

A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants

Toni Schmader · Jessica Whitehead · Vicki H. Wysocki

Published online: 17 August 2007
© Springer Science + Business Media, LLC 2007

Abstract Letters of recommendation are central to the hiring process. However, gender stereotypes could bias how recommenders describe female compared to male applicants. In the current study, text analysis software was used to examine 886 letters of recommendation written on behalf of 235 male and 42 female applicants for either a chemistry or biochemistry faculty position at a large U.S. research university. Results revealed more similarities than differences in letters written for male and female candidates. However, recommenders used significantly more standout adjectives to describe male as compared to female candidates. Letters containing more standout words also included more ability words and fewer grindstone words. Research is needed to explore how differences in language use affect perceivers' evaluations of female candidates.

Keywords Gender schemas · Sexism · Implicit biases · Hiring decisions · Chemistry

Introduction

Despite continuing efforts to increase the involvement of women in physical science, math, and engineering, men

This research was supported by National Institute of Mental Health Grant #1R01MH071749 to the first author and a University of Arizona grant from the College of Science Millennium Report Oversight Committee awarded to the third author.

T. Schmader (✉) · J. Whitehead
Department of Psychology, University of Arizona,
Tucson, AZ 85721, USA
e-mail: schmader@u.arizona.edu

V. H. Wysocki
Department of Chemistry, University of Arizona,
Tucson, AZ 85721, USA

continue to make up an overwhelming majority of the faculty in these programs. For example, in chemistry departments across the U.S., women comprise only 12% of the faculty overall and only 7.6% of full professors are women (Nelson 2005). Although this discrepancy might partly reflect a gender imbalance in who applies for faculty positions, data suggest that there should be many qualified women who earn their Ph.D.s in these disciplines. For example, between the years of 1993 and 2002, women received 31% of the Ph.D.s awarded in chemistry (Nelson 2005). Although there are likely to be many factors that contribute to the under-representation of women in the natural sciences, there has been a growing interest in recent years in the role of unconscious processes that could bias perception against women trying to succeed in domains that have been traditionally dominated by men (Heilman 1995; Vallian 1998).

The present research follows in this vein by investigating, in an actual job hiring context, whether there are differences in how recommenders describe male and female applicants for faculty positions in chemistry and biochemistry at a large research university in the United States. In the present study, job candidates' actual recommendation letters were transcribed and analyzed using text analysis software to compare language content such as usage of teaching and research related words, ability and grindstone adjectives, and standout adjectives. Understanding the blatant or subtle discrepancies in how recommenders describe male and female job candidates will allow search committees to conduct fairer and more successful searches for the most qualified candidate.

When search committees review job applications, recommendation letters are a critical part of the review process. Many applicants may look similar on objective criteria such as their number of publications, fellowships, and presentations. They are better distinguished in the review process by their

statements of research interests and teaching philosophy. However, in addition to the information provided by the applicants, recommendation letters provide a unique way for search committee members to get a better sense of the candidate as a scholar and colleague. Such letters call for subjective judgments of a candidate by those who have had the best opportunity to evaluate that candidate's work, personality, and potential for career success. However, research from social psychology suggests that even well-intentioned individuals can employ unconscious biases when evaluating those who are members of negatively stereotyped groups. Numerous studies have revealed that even when individuals are motivated to behave in egalitarian ways, they may still show bias at an implicit or unconscious level (Greenwald and Banaji 1995). These implicit biases, which might reflect years of exposure to cultural messages and could have little relation to one's consciously held attitudes and beliefs, tend to be elicited automatically and can manifest themselves in nonverbal behaviors, social judgments, and behavioral choices (Poehlman et al. 2007).

