A Meta-Analytic Test of the Token Effect on Job Performance

Conference Paper *in* Academy of Management Annual Meeting Proceedings · January 2015

CITATIONS
0

READS
144

4 authors, including:

Jeremy D. Mackey
Auburn University
54 PUBLICATIONS **1,080** CITATIONS
SEE PROFILE

Chad H Van Iddekinge
University of Iowa
77 PUBLICATIONS **2,287** CITATIONS
SEE PROFILE

Article

A Meta-Analysis of Gender Proportionality Effects on Job Performance

Group & Organization Management
2019, Vol. 44(3) 578–610

DOI: 10.1177/1059601117730519
journals.sagepub.com/home/gom
SAGE

Jeremy D. Mackey[1], Philip L. Roth[2], Chad H. Van Iddekinge[3], and Lynn A. McFarland[4]

Abstract

Critical mass theory and the tokenism hypothesis propose that females' job performance is adversely affected by perceptions and experiences that stem from females comprising a smaller proportion of organizations than males. Although belief in the gender token effect appears to be widely held, empirical evidence of this effect is relatively scarce; furthermore, the evidence that does exist is somewhat inconsistent. The purpose of the present study was to provide a meta-analytic test of the gender token effect by examining the extent to which the proportion of females in organizations relates to male–female differences in job performance. Meta-analytic results based on data from 158 independent studies (N = 101,071) reveal that (a) females tend to demonstrate higher job performance than males (d = −.10), and (b) this difference does not appear to vary based on the proportion of females in organizations. We found similar results for subjective task performance (e.g., supervisory ratings), organizational citizenship behaviors, and objective task performance (e.g., sales). Overall, this study's results demonstrate almost no

[1]Auburn University, Auburn, AL, USA
[2]Clemson University, Clemson, SC, USA
[3]Florida State University, Tallahassee, FL, USA
[4]University of South Carolina, Columbia, SC, USA

Corresponding Author:
Jeremy D. Mackey, Department of Management, Raymond J. Harbert College of Business, Auburn University, 405 West Magnolia Avenue, Auburn, AL 36849, USA.
Email: jmackey@auburn.edu

support for the gender token effect on job performance, which challenges the prevailing assumptions of critical mass theory and the tokenism hypothesis.

Keywords
gender, performance, token, meta-analysis, OCB

Researchers have been interested in the relationship between gender and organizational evaluations of employee behaviors for decades (e.g., Landy & Farr, 1980; McCarthy, Van Iddekinge, & Campion, 2010; Nieva & Gutek, 1980). Kanter (1977a) made a major theoretical contribution to this area when she wrote her highly influential book *Men and Women of the Corporation*. In her book, Kanter pioneered critical mass theory, which proposed that proportional representation (i.e., "tokenism") of a minority group (e.g., females) in organizations can harmfully affect minority individuals. Kanter's theory has inspired a great deal of discussion and research in diverse fields, including management, applied psychology, sociology, and criminal justice.

Although studies often provide rich qualitative accounts of the effects of tokenism, they tend to provide little quantitative data concerning the nature and magnitude of the relationship between gender and job performance. Furthermore, previous quantitative reviews in this area are limited in focus and yield conflicting results. For example, Sackett, Dubois, and Noe (1991) found support for the gender token effect, whereas Bowen, Swim, and Jacobs (2000) did not. Other researchers have conducted small meta-analyses of male–female differences in job performance or other outcomes (e.g., hiring recommendations) but did not address token effects (e.g., Davison & Burke, 2000; Roth, Purvis, & Bobko, 2012). Thus, we were motivated to conduct a meta-analysis of the gender token effect that could provide robust quantitative evidence of the nature and magnitude of the relationship between gender and job performance.

The purpose of the present study was to conduct a large-scale test of the gender token effect on performance in organizational settings. We searched a wide range of literatures to cumulate studies that reported female representation and one or more measures of job performance. We included studies that reported subjective and objective performance measures, as well as different dimensions of performance (i.e., task performance and organizational citizenship behaviors [OCBs]). Then, we used meta-analysis to assess the strength and consistency of gender token effects across a large number of studies.

