section of the letter provides a summary assessment of the candidate's academic abilities and recruitment prospects.

[5] See Appendix A for more details on how the ranking is constructed.

[6] The names of only 284 individuals fall below this threshold (5.9% of all letterwriters) and their gender has been determined using a manual search.

[7] If the country of birth was unknown, we have attributed it based on the location of the location of the institution that granted their undergraduate degree.

Table 1: Descriptive Statistics — candidate characteristics

| | Full Sample | | | | | Males | | Females | | Difference (M - F) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | N | Mean | SD | Min | Max | N | Mean | N | Mean | Estimate | p-value | sig. |
| **Characteristics of the candidates** | | | | | | | | | | | | |
| ***Gender*** | | | | | | | | | | | | |
| Female | 3721 | 0.291 | 0.454 | 0 | 1 | 2639 | 0 | 1082 | 1 | -1 | | |
| ***Ethnicity*** | | | | | | | | | | | | |
| Asian | 3721 | 0.316 | 0.465 | 0 | 1 | 2639 | 0.286 | 1082 | 0.389 | -0.103 | 0.000 | *** |
| Black | 3721 | 0.018 | 0.132 | 0 | 1 | 2639 | 0.022 | 1082 | 0.008 | 0.013 | 0.005 | *** |
| American Indian | 3721 | 0.005 | 0.067 | 0 | 1 | 2639 | 0.006 | 1082 | 0.002 | 0.004 | 0.115 | |
| Hispanic | 3721 | 0.091 | 0.287 | 0 | 1 | 2639 | 0.100 | 1082 | 0.068 | 0.031 | 0.003 | *** |
| Hispanic Withheld | 3721 | 0.133 | 0.340 | 0 | 1 | 2639 | 0.143 | 1082 | 0.109 | 0.034 | 0.006 | *** |
| White | 3721 | 0.435 | 0.496 | 0 | 1 | 2639 | 0.454 | 1082 | 0.389 | 0.065 | 0.000 | *** |
| ***PhD location*** | | | | | | | | | | | | |
| US-based Institution | 3721 | 0.504 | 0.500 | 0 | 1 | 2639 | 0.501 | 1082 | 0.510 | -0.009 | 0.610 | |
| ***Research field*** | | | | | | | | | | | | |
| Theory | 3721 | 0.237 | 0.425 | 0 | 1 | 2639 | 0.240 | 1082 | 0.228 | 0.012 | 0.436 | |
| Macro | 3721 | 0.262 | 0.440 | 0 | 1 | 2639 | 0.269 | 1082 | 0.247 | 0.022 | 0.168 | |
| Applied | 3721 | 0.243 | 0.429 | 0 | 1 | 2639 | 0.216 | 1082 | 0.310 | -0.094 | 0.000 | *** |
| Residual | 3721 | 0.238 | 0.427 | 0 | 1 | 2639 | 0.252 | 1082 | 0.202 | 0.050 | 0.001 | *** |
| ***PhD institution*** | | | | | | | | | | | | |
| RePEc Rank top-25 | 3721 | 0.183 | 0.387 | 0 | 1 | 2639 | 0.189 | 1082 | 0.166 | 0.023 | 0.098 | * |
| Rank 26-50 | 3721 | 0.133 | 0.339 | 0 | 1 | 2639 | 0.135 | 1082 | 0.127 | 0.009 | 0.480 | |
| Rank 51-100 | 3721 | 0.147 | 0.354 | 0 | 1 | 2639 | 0.143 | 1082 | 0.157 | -0.014 | 0.265 | |
| Rank 101-200 | 3721 | 0.192 | 0.394 | 0 | 1 | 2639 | 0.199 | 1082 | 0.175 | 0.025 | 0.083 | * |
| Rank 201-500 | 3721 | 0.222 | 0.416 | 0 | 1 | 2639 | 0.208 | 1082 | 0.256 | -0.048 | 0.002 | *** |
| Rank 500+ | 3721 | 0.123 | 0.329 | 0 | 1 | 2639 | 0.125 | 1082 | 0.119 | 0.005 | 0.646 | |
| Years since PhD | 3721 | 1.073 | 2.201 | 0 | 22 | 2639 | 1.144 | 1082 | 0.899 | 0.244 | 0.002 | *** |
| ***Publications (counts)*** | | | | | | | | | | | | |
| Total | 3721 | 1.156 | 2.057 | 0 | 18 | 2639 | 1.223 | 1082 | 0.991 | 0.232 | 0.002 | *** |
| Top-Five | 3721 | 0.013 | 0.121 | 0 | 3 | 2639 | 0.016 | 1082 | 0.007 | 0.008 | 0.062 | * |
| Top General Interest | 3721 | 0.020 | 0.148 | 0 | 2 | 2639 | 0.020 | 1082 | 0.018 | 0.002 | 0.765 | |
| Top Field | 3721 | 0.048 | 0.236 | 0 | 2 | 2639 | 0.055 | 1082 | 0.033 | 0.021 | 0.012 | ** |

**Notes:** Top-Field journals are JIE, JET, JoE, JME, JPubE, JLE, JDE, JEH, JFE, JF, Rand. Top-General Interest are the JEEA, REStat, EJ, IER, and all AEJs. See data section for additional information.

Table 2: Descriptive Statistics — letter and letter writer characteristics

| | Full Sample | | | | | Males | | Females | | Difference (M - F) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | N | Mean | SD | Min | Max | N | Mean | N | Mean | Estimate | p-value | sig. |
| **Panel (A) Characteristics of the letters** | | | | | | | | | | | | |
| Total Word count | 11846 | 1089 | 554 | 197 | 5000 | 8486 | 1092 | 3360 | 1081 | 11.820 | 0.295 | |
| *Word bags* | | | | | | | | | | | | |
| Ability | 11846 | 0.257 | 0.203 | 0 | 1.396 | 8486 | 0.258 | 3360 | 0.255 | 0.003 | 0.513 | |
| Grindstone | 11846 | 0.075 | 0.101 | 0 | 0.787 | 8486 | 0.073 | 3360 | 0.080 | -0.007 | 0.001 | *** |
| Recruitment | 11846 | 0.337 | 0.269 | 0 | 2.004 | 8486 | 0.340 | 3360 | 0.330 | 0.009 | 0.087 | * |
| Research | 11846 | 0.914 | 0.484 | 0 | 2.996 | 8486 | 0.921 | 3360 | 0.896 | 0.025 | 0.011 | ** |
| Standout | 11846 | 0.349 | 0.216 | 0 | 1.578 | 8486 | 0.350 | 3360 | 0.347 | 0.003 | 0.567 | |
| Teaching and Citizenship | 11846 | 0.368 | 0.333 | 0 | 1.949 | 8486 | 0.365 | 3360 | 0.376 | -0.012 | 0.086 | * |
| *Letter writers* | | | | | | | | | | | | |
| Female | 11846 | 0.148 | 0.355 | 0 | 1 | 8486 | 0.132 | 3360 | 0.188 | -0.056 | 0.000 | *** |
| RePEc Rank top-25 | 11846 | 0.188 | 0.391 | 0 | 1 | 8486 | 0.189 | 3360 | 0.183 | 0.006 | 0.439 | |
| Rank 26-50 | 11846 | 0.114 | 0.318 | 0 | 1 | 8486 | 0.116 | 3360 | 0.109 | 0.007 | 0.262 | |
| Rank 51-100 | 11846 | 0.149 | 0.356 | 0 | 1 | 8486 | 0.146 | 3360 | 0.157 | -0.010 | 0.161 | |
| Rank 101-200 | 11846 | 0.161 | 0.368 | 0 | 1 | 8486 | 0.168 | 3360 | 0.143 | 0.026 | 0.001 | *** |
| Rank 201-500 | 11846 | 0.186 | 0.389 | 0 | 1 | 8486 | 0.179 | 3360 | 0.205 | -0.026 | 0.001 | *** |
| Rank 500+ | 11846 | 0.202 | 0.401 | 0 | 1 | 8486 | 0.201 | 3360 | 0.204 | -0.003 | 0.746 | |

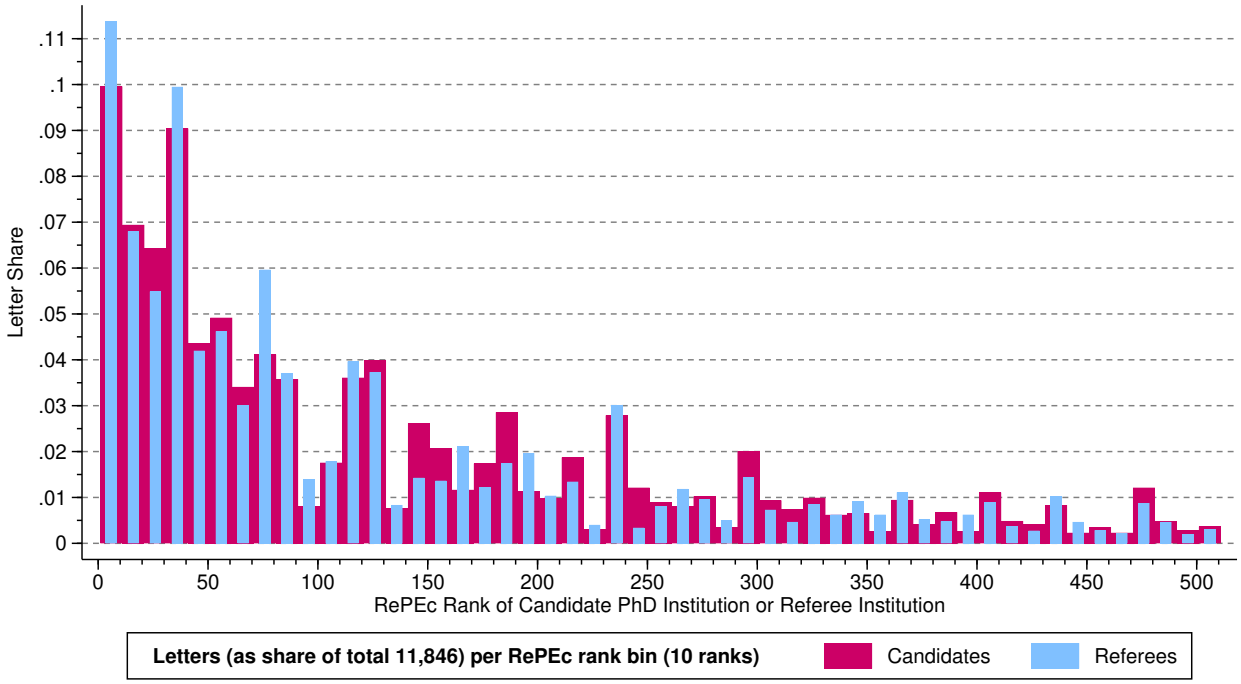
| | Full Sample | | | | | Male Writer | | Female Writer | | Difference (M - F) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | N | Mean | SD | Min | Max | N | Mean | N | Mean | Estimate | p-value | sig. |
| **Panel (B) Characteristics of the letterwriter** | | | | | | | | | | | | |
| Female Writer | 5655 | 0.174 | 0.379 | 0 | 1 | 4670 | 0 | 985 | 1 | -1 | | |
| Letters in the sample | 5655 | 2.140 | 1.951 | 1 | 23 | 4670 | 2.210 | 985 | 1.810 | 0.400 | 0.000 | *** |
| RePEc Rank top-25 | 5655 | 0.163 | 0.370 | 0 | 1 | 4670 | 0.166 | 985 | 0.153 | 0.012 | 0.346 | |
| Rank 26-50 | 5655 | 0.092 | 0.289 | 0 | 1 | 4670 | 0.090 | 985 | 0.098 | -0.008 | 0.423 | |
| Rank 51-100 | 5655 | 0.131 | 0.337 | 0 | 1 | 4670 | 0.132 | 985 | 0.125 | 0.007 | 0.540 | |
| Rank 101-200 | 5655 | 0.162 | 0.369 | 0 | 1 | 4670 | 0.162 | 985 | 0.163 | -0.002 | 0.903 | |
| Rank 201-500 | 5655 | 0.218 | 0.413 | 0 | 1 | 4670 | 0.219 | 985 | 0.210 | 0.009 | 0.519 | |
| Rank 500+ | 5655 | 0.234 | 0.423 | 0 | 1 | 4670 | 0.231 | 985 | 0.250 | -0.019 | 0.198 | |
| Assistant Professor | 5526 | 0.155 | 0.362 | 0 | 1 | 4568 | 0.143 | 957 | 0.215 | -0.072 | 0.000 | *** |
| Associate Professor | 5526 | 0.208 | 0.406 | 0 | 1 | 4568 | 0.195 | 957 | 0.271 | -0.076 | 0.000 | *** |
| Full Professor/Chair | 5526 | 0.637 | 0.481 | 0 | 1 | 4568 | 0.662 | 957 | 0.514 | 0.148 | 0.000 | *** |
| PhD Year | 5095 | 2000 | 11.430 | 1953 | 2021 | 4210 | 1999 | 885 | 2003 | -3.543 | 0.000 | *** |
| Prior to 2000 | 5095 | 0.420 | 0.494 | 0 | 1 | 4210 | 0.442 | 885 | 0.318 | 0.125 | 0.000 | *** |
| After (incl) 2000 | 5095 | 0.580 | 0.494 | 0 | 1 | 4210 | 0.558 | 885 | 0.682 | -0.125 | 0.000 | *** |

**Notes:** Institutional rankings are based on RePEc. See data section for additional information.

Since we are primarily interested in the way candidates are described, we focus our analysis on this end section. Section 3.1 explains how this section is extracted. A typical example of the information provided is given by the following quotation. Identifiable and sensitive characteristics have been redacted to protect privacy.

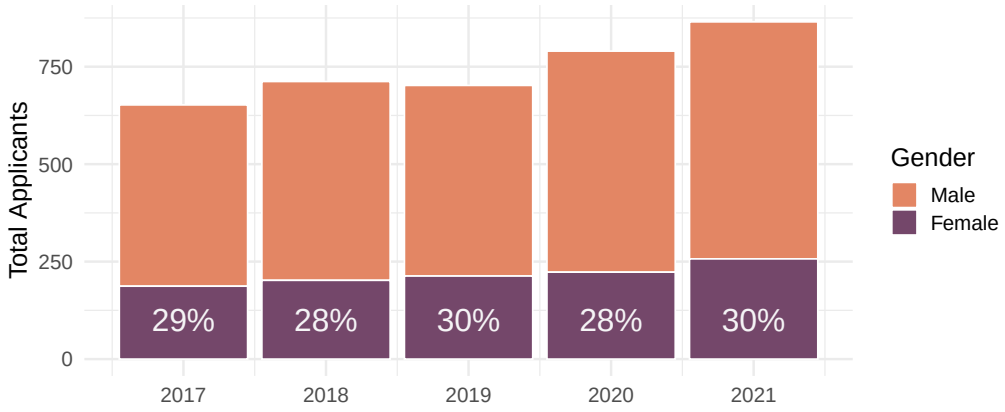
> "...working in this area. In terms of recent students coming out of [Institution X] that I have worked with, [Candidate $\alpha$] would be on a par of with a number of excellent recent placements such as [Candidate $\beta$] who went to [Institution Y], [Candidate $\gamma$], who went to [Institution Z] and [Candidate $\delta$] who went to [Institution W]. These economists are carving out excellent, innovative careers and I can see [Candidate $\alpha$] joining their ranks. What makes [Candidate $\alpha$] stand out from recent cohorts is [Candidate $\alpha$]'s ability to work with governments. [Candidate $\alpha$] has been central to the work that [Institution X] does in [Country A]. Precisely, [Candidate $\alpha$] has done such a good job starting up projects with the government and delivering answers to big, difficult to tackle questions. You can see this hallmark in all [Candidate $\alpha$]'s papers and I have a sense [Candidate $\alpha$] is going to be highly productive in [his/her] career for this reason. I therefore recommend that all top economics departments, business schools and public policy schools interested in hiring someone in [Field $\phi$] take a careful look at this application."

Figure 1: RePEc Rank of Candidate and Letter Writer Institution




**Notes:** The figure presents the frequency distribution of candidate and letter writer institutional rank (in bins of 10 institutions).

Figure 2: Gender distribution of applicants in the sample




**Notes:** The figure shows the total number of applicants per year and the share of female applicants each year

Table 3: Descriptive Statistics: Candidate Country

| ISO3 | Candidates | Percent | Cum | ISO3 | Candidates | Percent | Cum |
|---|---|---|---|---|---|---|---|
| USA | 1,874 | 50.5 | 51 | CHN | 8 | 0.2 | 99 |
| GBR | 503 | 13.6 | 64 | IRL | 8 | 0.2 | 99 |
| CAN | 191 | 5.2 | 69 | HUN | 7 | 0.2 | 99 |
| FRA | 190 | 5.1 | 74 | GRC | 6 | 0.2 | 99 |
| DEU | 164 | 4.4 | 79 | IND | 6 | 0.2 | 99 |
| ESP | 157 | 4.2 | 83 | BRA | 5 | 0.1 | 100 |
| ITA | 132 | 3.6 | 87 | RUS | 5 | 0.1 | 100 |
| NLD | 106 | 2.9 | 89 | PSE | 4 | 0.1 | 100 |
| SWE | 62 | 1.7 | 91 | TUR | 4 | 0.1 | 100 |
| AUS | 49 | 1.3 | 92 | IRN | 3 | 0.1 | 100 |
| CHE | 49 | 1.3 | 94 | ISR | 3 | 0.1 | 100 |
| BEL | 41 | 1.1 | 95 | JPN | 3 | 0.1 | 100 |
| HKG | 26 | 0.7 | 96 | MEX | 3 | 0.1 | 100 |
| DNK | 18 | 0.5 | 96 | CHL | 2 | 0.1 | 100 |
| NOR | 16 | 0.4 | 96 | BGR | 1 | 0.0 | 100 |
| SGP | 15 | 0.4 | 97 | CYP | 1 | 0.0 | 100 |
| n/a | 13 | 0.4 | 97 | GEO | 1 | 0.0 | 100 |
| PRT | 12 | 0.3 | 98 | KOR | 1 | 0.0 | 100 |
| AUT | 11 | 0.3 | 98 | MYS | 1 | 0.0 | 100 |
| CZE | 10 | 0.3 | 98 | NZL | 1 | 0.0 | 100 |
| FIN | 9 | 0.2 | 98 | | | | |
| | | | | **Total** | **3,721** | | |

**Notes:** 3-digit iso code for geographic location of the applicant (PhD institution, not nationality), in order of magnitude. Cum – cumulative sum (rounded).

# 3 Methods

## 3.1 Data processing

In this section, we explain the methods employed to transform our collection of letters into data.

Following standard procedure, we pre-process the text. First, we clean all punctuation and

clearly separate out the words. Next, we remove all common stop words such as articles or pronouns. Furthermore, we stem the words, i.e. we reduce the words to their common stem (or root). For instance, the words "published", "publishing", or "publishes", will all be collapsed to the stem "publish". Following these steps, we have converted each reference letter into a collection of (stemmed) words.

We then need to establish a measure of the importance of each word per letter. We compute the term-frequency-inverse-document-frequency (tf-idf) of each word using Python's Sklearn library.

We now define a few concepts to explain how we transform our collection of letters into data. Each letter is a *document*. Denote each document $d \in \{1, ..., D\}$. The corpus $D$ is the set of documents. Each document $d$ contains $N_d$ words $w_i(d),\ i \in \{1, ..., N_d\}$. Words are drawn from a set of terms $t \in \{1, ..., T\}$. The set of terms is the entire vocabulary present in the corpus.

We represent the corpus of letters with a matrix of dimension $D \times T$. Each row of this matrix represents a document, and each column represents a term. For each document, each cell refers to the term-frequency-inverse-document-frequency (tf-idf) of the term. The tf-idf is a common measure used to quantify the importance of a term in each document, compared to its prevalence in the corpus. The tf-idf is the product of the term frequency and the inverse-document frequency. The term frequency $\text{tf}(t, d)$ is the number of times term $t$ appears in document $d$:

$$\text{tf}(t, d) = \sum_{i}^{N_d} \mathbf{1}\{w_i = t\}. \tag{1}$$

The inverse-document frequency is the logarithmically scaled inverse fraction of the document frequency of $t$, $\text{df}(t)$, which is the number of documents that contain the term $t$:

$$\text{idf}(t) = \log \frac{1 + D}{1 + \text{df}(t)} \tag{2}$$

$$\text{with} \quad \text{df}(t) = \sum_{d} \mathbf{1}\{\text{tf}(t, d) > 0\}. \tag{3}$$

The term-frequency-inverse-document-frequency (tf-idf) is then:[8]

[8]By default, Python's Sklearn uses an L-2 normalization, which means that it normalises the final tf-idf with the vector's Euclidian norm. This is aimed at correcting for long versus short documents. Following standard procedure, we also drop terms that are either too common (i.e. that appear in more than 70% of documents) or too rare (less than 1% of documents).

$$\text{tfidf}(t,d) = \text{tf}(t,d) \times \text{idf}(t) = \log \frac{1+D}{1+\text{df}(t)} \sum_{i}^{N_d} \mathbf{1}\{w_i = t\}. \tag{4}$$

This approach is considered standard for text vectorization in natural language processing, and researchers have shown that this simple representation is sufficient to infer interesting properties from texts (Grimmer and Stewart, 2013). This approach has many advantages. First, it is easy to implement. Second, the tf-idf for each word has the simple interpretation of capturing the importance of each word in the document, relative to its frequency in the corpus. We can also measure the importance of specific attributes in each letter by summing the tf-idf for the groups of words in the attribute category for each letter.

This approach has two main shortcomings. First, the vector space grows linearly with the vocabulary, which can cause significant computational challenges. In our case, our sample size is not large enough for this to become an issue. The second shortcoming is that the relationships *between* words are not taken into account. More recent deep-learning techniques use word embedding representations resulting in a vector-space of low dimension. With word embeddings, terms represented with vectors that are close in space are semantically similar. Recent literature in law and economics has pioneered the implementation of word embeddings, for instance, to compare the similarity of different semantic fields inside a given corpus (Ash et al., 2022, 2021, among others). Many of these papers are interested in exploring whether different semantic fields are correlated in different corpora (e.g. whether 'female' words tend to be associated with 'career' words or 'family' words). Unfortunately, word embeddings may perform disappointingly compared to traditional methods in smaller samples (Shao et al., 2018; Ash et al., 2021), and our sample is much smaller than the those used in the new economics literature applying word embeddings.[9]

## 3.2 Separating ends

In most of our analysis, we concentrate on the end of the letter. The rationale behind this choice is that reference letters in economics follow a fairly rigid structure, and the end of the letter is where the referees summarize their opinion about the candidate, including their job market prospects.

We use a two-step procedure to separate the letter ends. First, we create a dictionary of commonly used closing phrases (e.g. "Yours sincerely"). These phrases flag the end of the letter, and permit cleaning out long signatures (with multiple affiliations, addresses, etc.). We then take the 200 words *before* the first closing phrase flagged, which roughly corresponds to the length of one large paragraph. With this approach, we cover more than 89% of the

[9]For instance, Ash et al.'s (2021) analysis of judge-specific corpora falls in the category of a "small" sample for word embeddings. Their analysis relies on corpora with at least 1.5 million tokens (pre-processed words). For comparison, our main sample of interest, which consists of the universe of end of letters, contains approximately 852,000 tokens.

letters. For letters without any identifiable closing phrase, we use the last 200 words of the document. We also consider 150 and 250 words cuts for the letter ends in the robustness section.

### 3.3 Language Categorisation

Reference letters for the economics job market tend to have an overwhelmingly positive tone. Therefore, a standard computational text analysis that aims at weighting positive terms against negative ones is not appropriate in this context. We build instead on the categorization proposed by Schmader et al. (2007) in their analysis of a smaller sample of applicants in chemistry ($n = 277$) for a large U.S. research university, which in turn builds on earlier qualitative work by Trix and Psenka (2003).