Furthermore, past research has shown that such biases can influence how job applicants are perceived (Heilman et al. 1988). For example, Biernat and Eidelman (2007) recently demonstrated that when people evaluate letters that use equivalent language to describe male and female students in a masculine domain, they translate those letters into less favorable judgments of qualifications when the applicant was female compared to male. Other research has similarly shown that unconsciously held gender stereotypes can systematically bias the judgments of male and female managers as well as applicants (Heilman 2001; Rudman and Glick 1999, 2001). Taken together, this research suggests that implicit gender biases can affect how applications are evaluated. The question that concerns us is whether these biases can also be reflected in the wording used by recommenders when they describe male and female job candidates.

In addition, most of the prior social psychological research examining the role of implicit biases on applicant evaluation has adopted a laboratory methodology that allows for control over potentially confounding factors. Results from these studies tell us that such biases *can* exist and affect evaluations of job candidates, but they do not demonstrate the degree to which these biases *do* exist in real world hiring contexts. For example, although research summarized above suggests that implicit gender biases lead to differences in how letter writers describe unknown female job applicants as compared to male job applicants, other research suggests that such stereotypes are less likely to bias judgment once individuals begin interacting with one another (Kunda et al. 2002; Neuberg and Fiske 1987). Since those writing letters for job candidates have had greater opportunity to get to know the candidate as an individual, their summary evaluations

might not contain strong evidence of gender bias. In any case, it becomes important to answer this question outside of a laboratory setting, where the letters being written have real world consequences. Thus, the present research focused on actual letters of recommendation written for male and female candidates applying for a faculty position.

There has been at least one previous study of gender biases in actual letters of recommendation written for male and female job candidates for faculty positions. Trix and Psenka (2003) examined a naturalistic set of recommendation letters for 62 female and 222 male applicants who had been hired at an American medical school over a 3 year period. Their analysis revealed that a higher percentage of recommendation letters written for women tended to be very short (fewer than 10 lines), and a higher percentage of letters written for men tended to be very long (over 50 lines). Trix and Psenka also discovered that 15% of letters written for female applicants (as compared to 6% for male candidates) could be termed letters of minimal assurance, in which the letters lacked a stated commitment to the applicant, detailed comments, or any evaluation of the applicants' traits or accomplishments.

In addition, Trix and Psenka also included a qualitative comparison of the content of letters written for male and female candidates. They observed that more letters written for females as compared to males included language related to gender (10 vs 5%), doubt (24 vs 12%), and what the authors called "grindstone adjectives" (e.g., hardworking; 34 vs 23%). There was also some suggestion that letters for male applicants included more reference to "his research," "his ability," or "his career," whereas letters for female applicants included more reference to "her teaching," or "her training." The researchers concluded that recommenders seemed to emphasize women's strong work ethic and portray them in terms of their training and teaching, whereas the focus in men's recommendations included greater confidence in their research and ability.

Although the findings of Trix and Psenka (2003) are provocative, one limitation of this study is that most of their comparisons were not statistically analyzed to provide information on the reliability of these differences. In the present study, we used text analysis software to compare letters written for male and female applicants for tenure track faculty positions in chemistry and biochemistry at a large Research I University. From the standard software coding scheme, seven categories were identified as having potential for revealing evidence of gender bias and allowed us to test the following hypotheses:

Hypothesis 1: Recommendation letters written for female as compared to male applicants would be shorter in length.

Hypothesis 2: Recommendation letters written for female as compared to male applicants would contain more negative and less positive language.

Hypothesis 3: Recommendation letters written for female as compared to male applicants would contain more tentative language and less certainty language.

Hypothesis 4: Recommendation letters written for female as compared to male applicants would make fewer references to achievement and more references to communication skills.

In addition, we created five user-defined categories to test the following hypotheses suggested by Trix and Psenka's (2003) data:

Hypothesis 5: Recommendation letters written for female as compared to male applicants would contain fewer standout adjectives.

Hypothesis 6: Recommendation letters written for female as compared to male applicants would contain fewer research-related words and more teaching-related words.

Hypothesis 7: Recommendation letters written for female as compared to male applicants would contain fewer ability-related words and more grindstone-related words.