We make several theoretical, empirical, and practical contributions in this study: First, our meta-analysis is important because endorsement of the gender token hypothesis appears to be widespread despite somewhat limited and inconsistent empirical evidence. It is important to clarify current evidence for or against the gender token hypothesis because an overreliance on the gender token hypothesis may have influenced equal employment legislation and organizational policies (Joshi, Son, & Roh, 2015). Thus, we focus on job performance as an important organizational outcome that could be affected by the gender token effect. We were motivated to meta-analytically examine job performance in this study because it has not received as much attention in the tokenism literature as other variables (e.g., laboratory study outcomes, hiring decisions). Also, evidence of token effects would be important to female employees whose performance, as well as any rewards tied to their performance, could be negatively affected.

Second, in addition to conducting overall tests of the gender token effect, we investigate variables that could moderate male–female differences in job performance and/or token effects. Specifically, we explore whether gender differences are more likely to be evident for subjective or objective measures of performance in civilian or military organizations, when single or multiple organizations are sampled, for managerial jobs versus nonmanagerial jobs, for gender stereotypes of the work across samples, when there is an administrative or research purpose for performance ratings, when studies are published versus unpublished, and according to when the studies were conducted.

Third, our study makes a theoretical contribution by testing the gender token effect, as described by the tenets of critical mass theory. Our findings have theoretical implications for how to revise the tenets of critical mass theory based on evidence regarding the conditions in which the gender token effect on performance likely exists. Overall, our study represents the largest empirical test of the gender token hypothesis to date, and has implications for both theory (e.g., testing propositions of critical mass theory and identifying boundary conditions of token effects) and practice (e.g., identifying whether minorities' performance may be disadvantaged within organizations).

Theoretical Bases Relevant to Gender Token Effects

Critical Mass Theory and the Tokenism Hypothesis

Critical mass theory and the tokenism hypothesis (Kanter, 1977a, 1977b) advocate that the proportion of males and females in an organization influences interactions among organizational members. Tokenism is an

organizational condition that can promote gender discrimination because it makes the scarcity of women salient (Heilman & Eagly, 2008). However, tokenism and gender discrimination are not synonymous because tokenism stems from the gender composition of the workplace, whereas gender discrimination stems from evaluative penalties that accompany employees "who do not possess the qualities that are stereotypical of their group" (p. 394). Thus, cultural beliefs about gender can engender discriminatory behaviors via sex categorization processes that activate descriptive and prescriptive gender stereotypes (Bobbitt-Zeher, 2011). Descriptive gender stereotypes indicate general beliefs about which traits each gender *will* possess, whereas prescriptive gender stereotypes indicate general beliefs about which traits each gender *should* possess.

According to critical mass theory, the percentage of females in an organization is viewed as a numerical characteristic of organizations, primarily at the organizational level of analysis (Gustafson, 2008; Mischel & Shoda, 2010). Kanter (1977a, 1977b) labeled work contexts in which less than 15% of workers are females as "skewed," and labeled organizational contexts with between 16% and 35% females as "tilted." The small proportion of females in skewed and tilted situations is thought to increase the use of stereotypes and contribute to the development of organizational cultures that are not supportive of females because it makes gender and gender composition of the workplace salient (Baron, 1984; Huffman, Cohen, & Pearlman, 2010; Ostroff & Atwater, 2003; Pfeffer & Davis-Blake, 1987). Specifically, skewed and tilted situations "increase evaluative bias because they exaggerate lack of fit with male sex-typed roles" (Heilman & Eagly, 2008, p. 396). This evaluative bias can affect raters' use of stereotypes in a manner that alters their subjective evaluations of females, especially in organizations where gender composition of the organization is salient because there are many more men than women (Bobbitt-Zeher, 2011).

The negative effects of gender discrimination likely downwardly bias females' performance evaluations due to mismatches between inferred female stereotypes and the requirements of male-gender-typed jobs. In contrast, critical mass theory and the tokenism hypothesis predict that gender discrimination increases as the percentage of women in organizations decreases because gender and gender stereotypes become salient. In these situations, majority group members are theorized to hold stereotypes about minorities' (e.g., females) social type, such that minorities are "entrapped" into stereotypical roles (Kanter, 1977b, p. 983). Also, in these situations, minorities not only stand out but they also may feel like they have to carry the weight of representing similar others. The aforementioned dynamics are thought to lead to substantial performance pressures for minorities (Kanter, 1977a). In fact,

researchers have long noted that the presence of token effects is "unavoidable" (e.g., Yoder, Adams, & Prince, 1983, p. 325).