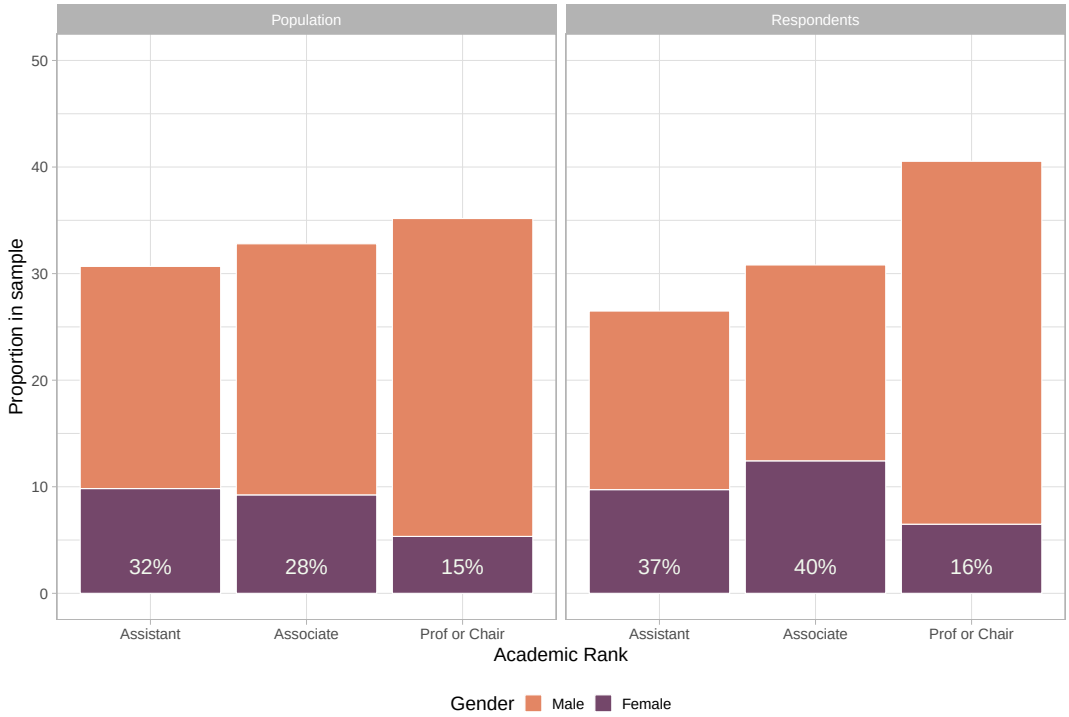
Schmader et al. (2007) propose five language categories that can be used to describe relevant features of an applicant, including *ability traits, grindstone traits, research terms, standout adjectives,* and *teaching and citizenship terms*. We add a category that refers to the *recruitment prospects* of the candidate. Ability traits involve language aimed at highlighting the applicant's suitability for the advertised position and include words such as *talent*, *intellectual*, *creative*, etc. Grindstone traits refer to language that, in the words of Trix and Psenka (2003, 207), resemble "putting one's shoulder to the grindstone". Words in this category include *hardworking*, *conscientious*, *diligent*, etc. Research terms are descriptors of the type of research carried out by the candidate and related matters e.g. *applied economics*, *game theory*, *public economics*, etc. Standout terms highlight especially desirable attributes of the applicant, like *excellent*, *top*, *strongest* etc. Teaching and citizenship is a broad category that refers both to the candidate's skills in the classroom, as well as their behavior with colleagues. Language in this group includes *good teacher*, *excellent colleague*, *friendly*, etc. The last category, recruitment prospects, has been added to identify words that, in the highly competitive and globalized labor market for fresh economics Ph.D.s, are widely used to describe the expected placement of the candidate. Words in this group include *highly recommended*, *top department*, *tenure track* etc. Appendix Figure B.2 shows word clouds for each of our language categories.

To corroborate our word classification, we carried out a survey of all faculty employed at U.K. economics departments which were submitted to the 2014 Research Excellence Framework (REF).[10] Each participant was shown a sample of 20 words and asked to classify them in one of the six categories listed above. The survey was run between the end of March and the beginning of April 2021, and a total of 1,205 individuals were contacted. Participants were incentivized with a lottery of Amazon vouchers worth £20 each. 195 took part in the survey, corresponding to 16 percent of the underlying population.

Figure 3 provides a breakdown of the population and of the survey respondents by level of seniority and gender. As can be seen, about one-third of the population (left panel) are as-

[10]The REF is a periodic, comprehensive assessment of the research carried out by UK universities. For more information, see De Fraja et al. (2019).

Figure 3: Population of academic surveyed compared to respondents



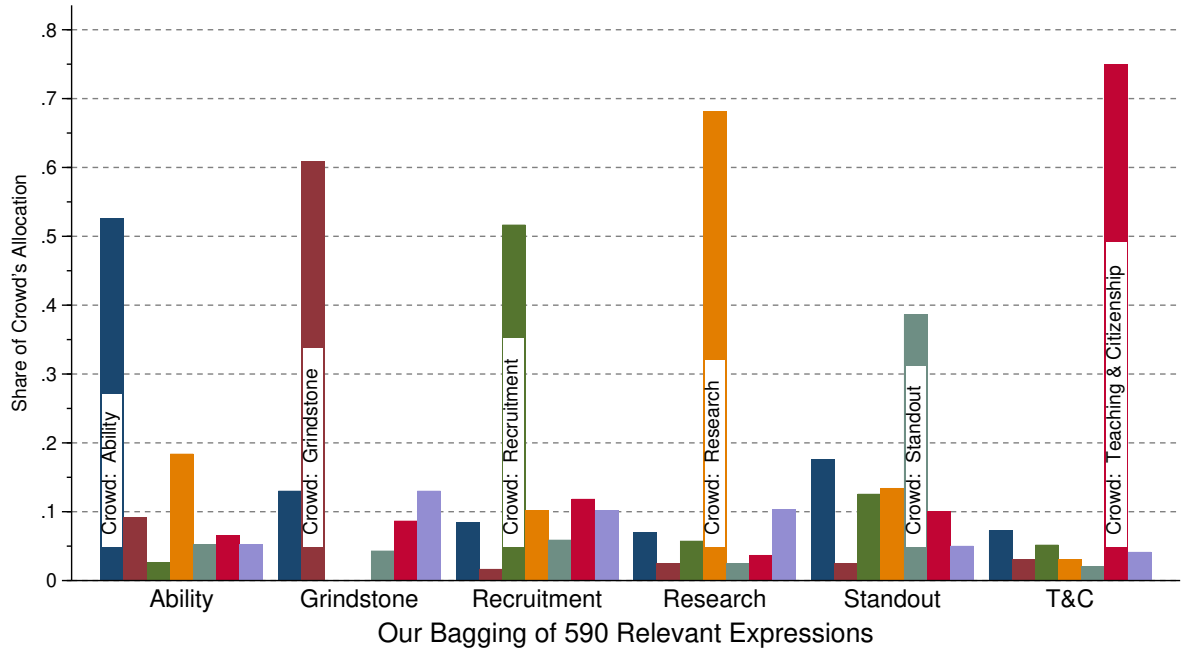

**Notes:** The figure compares the representation of women per academic rank between the total population surveyed and the respondents of the validation exercise. The percentages at the top of each bar are the share of women inside the category. The category 'others', is accounted for in the calculations, but excluded from the graphs because of its low representation (<2% of the sample in both the population and the validation sample).

sociate professors, with a slightly higher share represented by full professors, and a slightly smaller one by assistant professors. The share of females declines with seniority, representing 32% of staff at the assistant professor level, and only 15% at the most senior level. Turning to our sample (right panel), respondents are slightly more likely to be full professors, and slightly less likely to be assistant professors than in the underlying population. Not surprisingly, females are over-represented among respondents, especially at the intermediate level of seniority.

Figure 4 illustrates the extent to which our own assessment of an expression is shared by the academics who took part in our survey. For all expressions classified into a language category by the authors, we show the distribution of classifications chosen by the plurality of validators.[11] While there is variation across language categories, there is broad consensus between our categorization and that of the profession.

[11]See Appendix A for more details on how the figure is constructed.

Figure 4: Correspondence between authors' sentiment categories and the 'wisdom of the crowd'




**Notes:** This figure shows the correspondence between the authors' chosen classification for each expression and the classification chosen by validators. For any word validated, it is attributed to the category that was chosen by the plurality of validators who were shown that word. See more details in Appendix A.

# 4 Unsupervised Analysis

## 4.1 Methodology

As an initial unsupervised analysis, we ask whether specific terms used are more predictive of the gender of the candidate. To this end, we employ a least absolute shrinkage and selection operator (LASSO) to select the relevant set of terms. The LASSO estimator $\hat{\beta}$ solves the following problem:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2D} \sum_{d=1}^{D} (y_d - \boldsymbol{x}_d' \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} \omega_j |\beta_j| \right\}, \tag{5}$$

where $d$ is the letter. The gender of the candidate is the binary variable $y_d$. Vector $x_d$ is the collection of tfidf$(t, d)$ for the corpus. The second, penalty, term in equation (5) contains the 'tuning' parameters $\lambda$ and $\omega$ which are selected to reduce the number of non-zero but small coefficients. $p$ is the total number of terms.

We implement different LASSO estimators which vary in their treatment of the penalty function: for a 75% training sample, we consider a cross-validation (CV) LASSO, an adaptive LASSO, as well as an elastic net (enet) LASSO. These approaches differ in the way the optimal tuning parameters $(\lambda, \omega)$ are estimated or, in the case of the enet, by the specific form

the penalty function takes. Since female candidates make up only 30% of our sample, we also experiment with 'oversampling' females in the training sample. The final set of selected terms is not sensitive to the choice of LASSO method nor to the oversampling choice.

We only present results from the adaptive LASSO because across all specifications, it has higher predictive power than the enet and the CV.[12]

## 4.2 LASSO Results

A visualization of the results is presented in Figure 5. The Figure records the 289 predictors selected by the LASSO. We present the standardized beta coefficients of the linear probability model of candidate gender on tf-idf. Each line groups up to 6 predictors with similar coefficient magnitudes. The bars represent the range of the coefficient of the predictors listed in the line. Positive predictors are associated with female candidates, whereas negative ones are associated with males.

First, the figure reflects that women select across different research fields. Research on 'women', 'health', or 'environment(al)' tends to be disproportionately carried out by female candidates, whereas 'theory', 'history', or 'finance' appear to be associated with male candidates. This 'self-selection' mechanism is one that we also consider carefully in the remainder of the paper.

Second, qualitatively, it appears that certain personality traits are gender specific. While women are disproportionately more likely to be described as 'driven', 'determined', or 'hardworking', men are disproportionately seen as 'thinkers' or 'creative'. This is a pattern that will be confirmed in the next section. Also aligning with gender stereotyping, descriptives related to 'communal attributes' ('nice', 'pleasant person') are associated with women.

Finally, it is worth noting that traits such as youth ('young researcher') and shyness are reserved to women. This finding conforms to the stereotyping of women as naïve or child-like that has been documented in sociology (see for instance Goffman, 1979, pp. 5 and 50-51, and Gornick, 1979), and for which there is suggestive evidence that it may harm women's credibility in the workplace (for a review see MacArthur et al., 2020).

This exploratory analysis shows that even using an unsupervised method such as the LASSO, a portrait of women as 'determined' and 'hardworking' is drawn. This observation is consistent with the previous findings highlighting that female candidates are mostly praised on their 'grindstone' attributes (Trix and Psenka, 2003; Valian, 2005).

[12] We compare the areas under the receiving operator curve (AUROC). The ROC is a measure of predictive fit employed in the binary dependent variable literature, quantifying the correctly predicted 0s and correctly predicted 1s.

Figure 5: LASSO Visualization


**Notes:** This figure shows the terms selected in the LASSO exercise. In each line, the vertical bars illustrate the range of the standardized beta coefficient for all the words listed. The beta coefficient is the change in propensity that the candidate is female associated with a one standard deviation increase in the tf-idf of the term. This LASSO exercise is conducted with stemmed words. In this figure, we have attributed to each stem its most frequent corresponding word. 289 stems out of 1,425 are selected by the adaptive LASSO. $N = 11,846$, AUROC $= 0.739$.

# 5 Supervised Analysis with Dictionaries

In our supervised analysis, we employ the dictionaries related to 'ability', 'grindstone', 'research', 'recruitment', 'standout', and 'teaching & citizenship' discussed in Section 3.3 —we refer to these as 'sentiments' for ease of discussion.

## 5.1 Specification and implementation

We run regressions defined in equation (6) using ordinary least squares.

$$\text{Sentiment}_{diwt} = \alpha + \beta\,\text{Female}_i + \mathbf{X}_i'\gamma + \mathbf{W}_w'\lambda + \nu_t + \varepsilon_{diwt} \tag{6}$$

$\text{Sentiment}_{diwut}$ is the importance of each sentiment in letter $d$, written for candidate $i$ by letter writer $w$ in year $t$. For each sentiment ('ability', 'grindstone', etc.), $\text{Sentiment}_{diwut}$ is the sum of $\text{tfidf}(t, d)$ of all the terms in letter $d$ associated with that sentiment in our dictionaries. $\text{Female}_i$ is an indicator equal to 1 if the candidate is female, and $\beta$ is our coefficient of interest. $\mathbf{X}_i$ is a vector of candidate-level controls, $\mathbf{W}_w$ is a vector of letter-writer controls; both are described in more detail below. We further include recruitment cohort fixed effects $\nu_t$.

It is possible that attributes of candidates or letter writers that influence how a recommendation is written differ systematically between men and women. For instance, publication records may vary by gender, which in turn might affect the recommendation's strength (Hengel, 2022). Similarly, female candidates may not be represented in highly ranked institutions in the same way as males, etc. The variables included in the regression aim at accounting for these differences.

First, with regards to candidate attributes, all specifications include controls for their ethnicity, race, and the year they entered the job market. We sequentially add indicator variables accounting for the RePEc ranking band of the candidate's Ph.D.-awarding institution.[13] Finally, we control for the years since Ph.D. completion, for the broad field of specialization [14] and for the publication record. For the latter, we include the total number of publications and the number of articles published in top-field, top-5, and top general interest journals.[15]

Next, turning to the letterwriters' characteristics, we control for their gender, the RePEc ranking band of their institution, and the number of reference letters they provide in our sample. These controls proxy for the quality and prestige of the letter writer. Finally, we also account for the length of the letter (total word count).

Each empirical model is estimated using four different sets of standard errors: robust, clus-

[13] In particular we distinguish: top-25, top-26-50, top-51-100, top-101-200, top-201-500, and an indicator for institutions not included in our top-5% RePEc ranking in January 2021 (12% of the sample).

[14] Section 5.3 describes in greater detail how we define fields and the robustness of our results to alternative definitions.

[15] We define the following journals as top field: JDE, JEH, JET, JF, JFE, JIE, JME, JoE, JPubE, and RAND. Top general interest journals are: the four AEJs, EJ, IER, JEEA, and REStat.

tered by letterwriter, clustered by letter writer institution, and clustered by candidate Ph.D.-awarding institution.[16]

## 5.2 Main Results

**Baseline** Table 4 presents baseline results for the six outcomes using standard errors clustered by letter writer. In Figure 6 we visualize these results along those from a similar analysis carried out by further splitting the sample by letterwriter institutional ranking. Heterogeneity by institutional quality is a natural concern: familiarity with the job market might vary across institutions and in turn this might lead to different reference writing practices. Similarly, institutional culture, which may vary across the hierarchy of economic departments, can also shape language in references. For instance, Lundberg and Stearns (2019) highlight the hierarchical nature of the economics profession, in which a high fraction of potential letterwriters come from the most prestigious institutions. We address this here by focusing on top-25 or top-100 institutions and then probe this issue further in our analysis using institutional fixed effects below.

The standard errors are computed using the four types of clustering described at the end of section 5.1. The total number of outcomes, specifications and clusterings combine into a total of 504 regressions. To visualize all these results in Figure 6, a darker shading of the marker indicates more specifications yielding statistically significant estimates for $\hat{\beta}$ (see the figure's notes for more details). Fully filled symbols are significant at the 1% level across all possible standard error clusterings. Hollow symbols do not reach significance for any type of clustering. The coefficient magnitudes are the estimates from equation (6) normalized by the standard deviation of the respective dependent variable.

Figure 6 shows that no matter the institutional ranking, and across all specifications, female candidates are significantly more likely to be associated with 'grindstone' terms (from 5 to 12% of a standard deviation). These results confirm our interpretation of the unsupervised analysis (see section 4). We also observe that fewer terms related to research are used in letters supporting female candidates. Both of these results echo findings from other disciplines (Trix and Psenka, 2003; Valian, 2005).

Furthermore, in all subsamples, female candidates are on average associated weakly and insignificantly so with more 'teaching and citizenship' terms. We also find no statistically significant differences between female and male candidates for 'standout' terms —in contrast with Trix and Psenka (2003) and Schmader et al. (2007), who observe a higher frequency of these adjectives in letters supporting male applicants for academic positions in medicine, and chemistry and biochemistry, respectively.[17]

Finally, we find fewer terms related to 'ability' or 'recruitment' for female candidates, but the

[16] Exceptions here are naturally the candidate institution FE and writer FE specifications.

[17] We also experiment with separating the teaching and citizenship 'sentiments'. The coefficients remain insignificant and small in magnitude. Results are available in Appendix Figure E.1.

estimates are not statistically significant.

The magnitude of the estimates of interest does not differ greatly across specifications, even after controlling for proxies capturing determinants of language that correlate with gender. This stability provides some reassurance that other unobserved confounding determinants of language used in references are unlikely to explain away the results.

Table 4: Sentiments — End of Letters

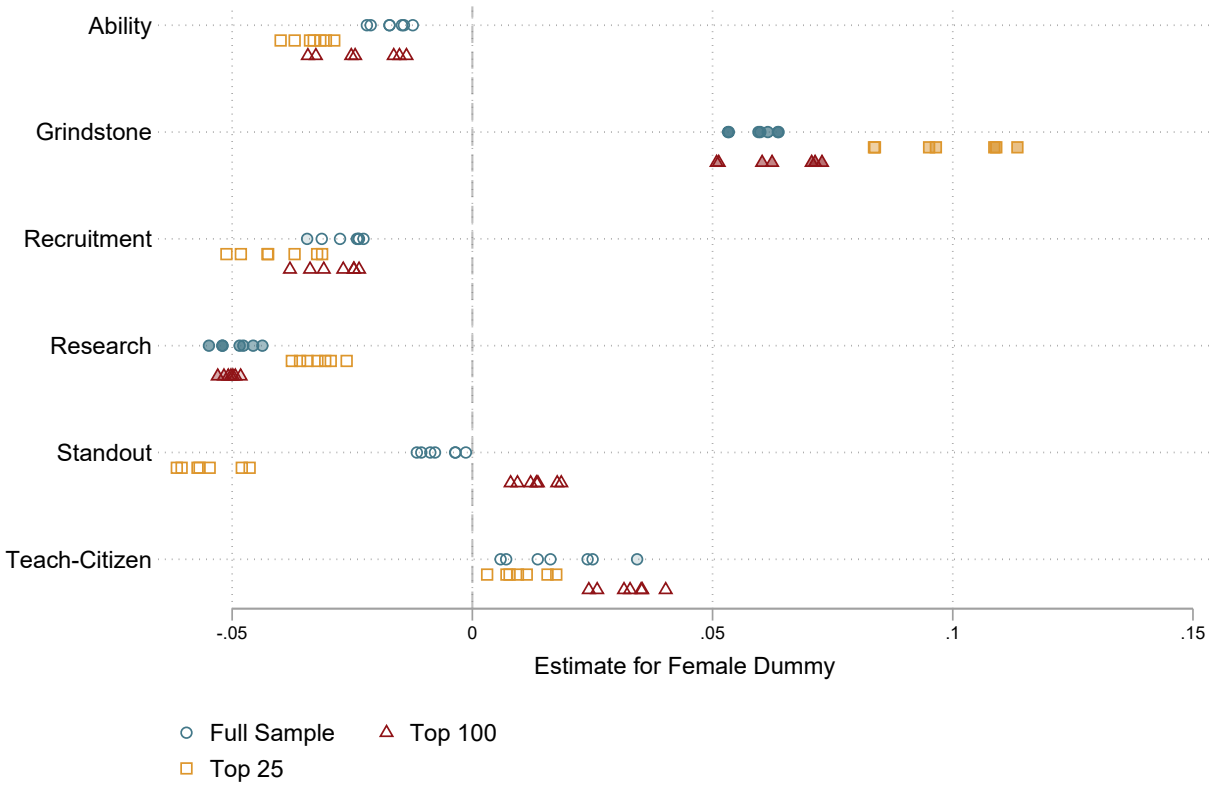
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Ability | -0.0147 | -0.0124 | -0.0142 | -0.0172 | -0.0172 | -0.0219 | -0.0212 |
| | (0.70) | (0.59) | (0.68) | (0.82) | (0.82) | (1.04) | (1.00) |
| Grindstone | 0.0637 | 0.0615 | 0.0636 | 0.0599 | 0.0596 | 0.0534 | 0.0532 |
| | (3.02)*** | (2.91)*** | (3.01)*** | (2.83)*** | (2.81)*** | (2.52)** | (2.52)** |
| Recruitment | -0.0236 | -0.0237 | -0.0227 | -0.0344 | -0.0313 | -0.0275 | -0.0240 |
| | (1.11) | (1.13) | (1.08) | (1.63) | (1.49) | (1.30) | (1.16) |
| Research | -0.0548 | -0.0520 | -0.0520 | -0.0484 | -0.0477 | -0.0437 | -0.0456 |
| | (2.66)*** | (2.52)** | (2.52)** | (2.35)** | (2.31)** | (2.11)** | (2.22)** |
| Standout | -0.0035 | -0.0013 | -0.0036 | -0.0115 | -0.0087 | -0.0106 | -0.0078 |
| | (0.17) | (0.06) | (0.17) | (0.55) | (0.42) | (0.51) | (0.37) |
| Teaching & Citizenship | 0.0343 | 0.0250 | 0.0240 | 0.0163 | 0.0136 | 0.0059 | 0.0070 |
| | (1.60) | (1.18) | (1.13) | (0.76) | (0.64) | (0.28) | (0.34) |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | 25 | 25 |
| Additional covariates | | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 11846 | 11846 | 11846 | 11846 | 11846 | 11846 | 11846 |
| dto for females | 3360 | 3360 | 3360 | 3360 | 3360 | 3360 | 3360 |
| Number of candidates | 3721 | 3721 | 3721 | 3721 | 3721 | 3721 | 3721 |
| dto female | 1082 | 1082 | 1082 | 1082 | 1082 | 1082 | 1082 |
| Number of writers | 5655 | 5655 | 5655 | 5655 | 5655 | 5655 | 5655 |
| dto female | 985 | 985 | 985 | 985 | 985 | 985 | 985 |
| Letters by fem writers | 1751 | 1751 | 1751 | 1751 | 1751 | 1751 | 1751 |
| Year FE | yes | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes | yes |
| Publications | no | no | no | no | yes | yes | yes |
| Writer characteristics | no | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. 'ability', 'grindstone', etc) on a female candidate indicator as well as controls mentioned in the text (letter ends with 200 words). The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

**Male and Female Writers** Figure 7 compares results by letterwriter gender.[18] The pattern uncovered for 'grindstone' words continues to hold when we separately consider male and

[18]There are 985 female letter writers (who have written 1,751 letters) in total, of whom only 156 are in the top-25 group (with 314 letters), and 382 in the top-100 group (735).

Figure 6: Regression results, all letter writers combined




**Notes:** This figure shows the coefficient estimates for the regressions specified in 6. We compare all seven specifications described in Table 4. The symbol's filling permits visualizing significance. Using four levels of possible standard error clustering (none, candidate's institution, letter-writer's institution, or letter writer), we flag significance at 3 different levels (10%, 5%, and 1%). We thus flag 12 possible significance indicators. For each level of clustering, the symbol in the graph is thus shaded with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any type of standard error.