Method

We obtained access to the complete set of recommendation letters written on behalf of job applicants for a tenure track faculty position in either chemistry (164 males, 21 females) or biochemistry (71 males, 21 females) at a large research

university. The letters were transcribed and then analyzed using Linguistic Inquiry Word Count (LIWC—Pennebaker et al. 2001). LIWC software has been widely used and extensively validated as a word-count based text analysis program (Mehl 2005). It operates by comparing all words of a given text to either an existing dictionary of word categories or to user-defined dictionary categories, counting the number of words from that dictionary that appear, and then dividing this number by the total number of words used in the letter. Of the 74 word categories in the LIWC2001 default dictionary, the following categories were of particular interest: achievement words (e.g., goal), communication words (e.g., talk), positive emotions (e.g., happy), negative emotions (e.g., worthless), tentative words (e.g., perhaps), and certainty words (e.g., always). In addition, based on research by Trix and Psenka (2003), we created five language categories including grindstone traits, ability traits, standout adjectives, research terms, and teaching terms. The complete list of words and word stems used is provided in the appendix. We also gathered information pertaining to the qualifications of the applicants including the number of publications, presentations, fellowships, and post-doctoral positions.

Results

Candidates had an average of 3.19 letters of recommendation (range=1 to 8). After conducting the word counts for each category on individual letters, we averaged these counts within language-use dimension and across letter writers to create aggregate variables for each candidate. We then conducted applicant sex x department ANOVA's on

Table 1 Average level of candidate qualifications by gender and academic department.

Parameters	Chemistry			Biochemistry			Total	
	Men <i>n</i> =164	Women <i>n</i> =21	Total <i>n</i> =186	Men <i>n</i> =71	Women <i>n</i> =21	Total <i>n</i> =92	Men <i>n</i> =235	Women <i>n</i> =42
Publications	25.84 (20.56)	20.88 (14.98)	25.20 (19.96)	15.97 (10.24)	15.29 (9.43)	15.82 (10.01)	22.83 (18.59)	18.27 (12.88)
Presentations	12.73 (17.34)	16.46 (19.24)	13.22 (17.59)	8.13 (11.22)	13.29 (13.34)	9.30 (11.86)	11.33 (15.85)	14.98 (16.64)
Fellowships	.31 (.64)	.38 (.77)	.32 (.66)	1.25 (1.13)	1.48 (1.08)	1.30 (1.12)	.60 (.93)	.89 (1.07)
Years in Ph.D.	4.81 (1.65)	4.25 (2.03)	4.74 (1.71)	5.37 (1.65)	5.38 (1.80)	5.37 (1.67)	4.98 (1.67)	4.78 (1.99)
Postdoctoral positions	.96 (.85)	.67 (.64)	.92 (.83)	1.38 (.72)	1.57 (.75)	1.42 (.73)	1.09 (.84)	1.09 (.82)
Years in postdoctoral	1.90 (1.77)	1.46 (1.69)	1.84 (1.76)	4.37 (2.32)	5.33 (2.31)	4.59 (2.34)	2.65 (2.26)	3.27 (2.78)

Numbers reflect the means for each specific qualification. SDs are represented in parentheses below the means.

each of the language-use dimensions. There were too few letters written by female recommenders to allow for an analysis by recommender sex.

Candidate Qualification

Table 1 reports average level of qualifications by applicant sex and department. Importantly, analyses of qualification variables revealed that there were no significant differences between male and female candidates in number of publications, presentations, fellowships, years in Ph.D., or post-doctoral positions, all p 's > .05. There were, however, departmental differences in these qualification variables. Candidates applying for a faculty position in chemistry had more publications ($M=25.20$) than those applying for a position in biochemistry ($M=15.82$), $p < .05$. Biochemistry applicants had a greater number of postdoctoral positions ($M=1.42$), longer postdoctoral positions ($M=4.59$), and had received a larger number of fellowships ($M=1.30$), as compared to the chemistry applicants, all p 's < .05. No interactions between applicant sex and department were observed, all p 's > .05.