Belief in tokenism appears to be fairly widespread. For example, Cleveland, Vescio, and Barnes-Farrell (2005, p. 158) noted that "stereotyping may also be particularly strong in situations where women are underrepresented (i.e., skewed environments; Kanter, 1977a)" (see also King, Hebl, George, & Matusik, 2010; Riordan, Schaffer, & Stewart, 2005; Thomas & Chrobot-Mason, 2005). Similarly, Dovido and Hebl (2005) noted that in skewed environments, minorities may encounter "a number of organizational repercussions such as decreased performance, less favorable performance appraisals, and fewer opportunities to advance" (p. 20). Thus, there is a belief that token status in organizations likely adversely affects others' subjective evaluations of performance.

Linear Token Effect

We present three theoretical perspectives, each of which argues for different expectations regarding the form and magnitude of gender token effects on performance: First, we draw from critical mass theory and the tokenism hypothesis to predict that minorities' performance increases as their representation in organizations increases. Specifically, the linear token effect perspective predicts that females' job performance increases as females comprise increasingly larger percentages of organizations because the saliency of gender, gender composition, and gender stereotypes decreases. This perspective suggests that females' job performance increases as gender composition increases from skewed to tilted positions because the mismatch between gender and gender stereotypes that is thought to downwardly bias evaluations of females' performance likely reduces. Eventually, the bias likely disappears when there are relatively even numbers of male and female employees. Then, the linear token effect perspective predicts that females' performance continues to increase to levels higher than males as the percentage of females in organizations increases to levels where females comprise larger percentages of organizations than males because jobs switch from being male gender typed to female gender typed.

Overall, the tenets of critical mass theory and the tokenism hypothesis suggest a linear relationship between percentage of females in organizations and male–female differences in outcomes (e.g., job performance). Specifically, the linear token effect perspective would be supported if females' performance increases as the percentage of females in organizations increases. This leads to the following hypothesis:

Hypothesis 1: There is a linear relationship between the proportion of females and male–female performance differences, such that females' performance relative to males' performance increases as the proportion of females in organizations increases.

Curvilinear Token Effect

Other researchers have suggested that the relationship between gender proportionality and workplace outcomes may be curvilinear (e.g., Frink et al., 2003; Laws, 1975; Pfeffer & Davis-Blake, 1987). For instance, researchers have argued that gender tokenism may occur as a result of societal norms that afford higher status to males than females (e.g., Baron, 1984; Joshi, Liao, & Jackson, 2006; Laws, 1975; Ott, 1989; Reskin, McBrier, & Kmec, 1999). Relatedly, some researchers (e.g., Kanter, 1977a) have argued for an "institutionalization" perspective of gender tokenism that contends that the norms and stereotypes of gender neutral or "women's work" present in organizations with a critical mass of females (i.e., nonskewed or tilted percentages of females) influence outcomes in a curvilinear manner as biases based on gender stereotypes begin to disappear.

The curvilinear token perspective (e.g., Laws, 1975) suggests that male–female differences in job performance will be largest (and in favor of males) in skewed situations (i.e., 1%-15% females), less in favor of males in tilted situations (16%-35% females), and begin to disappear when jobs are no longer male gender typed (i.e., not skewed or tilted). Curvilinear effects might occur because females in skewed conditions are controlled by the majority (i.e., males) and its organizational culture. Females in skewed conditions are thought to be treated as representatives of females in general rather than individual employees (Joecks, Pull, & Vetter, 2013; Kanter, 1977a), whereas females in tilted conditions are treated as individuals.

In summary, the curvilinear gender token perspective advocates that the relationship between gender proportionality and male–female differences in job performance will favor males in skewed and tilted conditions, but that gender token effects will disappear when organizations possess a critical mass of females (i.e., greater than 35% of employees in organizations; Joecks et al., 2013; Kanter, 1977a). The curvilinear token effect hypothesis would be supported if differences that favor male employees are largest in skewed settings, decrease in tilted settings, and then disappear when the proportion of females is similar to or exceeds that of males. Thus, this leads to the following competing hypothesis:

Hypothesis 2: There is a curvilinear relationship between the proportion of females and male–female differences in job performance, such that the differences favor male employees most strongly in skewed settings, decrease in tilted settings, then nullify and stabilize as the proportion of females is similar to or exceeds that of males.