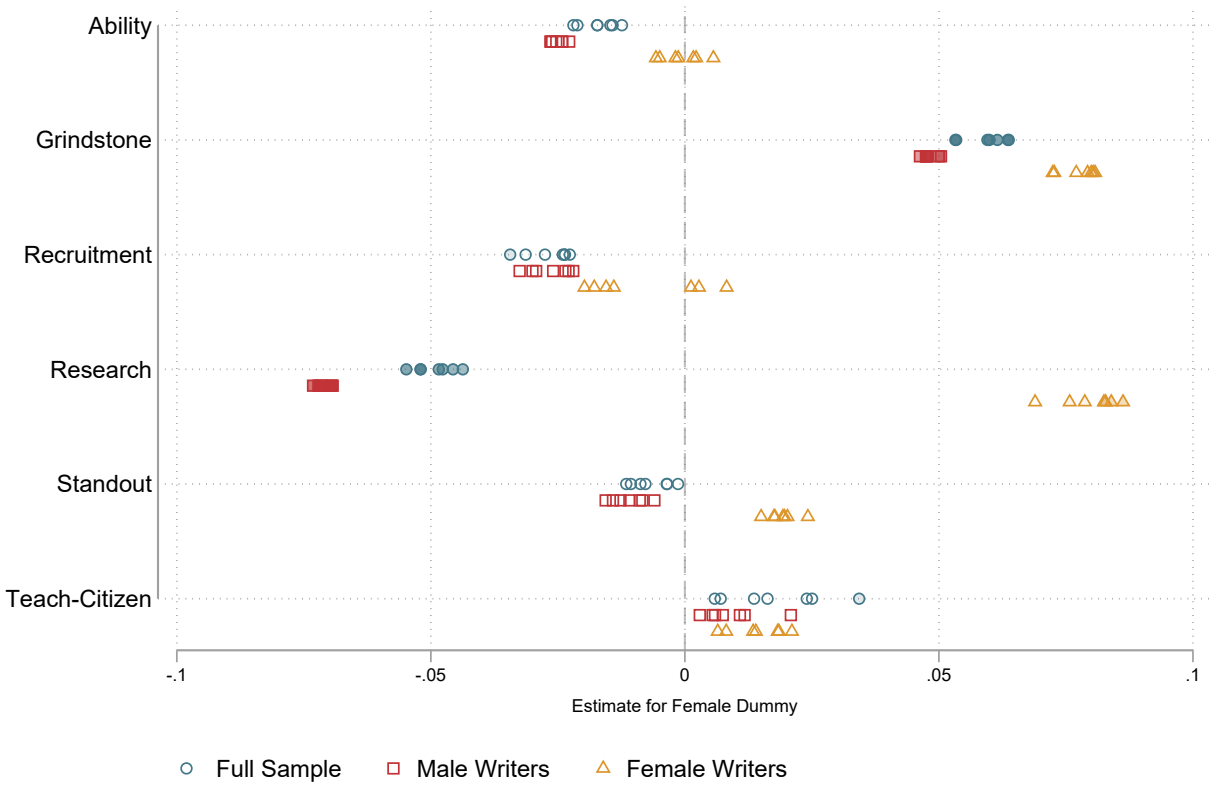
female referees. When it comes to research, it appears instead that the negative effect we have documented above is entirely driven by male writers. Female referees are actually *more* likely to use research terms for female candidates than for males.

In comparing male and female writers, we may worry that female referees cluster in departments with specific characteristics or that female writers attract different candidate types. We address these concerns later in this section by adding institution, writer or candidate fixed effects.

**Cultural Background** Gender norms differ across cultures and are highly persistent over time (see e.g. Alesina et al., 2013). Academic economists come from all over the world, and thus we can explore whether the effects we have uncovered so far are driven by writers born in countries with more traditional gender norms. To carry out this analysis we start by manually collecting, for each referee in our sample, information on their country of birth.[19]

[19]When this information was unavailable, we use the country of the institution granting their undergraduate degree as a proxy.

Figure 7: Regression results, by gender of letter writer




**Notes:** This figure shows the coefficient estimates for the regressions specified in 6, estimated separately for male and female letterwriters. We compare all seven specifications described in Table 4. The symbol's filling permits visualizing significance. Using 4 levels of possible standard error clustering (none, candidate's institution,letter-writer's institution, or letter writer), we flag significance at 3 different levels (10%, 5%, and 1%). We thus flag 12 possible significance indicators. Then, for each level of clustering, the symbol in the graph is shadowed with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error.

To measure gender norms, we then follow the literature and use data from the World Value Survey (WVS, Wave 7, 2017-2020).[20] In particular, we rely on whether the respondent agrees with the following statements: 'A pre-school child suffers with a working mother', 'University is more important for a boy than for a girl', and 'Men make better business executives than women do'. We consider a writer's country of birth as having more 'traditional' gender norms if the share of individuals agreeing/strongly agreeing with each statement is above the median for our sample.[21]

The results are reported in Figure 8. We observe that for all measures of gender norms, writers from all origins still tend to use more 'grindstone' terms for female candidates. Therefore, our results are not driven uniquely by referees born in countries with 'traditional' gender norms. However, we notice that the estimates are qualitatively larger for these writers, al-

[20] Due to lack of later data we use WVS results from 2010-14 for India.

[21] We average the shares by country across the three responses, akin to a first principal component, and take the median cut-off for this average response. Regression results for each of the three statements as well as the average are provided in Appendix Table C.4.

though the difference is not significant.[22]

Figure 8: Regression results, by gender norms of writer


**Notes:** This figure shows the coefficient estimates for the regressions specified in 6, estimated separately for letterwriters from countries with traditional versus liberal gender norms. We compare all seven specifications described in Table 4. The symbol's filling permits visualizing significance. Using four levels of possible standard error clustering (none, candidate's institution,letter-writer's institution, or letter writer), we flag significance at 3 different levels (10%, 5%, and 1%). We thus flag 12 possible significance indicators. Then, for each level of clustering, the symbol in the graph is shaded with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error.

**Specifications with Fixed Effects** We have uncovered systematic differences in the attributes highlighted for female and male candidates. Here we explore whether these differences are driven by the sorting of female candidates across institutions and/or letterwriters.

Boustan and Langan (2019) document that female representation is a persistent attribute of economics departments, and that it matters to promote women's careers. Hence, it is important to study whether institutional sorting drives our results. We run regressions including fixed effects for the candidate's institution. The results are reported in Figure 9. They suggest that among students from the same cohort, graduating from the same institution —who, for example, were admitted to PhD programs arguably applying the same entry requirements— women are still significantly more likely to be described with 'grindstone' terms.

We are still concerned that, even within the same graduate program, sorting across letter-

[22]The results for a fully interacted model are presented in Appendix Table C.4.

Figure 9: Regression results with candidate institution or writer fixed effects


**Notes:** This figure shows the coefficient estimates for the regressions specified in 6, estimated separately with candidate institution or letter-writer fixed effects. The symbol's filling permit visualizing significance. Using two levels of possible standard error clustering for each fixed effect: none or candidate's institution (resp. letter writer) for candidate's fixed effects (resp. letter-writers' fixed effects). We flag significance at three different levels (10%, 5%, and 1%). We thus flag six possible significance indicators. Then, for each level of clustering, the symbol in the graph is shaded with a 17% ($\approx 100/6$) opacity when it reaches significance at each possible level. The darker the symbol the more often they are significant. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error clustering. Additional information on the sample and results for the unclustered, robust standard errors are contained in Appendix Section C.

writers could explain our findings. To address this concern, in Figure 9 we plot also the estimates of a set of specifications including writer fixed effects.[23] Note that these models are identified from referees who have written two or more letters across all five sample years, with at least one for a female and one for a male candidate. This significantly reduces our sample (we can include only 18% of the letterwriters).

In the same figure we also separately consider the sample of referees who have less (more) experience with female candidates.[24] This analysis permits unveiling significant heterogeneity. The 'less experienced' group appears to have a significantly higher likelihood of using 'grindstone' terms for women *and* a lower likelihood of using 'ability' ones. These estimates

[23] In our analysis we drop the top-1% most prolific referees ($n = 12$), namely those with a dozen or more letters in the sample, since fixed effects estimates are sensitive to outliers. Leaving these referees in the sample leads to qualitatively similar results.

[24] Less (more) experience is defined as to balance the sample of letters across both groups. Writers who have written less (more) than a third of letters for women are considered less experienced. The 'less experienced' group accounts for 42% of referees in the subsample with two or more letters and at least one female candidate.

are also larger in magnitude compared to the baseline shown in Figure 6. Experience may matter for two main reasons. On the one hand, referees may vary in their perception of women, and female candidates could sort accordingly to avoid differential treatment. On the other hand, it could be that referees do not differ initially, but that their exposure to female candidates leads them to update preconceptions (a learning effect also observed for example by Beaman et al., 2009). Further research is needed to disentangle these two mechanisms.

Furthermore, it is worth pointing out that writers with more experience put less emphasis on 'teaching and citizenship' qualities for women compared to men, although not robustly significantly so.

Finally, fixed effects also allow a more subtle comparison of female and male *writers*. In particular we can contrast the language chosen by male and female writers for the same candidate by adding *candidate* fixed effects. For job applicants who have both male and female referees, we test whether female writers use different language in general and for female candidates in particular. Our results are presented in Appendix Table C.9. They suggest that, for the same candidate, female *writers* use different language. They rely more on 'grindstone' and 'teaching and citizenship' language, and less on 'recruitment'. Importantly, there is no difference in these patterns depending on the gender of the *candidate*, pointing towards the absence of 'same-sex preferences' of letterwriters.

## 5.3 Robustness Checks

**Alternative letter lengths** In our baseline analysis, we have defined the end of letter using the last 200 words before the 'polite' end phrase. In Appendix Figure E.2 we explore the sensitivity of the baseline results to this choice by experimenting with two alternative cutoffs, using 150 and 250 words. We also study the full reference letters (see Appendix Figure E.2).[25] Our findings are unaffected.

**Fields** We explore heterogeneity of the results according to the candidate's research field to assess possible sub-cultural differences in the profession. Grouping applicants into meaningful research areas is challenging. On the EJM platform, they typically choose a field, loosely based on JEL codes. Unfortunately, the EJM fields pool diverse subgroups of the profession, i.e. scholars that are unlikely to publish in the same journals or participate in the same events (conferences, seminars, etc.). For instance, the EJM field 'Development and Growth' includes both macroeconomists working on long-run growth and microeconomists carrying out field experiments in developing countries. Given these shortcomings, we employ an unsupervised data-driven approach to classify candidates into three broad research groups, namely Applied Economics, Theory, and Macroeconomics, and a residual category. Appendix sec-

[25] Appendix figures are accompanied by corresponding tables providing further details on specification and sample. For ease of presentation, we do not refer to these tables in the maintext.

tion D describes the procedure.[26]

One possible explanation for the association of women to 'grindstone' expressions is that they sort into research fields that require more industriousness than ability. This set of skills is often associated with empirical work. However, the results reported in Appendix Figure D.3 show that this association remains strong and significant within Applied Micro, casting doubt on such a hypothesis. We also uncover a strong negative effect for 'ability' and 'research' within Theory. This finding is worth highlighting as in this field raw talent is arguably valued very highly. This observation sheds new light on earlier findings by Leslie et al. (2015), according to whom academic fields that particularly emphasize the role of raw talent are characterized by lower female representation.

**Main Advisors** So far we have used all the letters that were submitted for each applicant, i.e. those which were written by the main advisor and those written by other faculty members familiar with the candidate's research. As the main advisor might have better knowledge of the applicant, it is important to investigate whether there are differences in the language he/she used compared to that of the other referees. We collect data on the identity of the letter writers for candidates who were in the job market up to three years after completing their Ph.D.[27] The results of our analysis are illustrated in Appendix Figure E.4, where we report our baseline estimates for the collected sample and those obtained focusing separately on the letters written by the main advisor and the other reviewers.

The findings indicate that the patterns for 'grindstone' terms are generally comparable, but accentuated for letterwriters who are *not* the main advisors. Moreover, there is notable divergence in case of 'ability' and 'standout'. Compared to main advisors, other referees use significantly fewer 'ability' and 'standout' terms. Overall, this analysis presents suggestive evidence that main advisors are writing more favorable letters for women compared to other referees. Main advisors arguably know the candidates much better and spend more time writing and polishing the letters[28] and through these lengthy processes some of their preconceptions may be toned down.[29]

**Location of PhD granting institution** The job market for economists is historically a U.S. institution, and faculty members based there may be better acquainted with the standards of

[26]The baseline analysis reported in Figure 6 employs research field fixed effects using these four clusters. In Figure E.3 we repeat the same exercise using instead the more detailed 145 field definitions from EJM. The findings are robust.

[27]This represents around 50% of the sample of candidates. Candidates who defended earlier were less likely to have a letter from their Ph.D. advisor and were also less likely to report that information on their CV.

[28]Letters from main advisors are on average 33% longer.

[29]The *Oxford English Dictionary* defines a stereotype as a 'widely held but fixed and oversimplified image or idea of a particular type of person or thing.' Describing women as hardworking conforms to the stereotype of women in science (Valian, 1999). Our results suggest that the most informed letterwriters —main advisors or writers with greater experience with female candidates— use language that is less in line with these stereotypes. These patterns align with the interpretation that less informed writers are 'stereotyping', in the sense of using an 'oversimplified image' as a shortcut.

reference writing. We investigate whether our results are driven by letterwriters outside the U.S., in which case our findings might result from lower levels of experience in the process. Figure E.5 presents the results. Overall, we do not uncover significant differences between the two groups, with the exception of 'research' terms. Referees based outside the U.S. use significantly fewer research-related words for female candidates compared to their U.S.-based counterparts.

**Candidate's visibility** So far our analysis has accounted for the underlying potential of the candidate by controlling for the number and quality of their publications and the ranking of their institution. Additionally, we have shown that our results continue to hold when we compare candidates within the same institution. As a further robustness check, we also account for the circulation of the candidate's job market paper at the time they are on the market. This proxies for the candidate's visibility and/or the extent of networking carried out in this period. We do so by manually collecting information from the job market paper acknowledgements. Using the Stanford Name Entity Recognition Tagger, we separate out *people* thanked and *institutions* mentioned. We also compute the length of this note and flag whether the job market paper is single-authored. Appendix Figure E.3 shows that results with these controls remain unchanged compared to the baseline.

**Postdocs** Female candidates — who may be conscious of potential gender stereotyping — may change their behavior during their career to make stereotypical traits less salient (e.g. Hengel, 2022, highlights that women improve their writing throughout their careers, whereas men do not). To assess this possibility, we contrast the estimates for candidates fresh out of Ph.D. programs and those who have been out for 1-3 years ('post-docs'). Results are reported in Appendix Figure E.6. Overall, the estimates for the postdoc sample are noisier, as expected due to smaller sample sizes. For 'grindstone' language, the effects remain generally stable. We do however observe an increase in the language related to 'ability' for female postdocs compared to their male counterparts. Further research is needed to establish whether this effect is driven by learning, as found by Hengel (2022), or by differential selection into postdocs.

**Letterwriter Seniority/Rank** Results with fixed effects in Figure 9 indicate that writers with less experience with female candidates use more stereotypical language. One alternative explanation for this finding could be that such practice declines with academic experience *per se*. Using manually collected information on the year the letterwriter graduated from their Ph.D. or their academic rank (assistant, associate, full professor), we illustrate that this is not the case. In Appendix Figures E.7 and E.8, we observe that the most senior letterwriters use gender stereotypical language much more often and that they drive the 'grindstone' results.[30]

[30]We conduct additional analysis (not reported) splitting the Ph.D. cohorts into five rather than two groups and obtain qualitatively identical findings.

## 5.4 Additional Results

In this section, we shift our attention away from the analysis of the 'sentiments' expressed in letters ('ability', 'grindstone', etc.) and consider alternative attributes that speak to the quality of the candidate or how the letter is written.

**Placement Qualifiers** Many letter ends carry explicit signals about the candidate, which can be positive or negative, as well as comparisons with placements of recent graduates —see our earlier discussion in Section 2. To analyse potentially gendered patterns in the prevalence of these signals, we compile a dictionary of over 1,000 placement qualifiers. Examples include (for negatives) "except maybe from those in the top 10/20/30" or "apart from the very best"; (for positives) "great hire", "a star candidate", "including institutions at the very top"; (for comparatives) "compared to" or "on par." 24% of letters in our sample include at least one positive signal, 13% a negative one and 6% of all letters include a comparative term.

In the baseline specification, we replace the dependent variable with outcomes related to these qualifiers. The first three lines in Figure 10 show that letters written in support of women tend to have significantly fewer positive signals, no significant difference in terms of negative ones, and a net negative signal.[31] These results also hold when we consider instead binary variables flagging the presence of either positive or negative signals, or the sign of the net signal.[32] The effects are sizeable. For instance, a letter in support of a female candidate has a 3 percentage point lower probability of containing a positive signal, to be compared with the fact that only 24% of letters contain one.

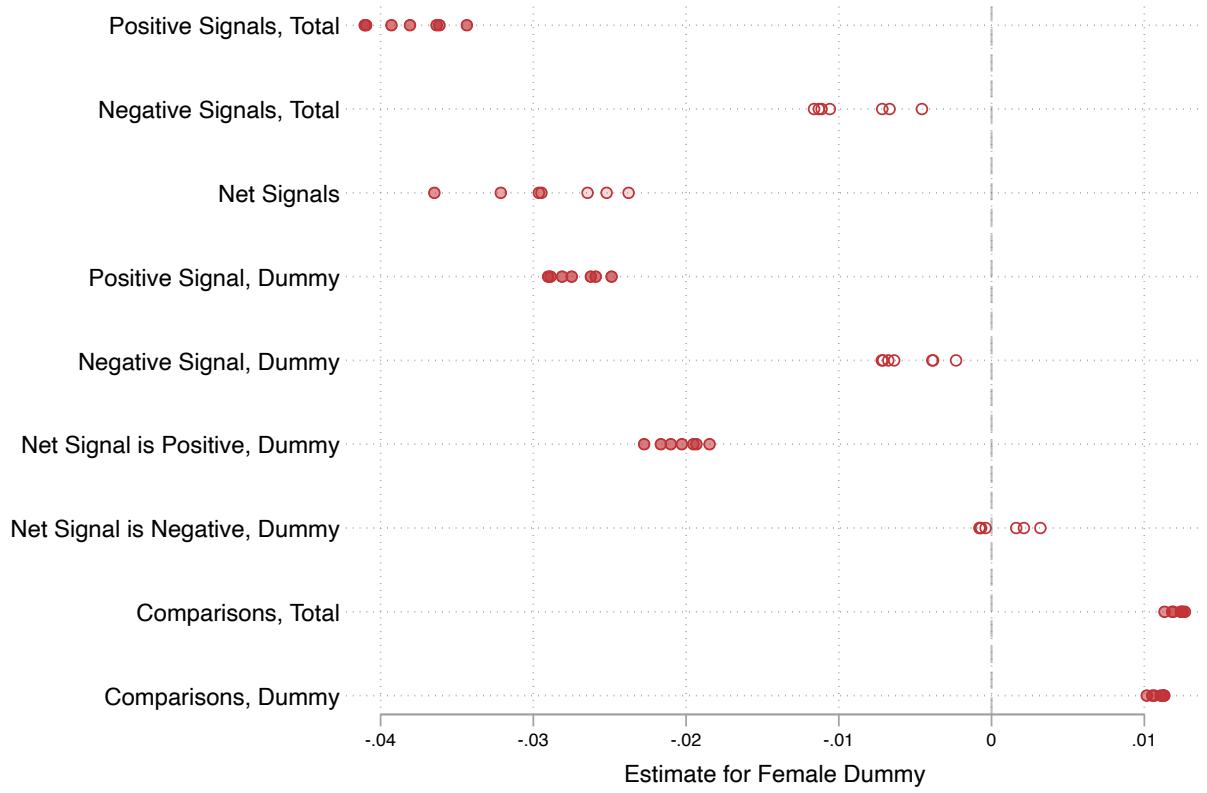
Finally, we study comparative terms using total counts or an indicator for their presence. This analysis suggests that letters written in support of female candidates have a 1 percentage point higher likelihood of carrying a comparison, a sizeable effect given that only 6% of letters contain one.

Overall, this analysis suggests that women are not shown in a more negative light (in contrast to findings in the literature about 'doubt raisers', e.g. Trix and Psenka, 2003; Madera et al., 2009). However, they obtain less outright praise, which is consistent with the work of Dutt et al. (2016), who find that women in geosciences are less likely to receive 'excellent' letters. The higher prevalence of comparative terms suggests though that the information provided for female candidates might be more 'precise'.

**Letter Length and Readability** A standard finding in the literature suggests that letters for female candidates are shorter (Trix and Psenka, 2003). We also investigate proxies for letterwriter effort by looking at letter length and writing clarity in the full reference letter as well as on the discussion of the candidate's Job Market Paper.[33] Our results, reported in

[31] The net signal simply subtracts the count of negative from that of the positive signals.
[32] The dummy for net negative (positive) signal is 1 if the net signal is negative (positive) and 0 otherwise.
[33] This corresponds to the 'research slice' — see Appendix section D for more details.

Figure 10: Regression results, placement signals as outcomes




**Notes:** This figure shows the coefficient estimates for the regressions specified in 6 when outcomes are proxies for academic placement. Rows 1-3 are for counts of positive, negative and net signals; rows 4-7 adopt binary variables for the same outcomes; the final two rows are counts and dummy for comparative statements in the letter end. The symbol's filling permit visualizing significance. Using four levels of possible standard error clustering (none, candidate's institution, letter-writer's institution, or letter writer), we flag significance at three different levels (10%, 5%, and 1%). We thus flag twelve possible significance indicators. Then, for each level of clustering, the symbol in the graph is shaded with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol the more often they are significant. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error clustering.

Appendix Figure F.1 show that female candidates in economics do not receive shorter letters than their male peers. However, the analysis of readability, using the Flesch Reading Ease score suggests that letters for female candidates are harder to read.[34] The same pattern holds true for the Dale-Chall Readability score, even if the results are not statistically significant. Finally, our investigation of the 'research slice' provides no clear evidence of a bias in the discussion of the candidate's JMP.

**Timing of the Reference Letter** Work by Baltrunaite et al. (2022) finds that, for their sample of references submitted to two Italian institutions, letterwriters are significantly less likely to send references for female candidates by the deadline stated on the EJM platform.

We test whether this finding also holds in our sample, using two alternative measures of

[34]See Hengel (2022, Table 1) for exact definitions. For the Flesch index a higher score means the text is easier to read, for the Dale-Chall the reverse is the case.

timeliness. First, we construct a binary variable equal to 1 if the application package contains strictly fewer than 3 letters, the required number in our application process. The results, shown in Appendix Figure F.1, confirm that female candidates are significantly more likely to have an incomplete set of references.

Second, we exploit information on the date stated on the reference letter, which is a proxy for the timeliness of the referee given that by default the same letter is automatically submitted by the EJM platform to all institutions.[35] Our results, shown in the same figure, indicate that letters for female candidates tend to be written earlier (between 5 to 10 days earlier, depending on the specification), but this effect is imprecisely estimated and not robust to the inclusion of a full set of controls.

# 6 Placement

So far our analysis has focused on gendered patterns in the reference letters. We have documented differences in the language chosen for female and male candidates. In line with previous literature, we observe that women are described with more 'grindstone' attributes and at times fewer 'ability' and 'research' ones (Bourdieu and Passeron, 1977; Trix and Psenka, 2003; Schmader et al., 2007). The obvious corollary question is whether being described as a 'grindstone' candidate actually affects job prospects.

To answer this question, we have carried out a systematic collection of the first professional placement of each candidate in the year following their appearance in our sample. Combining information from personal websites, academic departments' placement records, and LinkedIn profiles, we establish whether the candidate placed in academia or elsewhere. For academic placements, we also link the name of the institution to its RePEc rankings to proxy for the prestige of the job.

It is important to be mindful that reference letters are only one input in the recruitment process. They typically play an important role in enabling candidates to secure interviews at the job market meetings. Many other factors such as presentation skills, research agenda, or departmental politics determine whether a candidate actually receives a job offer. Therefore any result linking letters to actual placement needs to be interpreted with caution.

To study the relation between letter sentiment and placement, we estimate the following

[35] We focus only on recent Ph.D.s for whom letters are presumably written for the first time. We rely on the package `ctparse` to detect and parse dates in the beginning of the letters, which we also double-check manually.

regression model:

$$\begin{aligned} \text{Placement}_{diwt} &= \alpha + \beta\,\text{Female}_i \\ &+ \sum_{k=1}^{6} (\theta_k\,\text{Sentiment}_{diwt} + \kappa_k\,\text{Sentiment}_{diwt} \times \text{Female}_i) \\ &+ \mathbf{X}_i'\gamma + \mathbf{W}_w'\lambda + \nu_t + \varepsilon_{diwt}, \end{aligned} \tag{7}$$

where $\text{Placement}_{diwt}$ is the job market outcome of individual $i$, for whom letter $d$ was written by writer $w$ in year $t$. For each sentiment $k$ — 'ability', 'grindstone', etc. — we are interested in its impact on placement and whether this impact varies with the gender of the candidate. Controls are the same as defined in section 5.

Our results are reported in Table 5. We consider three measures of placement.[36] The first is a binary variable flagging whether a candidate obtained an academic job and takes into consideration all candidates for whom we found job market outcomes. The second and third measures focus on those who embarked on an academic career, and study the 'quality' of the academic placement. Our first proxy is the RePEc institution score, a continuous variable, which we rescale for ease of interpretation so that a positive coefficient indicates a more prestigious institution. The other is a binary variable indicating whether the candidate placed among the top-200 institutions in RePEc.[37] Odd columns report a parsimonious model, with only the sentiments and cohort fixed effects as controls, whereas even ones report estimates accounting for the full set of controls.

Overall, we observe that women are more likely to place in academia (columns 1, 2), and conditional on embarking on an academic career, they land jobs in more prestigious institutions (columns 4 and 6). These results are compatible with both positive selection of women in academia and 'positive discrimination'. Evidence of both phenomena has been uncovered by recent literature (for positive selection, see Iaria et al., 2022, for 'positive discrimination' see Card et al., 2022).