LIWC Default Dimensions

We next analyzed the LIWC generated language-use dimensions using a series of applicant sex x department ANOVAs (See Table 2). Female candidates ($M=3.38$) tended to have somewhat more recommendation letters than male candidates ($M=3.16$), $F(1, 273) = 2.93, p = .08$. However, in contrast to Hypotheses 1, 2, 3, and part of 4, no significant gender differences emerged for any of the following LIWC generated language-use dimensions: length of letters, negative feeling words, positive feeling words, tentative words, certainty words, or achievement words, p 's > .05. Providing partial support for Hypothesis 4, letters for female candidates ($M=1.13$) included marginally more words related to communication than did letters for male candidates ($M=.98$), $F(1, 273) = 3.06, p = .08$. In addition, there were significant department differences for communication words, $F(1, 273)=5.84, p = .02$; negative feeling words, $F(1, 273)=5.07, p = .02$; and positive feeling words, $F(1, 273)=4.10, p = .04$. Letters for biochemistry candidates included more communication words ($M=1.12$), more negative feeling words ($M=.41$) and fewer positive

Table 2 Content of recommendation letter as a function of gender and academic department.

Parameters	Chemistry			Biochemistry			Total	
	Men n=164	Women n=21	Total n=185	Men n=71	Women n=21	Total n=92	Men n=235	Women n=42
Number of letters	3.29 (.10)	3.81 (1.33)	3.35 (1.05)	2.86 (1.05)	2.95 (1.16)	2.28 (1.07)	3.16 (1.03)	3.38 (1.31)
Length of letters (in words)	531 (194)	596 (136)	539 (190)	611 (213)	579 (178)	604 (205)	555 (203)	588 (157)
Achievement	2.12 (.57)	2.22 (.65)	2.13 (.58)	2.05 (.61)	2.14 (.52)	2.07 (.59)	2.10 (.58)	2.18 (.58)
Communication	.94 (.40)	1.04 (.39)	.95 (.40)	1.09 (.42)	1.23 (.46)	1.12 (.43)	.98 (.41)	1.13 (.44)
Positive feelings	.34 (.18)	.33 (.16)	.34 (.18)	.27 (.14)	.29 (.19)	.28 (.15)	.32 (.17)	.31 (.17)
Negative feelings	.31 (.21)	.33 (.15)	.31 (.20)	.41 (.31)	.41 (.25)	.41 (.29)	.34 (.24)	.37 (.21)
Tentative words	.72 (.29)	.78 (.38)	.73 (.30)	.76 (.35)	.68 (.27)	.74 (.33)	.73 (.31)	.73 (.33)
Certainty words	1.08 (.39)	1.06 (.33)	1.08 (.38)	1.11 (.48)	1.07 (.39)	1.10 (.46)	1.09 (.42)	1.06 (.35)
Standout words	.70 (.30)	.57 (.28)	.68 (.30)	.71 (.36)	.63 (.26)	.69 (.34)	.70 (.32)	.60 (.27)
Ability words	.76 (.35)	.73 (.31)	.76 (.34)	.78 (.36)	.84 (.29)	.79 (.34)	.76 (.35)	.78 (.30)
Grindstone words	.43 (.25)	.48 (.34)	.44 (.26)	.39 (.21)	.39 (.19)	.39 (.20)	.42 (.24)	.44 (.28)
Teaching words	1.31 (.93)	1.29 (.71)	1.31 (.90)	.97 (.53)	1.08 (.53)	1.00 (.53)	1.21 (.84)	1.20 (.64)
Research words	2.86 (.86)	2.76 (.74)	2.84 (.85)	2.85 (.75)	2.90 (.87)	2.86 (.78)	2.85 (.83)	2.82 (.80)

Numbers reflect the mean percentage of words from the recommendation letter that fall within a given category. SDs are represented in parentheses below the means.

feeling words ($M=.28$) as compared to letters for chemistry candidates ($M=.95$, $M=.31$, $M=.34$, respectively). No department effects emerged for length of letters, achievement words, tentative words, or certainty words and no interaction effects were significant, p 's $>.05$.