No Token Effect

A third theoretical perspective suggests that the proportion of females in organizations does not affect the job performance of female employees. For example, Stone, Foster, Webster, Harrison, and Jawahar (2016) found that gender, gender proportionality, and job type failed to predict overall performance. Stone et al.'s findings challenged Heilman's (1983) "lack of fit" model, which predicts that performance ratings decline as employees' perceptions of job fit decrease. It is possible that we will obtain similarly small effect sizes in our study, which would challenge the gender token hypothesis and the "lack of fit" model.

For the current study, two theoretical positions are particularly relevant for predicting the no gender token effect: First, researchers have noted the importance of *individuating information* in organizations, which is defined as "job related behaviors and attributes of a specific individual . . . it includes, but is not limited to, knowledge, skills, abilities, personality traits, and behaviors" (McCarthy et al., 2010, p. 335). The presence of individuating information may mitigate token effects by reducing reliance on stereotypes and other forms of bias (Glick, Zion, & Nelson, 1988; Landy, 2009; Olian, Schwab, & Haberfield, 1988; Reskin, 2000). Also, individuating information can overwhelm other sources of variance related to gender (e.g., judgments of assertiveness; Locksley, Borgida, Brekke, & Hepburn, 1980).

Second, theory on diffuse and direct status cues (e.g., Berger & Webster, 2006) suggests an absence of gender token effects. *Diffuse status cues* are related to a variety of judgments because they involve distinct states (e.g., gender, ethnicity, age) with specific sets of expectations. For example, a job applicant could walk into an interview and react to the diffuse cue "female," and expect the female interviewer to be more caring and nurturing than a male interviewer. In contrast, *direct status cues* are job-relevant pieces of information. For example, the job applicant may see that the female interviewer has a MBA from Wharton, which likely cues expectations about levels of job knowledge, status, and power in the organization. Importantly, direct status cues typically outweigh diffuse cues when both are present (Shore & Thornton, 1986; W. Wood & Karten, 1986).

In summary, theory and research on individuating information and diffuse versus direct status cues predict the absence of a gender token effect. The no token effect hypothesis would be supported if the proportion of females in organizations is not associated with male–female differences in job performance. This leads to the following competing hypothesis:

Hypothesis 3: The proportion of females in organizations is not associated with male–female differences in job performance.

Potential Moderators

Studies on gender differences in performance have been conducted across a variety of contexts. It is possible that the nature of these contexts contributes to some of the discrepancies in the findings across studies. Therefore, we explore several potential moderators of the relationship between gender and job performance.

Task Performance Versus OCBs

First, we examine whether gender differences vary by performance dimension. Overall job performance is defined as "the aggregated value to the organization of the discrete behavioral episodes that an individual performs over a standard interval of time" (Motowidlo, Borman, & Schmit, 1997, p. 71). Oftentimes, overall job performance consists of both task performance and OCBs. Task performance involves behaviors that directly affect organizations' technical cores via technical processes, whereas OCBs are discretionary behaviors that improve the organizational, psychological, and/or social environments in which organizations' technical cores function (Motowidlo & Van Scotter, 1994). Thus, task performance involves job proficiency, whereas OCBs are contextual behaviors that contribute to effective functioning without being formally required by organizations (Organ, 1988). In our study, we explore whether the type of performance affects our conclusions about the gender token effect. Specifically, we conduct separate meta-analyses for (a) overall job performance (i.e., overall organizational contribution), (b) task performance, and (c) OCBs. We conduct separate analyses for task performance and OCBs due to the conceptual distinctions between these constructs. We included studies in the overall job performance analyses if they (a) reported results for overall job performance, or (b) included both task performance and OCBs, so that we could create a composite overall job performance variable.

Subjective Versus Objective Performance Measures

Researchers have noted that gender differences in job performance might be more prominent for subjective measures of performance than for objective measures (e.g., Reskin et al., 1999). Kanter (1977a) noted that stereotyping is likely to be prevalent in situations with low female representation because boundaries between groups are exaggerated. We theorize that the gender proportionality effects in skewed and tilted work contexts activate stereotypes that result in biased performance evaluations because rater discretion, biases, and stereotypes influence subjective evaluations of male and female employees. In contrast, these issues may be less likely to bias objective measures such as sales and productivity rates. Thus, we explore the possibility that we may find a different pattern of results for subjective and objective measures of performance.