We turn now to the analysis of the effect of 'sentiment' on placement. 'Standout' and 'teaching and citizenship' terms are the only ones to significantly affect the probability of placing in academia. Columns 1 and 2 show that a one standard deviation increase in the usage of 'standout' terms is associated with a 2 percentage point higher likelihood of an academic placement. For female candidates the aggregate effect is effectively nil instead. For 'teaching

[36] Models 1 and 2 contain letters for all 2,588 candidates for whom placement information was found, the dependent variable is 1 for an academic placement (AP position or postdoc) and 0 otherwise (international organisations, central banks, or private sector). Models 3 and 4 contain letters for 957 candidates who placed among the top-500 institutions in RePEc, the only ones for which a RePEc score is computed. The RePEc score is a continuous variable (e.g. third-placed UC Berkeley has a score of 7.12, first-placed Harvard of 1.96). Models 5 and 6 include letters for 1,865 candidates who placed in academia as either AP or postdocs. In this sample we can include all academic institutions (we are not constrained by the availability of a RePEc score). Teaching fellows are included in academic placements in 1 and 2, but results remain identical if we exclude them.

[37] Our results are robust to alternative measures of placement quality, please refer to Appendix Table F.3.

and citizenship' we find no effect for male candidates, but a positive one for women (amounting to a 1.7 percentage point increase in the likelihood of academic placement).

In columns 3 and 4 three 'sentiments' stand out: 'grindstone', 'standout', and 'teaching and citizenship'. The results for 'grindstone' indicate no statistically significant effect for men, whereas for women a one standard deviation increase in this 'sentiment' is associated with a large and significant (10 points in the rank score) decrease in the ranking of the institutions where they place.[38] Moreover, women benefit more from standout terminology (6.6 to 7.1 increase in the rank score). Finally, men who receive letters emphasizing 'teaching and citizenship' get jobs in higher-ranked institutions (1.5 to 4.3 points increase in rank score), whereas the effect is reversed for women (6.2 to 7.4 points decrease in rank score).

When considering the likelihood of placing in a top-200 institution (columns 5 and 6), the 'grindstone' sentiment plays again an important role, especially for women. A one standard deviation increase in 'grindstone' terms is associated with a negligible effect for male candidates but a large negative effect for women, in the order of a 2.5 percentage points decrease in the probability of obtaining a job in a higher-ranked institution. Letters with more 'standout' and 'recruitment' terms are associated with better placement for both men and women, although significance and magnitudes drop for men when adding controls. Results for other sentiments are not statistically significant.

In this analysis, we do not account for the presence of placement qualifiers, which also exhibit gendered patterns as shown in Figure 10. The reason is that many placement terms are already included in the 'standout', 'ability', or 'recruitment' sentiments, although without differentiating between positives and negatives. As a robustness check, we present in Appendix Table F.4 the results when including the binary variables for placement qualifiers. The results for sentiments commented on above remain unchanged. The analysis also suggests that positive signals only improve placement for men, and comparatives mostly worsen outcomes for women.

This discussion indicates that the gendered sentiments expressed in job market reference letters are associated with initial placement patterns. Our RePEc score/ranking analysis indicates that 'grindstone' terminology hurts female placement, whether we consider a continuous measure of institutional quality or a binary identifier.[39] Although further research is needed to establish causality, our results indicate that the language in reference letters can play an important role in the first step of the academic career. These results are consistent with findings by Baltrunaite et al. (2022) on longer-term career outcomes.

[38] At the median rank score of 144, a 10 point decline in the score represents a drop of approximately 10 positions in the rank.

[39] Note that more work is needed to establish clear-cut implications in terms of discrimination. On the one hand, one may argue that if employers are seeking to have a balanced workforce in terms of 'grindstone' and other attributes, then penalising 'grindstone' women could just compensate for their greater propensity to exhibit those traits. On the other hand, a fully-fledged model of the job market should account for the fact that letterwriters could strategically adjust and choose fewer 'grindstone' attributes for their female candidates in order to increase the chances of securing a better placement.

Table 5: Letter Sentiment and Placement

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dependent Variable | Academia (dummy) | | Inst. RePEc Score | | Top-200 RePEc Inst. | |
| Sample | All Placements | | Academic Placements | | AP & Postdoc | |
| Controls | Sentiment | All | Sentiment | All | Sentiment | All |
| Female Candidate | 8.9056 | 7.7799 | 12.3080 | 19.5029 | 7.2032 | 10.8623 |
| | (2.37)** | (2.10)** | (0.84) | (1.36) | (1.56) | (2.43)** |
| Ability | 0.1695 | 0.0250 | 0.7169 | 0.5676 | 0.9848 | 0.8568 |
| | (0.28) | (0.04) | (0.29) | (0.23) | (1.39) | (1.26) |
| Ability × | -0.0861 | -0.1182 | 2.5162 | 1.5801 | 0.3787 | 0.2476 |
| Female Candidate | (0.08) | (0.11) | (0.59) | (0.37) | (0.29) | (0.19) |
| Grindstone | -0.4843 | -0.6444 | -2.6959 | 1.7869 | -0.3714 | -0.1503 |
| | (0.82) | (1.10) | (1.12) | (0.77) | (0.51) | (0.22) |
| Grindstone × | 0.1153 | 0.4257 | -10.0826 | -10.1151 | -2.4636 | -2.5101 |
| Female Candidate | (0.11) | (0.43) | (2.38)** | (2.41)** | (2.02)** | (2.09)** |
| Recruitment | 0.7008 | 0.8262 | 3.1951 | 2.1187 | 1.8494 | 0.8343 |
| | (1.14) | (1.34) | (1.35) | (0.90) | (2.58)*** | (1.23) |
| Recruitment × | 0.4961 | 0.2396 | -3.2295 | -2.8548 | -0.4627 | -0.3362 |
| Female Candidate | (0.48) | (0.23) | (0.80) | (0.71) | (0.37) | (0.27) |
| Research | -0.4928 | -0.1069 | 3.2833 | 4.3560 | 0.6711 | 0.9642 |
| | (0.80) | (0.18) | (1.39) | (1.89)* | (0.95) | (1.43) |
| Research × | -1.7432 | -1.5594 | 0.8732 | -0.4482 | -0.6655 | -0.9822 |
| Female Candidate | (1.63) | (1.47) | (0.21) | (0.11) | (0.51) | (0.77) |
| Standout | 1.9208 | 1.9660 | -0.0326 | -0.8585 | 1.2065 | 0.2830 |
| | (3.28)*** | (3.39)*** | (0.01) | (0.35) | (1.74)* | (0.43) |
| Standout × | -1.9979 | -1.8693 | 7.1538 | 6.6114 | 2.5831 | 2.2876 |
| Female Candidate | (1.85)* | (1.76)* | (1.72)* | (1.65)* | (1.94)* | (1.77)* |
| Teaching and | 0.1599 | -0.2112 | 1.5916 | 4.3727 | -1.1806 | 0.4990 |
| Citizenship | (0.26) | (0.35) | (0.68) | (1.86)* | (1.63) | (0.72) |
| T&C × | 1.6494 | 1.7722 | -6.2089 | -7.4446 | -3.4235 | -4.0059 |
| Female Candidate | (1.61) | (1.75)* | (1.55) | (1.88)* | (2.75)*** | (3.29)*** |
| FE absorbed | 5 | 25 | 5 | 25 | 5 | 25 |
| Add. covariates | 0 | 6 | 0 | 6 | 0 | 6 |
| Number of Letters | 8760 | 8760 | 3119 | 3119 | 6008 | 6008 |
| dto for females | 2588 | 2588 | 991 | 991 | 1872 | 1872 |
| Number of candidates | 2738 | 2738 | 957 | 957 | 1865 | 1865 |
| dto female | 830 | 830 | 313 | 313 | 596 | 596 |
| Number of writers | 4461 | 4461 | 2091 | 2091 | 3453 | 3453 |
| dto female | 774 | 774 | 324 | 324 | 586 | 586 |
| Letters by fem writers | 1339 | 1339 | 445 | 445 | 910 | 910 |
| Year FE | yes | yes | yes | yes | yes | yes |
| Letter Sentiments | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | no | yes | no | yes | no | yes |
| Institution Rank FE | no | yes | no | yes | no | yes |
| Years since PhD | no | yes | no | yes | no | yes |
| Research Field FE | no | yes | no | yes | no | yes |
| Publications | no | yes | no | yes | no | yes |
| Writer Chars | no | yes | no | yes | no | yes |
| Letter length | no | yes | no | yes | no | yes |

**Notes:** The table shows OLS regression results of placement outcomes on the letter-specific sum of tf-idf statistics related to the bag of expressions mentioned in the row label and its interaction with a female candidate dummy as well as the additional controls as indicated. We report the absolute $t$-statistics in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Results are in percentage points except in (3) and (4) where they are in percent. Further, signs in (3) and (4) are reversed for consistency with the other two placement outcomes. Additional results are presented in Appendix Tables F.3 and F.4.

# 7 Concluding Remarks

In this paper, we carried out what is to the best of our knowledge the first systematic analysis of recommendation letters in the junior academic job market in economics. Using both supervised and unsupervised methods, we have documented the presence of important differences in the language used to describe female applicants. Women are more often described with terms praising their 'hard work' or 'dedication' than men. This pattern is robust to alternative specifications and holds across many subsamples of the data. Similarly, we uncover evidence of a lower emphasis on 'ability', especially when comparing individuals within the same institution or for those sharing the same referee.

Sociologists characterise these systematic language patterns as possibly resulting from stereotyping, and highlight their potential negative connotations as a strong emphasis on diligence may imply a lack of 'brilliance' (Bourdieu and Passeron, 1977; Valian, 1999). We illustrate the salience of language in reference letters for job market placement by documenting that women receiving letters emphasizing that they 'work hard' obtain less prestigious academic positions, while the same is not true for men. On the contrary, those whose letters highlight 'standout' attributes benefit from improved academic placement. Although further evidence is needed, our results thus suggest that for letterwriters who are pushing their candidates towards research-intensive academic employment, using a language emphasizes less 'grindstone' and more 'standout' attributes increases the chances of achieving the desired objective.

As academics, we know how much time is spent writing and polishing reference letters for job market candidates. This is an occasion where we try our best to promote our students. Therefore, it is unlikely that, on average, we are willingly undermining female students by emphasizing less desirable attributes. On a positive note, recent research has shown that unconscious biases can be addressed by providing the actors involved with evidence of the existence of such biases (Boring and Philippe, 2021). By shedding light on these patterns, we hope this research will be a first step towards increasing awareness of our biases and thereby reducing possible stereotyping in the job markets.

# References


Sule Alan, Teodora Boneva, and Seda Ertac. Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*, 134 (3):1121–1162, 2019.

Alberto Alesina, Paola Giuliano, and Nathan Nunn. On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, 128(2):469–530, 2013.

Elliott Ash, Daniel L Chen, and Arianna Ornaghi. Gender attitudes in the judiciary: Evidence from u.s. circuit courts. Quantitative and Analytical Political Economy Research Centre (QAPEC Discussion Papers) 08, 2021.

Elliott Ash, Daniel L Chen, and Suresh Naidu. Ideas have consequences: The impact of law and economics on american justice. NBER Working Paper 29788, 2022.

Audinga Baltrunaite, Alessandra Casarico, and Lucia Rizzica. Women in economics: the role of gendered references at entry in the profession. CEPR Discussion Paper 17474, 2022.

Amanda Bayer and Cecilia Elena Rouse. Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives*, 30(4):221–42, 2016.

Lori Beaman, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. Powerful Women: Does Exposure Reduce Bias? *The Quarterly Journal of Economics*, 124(4): 1497–1540, 2009.

Anne Boring. Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145:27–41, 2017.

Anne Boring and Arnaud Philippe. Reducing discrimination in the field: Evidence from an awareness raising intervention targeting gender biases in student evaluations of teaching. *Journal of Public Economics*, 193:104323, 2021.

Clément Bosquet, Pierre-Philippe Combes, and Cecilia García-Peñalosa. Gender and promotions: evidence from academic economists in france. *The Scandinavian Journal of Economics*, 121(3):1020–1053, 2019.

Pierre Bourdieu and Jean-Calude Passeron. *Reproduction in Education, Society, and Culture*. Sage, 1977.

Leah Boustan and Andrew Langan. Variation in women's success across PhD programs in economics. *Journal of Economic Perspectives*, 33(1):23–42, 2019.

David Card, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. Are referees and editors in economics gender neutral? *The Quarterly Journal of Economics*, 135(1):269–327, 2020.

David Card, Stefano DellaVigna, Patricia Funk, and Naorre Iriberri. Gender differences in peer recognition by economists. *Econometrica*, 90(5):1937–1971, 2022.

Stephen J Ceci and Wendy M Williams. *The mathematics of sex: How biology and society conspire to limit talented women and girls*. Oxford University Press, 2009.

Peter Coles, John Cawley, Pillip B. Levine, Muriel Niederle, Alvin E. Roth, and John J. Siegfried. The job market for new econonomists: A market design perspective. *Journal of Economic Perspectives*, 24(4):187– 205, 2010.

Gianni De Fraja, Giovanni Facchini, and John Gathergood. Academic salaries and public evaluation of university research: Evidence from the UK Research Excellence Framework. *Economic Policy*, 34:523–583, 2019.

Pascaline Dupas, Alice Sasser Modestino, Muriel Niederle, Justin Wolfers, and The Seminar Dynamics Collective. Gender and the Dynamics of Economics Seminars. NBER Working Paper 28494, 2021.

Kuheli Dutt, Danielle L Pfaff, Ariel F Bernstein, Joseph S Dillard, and Caryn J Block. Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nature Geoscience*, 9(11):805–808, 2016.

Y Fan, LJ Shepherd, E Slavich, D Waters, M Stone, R Abel, and EL Johnston. Gender and cultural bias in student evaluations: Why representation matters. *PloS One*, 14(2):e0209749, 2019.

Patricia Funk, Nagore Iriberri, and Giulia Savio. Does scarcity of female instructors create demand for diversity among students? evidence from observational and experimental data. CEPR Discussion Paper 14190, 2019.

Donna K Ginther and Shulamit Kahn. Women in economics: moving up or falling off the academic career ladder? *Journal of Economic Perspectives*, 18(3):193–214, 2004.

Erving Goffman. *Gender Advertisements*. Harper and Row, New York, 1 edition, 1979.

Vivian Gornick. Introduction to 'Gender Advertisements' by Erving Goffman. In *Gender Advertisement*, pages vii–ix. Harper and Row, 1979.

Justin Grimmer and Brandon M Stewart. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3):267–297, 2013.

Shoshana Grossbard, Tansel Yilmazer, and Lingrui Zhang. The gender gap in citations of articles published in two demographic economics journals. *Review of Economics of the Household*, 19(3):677–697, 2021.

Mikki Hebl, Christine Nittrouer, Abigail Corrington, and Juan Madera. How we describe male and female job applicants differently. *Harvard Business Review*, 27, 2018.

Erin Hengel. Publishing while female. *Economic Journal*, forthcoming, 2022.

Laura Hospido and Carlos Sanz. Gender gaps in the evaluation of research: evidence from submissions to economics conferences. *Oxford Bulletin of Economics and Statistics*, 83(3): 590–618, 2021.

Alessandro Iaria, Carlo Schwarz, and Fabian Waldinger. Gender gaps in academia: Global evidence over the twentieth century. CEPR Discussion Paper 17422, 2022.

Shulamit Kahn and Donna Ginther. Women and stem. NBER Working Paper 23525, 2017.

Marlène Koffi. Innovative ideas and gender inequality. CLEF Working Paper Series 35, Canadian Labor Economics Forum, 2021.

Sarah-Jane Leslie, Andrei Cimpian, Meredith Meyer, and Edward Freeland. Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219): 262–265, 2015.

Shelly Lundberg, editor. *Women in Economics*. CEPR, 2020.

Shelly Lundberg and Jenna Stearns. Women in economics: Stalled progress. *Journal of Economic Perspectives*, 33(1):3–22, 2019.

Heather J. MacArthur, Jessica L. Cundiff, and Matthias R. Mehl. Estimating the Prevalence of Gender-Biased Language in Undergraduates' Everyday Speech. *Sex Roles*, 82(1-2):81–93, 2020.

Lillian MacNell, Adam Driscoll, and Andrea N Hunt. What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4):291–303, 2015.

Juan M Madera, Michelle R Hebl, and Randi C Martin. Gender and letters of recommendation for academia: agentic and communal differences. *Journal of Applied Psychology*, 94(6): 1591, 2009.

Juan M Madera, Michelle R Hebl, Heather Dial, Randi Martin, and Virgina Valian. Raising doubt in letters of recommendation for academia: Gender differences and their impact. *Journal of Business and Psychology*, 34(3):287–303, 2019.

Friederike Mengel, Jan Sauermann, and Ulf Zölitz. Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2):535–566, 2019.

Christine L. Nittrouer, Michelle R. Hebl, Leslie Ashburn-Nardo, Rachel C.E. Trump-Steele, David M. Lane, and Virginia Valian. Gender disparities in colloquium speakers at top universities. *Proceedings of the National Academy of Sciences*, (1):104–108, 2018.

Heather Sarsons. Recognition for group work: Gender differences in academia. *American Economic Review*, 107(5):141–45, 2017.

Toni Schmader, Jessica Whitehead, and Vicki H. Wysocki. A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles*, 57(7-8):509–514, 2007.

Yijun Shao, Stephanie Taylor, Nell Marshall, Craig Morioka, and Qing Zeng-Treitler. Clinical Text Classification with Word Embedding Features vs. Bag-of-Words Features. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 2874–2878, 2018.

Frances Trix and Carolyn Psenka. Exploring the color of glass: letters of recommendation for female and male faculty. *Discourse & Society*, 14:191–220, 2003.

Virginia Valian. *Why so slow? The advancement of women*. MIT press, 1999.

Virginia Valian. Beyond Gender Schemas: Improving the Advancement of Women in Academia. *Hypatia*, 20(3):198–213, 2005.

Alice H Wu. Gendered language on the economics job market rumors forum. *American Economic Review Papers & Proceedings*, 108(5):175–79, 2018.

# A Variable and methods description

## Validation exercise

To construct Figure 4 we assess the correspondence between the validators' chosen categories and ours as follows. Within each of the authors' chosen categories, for each word, we identify the category chosen by a plurality of validators. In the case of ties (e.g. "diligent", which the authors classified as "grindstone", was classified by 28.5% of validators as "ability", and 28.5% as "grindstone"), we attribute that word to both categories ("diligent" is attributed both to "ability" and "grindstone"). For each of our chosen categories, Figure 4 presents the distribution of winning categories. Words for which there are two winning categories count twice in the total, so that the sum of the bars is equal to 1.

Table A.1: Summary Statistics of words in each category

| Category | Av. Doc Freq | Av. TF-IDF (x 1000) | N Words | Av. Validators per Word |
|---|---|---|---|---|
| Ability | 8896.56 | 5.77 | 57 | 6.98 |
| Grindstone | 8991.60 | 4.99 | 20 | 6.56 |
| Recruitment | 9011.69 | 4.32 | 118 | 6.72 |
| Research | 9038.12 | 4.77 | 210 | 6.46 |
| Standout | 8958.87 | 5.02 | 106 | 6.70 |
| Teach-Citizen | 8971.88 | 5.06 | 94 | 6.81 |

**Notes:** This table shows summary statistics of words in each category. The First column gives the categories. The second (third) columns give the average TF-IDF (document frequency) of words in each category. The fourth column gives the number of words in each category. The fifth column gives the average number of validators who cross-validated our categorisation for each word.

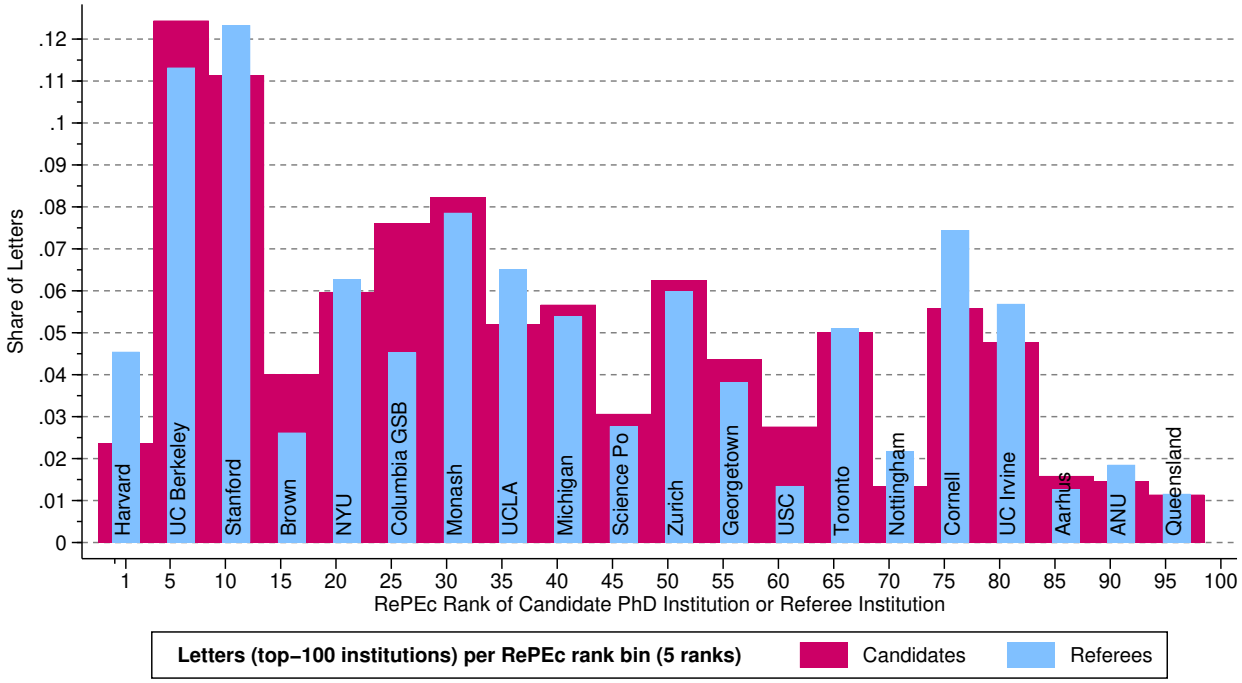
## Institutional Ranking

We used the Research Papers in Economics (RePEc) ranking for the top 5% of economic institutions as our guide to rank writer and candidate institutions.[40] We drop three research organisations (NBER, IZA, CEPR) but keep international institutions like the IMF as well as the Federal Reserve Banks in the rankings since referees from these institutions are not uncommon. Writer institutional affiliation is collected from their CV via manual internet search and manually matched to the RePEc institutions. We categorise writers into bands on the basis of their institutional ranking: 1-25, 26-50, 51-100, 101-200, 201-500, and higher (omitted category in our regressions). We are missing RePEc-listed affiliation and hence rankings for around 16% of writers, but these only account for 12% of our sample of reference letters. The rank of candidate PhD-institutions has been similarly constructed.

[40] Version January 2021, see https://ideas.repec.org/top/top.inst.all.html for the current version. The RePEc ranking refers to the top 10% but only the top 5% are ranked, the remainder are unranked within the percentile (all those within the 6th percentile, all those within the 7th percentile, etc).

# B Additional Descriptive Statistics

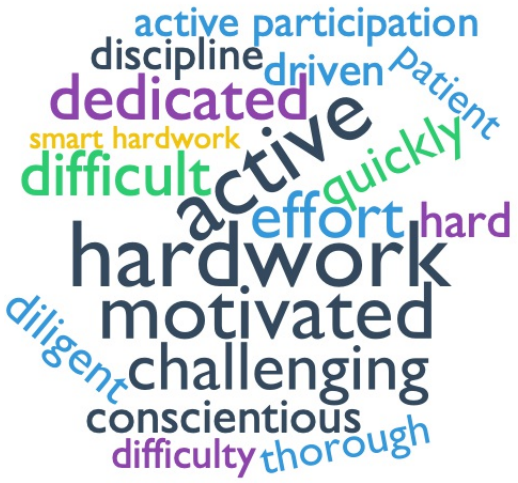
Figure B.1: RePEc Rank of Candidate and Letter Writer Institution, Zooming into Top-100 institutions




**Notes**: The figure presents the frequency distribution of candidate and letter writer institution rank, zooming in on the top-100 (bin width 5 ranks), highlighting one institutions for each bin.