We next conducted a series of sex \times department ANOVAs to analyze the language-use dimensions that were created to address the specific goals of this research. In line with Hypothesis 5, results revealed a significant gender difference in how many standout adjectives (e.g. outstanding, unique, and exceptional) the recommender used to describe the candidate, $F(1, 278)=3.95$, $p=.05$. Consistent with the notion that implicit biases can influence how letter writers describe female candidates, recommenders described male candidates ($M=.70$) with significantly more standout adjectives compared to female candidates ($M=.60$). To address the possibility that this difference could be accounted for by differences in the qualifications of male and female candidates, we conducted an ANCOVA that included number of publications, presentations, fellowships, post-doctoral positions, and number letters of recommendation as covariates. Even after removing variance in standout language due to any and all of these variables, the gender difference remained significant, $p=.04$. There were no differences between departments in how many standout adjectives candidates' letters included.

Contrary to Hypotheses 6 and 7, there were no significant gender differences in the number of grindstone traits (e.g. hardworking, conscientious), ability traits (e.g. talented, smart), research terms (e.g. manuscript, theory), or teaching terms (e.g. adviser, colleague) used to describe candidates, p 's $>.05$. There was however, a significant main effect of department on the number of teaching terms used to describe the candidates, $F(1, 278) = 4.38$, $p<.05$. Letters written to describe chemistry candidates ($M=1.31$) included more language about teaching as compared to letters written to describe biochemistry candidates ($M=1.00$). No other department effects or interactions emerged on the language-use dimensions described above.

Supplementary analyses were conducted to further understand the significance of the finding that male candidates were more likely to be described with standout adjectives. These analyses explored the possible covariation of using standout words to describe an applicant and focusing on ability and research skill when describing that candidate. They revealed both a significant positive correlation between using standout words and ability words, $r=.14$, $p<.05$, and a significant negative correlation between using standout words and grindstone words, $r=-.17$, $p<.01$ (regardless of the gender of the applicant). In other words, recommenders who use superlatives to describe candidates were also more likely to focus on a candidate as having intrinsic ability as opposed to being a

conscientious and hard worker. Thus, even though men were not more likely to be described in terms of their inherent abilities in general, the covariation of this language with standout words might suggest that men were more likely to be described as having a superlative amount of natural ability.

Discussion

Overall, the results of the current study revealed more similarity in the letters written for male and female job candidates than differences. Male and female candidates had similar levels of qualifications and this was reflected in their letters of recommendation. Letters written for women included language that was just as positive and placed equivalent emphasis on ability, achievement, and research. Thus, in contrast to the findings of Trix and Psenka (2003), letters for female candidates to jobs in chemistry and biochemistry did not contain significantly more tentative language and did not overemphasize teaching and hard work over research and ability.

However, it is notable that recommenders used significantly more standout adjectives to describe male candidates as compared to female candidates, even though objective criteria showed no gender differences in qualifications. It is likely that evaluators place higher weight on letters that describe a candidate as the *most* gifted, *best* qualified, or a rising *star*. This could mean that even a small difference in the proportion of standout adjectives used in describing female candidates could translate into much larger evaluative effects. Interestingly, the data also revealed that letters that contained more standout words also included more ability related terms and fewer grindstone words. Even though no sex differences were found in these latter categories, the use of standout adjectives in combination with ability language could also have the effect of amplifying the weight that search committees place on ability when evaluating a given application. More research is needed to understand how these seemingly small differences in language use affect the overall evaluations made by social perceivers.