Civilian Versus Military Organizations

Eagly, Karau, and Makhijani (1995) noted that there may be stronger male–female differences in military organizations than civilian organizations due to stronger gender-based stereotypes in military settings. For example, many military jobs are thought to be stereotypically male (e.g., combat, vehicle repair). In addition to gender stereotypes associated with many military jobs, some military jobs (e.g., combat-related positions) require physical abilities on which there are true sex differences that favor males (Courtright, McCormick, Postlethwaite, Reeves, & Mount, 2013). Furthermore, both Eagly et al. (1995) and Bowen et al. (2000) found that male–female differences tend to be larger when raters are primarily or entirely comprised of males, which is often the case in some military contexts. Taken together, we expect that male–female differences in job performance may differ in military organizations and civilian organizations.

Single Versus Multiple Organizations

Some studies report results from respondents in a single organization, whereas others report results from respondents across multiple organizations that may have varying gender representations. Thus, we examine whether our results differ for studies that include single versus multiple organizations.

Managerial Jobs Versus Nonmanagerial Jobs

Some prior meta-analyses of gender effects on performance did not include many (or any) managerial jobs (e.g., Sackett et al., 1991), whereas others

have demonstrated that job complexity can affect performance evaluations and rewards (e.g., Joshi, Son, & Roh, 2015). Thus, we explore whether performance differs for managerial and nonmanagerial jobs.

Stereotypes of Men's Work Versus Women's Work

Stereotyping is likely enhanced in situations with low female representation because boundaries between groups are exaggerated and members of the "in-groups" protect their status (Heilman, Wallen, Fuchs, & Tamkins, 2004; Kanter, 1977a). For example, achievements of minority individuals can be taken for granted, or individuals may have to expend extra effort to make technical skills known or have them fairly assessed to overcome stereotypes (Heilman & Okimoto, 2007; Kanter, 1977a). Stereotypes may work in a multiple-step process whereby (a) individuals are categorized by noticeable demographic characteristics (e.g., sex), (b) individuals are assigned higher or lower levels of characteristics based on their demographic category (e.g., cooperation/noncompetitiveness for females), and (c) the characteristics of the job in question are then "matched" to the individual (Davison & Burke, 2000). Stereotypes emphasize the importance of the sex stereotype of jobs, so it is possible that job stereotypes could influence our results. For example, females who work male-gender-typed jobs (e.g., construction) may be evaluated differently than females who work female-gender-typed jobs (e.g., nursing). Thus, we examine whether the stereotypes for respondents' jobs (i.e., men's work vs. women's work; Reskin & Hartmann, 1986) affect the results.

Administrative Versus Research Purpose for Performance Ratings

Prior research suggests that performance ratings obtained for administrative purposes (e.g., pay increases, promotion opportunities, retention) tend to be more lenient and potentially less accurate than performance ratings obtained for research purposes (Jawahar & Williams, 1997; Taylor & Wherry, 1951). We explore whether gender differences in performance differ based on whether the performance ratings were obtained for administrative versus research purposes.

Published Versus Unpublished Research

We examine whether conclusions about male–female performance differences and/or gender token effects vary between published and unpublished

studies because checking for publication bias (McDaniel, Rothstein, & Whetzel, 2006) in meta-analysis is a recommended practice (Aytug, Rothstein, Zhou, & Kern, 2012; Dalton, Aguinis, Dalton, Bosco, & Pierce, 2012; Kepes, Banks, McDaniel, & Whetzel, 2012). We explore whether conclusions about male–female performance differences vary between published and unpublished studies.

Year of Publication

Finally, prior meta-analytic research has examined whether the year the study conducted influenced the strength of gender effects on performance (e.g., Joshi, Son, & Roh, 2015). It is important to examine temporal effects in gender research because legislation and humanitarian efforts (e.g., campaigns for gender equality) have led to an evolution of gender equality and its manifestation in organizations over time (Joshi, Neely, Emrich, Griffiths, & George, 2015). Thus, we examine whether the decade in which the study was published is related to gender differences in performance.