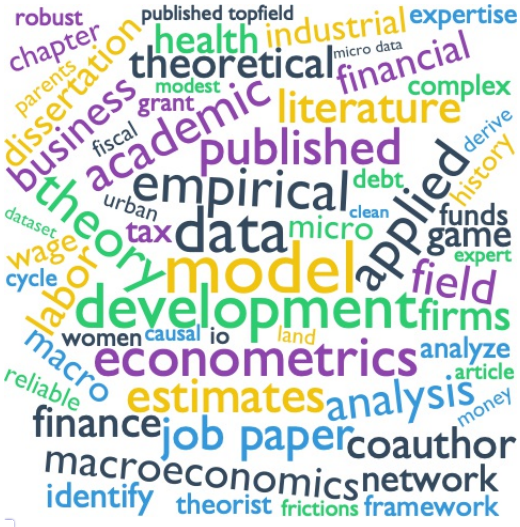
Figure B.2: Word clouds for each sentiment


(a) Ability



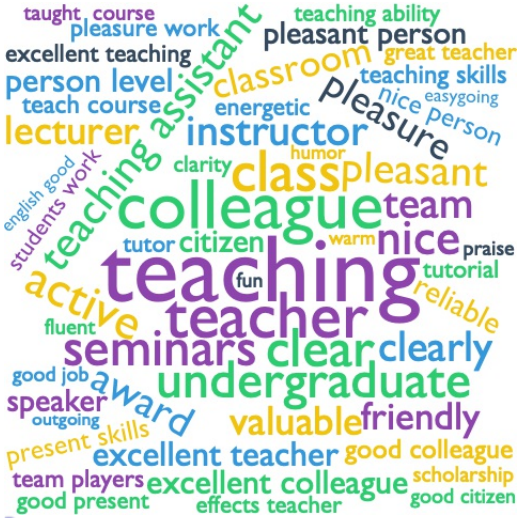

(b) Grindstone


(c) Recruitment




(d) Research


(e) Standout




(f) Teach and Citizen

**Notes:** The word clouds depict the expressions attributed to each sentiment. The size of the word is illustrative of its document frequency. Within each cloud, larger words are more common in the corpus. The size of the words should not be compared across wordclouds, as the font sizes are adjusted to improve legibility.

# C Results Tables

## Baseline results

Table C.1: Baseline Results by Writer Institutional Rank

| Writer RePEc Rank | (1) All | (2) Top-25 | (3) Top-100 |
|---|---|---|---|
| Ability | -0.0212 | -0.0370 | -0.0326 |
| | (1.00) | (0.75) | (1.02) |
| Grindstone | 0.0532 | 0.0836 | 0.0509 |
| | (2.52)** | (1.72)* | (1.62) |
| Recruitment | -0.0240 | -0.0425 | -0.0269 |
| | (1.16) | (0.87) | (0.86) |
| Research | -0.0456 | -0.0376 | -0.0503 |
| | (2.22)** | (0.77) | (1.64) |
| Standout | -0.0078 | -0.0547 | 0.0134 |
| | (0.37) | (1.11) | (0.43) |
| Teaching & Citizenship | 0.0070 | 0.0071 | 0.0260 |
| | (0.34) | (0.14) | (0.83) |
| FE/Variables absorbed | 25 | 20 | 23 |
| Additional covariates | 7 | 7 | 7 |
| Number of Letters | 11846 | 2224 | 5344 |
| dto for females | 3360 | 616 | 1508 |
| Number of candidates | 3721 | 1111 | 2301 |
| dto female | 1082 | 318 | 664 |
| Number of writers | 5655 | 969 | 2285 |
| dto female | 985 | 156 | 382 |
| Letters by fem writers | 1751 | 314 | 735 |
| Year FE | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes |
| Institution Rank FE | yes | yes | yes |
| Years since PhD | yes | yes | yes |
| Research Field FE | yes | yes | yes |
| Publications | yes | yes | yes |
| Writer characteristics | yes | yes | yes |
| Letter length | yes | yes | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator in the full sample, column 1, and for writer institutions in the top-25 and top-100 in columns 2 and 3, respectively. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.
Return to Figure 6 in the maintext.

## Male and Female Writers

Table C.2: Male Writers

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Ability | -0.0241 | -0.0228 | -0.0242 | -0.0265 | -0.0263 | -0.0261 | -0.0253 |
| | (1.04) | (0.98) | (1.05) | (1.14) | (1.13) | (1.12) | (1.09) |
| Grindstone | 0.0504 | 0.0477 | 0.0500 | 0.0475 | 0.0463 | 0.0476 | 0.0475 |
| | (2.20)** | (2.08)** | (2.18)** | (2.06)** | (2.00)** | (2.06)** | (2.05)** |
| Recruitment | -0.0235 | -0.0229 | -0.0220 | -0.0325 | -0.0293 | -0.0300 | -0.0259 |
| | (1.00) | (0.98) | (0.94) | (1.40) | (1.25) | (1.29) | (1.14) |
| Research | -0.0732 | -0.0718 | -0.0715 | -0.0696 | -0.0693 | -0.0696 | -0.0721 |
| | (3.25)*** | (3.19)*** | (3.17)*** | (3.09)*** | (3.07)*** | (3.08)*** | (3.21)*** |
| Standout | -0.0089 | -0.0060 | -0.0084 | -0.0156 | -0.0127 | -0.0142 | -0.0109 |
| | (0.38) | (0.26) | (0.36) | (0.67) | (0.54) | (0.61) | (0.47) |
| Teaching & | 0.0209 | 0.0118 | 0.0108 | 0.0055 | 0.0029 | 0.0059 | 0.0075 |
| Citizenship | (0.89) | (0.51) | (0.47) | (0.23) | (0.13) | (0.26) | (0.32) |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | 24 | 24 |
| Additional covariates | | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 10095 | 10095 | 10095 | 10095 | 10095 | 10095 | 10095 |
| dto for females | 2729 | 2729 | 2729 | 2729 | 2729 | 2729 | 2729 |
| Number of candidates | 3683 | 3683 | 3683 | 3683 | 3683 | 3683 | 3683 |
| dto female | 1057 | 1057 | 1057 | 1057 | 1057 | 1057 | 1057 |
| Number of writers | 4670 | 4670 | 4670 | 4670 | 4670 | 4670 | 4670 |
| dto female | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Letters by fem writers | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Year FE | yes | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes | yes |
| Publications | no | no | no | no | yes | yes | yes |
| Writer characteristics | no | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | no | yes |

**Notes:** The sample is restricted to male letter writers. The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute t-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Return to Figure 7 in the maintext.

Table C.3: Female Writers

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Ability | 0.0022 | 0.0056 | 0.0017 | -0.0013 | -0.0019 | -0.0057 | -0.0050 |
| | (0.04) | (0.11) | (0.03) | (0.02) | (0.04) | (0.11) | (0.10) |
| Grindstone | 0.0804 | 0.0793 | 0.0807 | 0.0770 | 0.0800 | 0.0727 | 0.0724 |
| | (1.52) | (1.52) | (1.54) | (1.47) | (1.52) | (1.39) | (1.39) |
| Recruitment | 0.0082 | 0.0012 | 0.0028 | -0.0155 | -0.0140 | -0.0198 | -0.0178 |
| | (0.16) | (0.02) | (0.06) | (0.31) | (0.28) | (0.39) | (0.35) |
| Research | 0.0689 | 0.0787 | 0.0757 | 0.0862 | 0.0839 | 0.0828 | 0.0825 |
| | (1.33) | (1.53) | (1.47) | (1.66)* | (1.62) | (1.60) | (1.60) |
| Standout | 0.0242 | 0.0202 | 0.0193 | 0.0150 | 0.0175 | 0.0177 | 0.0196 |
| | (0.49) | (0.41) | (0.39) | (0.30) | (0.36) | (0.36) | (0.41) |
| Teaching & | 0.0211 | 0.0140 | 0.0134 | 0.0065 | 0.0081 | 0.0183 | 0.0185 |
| Citizenship | (0.40) | (0.27) | (0.26) | (0.12) | (0.16) | (0.36) | (0.36) |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | 24 | 24 |
| Additional covariates | | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 1751 | 1751 | 1751 | 1751 | 1751 | 1751 | 1751 |
| dto for females | 631 | 631 | 631 | 631 | 631 | 631 | 631 |
| Number of candidates | 1414 | 1414 | 1414 | 1414 | 1414 | 1414 | 1414 |
| dto female | 482 | 482 | 482 | 482 | 482 | 482 | 482 |
| Number of writers | 985 | 985 | 985 | 985 | 985 | 985 | 985 |
| dto female | 985 | 985 | 985 | 985 | 985 | 985 | 985 |
| Letters by fem writers | 1751 | 1751 | 1751 | 1751 | 1751 | 1751 | 1751 |
| Year FE | yes | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes | yes |
| Publications | no | no | no | no | yes | yes | yes |
| Writer characteristics | no | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | no | yes |

**Notes:** The sample is restricted to female letter writers. The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute t-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Return to Figure 7 in the maintext.

## Cultural Background

Table C.4: Cultural Background

| 'Traditional Norms' | (1) | (2) Pre-Schooler | (3) Uni Boys | (4) Male Execs | (5) Average |
|---|---|---|---|---|---|
| Ability | -0.0197 | -0.0095 | 0.0009 | -0.0113 | -0.0124 |
| | (0.88) | (0.32) | (0.03) | (0.42) | (0.43) |
| Interaction: | | -0.0120 | -0.0335 | -0.0163 | -0.0110 |
| Traditional Norms | | (0.27) | (0.73) | (0.34) | (0.24) |
| Grindstone | 0.0573 | 0.0634 | 0.0364 | 0.0479 | 0.0482 |
| | (2.56)** | (2.10)** | (1.06) | (1.80)* | (1.66)* |
| Interaction: | | -0.0062 | 0.0359 | 0.0358 | 0.0240 |
| Traditional Norms | | (0.14) | (0.79) | (0.72) | (0.53) |
| Recruitment | -0.0333 | -0.0448 | -0.0407 | -0.0431 | -0.0476 |
| | (1.52) | (1.49) | (1.14) | (1.65)* | (1.64) |
| Interaction: | | 0.0303 | 0.0082 | 0.0344 | 0.0373 |
| Traditional Norms | | (0.69) | (0.18) | (0.72) | (0.84) |
| Research | -0.0377 | -0.0015 | -0.0499 | -0.0256 | -0.0080 |
| | (1.74)* | (0.05) | (1.49) | (0.99) | (0.28) |
| Interaction: | | -0.0739 | 0.0175 | -0.0328 | -0.0643 |
| Traditional Norms | | (1.71)* | (0.40) | (0.70) | (1.47) |
| Standout | 0.0129 | 0.0029 | 0.0261 | 0.0029 | -0.0016 |
| | (0.59) | (0.10) | (0.73) | (0.11) | (0.06) |
| Interaction: | | 0.0230 | -0.0198 | 0.0317 | 0.0342 |
| Traditional Norms | | (0.52) | (0.44) | (0.67) | (0.76) |
| Teaching & Citizenship | 0.0118 | 0.0075 | -0.0271 | 0.0020 | 0.0027 |
| | (0.53) | (0.25) | (0.81) | (0.07) | (0.09) |
| Interaction: | | 0.0091 | 0.0681 | 0.0245 | 0.0208 |
| Traditional Norms | | (0.20) | (1.52) | (0.50) | (0.46) |
| FE/Variables absorbed | 25 | 25+ | 25+ | 25+ | 25+ |
| Additional covariates | 6 | 6 | 6 | 6 | 6 |
| Number of Letters | 10542 | 10542 | 10542 | 10542 | 10542 |
| dto for females | 3004 | 3004 | 3004 | 3004 | 3004 |
| Number of candidates | 3674 | 3674 | 3674 | 3674 | 3674 |
| dto female | 1066 | 1066 | 1066 | 1066 | 1066 |
| Number of writers | 4907 | 4907 | 4907 | 4907 | 4907 |
| dto female | 828 | 828 | 828 | 828 | 828 |
| Letters by fem writers | 1506 | 1506 | 1506 | 1506 | 1506 |
| Year FE | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes |
| Institution Rank FE | yes | yes | yes | yes | yes |
| Years since PhD | yes | yes | yes | yes | yes |
| Research Field FE | yes | yes | yes | yes | yes |
| Publications | yes | yes | yes | yes | yes |
| Writer characteristics | yes | yes | yes | yes | yes |
| Letter length | yes | yes | yes | yes | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator interacted with a binary variable indicating whether the writer's country of birth has traditional gender norms and the full set of controls. Column (5) reports estimates when the average of the 3 WVS questions is used to construct the 'gender norms' indicator, whereas columns (2)-(4) use each question separately. The questions are stated fully in the main text. Column (1) reports the benchmark result for this reduced sample (birth or UG countries could be identified for 87% of referees). The table reports the estimate for the female indicator and the interaction term. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute t-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Return to Figure 8 in the main text.

## Specifications with Fixed Effects

Table C.5: Candidate Institution FE

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Ability | -0.0147 | -0.0183 | -0.0233 | -0.0228 | -0.0280 | -0.0265 |
| | (0.66) | (0.82) | (1.04) | (1.02) | (1.25) | (1.18) |
| Grindstone | 0.0526 | 0.0537 | 0.0494 | 0.0494 | 0.0443 | 0.0443 |
| | (2.35)** | (2.39)** | (2.19)** | (2.19)** | (1.96)* | (1.96)* |
| Recruitment | -0.0213 | -0.0214 | -0.0321 | -0.0308 | -0.0284 | -0.0237 |
| | (0.98) | (0.98) | (1.46) | (1.40) | (1.29) | (1.10) |
| Research | -0.0330 | -0.0335 | -0.0316 | -0.0304 | -0.0276 | -0.0299 |
| | (1.50) | (1.52) | (1.43) | (1.37) | (1.24) | (1.35) |
| Standout | 0.0015 | -0.0007 | -0.0098 | -0.0080 | -0.0103 | -0.0064 |
| | (0.07) | (0.03) | (0.44) | (0.36) | (0.47) | (0.29) |
| Teaching & | 0.0065 | 0.0074 | 0.0044 | 0.0027 | -0.0044 | -0.0028 |
| Citizenship | (0.30) | (0.34) | (0.20) | (0.13) | (0.20) | (0.13) |
| FE/Variables absorbed | 237 | 237 | 241 | 241 | 247 | 247 |
| Additional covariates | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 10604 | 10604 | 10604 | 10604 | 10604 | 10604 |
| dto for females | 3158 | 3158 | 3158 | 3158 | 3158 | 3158 |
| Number of candidates | 3309 | 3309 | 3309 | 3309 | 3309 | 3309 |
| dto female | 1010 | 1010 | 1010 | 1010 | 1010 | 1010 |
| Number of writers | 4918 | 4918 | 4918 | 4918 | 4918 | 4918 |
| dto female | 853 | 853 | 853 | 853 | 853 | 853 |
| Letters by fem writers | 1553 | 1553 | 1553 | 1553 | 1553 | 1553 |
| Year FE | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes |
| Institution FE | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | yes | yes | yes | yes | yes |
| Research Field FE | no | no | yes | yes | yes | yes |
| Publications | no | no | no | yes | yes | yes |
| Writer Characteristics | yes | yes | yes | no | yes | yes |
| Letter length | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. These specifications include FE for the candidate's institution. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.
Return to Figure 9 in the maintext.

Table C.6: Writer FE

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Ability | -0.0391 | -0.0367 | -0.0381 | -0.0375 | -0.0381 | -0.0358 |
| | (1.39) | (1.30) | (1.35) | (1.33) | (1.35) | (1.27) |
| Grindstone | 0.0209 | 0.0191 | 0.0179 | 0.0174 | 0.0181 | 0.0195 |
| | (0.73) | (0.66) | (0.62) | (0.60) | (0.63) | (0.67) |
| Recruitment | -0.0056 | -0.0049 | -0.0044 | -0.0057 | -0.0029 | 0.0003 |
| | (0.22) | (0.19) | (0.18) | (0.23) | (0.12) | (0.01) |
| Research | -0.0273 | -0.0248 | -0.0236 | -0.0226 | -0.0231 | -0.0265 |
| | (1.00) | (0.91) | (0.86) | (0.82) | (0.84) | (0.97) |
| Standout | -0.0181 | -0.0187 | -0.0208 | -0.0206 | -0.0173 | -0.0114 |
| | (0.65) | (0.67) | (0.74) | (0.73) | (0.61) | (0.41) |
| Teaching & | -0.0244 | -0.0254 | -0.0279 | -0.0280 | -0.0309 | -0.0286 |
| Citizenship | (0.93) | (0.97) | (1.06) | (1.06) | (1.17) | (1.09) |
| FE/Variables absorbed | 1319 | 1324 | 1324 | 1328 | 1328 | 1328 |
| Additional covariates | | | 1 | 1 | 5 | 6 |
| Number of Letters | 5226 | 5226 | 5226 | 5226 | 5226 | 5226 |
| dto for females | 1997 | 1997 | 1997 | 1997 | 1997 | 1997 |
| Number of candidates | 2774 | 2774 | 2774 | 2774 | 2774 | 2774 |
| dto female | 924 | 924 | 924 | 924 | 924 | 924 |
| Number of writers | 1314 | 1314 | 1314 | 1314 | 1314 | 1314 |
| dto female | 197 | 197 | 197 | 197 | 197 | 197 |
| Letters by fem writers | 699 | 699 | 699 | 699 | 699 | 699 |
| Writer FE | yes | yes | yes | yes | yes | yes |
| Year FE | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes |
| Publications | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. These specifications include letterwriter fixed effects. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.
The sample includes only those letters from writers with two or more references for at least one male and one female candidate (gender mix). Return to Figure 9 in the maintext.

Table C.7: Writer FE, writers 'more familiar' with female candidates

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Ability | 0.0057 | 0.0085 | 0.0075 | 0.0083 | 0.0059 | 0.0074 |
| | (0.15) | (0.22) | (0.19) | (0.21) | (0.15) | (0.19) |
| Grindstone | -0.0278 | -0.0313 | -0.0311 | -0.0306 | -0.0277 | -0.0274 |
| | (0.70) | (0.79) | (0.79) | (0.77) | (0.69) | (0.69) |
| Recruitment | -0.0134 | -0.0118 | -0.0121 | -0.0130 | -0.0056 | -0.0034 |
| | (0.38) | (0.33) | (0.34) | (0.37) | (0.16) | (0.10) |
| Research | -0.0408 | -0.0344 | -0.0328 | -0.0325 | -0.0337 | -0.0357 |
| | (1.09) | (0.92) | (0.87) | (0.86) | (0.89) | (0.95) |
| Standout | -0.0558 | -0.0548 | -0.0564 | -0.0558 | -0.0524 | -0.0485 |
| | (1.42) | (1.40) | (1.44) | (1.42) | (1.32) | (1.24) |
| Teaching & | -0.0488 | -0.0497 | -0.0513 | -0.0513 | -0.0585 | -0.0568 |
| Citizenship | (1.35) | (1.38) | (1.42) | (1.41) | (1.61) | (1.57) |
| FE/Variables absorbed | 754 | 759 | 759 | 763 | 763 | 763 |
| Additional covariates | | | 1 | 1 | 5 | 6 |
| Number of Letters | 2512 | 2512 | 2512 | 2512 | 2512 | 2512 |
| dto for females | 1334 | 1334 | 1334 | 1334 | 1334 | 1334 |
| Number of candidates | 1682 | 1682 | 1682 | 1682 | 1682 | 1682 |
| dto female | 793 | 793 | 793 | 793 | 793 | 793 |
| Number of writers | 749 | 749 | 749 | 749 | 749 | 749 |
| dto female | 129 | 129 | 129 | 129 | 129 | 129 |
| Letters by fem writers | 408 | 408 | 408 | 408 | 408 | 408 |
| Writer FE | yes | yes | yes | yes | yes | yes |
| Year FE | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes |
| Publications | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. These specifications include letterwriter FE. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Sample includes only those letters from writers with two or more references for at least one male and one female candidate (gender mix), and who have had more than 1/3 of female Ph.D. students. Return to Figure 9 in the main text.

Table C.8: Writer FE, writers 'less familiar' with female candidates

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Ability | -0.0961 | -0.0953 | -0.0973 | -0.0973 | -0.0961 | -0.0931 |
| | (2.35)** | (2.33)** | (2.38)** | (2.37)** | (2.34)** | (2.27)** |
| Grindstone | 0.0840 | 0.0842 | 0.0817 | 0.0815 | 0.0843 | 0.0875 |
| | (1.98)** | (1.98)** | (1.92)* | (1.90)* | (1.97)** | (2.05)** |
| Recruitment | 0.0033 | 0.0009 | 0.0024 | 0.0006 | 0.0005 | 0.0051 |
| | (0.09) | (0.02) | (0.06) | (0.02) | (0.01) | (0.14) |
| Research | -0.0135 | -0.0129 | -0.0121 | -0.0099 | -0.0122 | -0.0174 |
| | (0.33) | (0.32) | (0.30) | (0.24) | (0.30) | (0.43) |
| Standout | 0.0342 | 0.0331 | 0.0303 | 0.0282 | 0.0298 | 0.0379 |
| | (0.84) | (0.81) | (0.74) | (0.69) | (0.73) | (0.94) |
| Teaching & | 0.0014 | -0.0015 | -0.0046 | -0.0038 | -0.0025 | 0.0001 |
| Citizenship | (0.03) | (0.04) | (0.12) | (0.10) | (0.06) | (0.00) |
| FE/Variables absorbed | 570 | 575 | 575 | 579 | 579 | 579 |
| Additional covariates | | | 1 | 1 | 5 | 6 |
| Number of Letters | 2714 | 2714 | 2714 | 2714 | 2714 | 2714 |
| dto for females | 663 | 663 | 663 | 663 | 663 | 663 |
| Number of candidates | 1905 | 1905 | 1905 | 1905 | 1905 | 1905 |
| dto female | 478 | 478 | 478 | 478 | 478 | 478 |
| Number of writers | 565 | 565 | 565 | 565 | 565 | 565 |
| dto female | 68 | 68 | 68 | 68 | 68 | 68 |
| Letters by fem writers | 291 | 291 | 291 | 291 | 291 | 291 |
| Writer FE | yes | yes | yes | yes | yes | yes |
| Year FE | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes |
| Publications | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. These specifications include letterwriter FE. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Sample includes only those letters from writers with two or more references for at least one male and one female candidate (gender mix), and who have had less than 1/3 of female Ph.D. students. Return to Figure 9 in the main text.

Table C.9: Candidate Fixed Effects

| | (1) | (2) | (3) |
|---|---|---|---|
| Ability: Female writer | 0.0789 | 0.0790 | 0.0814 |
| | (1.48) | (1.46) | (1.51) |
| Female writer × | 0.0265 | 0.0279 | 0.0272 |
| Female candidate | (0.29) | (0.31) | (0.30) |
| Grindstone: Female writer | 0.1724 | 0.1513 | 0.1535 |
| | (3.32)*** | (2.88)*** | (2.92)*** |
| Female writer × | -0.0708 | -0.0788 | -0.0794 |
| Female candidate | (0.76) | (0.84) | (0.85) |
| Recruitment: Female writer | -0.1167 | -0.1130 | -0.1025 |
| | (2.28)** | (2.19)** | (2.00)** |
| Female writer × | -0.0145 | -0.0259 | -0.0288 |
| Female candidate | (0.16) | (0.29) | (0.33) |
| Research: Female writer | -0.0594 | -0.0524 | -0.0542 |
| | (1.18) | (1.03) | (1.07) |
| Female writer × | 0.0536 | 0.0563 | 0.0567 |
| Female candidate | (0.62) | (0.65) | (0.66) |
| Standout: Female writer | -0.0851 | -0.0793 | -0.0731 |
| | (1.65)* | (1.52) | (1.41) |
| Female writer × | 0.1089 | 0.1078 | 0.1061 |
| Female candidate | (1.28) | (1.27) | (1.26) |
| T&C: Female Writer | 0.1081 | 0.1074 | 0.1104 |
| | (2.26)** | (2.22)** | (2.29)** |
| Female writer × | 0.0667 | 0.0673 | 0.0665 |
| Female candidate | (0.78) | (0.78) | (0.77) |
| FE/Variables absorbed | 822 | 827 | 827 |
| Additional covariates | | 1 | 2 |
| Number of Letters | 2335 | 2335 | 2335 |
| dto for females | 778 | 778 | 778 |
| Number of candidates | 822 | 822 | 822 |
| dto female | 274 | 274 | 274 |
| Number of writers | 1204 | 1204 | 1204 |
| dto female | 348 | 348 | 348 |
| Letters by fem writers | 930 | 930 | 930 |
| Year FE | no | no | no |
| Candidate FE | yes | yes | yes |
| Writer Characteristics | no | yes | yes |
| Letter length | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. These specifications include candidate FE interacted with the gender of the letterwriter. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute t-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Return to Figure 7 in the maintext.