Along those same lines, it is important to take into account research showing that applicants with similar objective skills and qualifications can still be perceived differently by those reviewing their applications (Biernat and Eidelman 2007). In other words, when judging equivalent letters for a male and female candidate, a perceiver who is making a judgment based on minimal information could still experience the influence of unconscious gender biases that could lead them to evaluate the male candidate more positively. This same bias might be somewhat less likely to reveal itself in letters of recom-

mendation given that the effects of stereotypes on evaluations tend to be muted with further interpersonal contact (Kunda et al. 2002). In other words, just because there were relatively few gender differences in the letters of recommendations analyzed in this study, we cannot infer that gender stereotypes do not still play a role in how applicants are evaluated as part of the search process.

The present study reveals that even as individuals continue to work towards egalitarian treatment, gender biases may still reveal themselves in subtle forms. Future research must examine the specific processes that contribute to both gender discrepancies and inequalities in science-related disciplines. However, alerting recommenders and search committees to the role of implicit biases in evaluation can allow them to begin to police their own behavior and will help to ensure fair and successful searches for the most qualified male and female candidates. In addition, ensuring a fair hiring process may eventually encourage more women to enter into male-dominated fields, helping to eliminate the under-representation of women in science.

Appendix

Study-Defined Dimension Dictionaries

Standout words: excellen*, superb, outstanding, unique, exceptional, unparalleled, *est, most, wonderful, terrific*, fabulous, magnificent, remarkable, extraordinary*, amazing, supreme*, unmatched

Ability words: talent*, intell*, smart*, skill*, ability, genius, brilliant*, bright*, brain*, aptitude, gift*, capacity, propensity, innate, flair, knack, clever*, expert*, proficient*, capable, adept*, able, competent, natural*, inherent*, instinct*, adroit*, creative*, insight*, analytical

Grindstone words: hardworking, conscientious, depend*, meticulous, thorough, diligen*, dedicate, careful, reliab*, effort*, assiduous, trust*, responsib*, methodical, industrious, busy, work*, persist*, organiz*, disciplined

Teaching words: teach, instruct, educat*, train*, mentor, supervis*, adviser, counselor, syllabus, syllabus, course*, class, service, colleague, citizen, communicate*, lectur*, student*, present*, rapport

Research words: research*, data, study, studies, experiment*, scholarship, test*, result*, finding*, publication*,

publish*, vita*, method*, scien*, grant*, fund*, manuscript*, project*, journal*, theor*, discover*, contribution*

Note. * indicates that any word containing the letter string that precedes or follows the asterisk should be counted.

References

- Biernat, M., & Eidelman, S. (2007). Translating subjective language in letters of recommendation: The case of the sexist professor. *European Journal of Social Psychology* (in press).
- Greenwald, A., & Banaji, M. (1995). Implicit cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.
- Heilman, M. E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. *Journal of Social Behavior and Personality*, *10*, 3–26.
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*, 657–674.
- Heilman, M. E., Martell, R. F., & Simon, M. C. (1988). The vagaries of sex bias: Conditions regulating the undervaluation, equivalence, and overvaluation of female job applicants. *Organizational Behavior and Human Decision Processes*, *41*, 98–110.
- Kunda, Z., Davies, P., Adams, B., & Spencer, S. (2002). The dynamic time course of stereotype activation: Activation, dissipation, and resurrection. *Journal of Personality and Social Psychology*, *82*, 283–299.
- Mehl, M. R. (2005). Quantitative text analysis. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 141–156). Washington, DC: American Psychological Association.
- Nelson, D. (2005). *A national analysis of diversity in science and engineering faculties at research universities*. Retrieved January 15, 2007, from <http://cheminfo.chem.ou.edu>.
- Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology*, *53*, 431–444.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count: LIWC2001*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Poehlman, T. A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2007). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Unpublished manuscript*.
- Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle manager. *Journal of Personality and Social Psychology*, *77*, 1004–1010.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. In L. L. Carli & A. H. Eagly (Eds.), *Journal of Social Issues*, *57*, 743–762.
- Trix, F., & Psenka C. (2003). Exploring the color of glass: letters of recommendation for female and male medical faculty. *Discourse and Society*, *14*, 191–220.
- Valian, V. (1998). *Why so slow? The advancement of women*. Cambridge, MA: M.I.T. Press.