# D Candidate Research Fields

In this section, we describe the procedure to establish candidates' fields using an unsupervised approach.

From the recommendation letters we extract the text slice that is most likely to discuss the candidates' job market paper. To do so we flag the first instance of the term 'job market paper' or 'dissertation'. We then slice the subsequent 400 words and assemble the research slices from all the recommendation letters written for the same candidate into a single text.[41] We process these texts as described in section 3.1 and cluster them into four groups using an unsupervised k-means clustering approach.

Given that the objective of this procedure is to group texts that use similar terms, we deploy a different approach when transforming the text into a database. Instead of computing the tfidf, which would give more weight to terms that are more frequently used in a document compared to the rest of the corpus, we just use a binary representation in which a term is given a value equal to one if it appears in the text. This approach allows us to more easily identify the research texts that contain broad terms that could characterise a field (e.g. 'macro', 'Nash equilibrium', 'causality'), rather than singling out terms used multiple times to describe the job market paper, but that could be very specific to a particular piece of research (e.g. 'assortative matching', 'babbling equilibiria'). Finally, following common recommendations for k-means clustering, we reduce the dimensionality of the problem by carrying out a PCA.

Figure D.1 shows the SSE of the k-means clustering procedure as a function of the number of clusters chosen. We identify a kink at four clusters, hence, using the 'elbow method', that is the final number of clusters we select in our analysis.

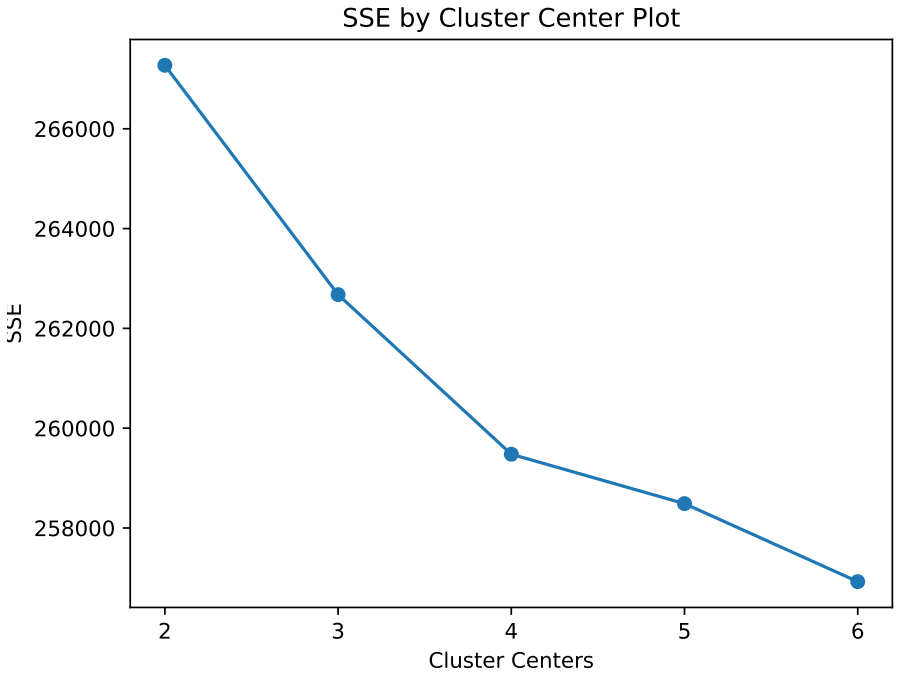
We validate these groupings by highlighting the mapping between them and the self-reported, unstructured primary research field that candidates add to their CV.[42] The word clouds in Figure D.2 show the frequency of the reported main fields for each of the candidates in each broad category. Three clearly identified broad groups emerge: macro, applied, and theory. 47% of candidates report 'Macro' as their main field in panel (a). Similarly, applicants listing 'Labor', 'Development', 'Public' or 'Applied Micro' make up 45% of those in panel (b); and those indicating 'Micro Theory', 'Industrial Organization', 'Econometrics', 'Behavioral', 'Applied Theory', 'Game Theory' or 'Economic Theory' represent 44% of the individuals in panel (c). The clustering procedure also creates a fourth category which we cannot credibly assign to a specific broad area and which as a result has been treated as residual.[43]

[41] 84% are sliced based on the word 'job market paper' and 16% on 'dissertation'.

[42] 222 distinct fields are reported. While these fields do not necessarily map precisely into an existing JEL code, they are typically highly informative when it comes to the actual content of research pursued by the candidates. Moreover, not all candidates report a main field of specialization.

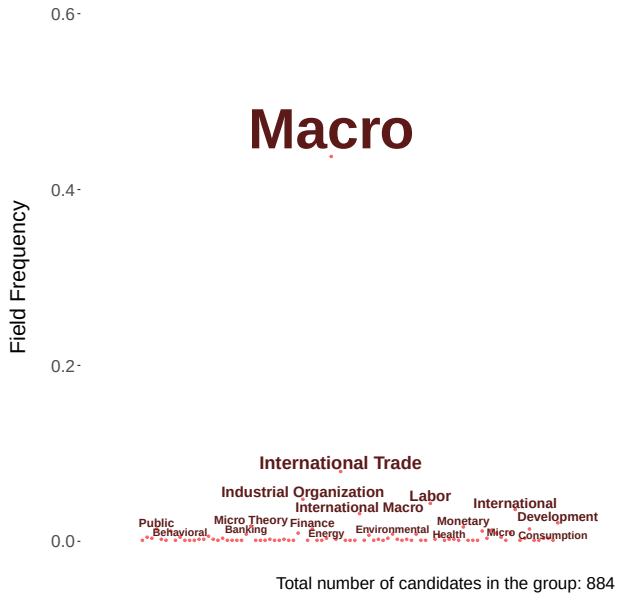
[43] We experiment with alternative definitions of research fields as controls in the baseline regressions in Section 5.3.

Figure D.1: SSE per cluster number

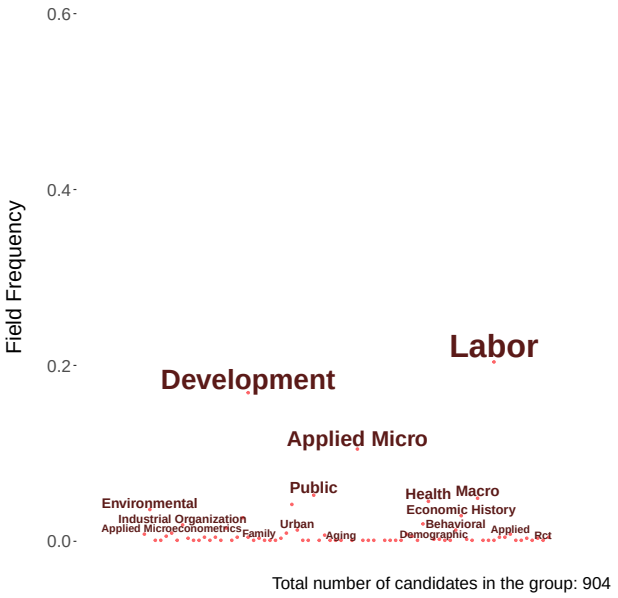



**Notes:** This figure presents the SSE of the k-means clustering procedure as a function of the number of clusters used to group candidates into research fields.

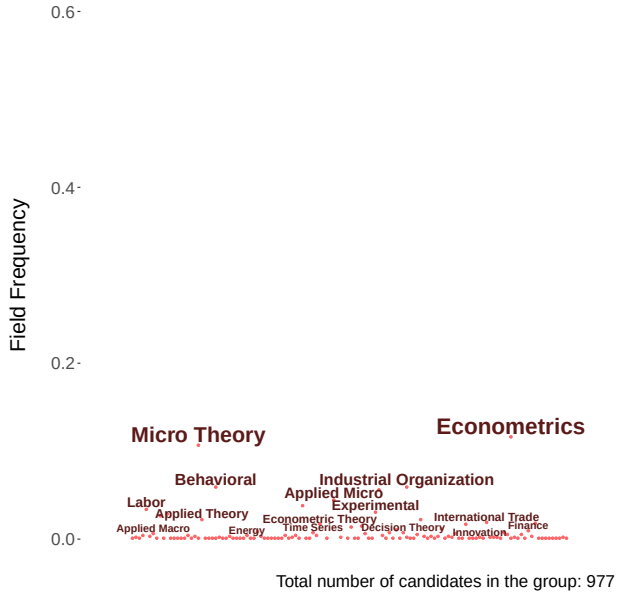
Figure D.2: Word clouds for Fields



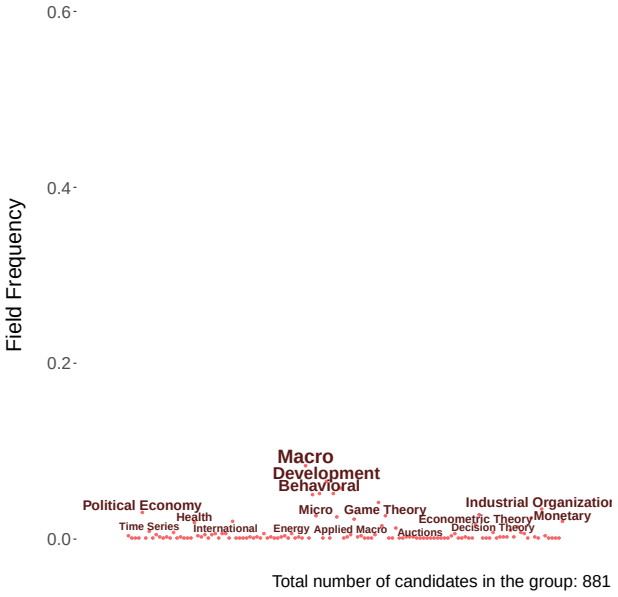

(a) Macro




(b) Applied



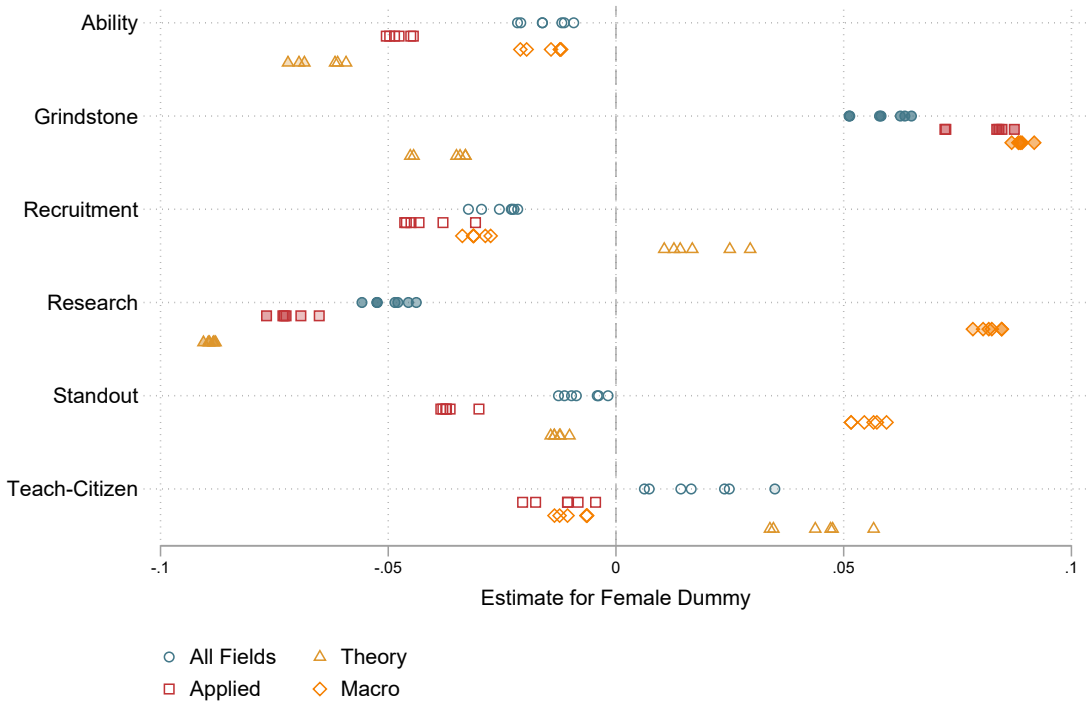

(c) Theory




(d) Residual

**Notes:** The word clouds depict the research fields freely written by candidates for each of the categories. For each category, the $y$-axis and the font size of the fields reflects its frequency as a primary field in the CVs of candidates that reported them. The fields are randomly distributed across the $x$-axis.

Figure D.3: Regression results, different candidate research fields




**Notes:** This figure shows the coefficient estimates for the regressions specified in 6, estimated separately for different (aggregated) research field clusters. We show the three most demanding specifications. The symbol's filling permit visualizing significance. The symbol's filling permits visualizing significance. Using 4 levels of possible standard error clustering (none, candidate's institution,letter-writer's institution, or letter writer), we flag significance at 3 different levels (10%, 5%, and 1%). We thus flag 12 possible significance indicators. Then, for each level of clustering, the symbol in the graph is shadowed with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error clustering. Additional information on the sample and results for the clustered standard errors by letter-writer are contained in Appendix Section D.

Table D.1: By Candidate Research Field

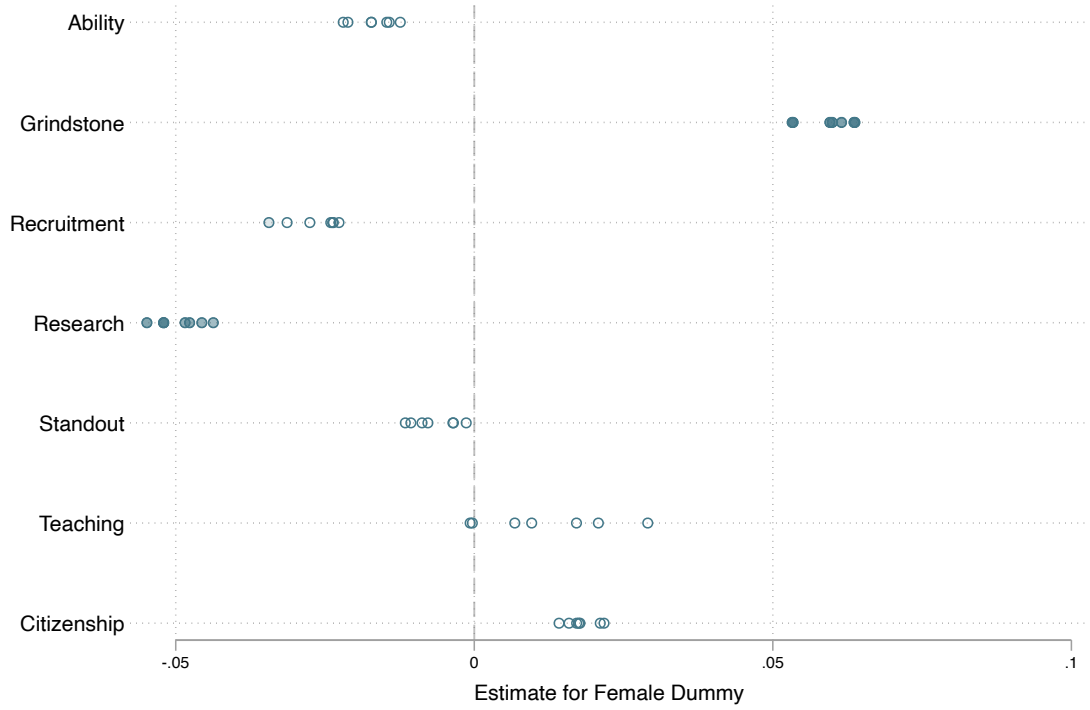
| Broad Research Fields | (1) All Fields | (2) Macro | (3) Theory | (4) Applied Micro |
|---|---|---|---|---|
| Ability | -0.0209 | -0.0210 | -0.0684 | -0.0497 |
| | (0.98) | (0.46) | (1.65)* | (1.31) |
| Grindstone | 0.0512 | 0.0888 | -0.0452 | 0.0722 |
| | (2.41)** | (1.91)* | (1.11) | (1.84)* |
| Recruitment | -0.0425 | -0.0946 | -0.0269 | -0.0253 |
| | (0.87) | (2.43)** | (0.86) | (0.63) |
| Research | -0.0376 | -0.0301 | -0.0503 | -0.0633 |
| | (0.77) | (0.80) | (1.64) | (1.60) |
| Standout | -0.0547 | -0.0299 | 0.0134 | 0.0588 |
| | (1.11) | (0.77) | (0.43) | (1.49) |
| Teaching & Citizenship | 0.0071 | 0.0245 | 0.0260 | 0.0441 |
| | (0.14) | (0.63) | (0.83) | (1.09) |
| FE/Variables absorbed | 24 | 21 | 21 | 21 |
| Additional covariates | 7 | 7 | 7 | 7 |
| Number of Letters | 11638 | 2832 | 3188 | 3007 |
| dto for females | 3328 | 699 | 838 | 1089 |
| Number of candidates | 3645 | 884 | 976 | 904 |
| dto female | 1068 | 219 | 267 | 335 |
| Number of writers | 5523 | 1492 | 2135 | 1905 |
| dto female | 965 | 215 | 284 | 412 |
| Letters by fem writers | 1723 | 360 | 386 | 633 |
| Year FE | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes |
| Institution Rank FE | yes | yes | yes | yes |
| Years since PhD | yes | yes | yes | yes |
| Research Field FE | yes | n/a | n/a | n/a |
| Publications | yes | yes | yes | yes |
| Writer characteristics | yes | yes | yes | yes |
| Letter length | yes | yes | yes | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. The regressions are run separately for candidates in each research field. See Section D for details on how the fields are constructed.
Return to Section 5.3 in the maintext.

# E Robustness checks

## Splitting the Teaching and Citizenship 'Sentiment'

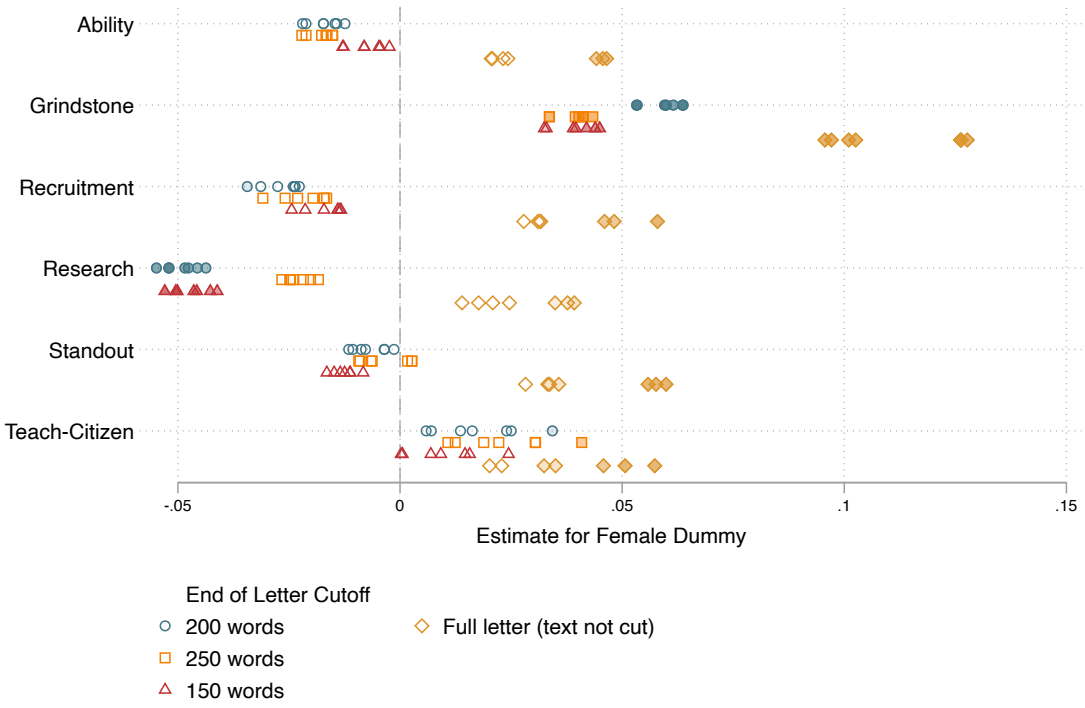
Figure E.1: Regression results, separating teaching and citizenship




**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Return to Section 5.3 in the maintext.

## Different letter end lengths

Figure E.2: Regression results, different end of letter lengths and full letter




**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Regressions are estimated separately for the full letter and samples where the end letter is defined using 150, 200, or 250 words. Return to Section 5.3 in the maintext.

Table E.1: Different end of letter lengths and full letter

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | 150 words | | 200 words | | 250 words | |
| Ability | -0.0048 | -0.0127 | -0.0147 | -0.0212 | -0.0164 | -0.0212 |
| | (0.23) | (0.60) | (0.70) | (1.00) | (0.79) | (1.00) |
| Grindstone | 0.0449 | 0.0325 | 0.0637 | 0.0532 | 0.0434 | 0.0336 |
| | (2.10)** | (1.52) | (3.02)*** | (2.52)** | (2.08)** | (1.61) |
| Recruitment | -0.0140 | -0.0140 | -0.0236 | -0.0240 | -0.0165 | -0.0196 |
| | (0.65) | (0.67) | (1.11) | (1.16) | (0.77) | (0.94) |
| Research | -0.0530 | -0.0427 | -0.0548 | -0.0456 | -0.0266 | -0.0201 |
| | (2.56)** | (2.05)** | (2.66)*** | (2.22)** | (1.28) | (0.97) |
| Standout | -0.0112 | -0.0125 | -0.0035 | -0.0078 | 0.0017 | -0.0063 |
| | (0.53) | (0.59) | (0.17) | (0.37) | (0.08) | (0.30) |
| Teaching & | 0.0244 | 0.0005 | 0.0343 | 0.0070 | 0.0409 | 0.0124 |
| Citizenship | (1.15) | (0.03) | (1.60) | (0.34) | (1.90)* | (0.59) |
| FE/Variables absorbed | 10 | 25 | 10 | 25 | 10 | 25 |
| Additional covariates | 1 | 7 | 1 | 7 | 1 | 7 |
| Number of Letters | 11814 | 11814 | 11846 | 11846 | 11794 | 11794 |
| dto for females | 3355 | 3355 | 3360 | 3360 | 3344 | 3344 |
| Number of candidates | 3722 | 3722 | 3721 | 3721 | 3718 | 3718 |
| dto female | 1082 | 1082 | 1082 | 1082 | 1079 | 1079 |
| Number of writers | 5652 | 5652 | 5655 | 5655 | 5617 | 5617 |
| dto female | 981 | 981 | 985 | 985 | 981 | 981 |
| Letters by fem writers | 1746 | 1746 | 1751 | 1751 | 1745 | 1745 |
| Year FE | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | no | yes | no | yes |
| Years since PhD | no | yes | no | yes | no | yes |
| Research Field FE | no | yes | no | yes | no | yes |
| Publications | no | yes | no | yes | no | yes |
| Writer characteristics | no | yes | no | yes | no | yes |
| Letter length | no | yes | no | yes | no | yes |

**Notes:** This table presents results for the analysis of three different letter end cut-offs: 150 words, 200 words or 250 words. For each category, we present the most parsimonious and the most elaborate regression model. Return to Section 5.3 in the maintext.

## Full Letter

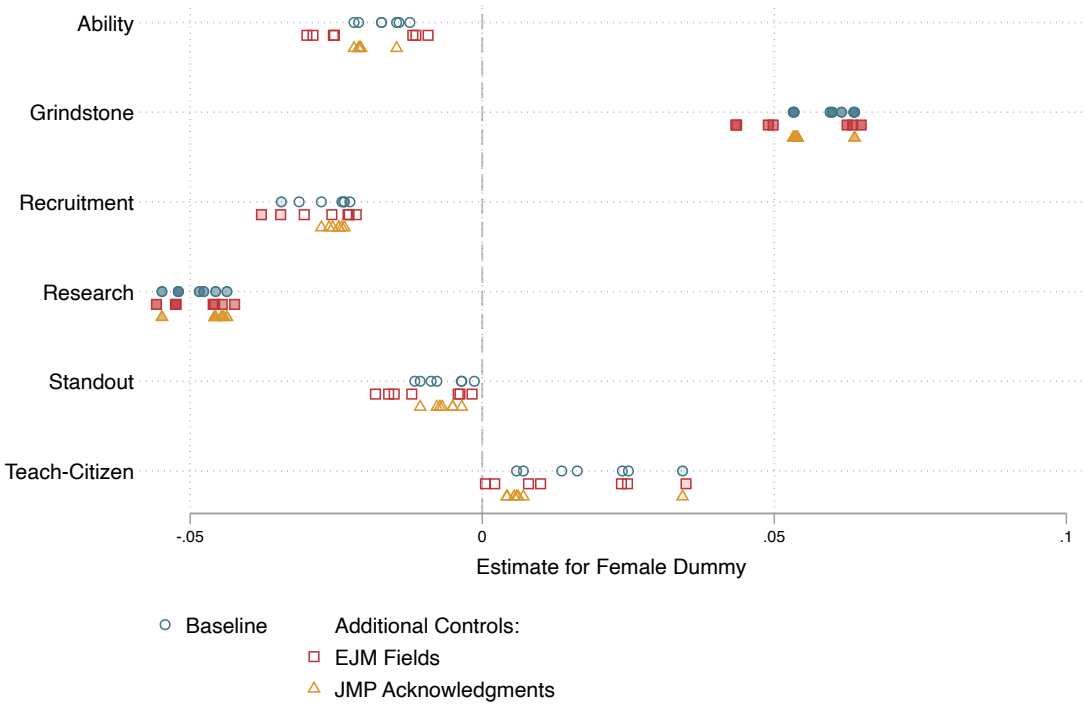
Table E.2: Full Letters — By Writer Institutional Rank

| | (1) | (2) | (3) |
|---|---|---|---|
| Writer RePEc Rank | All | Top-25 | Top-100 |
| Ability | 0.0205 | 0.0410 | 0.0177 |
| | (0.97) | (0.85) | (0.55) |
| Grindstone | 0.0971 | 0.1462 | 0.0880 |
| | (4.64)*** | (2.98)*** | (2.87)*** |
| Recruitment | 0.0310 | 0.1094 | 0.0760 |
| | (1.48) | (2.16)** | (2.42)** |
| Research | 0.0246 | 0.0832 | 0.0454 |
| | (1.19) | (1.64) | (1.45) |
| Standout | 0.0357 | 0.0590 | 0.0500 |
| | (1.69)* | (1.20) | (1.60) |
| Teaching & | 0.0202 | 0.0578 | 0.0520 |
| Citizenship | (1.02) | (1.17) | (1.72)* |
| FE/Variables absorbed | 25 | 20 | 23 |
| Additional covariates | 7 | 7 | 7 |
| Number of Letters | 11898 | 2228 | 5371 |
| dto for females | 3367 | 616 | 1513 |
| Number of candidates | 3721 | 1111 | 2304 |
| dto female | 1082 | 318 | 667 |
| Number of writers | 5670 | 971 | 2292 |
| dto female | 986 | 156 | 382 |
| Letters by fem writers | 1756 | 314 | 737 |
| Year FE | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes |
| Institution Rank FE | yes | yes | yes |
| Years since PhD | yes | yes | yes |
| Research Field FE | yes | yes | yes |
| Publications | yes | yes | yes |
| Writer characteristics | yes | yes | yes |
| Letter length | yes | yes | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.
The regressions are run for the full letter.
Return to Section 5.3 in the maintext.

# Additional Controls

Figure E.3: Regression results, EJM fields and JMP acknowledgements as controls




**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. These specifications also include two sets of additional controls, either proxies of candidate's visibility using JMP acknowledgements, or alternative definitions of research fields using EJM Fields. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Return to Section 5.3 in the maintext.

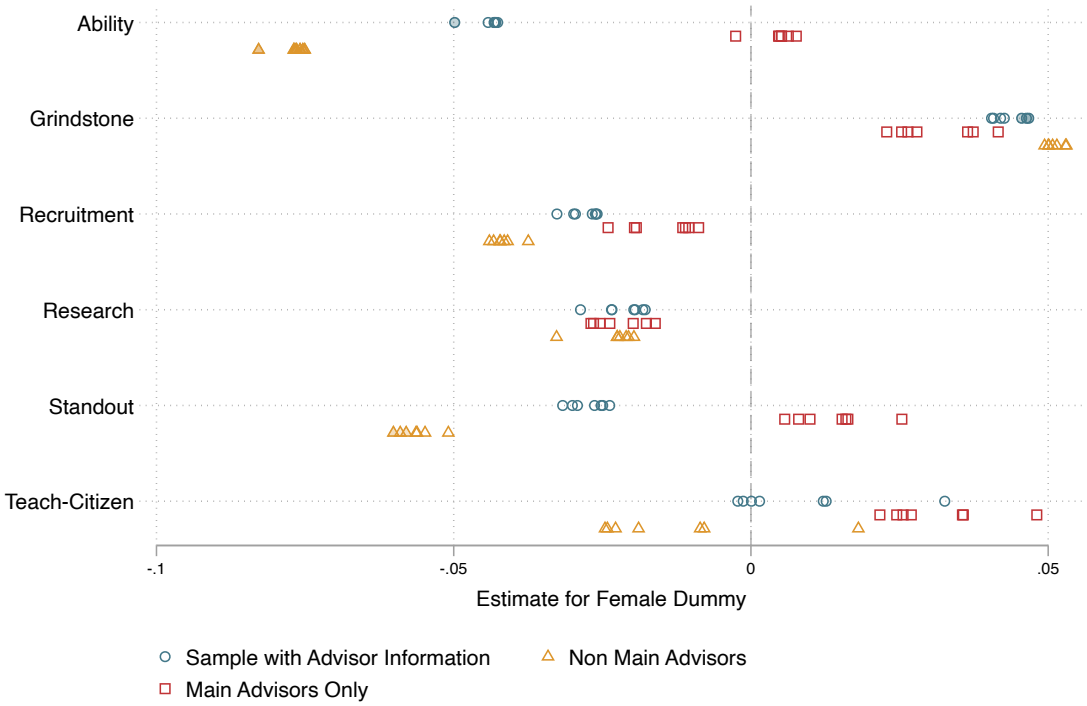
Table E.3: EJM Fields as Controls

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Ability | -0.0119 | -0.0093 | -0.0113 | -0.0253 | -0.0255 | -0.0300 | -0.0290 |
| | (0.56) | (0.44) | (0.54) | (1.16) | (1.17) | (1.38) | (1.33) |
| Grindstone | 0.0649 | 0.0625 | 0.0634 | 0.0498 | 0.0490 | 0.0435 | 0.0434 |
| | (3.05)*** | (2.94)*** | (2.99)*** | (2.25)** | (2.21)** | (1.97)** | (1.97)** |
| Recruitment | -0.0230 | -0.0228 | -0.0216 | -0.0378 | -0.0345 | -0.0305 | -0.0257 |
| | (1.08) | (1.08) | (1.02) | (1.75)* | (1.59) | (1.40) | (1.21) |
| Research | -0.0558 | -0.0525 | -0.0524 | -0.0461 | -0.0458 | -0.0424 | -0.0445 |
| | (2.69)*** | (2.53)** | (2.53)** | (2.15)** | (2.14)** | (1.97)** | (2.08)** |
| Standout | -0.0041 | -0.0017 | -0.0038 | -0.0183 | -0.0151 | -0.0160 | -0.0121 |
| | (0.19) | (0.08) | (0.18) | (0.84) | (0.69) | (0.74) | (0.56) |
| Teaching & | 0.0349 | 0.0249 | 0.0238 | 0.0100 | 0.0079 | 0.0006 | 0.0021 |
| Citizenship | (1.62) | (1.16) | (1.11) | (0.45) | (0.35) | (0.02) | (0.10) |
| FE/Variables absorbed | 10 | 15 | 15 | 159 | 159 | 165 | 165 |
| Additional covariates | | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 11638 | 11638 | 11638 | 11482 | 11482 | 11482 | 11482 |
| dto for females | 3328 | 3328 | 3328 | 3268 | 3268 | 3268 | 3268 |
| Number of candidates | 3645 | 3645 | 3645 | 3591 | 3591 | 3591 | 3591 |
| dto female | 1068 | 1068 | 1068 | 1048 | 1048 | 1048 | 1048 |
| Number of writers | 5523 | 5523 | 5523 | 5466 | 5466 | 5466 | 5466 |
| dto female | 965 | 965 | 965 | 956 | 956 | 956 | 956 |
| Letters by fem writers | 1723 | 1723 | 1723 | 1704 | 1704 | 1704 | 1704 |
| Year FE | yes | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes | yes |
| EJM Research Field FE | no | no | no | yes | yes | yes | yes |
| Publications | no | no | no | no | yes | yes | yes |
| Writer characteristics | no | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. These specifications also include as controls alternative definitions of research fields using EJM Fields. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Return to Section 5.3 in the maintext.

## Main Advisor vs Other Letter Writers

Figure E.4: Regression results, main advisor vs other letter writers




**Notes:** This figure shows the coefficient estimates for the regressions specified in 6, estimated separately for letters written by the main advisor and by others. We show the three most demanding specifications. The symbol's filling permit visualizing significance. Using 4 levels of possible standard error clustering (none, candidate's institution, letter-writer's institutions, and field), we flag significance at 3 different levels (10%, 5%, and 1%). We thus flag 12 possible significance indicators. Then, for each level of clustering, the symbol in the graph is shadowed with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol the more often they are significant. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error clustering. See overleaf for information on sample and results tables for clustering by letter writer. Return to Section 5.3 in the maintext.

Table E.4: Main Advisors Only

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Ability | -0.0026 | 0.0052 | 0.0051 | 0.0048 | 0.0063 | 0.0046 | 0.0076 |
| | (0.05) | (0.11) | (0.11) | (0.10) | (0.13) | (0.10) | (0.16) |
| Grindstone | 0.0416 | 0.0364 | 0.0374 | 0.0228 | 0.0279 | 0.0264 | 0.0253 |
| | (0.91) | (0.80) | (0.82) | (0.51) | (0.62) | (0.59) | (0.56) |
| Recruitment | -0.0088 | -0.0105 | -0.0115 | -0.0240 | -0.0193 | -0.0196 | -0.0110 |
| | (0.19) | (0.22) | (0.25) | (0.51) | (0.41) | (0.42) | (0.24) |
| Research | -0.0254 | -0.0265 | -0.0269 | -0.0198 | -0.0176 | -0.0161 | -0.0237 |
| | (0.57) | (0.59) | (0.60) | (0.44) | (0.39) | (0.36) | (0.53) |
| Standout | 0.0254 | 0.0163 | 0.0160 | 0.0057 | 0.0080 | 0.0099 | 0.0153 |
| | (0.54) | (0.35) | (0.34) | (0.12) | (0.17) | (0.21) | (0.33) |
| Teaching & | 0.0481 | 0.0356 | 0.0357 | 0.0245 | 0.0270 | 0.0217 | 0.0256 |
| Citizenship | (1.03) | (0.77) | (0.77) | (0.53) | (0.58) | (0.47) | (0.55) |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | 24 | 24 |
| Additional covariates | | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 2348 | 2348 | 2348 | 2348 | 2348 | 2348 | 2348 |
| dto for females | 683 | 683 | 683 | 683 | 683 | 683 | 683 |
| Number of candidates | 1875 | 1875 | 1875 | 1875 | 1875 | 1875 | 1875 |
| dto female | 536 | 536 | 536 | 536 | 536 | 536 | 536 |
| Number of writers | 1523 | 1523 | 1523 | 1523 | 1523 | 1523 | 1523 |
| dto female | 217 | 217 | 217 | 217 | 217 | 217 | 217 |
| Letters by fem writers | 298 | 298 | 298 | 298 | 298 | 298 | 298 |
| Year FE | yes | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes | yes |
| Publications | no | no | no | no | yes | yes | yes |
| Writer characteristics | no | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. The sample includes letters written by the main advisors, for candidates for whom that information was available and who obtained their Ph.D. 0-3 years before they enter our sample. Return to Section 5.3 in the maintext.

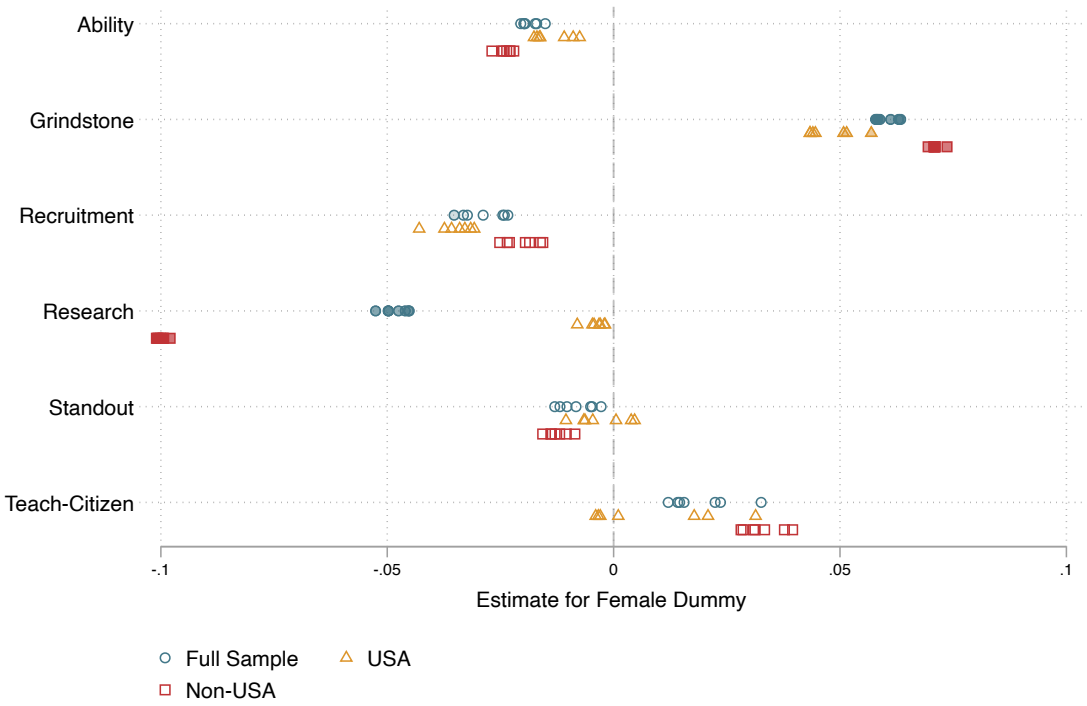
Table E.5: Exclude Main Advisors

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Ability | -0.0828 | -0.0769 | -0.0767 | -0.0764 | -0.0754 | -0.0758 | -0.0751 |
| | (2.23)** | (2.07)** | (2.06)** | (2.04)** | (2.01)** | (2.02)** | (1.99)** |
| Grindstone | 0.0501 | 0.0507 | 0.0515 | 0.0529 | 0.0531 | 0.0500 | 0.0494 |
| | (1.32) | (1.33) | (1.35) | (1.38) | (1.38) | (1.30) | (1.29) |
| Recruitment | -0.0433 | -0.0415 | -0.0409 | -0.0440 | -0.0421 | -0.0422 | -0.0374 |
| | (1.11) | (1.07) | (1.06) | (1.14) | (1.09) | (1.09) | (0.99) |
| Research | -0.0327 | -0.0225 | -0.0225 | -0.0210 | -0.0197 | -0.0206 | -0.0220 |
| | (0.86) | (0.59) | (0.59) | (0.55) | (0.51) | (0.54) | (0.58) |
| Standout | -0.0590 | -0.0562 | -0.0563 | -0.0601 | -0.0580 | -0.0548 | -0.0509 |
| | (1.60) | (1.52) | (1.52) | (1.61) | (1.56) | (1.48) | (1.40) |
| Teaching & | 0.0181 | -0.0079 | -0.0085 | -0.0189 | -0.0228 | -0.0241 | -0.0245 |
| Citizenship | (0.47) | (0.21) | (0.23) | (0.50) | (0.61) | (0.65) | (0.66) |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | 24 | 24 |
| Additional covariates | | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 3733 | 3733 | 3733 | 3733 | 3733 | 3733 | 3733 |
| dto for females | 1030 | 1030 | 1030 | 1030 | 1030 | 1030 | 1030 |
| Number of candidates | 1850 | 1850 | 1850 | 1850 | 1850 | 1850 | 1850 |
| dto female | 524 | 524 | 524 | 524 | 524 | 524 | 524 |
| Number of writers | 2522 | 2522 | 2522 | 2522 | 2522 | 2522 | 2522 |
| dto female | 434 | 434 | 434 | 434 | 434 | 434 | 434 |
| Letters by fem writers | 628 | 628 | 628 | 628 | 628 | 628 | 628 |
| Year FE | yes | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes | yes |
| Publications | no | no | no | no | yes | yes | yes |
| Writer characteristics | no | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. The sample includes the letters written by referees who are not the main advisor, for candidates for whom this information was available and who obtained their Ph.D. 0-3 years before they appear in our data. Return to Section 5.3 in the maintext.

## Location of PhD-granting institution

Figure E.5: Regression results, by location of letter writer institution




**Notes:** This figure shows the coefficient estimates for the regressions specified in 6, estimated separately for letter writers based in the US and in all other countries. We show the three most demanding specifications. The symbol's filling permit visualizing significance. Using 4 levels of possible standard error clustering (none, candidate's institution, letter-writer's institutions, and field), we flag significance at 3 different levels (10%, 5%, and 1%). We thus flag 12 possible significance indicators. Then, for each level of clustering, the symbol in the graph is shadowed with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol the more often they are significant. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error clustering. Return to Section 5.3 in the maintext.

Table E.6: US-based candidates

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Ability | -0.0089 | -0.0075 | -0.0109 | -0.0176 | -0.0162 | -0.0169 | -0.0164 |
| | (0.31) | (0.26) | (0.38) | (0.61) | (0.56) | (0.59) | (0.57) |
| Grindstone | 0.0569 | 0.0515 | 0.0508 | 0.0446 | 0.0440 | 0.0434 | 0.0434 |
| | (1.91)* | (1.73)* | (1.70)* | (1.47) | (1.45) | (1.43) | (1.43) |
| Recruitment | -0.0308 | -0.0329 | -0.0316 | -0.0429 | -0.0374 | -0.0358 | -0.0340 |
| | (1.04) | (1.12) | (1.08) | (1.46) | (1.27) | (1.22) | (1.18) |
| Research | -0.0080 | -0.0032 | -0.0047 | -0.0043 | -0.0019 | -0.0021 | -0.0030 |
| | (0.28) | (0.11) | (0.16) | (0.15) | (0.07) | (0.07) | (0.10) |
| Standout | 0.0046 | 0.0039 | 0.0005 | -0.0105 | -0.0066 | -0.0063 | -0.0046 |
| | (0.16) | (0.13) | (0.02) | (0.37) | (0.23) | (0.22) | (0.16) |
| Teaching & | 0.0314 | 0.0209 | 0.0178 | 0.0010 | -0.0029 | -0.0040 | -0.0034 |
| Citizenship | (1.05) | (0.71) | (0.60) | (0.04) | (0.10) | (0.14) | (0.12) |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | 24 | 24 |
| Additional covariates | | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 5969 | 5969 | 5969 | 5969 | 5969 | 5969 | 5969 |
| dto for females | 1716 | 1716 | 1716 | 1716 | 1716 | 1716 | 1716 |
| Number of candidates | 1874 | 1874 | 1874 | 1874 | 1874 | 1874 | 1874 |
| dto female | 552 | 552 | 552 | 552 | 552 | 552 | 552 |
| Number of writers | 2749 | 2749 | 2749 | 2749 | 2749 | 2749 | 2749 |
| dto female | 521 | 521 | 521 | 521 | 521 | 521 | 521 |
| Letters by fem writers | 981 | 981 | 981 | 981 | 981 | 981 | 981 |
| Year FE | yes | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes | yes |
| Publications | no | no | no | no | yes | yes | yes |
| Writer characteristics | no | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. The sample includes letters in support of candidates who obtained their Ph.D. in institutions in the U.S.. Return to Section 5.3 in the maintext.

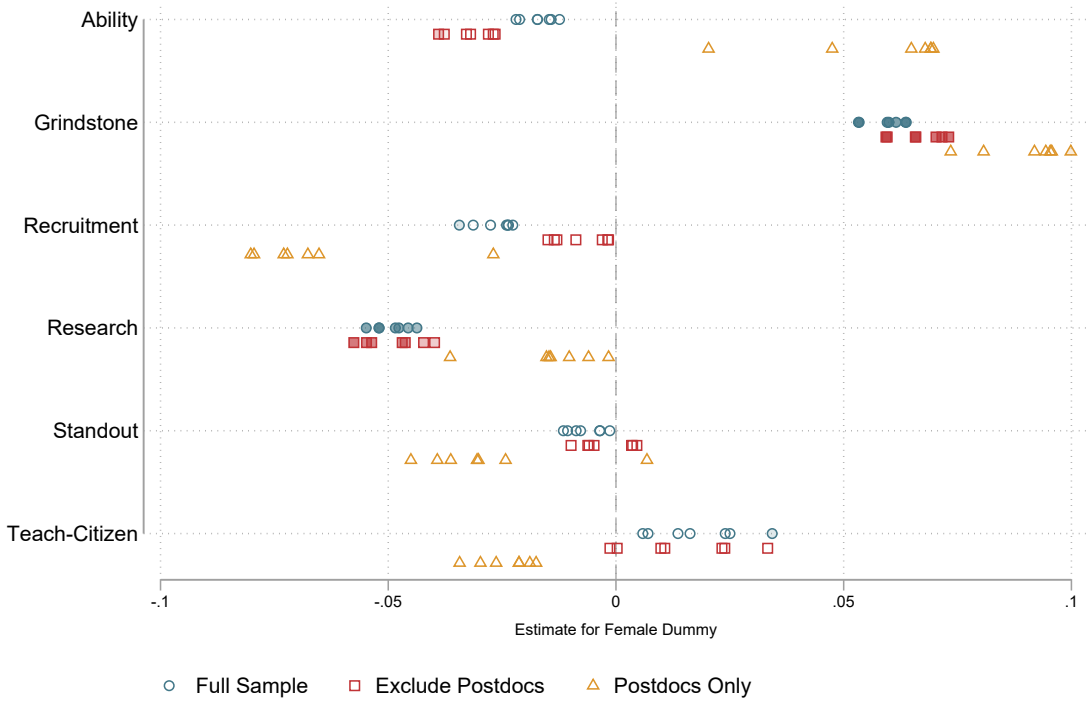
Table E.7: Non US-based candidates

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Ability | -0.0269 | -0.0229 | -0.0243 | -0.0247 | -0.0236 | -0.0230 | -0.0221 |
| | (0.88) | (0.75) | (0.80) | (0.80) | (0.77) | (0.75) | (0.72) |
| Grindstone | 0.0695 | 0.0710 | 0.0736 | 0.0709 | 0.0709 | 0.0710 | 0.0707 |
| | (2.32)** | (2.38)** | (2.47)** | (2.38)** | (2.38)** | (2.38)** | (2.37)** |
| Recruitment | -0.0186 | -0.0156 | -0.0162 | -0.0235 | -0.0230 | -0.0251 | -0.0195 |
| | (0.61) | (0.52) | (0.53) | (0.78) | (0.76) | (0.83) | (0.65) |
| Research | -0.1008 | -0.1005 | -0.0994 | -0.0996 | -0.1001 | -0.0980 | -0.1010 |
| | (3.43)*** | (3.40)*** | (3.36)*** | (3.35)*** | (3.36)*** | (3.29)*** | (3.41)*** |
| Standout | -0.0129 | -0.0086 | -0.0105 | -0.0137 | -0.0139 | -0.0157 | -0.0119 |
| | (0.42) | (0.28) | (0.34) | (0.44) | (0.45) | (0.51) | (0.39) |
| Teaching & | 0.0333 | 0.0282 | 0.0286 | 0.0309 | 0.0312 | 0.0377 | 0.0395 |
| Citizenship | (1.10) | (0.93) | (0.94) | (1.01) | (1.02) | (1.24) | (1.30) |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | 24 | 24 |
| Additional covariates | | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 5836 | 5836 | 5836 | 5836 | 5836 | 5836 | 5836 |
| dto for females | 1630 | 1630 | 1630 | 1630 | 1630 | 1630 | 1630 |
| Number of candidates | 1834 | 1834 | 1834 | 1834 | 1834 | 1834 | 1834 |
| dto female | 526 | 526 | 526 | 526 | 526 | 526 | 526 |
| Number of writers | 3269 | 3269 | 3269 | 3269 | 3269 | 3269 | 3269 |
| dto female | 497 | 497 | 497 | 497 | 497 | 497 | 497 |
| Letters by fem writers | 759 | 759 | 759 | 759 | 759 | 759 | 759 |
| Year FE | yes | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes | yes |
| Publications | no | no | no | no | yes | yes | yes |
| Writer characteristics | no | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of each 'sentiment' (e.g. ability, grindstone, etc) on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. The sample includes letters in support of candidates who obtained their Ph.D. in institutions outside the U.S.. Return to Section 5.3 in the maintext.

## Postdocs vs Others

Figure E.6: Regression results, Postdocs and Candidates Freshly out of Ph.D.



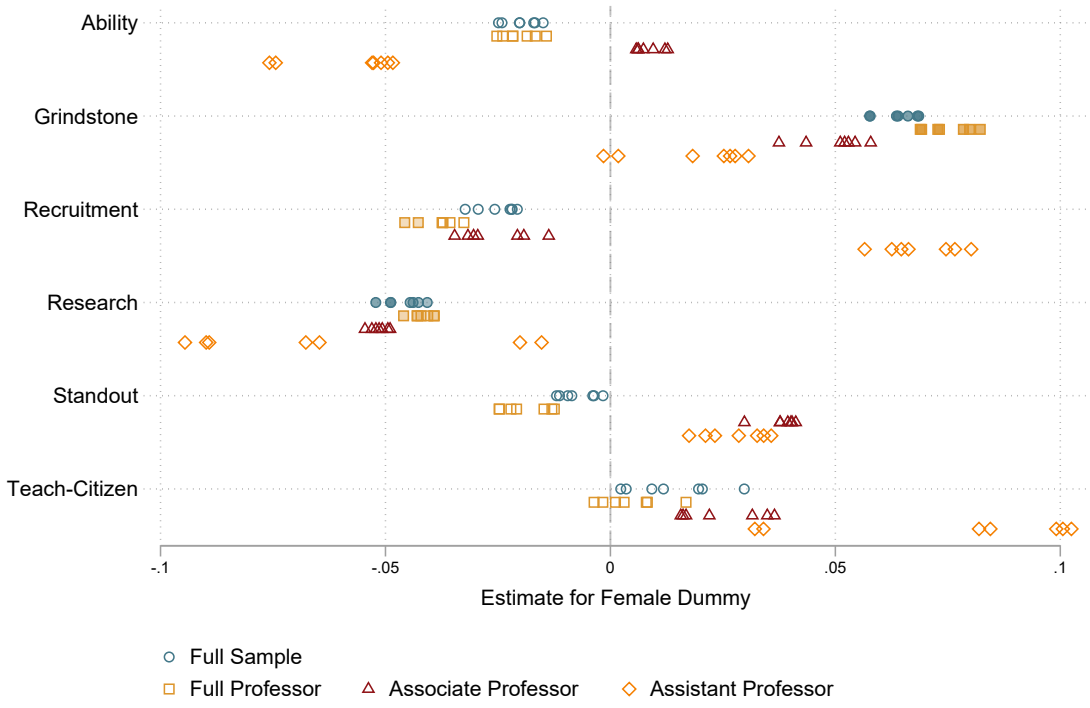

**Notes:** This figure shows the coefficient estimates for the regressions specified in 6, estimated separately for postdocs and those who are freshly out of PhD programs. We show the three most demanding specifications. The symbol's filling permit visualizing significance. Using 4 levels of possible standard error clustering (none, candidate's institution, letter-writer's institutions, and field), we flag significance at 3 different levels (10%, 5%, and 1%). We thus flag 12 possible significance indicators. Then, for each level of clustering, the symbol in the graph is shadowed with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol the more often they are significant. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error clustering. See overleaf for information on sample and results tables for clustering by letter writer. Return to Section 5.3 in the maintext.

## Seniority of Letterwriter

Figure E.7: Regression results, by year of PhD for letterwriter


**Notes:** This figure shows the coefficient estimates for the regressions specified in 6, estimated separately for letters written by the advisors who obtained their PhDs before or after 2000. We show the three most demanding specifications. The symbol's filling permit visualizing significance. Using 4 levels of possible standard error clustering (none, candidate's institution, letter-writer's institutions, and field), we flag significance at 3 different levels (10%, 5%, and 1%). We thus flag 12 possible significance indicators. Then, for each level of clustering, the symbol in the graph is shadowed with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol the more often they are significant. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error clustering. See overleaf for information on sample and results tables for clustering by letter writer. Return to Section 5.3 in the maintext.

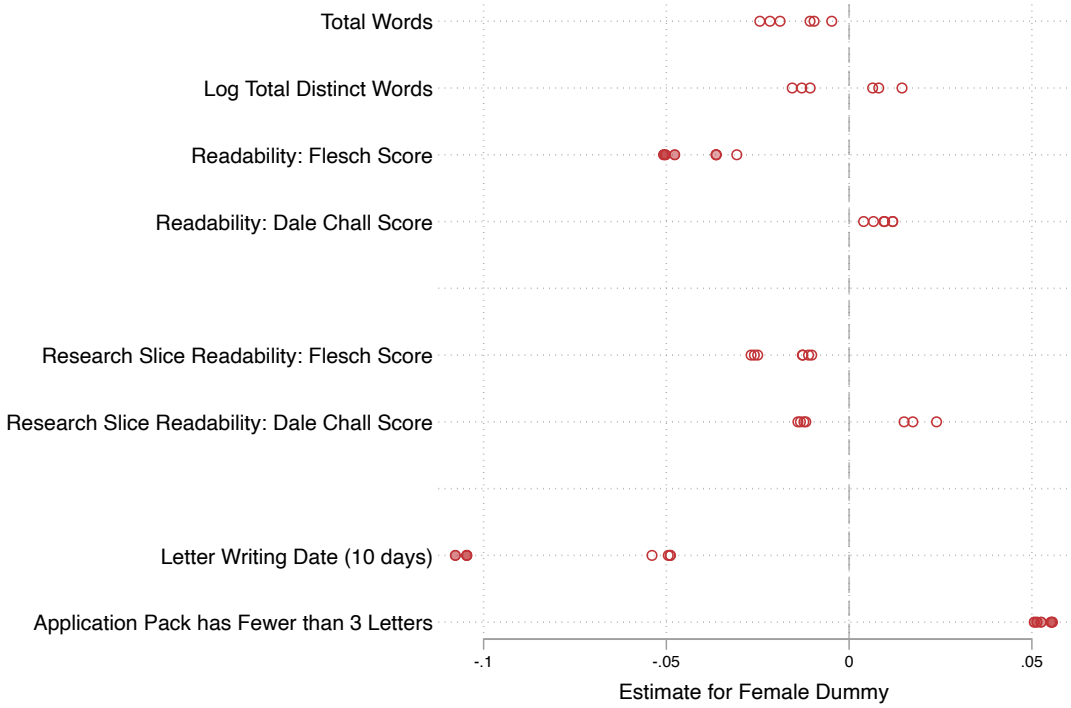
Figure E.8: Regression results, by Academic Rank of Letterwriter




**Notes:** This figure shows the coefficient estimates for the regressions specified in 6, estimated separately for letters written by the advisors who obtained their PhDs before or after 2000. We show the three most demanding specifications. The symbol's filling permit visualizing significance. Using 4 levels of possible standard error clustering (none, candidate's institution, letter-writer's institutions, and field), we flag significance at 3 different levels (10%, 5%, and 1%). We thus flag 12 possible significance indicators. Then, for each level of clustering, the symbol in the graph is shadowed with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The darker the symbol the more often they are significant. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error clustering. See overleaf for information on sample and results tables for clustering by letter writer. Return to Section 5.3 in the maintext.

# F Additional Results

Figure F.1: Regression Results Length, Readability and Timeliness




**Notes:** This figure shows the coefficient estimates for the regressions specified in 6 when outcomes are proxies for length and readability of the letter (first four rows), readability of the research slice (next two rows), and the letter date (final row). The symbol's filling permit visualizing significance. In the first to seventh line, we use four levels of possible standard error clustering (none, candidate's institution, letter-writer's institution, or letter writer), flag significance at three different levels (10%, 5%, and 1%). Then, for each level of clustering, the symbol in the graph is shaded with a 9% ($\approx 100/12$) opacity when it reaches significance at each possible level. The regression reported in the eighth line is conducted at the *candidate* level, hence only two clustering levels are used (none, and letter writer institution). The symbols are then shaded accordingly. The darker the symbol the more often they are significant. The darker the symbol, the more often it is significant. Fully filled symbols are significant at 1% level across all possible clustering. Hollow symbols do not reach significance for any level of standard error clustering. Return to Section 5.4 in the maintext.

Table F.1: Readability

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Panel A**: Full Letter | | | | | | |
| *(a) Word Counts as Dependent Variable* | | | | | | |
| Number of words | -0.0096 | -0.0048 | -0.0107 | -0.0244 | -0.0217 | -0.0189 |
| | (0.44) | (0.22) | (0.51) | (1.18) | (1.05) | (0.91) |
| Log (Number of words) | 0.0081 | 0.0145 | 0.0064 | -0.0156 | -0.0129 | -0.0107 |
| | (0.38) | (0.70) | (0.31) | (0.78) | (0.65) | (0.54) |
| *(b) Writing Quality Measures as Dependent Variable* | | | | | | |
| Flesch Readability | -0.0477 | -0.0502 | -0.0507 | -0.0363 | -0.0364 | -0.0307 |
| (higher=easier) | (2.28)** | (2.39)** | (2.41)** | (1.74)* | (1.74)* | (1.47) |
| Dale-Chall Readability | 0.0095 | 0.0040 | 0.0119 | 0.0119 | 0.0097 | 0.0067 |
| (higher=harder) | (0.45) | (0.19) | (0.57) | (0.58) | (0.47) | (0.33) |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | 25 |
| Additional covariates | | | 1 | 1 | 5 | 6 |
| Number of Letters | 11846 | 11846 | 11846 | 11846 | 11846 | 11846 |
| dto for females | 3360 | 3360 | 3360 | 3360 | 3360 | 3360 |
| Number of candidates | 3721 | 3721 | 3721 | 3721 | 3721 | 3721 |
| dto female | 1082 | 1082 | 1082 | 1082 | 1082 | 1082 |
| Number of writers | 5655 | 5655 | 5655 | 5655 | 5655 | 5655 |
| dto female | 985 | 985 | 985 | 985 | 985 | 985 |
| Letters by fem writers | 1751 | 1751 | 1751 | 1751 | 1751 | 1751 |
| **Panel B**: Research 'Slice' | | | | | | |
| *Writing Quality Measures as Dependent Variable* | | | | | | |
| Flesch Readability | -0.0268 | -0.0259 | -0.0250 | -0.0128 | -0.0127 | -0.0102 |
| (higher = easier) | (1.03) | (0.99) | (0.96) | (0.49) | (0.49) | (0.40) |
| Dale-Chall Readability | -0.0268 | -0.0259 | -0.0250 | -0.0128 | -0.0127 | -0.0102 |
| (higher = harder) | (1.03) | (0.99) | (0.96) | (0.49) | (0.49) | (0.40) |
| FE/Variables absorbed | 10 | 15 | 15 | 18 | 18 | 24 |
| Additional covariates | | | 1 | 1 | 5 | 6 |
| Number of Letters | 8010 | 8010 | 8010 | 8010 | 8010 | 8010 |
| dto for females | 2348 | 2348 | 2348 | 2348 | 2348 | 2348 |
| Number of candidates | 3203 | 3203 | 3203 | 3203 | 3203 | 3203 |
| dto female | 934 | 934 | 934 | 934 | 934 | 934 |
| Number of writers | 3834 | 3834 | 3834 | 3834 | 3834 | 3834 |
| dto female | 659 | 659 | 659 | 659 | 659 | 659 |
| Letters by fem writers | 1199 | 1199 | 1199 | 1199 | 1199 | 1199 |
| Year FE | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes |
| Publications | no | no | no | no | yes | yes |
| Writer | no | no | no | no | no | yes |

**Notes:** The table shows results of the OLS regression of proxies of length and readability of the letter (Panel A), readability of the research slice (Panel B), and the letter date (Panel C), on a female candidate indicator as well as controls mentioned in the text. The table reports the estimate for the female indicator. Each row reports a different outcome, whereas each column reports a different specification. Standard errors are clustered at the letter writer level, we report the absolute $t$-statistics in parentheses. The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. The sample includes letters in support of candidates who obtained their Ph.D. in institutions outside the U.S. Return to Section 5.4 in the maintext.

Table F.2: Timing of the Reference Letter; Incomplete Set of References

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| **Panel A**: Letter dates | | | | | | | |
| Female candidate | -1.0772 | -1.0461 | -1.0461 | -0.4947 | -0.5393 | -0.4889 | -0.4894 |
| | (2.47)** | (2.39)** | (2.39)** | (1.14) | (1.24) | (1.13) | (1.13) |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | 25 | 25 |
| Additional covariates | | | 1 | 1 | 5 | 6 | 7 |
| Number of Letters | 6335 | 6335 | 6335 | 6335 | 6335 | 6335 | 6335 |
| dto for females | 1921 | 1921 | 1921 | 1921 | 1921 | 1921 | 1921 |
| Number of candidates | 2518 | 2518 | 2518 | 2518 | 2518 | 2518 | 2518 |
| dto female | 766 | 766 | 766 | 766 | 766 | 766 | 766 |
| Number of writers | 3362 | 3362 | 3362 | 3362 | 3362 | 3362 | 3362 |
| dto female | 571 | 571 | 571 | 571 | 571 | 571 | 571 |
| Letters by fem writers | 962 | 962 | 962 | 962 | 962 | 962 | 962 |
| Year FE | yes | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | yes | yes |
| Institution Rank FE | no | yes | yes | yes | yes | yes | yes |
| Years since PhD | no | no | yes | yes | yes | yes | yes |
| Research Field FE | no | no | no | yes | yes | yes | yes |
| Publications | no | no | no | no | yes | yes | yes |
| Writer | no | no | no | no | no | yes | yes |
| Letter length | no | no | no | no | no | no | yes |
| **Panel B**: Missing Letters | | | | | | | |
| Female candidate | 0.0514 | 0.0507 | 0.0525 | 0.0556 | 0.0552 | | |
| | (4.75)*** | (4.74)*** | (4.93)*** | (5.27)*** | (5.23)*** | | |
| FE/Variables absorbed | 10 | 15 | 15 | 19 | 19 | | |
| Additional covariates | | | 1 | 1 | 5 | | |
| Number of candidates | 3599 | 3599 | 3599 | 3599 | 3599 | | |
| dto female | 1014 | 1014 | 1014 | 1014 | 1014 | | |
| Year FE | yes | yes | yes | yes | yes | | |
| Ethnicity/Race FE | yes | yes | yes | yes | yes | | |
| Institution Rank FE | no | yes | yes | yes | yes | | |
| Years since PhD | no | no | yes | yes | yes | | |
| Research Field FE | no | no | no | yes | yes | | |
| Publications | no | no | no | no | yes | | |

**Notes:** The table shows two sets of OLS regression results: in Panel (A), we provide results for the date of creation mentioned in reference letter (analysis at the *letter* level; not all letters carry a date); in Panel (B), we provide results for a dummy variable indicating candidates which received fewer than three reference letters (analysis at the *candidate* level). In both cases, the dependent variable is regressed on a female candidate indicator as well as controls as indicated. Standard errors are clustered at the letterwriter level in Panel (A) and at the candidate institution level in Panel (B), we report the absolute $t$-statistics in parentheses. Results can be interpreted as follows: in Panel (A) in days relative to letters for male candidates; in Panel (B) as percentage differences in the propensity of having fewer than 3 letters for women relative to men (unconditional propensity: 4%). The coefficients are reported in terms of standard deviations of the dependent variable. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Return to Section 5.4 in the maintext.

Table F.3: Letter Sentiment and Placement (Robustness)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable | Inst. RePEc Score (log) | | Top-100 RePEc Inst. | |
| Sample | Academic Placements | | AP & Postdoc | |
| Controls | Sentiment | All | Sentiment | All |
| Female Candidate | 9.3252 | 16.1753 | 7.5799 | 11.0226 |
| | (0.58) | (1.05) | (1.91)* | (2.88)*** |
| Ability | 1.6531 | 1.6181 | 0.5855 | 0.5340 |
| | (0.61) | (0.60) | (0.97) | (0.91) |
| Ability × | 2.7112 | 1.6501 | -0.1985 | -0.3874 |
| Female Candidate | (0.58) | (0.36) | (0.18) | (0.35) |
| Grindstone | -2.9895 | 2.2784 | -0.1255 | 0.2417 |
| | (1.28) | (1.00) | (0.20) | (0.41) |
| Grindstone × | -7.5681 | -7.3065 | -2.4693 | -2.6931 |
| Female Candidate | (1.71)* | (1.67)* | (2.31)** | (2.61)*** |
| Recruitment | 2.9000 | 1.2626 | 1.7340 | 0.6485 |
| | (1.15) | (0.50) | (2.84)*** | (1.07) |
| Recruitment × | -1.5203 | -1.1223 | -1.5329 | -1.4431 |
| Female Candidate | (0.34) | (0.26) | (1.44) | (1.37) |
| Research | 0.5148 | 2.1020 | 0.4740 | 0.8660 |
| | (0.21) | (0.88) | (0.79) | (1.51) |
| Research × | 1.4390 | -0.4129 | -0.6586 | -0.9651 |
| Female Candidate | (0.32) | (0.10) | (0.58) | (0.87) |
| Standout | 0.9753 | -0.5659 | 1.0028 | -0.0134 |
| | (0.37) | (0.22) | (1.72)* | (0.02) |
| Standout × | 5.7513 | 5.6810 | 1.0984 | 0.7742 |
| Female Candidate | (1.25) | (1.29) | (0.96) | (0.70) |
| Teaching and | 2.1171 | 4.6536 | -0.8781 | 0.3622 |
| Citizenship | (0.84) | (1.87)* | (1.45) | (0.62) |
| T&C × | -6.0157 | -7.1823 | -1.1095 | -1.4229 |
| Female Candidate | (1.42) | (1.73)* | (1.05) | (1.39) |
| FE absorbed | 5 | 25 | 5 | 25 |
| Add. covariates | 0 | 6 | 0 | 6 |
| Number of Letters | 3119 | 3119 | 6008 | 6008 |
| dto for females | 991 | 991 | 1872 | 1872 |
| Number of candidates | 957 | 957 | 1865 | 1865 |
| dto female | 313 | 313 | 596 | 596 |
| Number of writers | 2091 | 2091 | 3453 | 3453 |
| dto female | 324 | 324 | 586 | 586 |
| Letters by fem writers | 445 | 445 | 910 | 910 |
| Year FE | yes | yes | yes | yes |
| Letter Sentiments | yes | yes | yes | yes |
| Ethnicity/Race FE | no | yes | no | yes |
| Institution Rank FE | no | yes | no | yes |
| Years since PhD | no | yes | no | yes |
| Research Field FE | no | yes | no | yes |
| Publications | no | yes | no | yes |
| Writer Chars | no | yes | no | yes |
| Letter length | no | yes | no | yes |

**Notes:** This table presents alternative definitions of placement outcomes (rank score in logs and top-100 institutions). See footnote in Table 5. Return to Section 6 in the maintext.

Table F.4: Letter Sentiment and Placement (including Signals)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dependent Variable | Academia (dummy) | | Inst. RePEc Score | | Top-200 RePEc Inst. | |
| Sample | All Placements | | Academic Placements | | AP & Postdoc | |
| Controls | Sentiment | All | Sentiment | All | Sentiment | All |
| Female Candidate | 8.6237 | 7.3242 | 16.8397 | 23.9690 | 7.5741 | 11.2950 |
| | (2.29)** | (1.96)** | (1.16) | (1.67)* | (1.65)* | (2.53)** |
| Ability | 0.1386 | -0.0015 | 0.9225 | 0.4728 | 0.9198 | 0.7653 |
| | (0.23) | (0.00) | (0.38) | (0.19) | (1.30) | (1.13) |
| Ability × | 0.0067 | -0.0159 | 1.7861 | 1.1437 | 0.3317 | 0.2108 |
| Female Candidate | (0.01) | (0.02) | (0.43) | (0.27) | (0.26) | (0.17) |
| Grindstone | -0.4505 | -0.6308 | -2.4371 | -1.3932 | -0.2880 | -0.0419 |
| | (0.76) | (1.08) | (1.03) | (0.60) | (0.40) | (0.06) |
| Grindstone × | 0.1612 | 0.4785 | -10.1305 | -10.2279 | -2.3120 | -2.3884 |
| Female Candidate | (0.16) | (0.48) | (2.42)** | (2.46)** | (1.91)* | (2.00)** |
| Recruitment | 0.5132 | 0.6416 | 0.9931 | -0.1337 | 0.8063 | -0.0934 |
| | (0.75) | (0.95) | (0.37) | (0.05) | (1.02) | (0.12) |
| Recruitment × | -0.4642 | -0.7090 | -1.7615 | -1.4621 | -0.3591 | -0.2011 |
| Female Candidate | (0.40) | (0.61) | (0.39) | (0.33) | (0.26) | (0.15) |
| Research | -0.5320 | -0.1616 | 3.1107 | 4.6681 | 0.5463 | 1.0494 |
| | (0.86) | (0.27) | (1.32) | (2.02)** | (0.78) | (1.55) |
| Research × | -1.7531 | -1.5503 | 0.2748 | -0.9568 | -0.6939 | -0.9847 |
| Female Candidate | (1.64) | (1.46) | (0.07) | (0.24) | (0.54) | (0.78) |
| Standout | 1.7139 | 1.8050 | -1.7321 | -2.3437 | 0.4953 | -0.3333 |
| | (2.90)*** | (3.09)*** | (0.71) | (0.97) | (0.72) | (0.51) |
| Standout × | -1.9051 | -1.7516 | 8.0671 | 6.9163 | 2.7592 | 2.2839 |
| Female Candidate | (1.75)* | (1.64) | (1.96)** | (1.72)* | (2.06)** | (1.76)* |
| Teaching and | 0.2128 | -0.1942 | 2.4058 | 4.5869 | -0.9439 | 0.4634 |
| Citizenship | (0.35) | (0.32) | (1.02) | (1.95)* | (1.31) | (0.67) |
| T&C × | 1.7578 | 1.9019 | -7.5156 | -8.4242 | -3.6498 | -4.1560 |
| Female Candidate | (1.71)* | (1.87)* | (1.87)* | (2.13)** | (2.92)*** | (3.41)*** |
| Positive Signal | 3.6735 | 3.7462 | 22.1071 | 15.3801 | 12.1105 | 7.3623 |
| | (2.66)*** | (2.74)*** | (4.15)*** | (2.92)*** | (6.85)*** | (4.35)*** |
| Positive Signal × | 0.2141 | 0.1637 | -3.0156 | -3.7103 | -1.0424 | -1.3835 |
| Female Candidate | (0.09) | (0.07) | (0.32) | (0.40) | (0.33) | (0.44) |
| Negative Signal | -3.9254 | -3.5149 | -15.5574 | -9.4118 | -8.9975 | -5.7669 |
| | (2.13)** | (1.94)* | (1.98)** | (1.22) | (4.24)*** | (2.83)*** |
| Negative Signal × | 8.1602 | 7.8531 | -4.5804 | -7.8088 | 3.9169 | 2.0739 |
| Female Candidate | (2.57)** | (2.47)** | (0.33) | (0.57) | (1.00) | (0.54) |
| Comparison | 2.0270 | 1.7199 | 15.6283 | 14.8671 | 4.3365 | 3.4365 |
| | (0.84) | (0.71) | (1.79)* | (1.73)* | (1.41) | (1.20) |
| Comparison × | 1.1538 | 2.2788 | -24.6483 | -21.5614 | -5.5608 | -4.7815 |
| Female Candidate | (0.28) | (0.57) | (1.53) | (1.34) | (1.11) | (0.97) |
| FE absorbed | 5 | 25 | 5 | 25 | 5 | 25 |
| Add. covariates | 0 | 6 | 0 | 6 | 0 | 6 |
| Number of Letters | 8760 | 8760 | 3119 | 3119 | 6008 | 6008 |
| dto for females | 2588 | 2588 | 991 | 991 | 1872 | 1872 |
| Number of candidates | 2738 | 2738 | 957 | 957 | 1865 | 1865 |
| dto female | 830 | 830 | 313 | 313 | 596 | 596 |
| Number of writers | 4461 | 4461 | 2091 | 2091 | 3453 | 3453 |
| dto female | 774 | 774 | 324 | 324 | 586 | 586 |
| Letters by fem writers | 1339 | 1339 | 445 | 445 | 910 | 910 |
| Year FE | yes | yes | yes | yes | yes | yes |
| Letter Sentiments | yes | yes | yes | yes | yes | yes |
| Ethnicity/Race FE | no | yes | no | yes | no | yes |
| Institution Rank FE | no | yes | no | yes | no | yes |
| Years since PhD | no | yes | no | yes | no | yes |
| Research Field FE | no | yes | no | yes | no | yes |
| Publications | no | yes | no | yes | no | yes |
| Writer Chars | no | yes | no | yes | no | yes |
| Letter length | no | yes | no | yes | no | yes |



**Notes:** See footnote in Table 5. Return to Section 6 in the maintext.