Method

Literature Search

We searched numerous databases for potentially relevant studies, including PsycINFO, Business Source Complete, General Business File, Educational Resources Information Center (ERIC), Sociological Index with Full Text (SIFT), ProQuest, APA PsycNET, ScienceDirect, Dissertation Abstracts International, and Defense Technical Information Center (DTIC) Online. Also, we searched recent convention programs of the Academy of Management and the Society for Industrial and Organizational Psychology for relevant studies. Next, we checked the reference list of studies that contained data and examined the sources used by previous meta-analysts (e.g., Bowen et al., 2000; Roth et al., 2012).

Then, we searched electronically available published and in press articles on individual journals' websites (i.e., *Academy of Management Journal, Group & Organization Management, Human Performance, Human Relations, Journal of Applied Psychology, Journal of Applied Social Psychology, Journal of Business and Psychology, Journal of Business Ethics, Journal of Leadership and Organizational Studies, Journal of Organizational Behavior, Journal of Occupational Health Psychology, Journal of Management, Journal of Occupational and Organizational Psychology, Journal of Managerial Psychology, Journal of Vocational Behavior, Organizational*

Behavior and Human Decision Processes, Personnel Psychology, and *The Leadership Quarterly*). Finally, we wrote to other researchers in the area for unpublished data (e.g., McKay & McDaniel, 2006), and searched individual researchers' websites for studies.

Throughout our searches, we followed Aytug et al.'s (2012) recommendation to use broad search terms. For gender, we used the following search terms: gender, sex, female, male, token, and tokenism. For performance, we used the following terms: performance, job performance, task performance, citizenship, citizenship performance, contextual performance, extra-role, extra role, organizational citizenship behavior, OCB, effectiveness, productivity, sales, supervisor ratings, and performance evaluation.

Inclusion Criteria

We used four criteria to determine whether studies should be included in our meta-analysis: First, we only included studies that measured job performance using (a) supervisory ratings, (b) peer ratings, (c) objective performance ratings from company records, or (d) composite measures that included supervisor ratings and work samples (e.g., Shoss, Eisenberger, Restubog, & Zagenczyk, 2013). We did not include self-report ratings of job performance (e.g., Shore & Thornton, 1986), subordinate ratings of performance, or student ratings of instructors (e.g., Bowen et al., 2000).

Second, we only included studies that measured on-the-job performance. For example, we did not include experimental studies that included college students as participants (e.g., Davison & Burke, 2000; Koch, D'Mello, & Sackett, 2015). Furthermore, we did not include studies of hypothetical employees (e.g., Fogarty, Parker, & Robinson, 1998).

Third, consistent with previous research (e.g., Bowen et al., 2000; Pfeffer & Davis-Blake, 1987), we required that studies reported the percentage of females in the sample. Many of the studies we reviewed did not report the percentage of females in the sample and/or the correlation between gender and performance, so we emailed the authors of 169 studies to see if this information was available and could be provided to us. We excluded studies that "forced" the percentages of males and females to be equal (e.g., Helmick, 1987). Furthermore, we only included studies that oversampled females if the study reported the actual percentage of females within the sample (e.g., Tsui & Gutek, 1984).

Finally, we required that data came from independent samples (J. Wood, 2008). When studies reported multiple effect sizes, we used Mosier's (1943) Equation 8 to estimate a composite variable and calculate its reliability. Specifically, we formed a unit-weighted composite of overall job performance if studies reported separate effect sizes for task performance and OCBs.

Coding and Analyses

The first three authors pilot tested several iterations of the coding form to ensure that it captured all the variables as clearly as possible. Then, two authors independently coded every study. For each sample, we coded how job performance was measured, the type of organization (civilian or military), the number of organizations that were sampled (single or multiple), whether respondents held managerial jobs or not (i.e., whether respondents had supervisory responsibilities according to O*NET codes; http://www.onetonline.org; Stone et al., 2016), gender stereotypes of the respondents' occupations (men's work, women's work, or neutral/unclear per O*NET codes; Stone et al., 2016), the purpose of performance ratings (administrative, research, or unclear), whether the study was published or unpublished, and the year of each study was published. Next, we recorded information (e.g., means, standard deviations, correlations) needed to calculate male–female differences in job performance. Many studies did not report this information, so we often had to contact the studies' authors to request the information.

Finally, we measured gender proportionality using the percentage of females in the sample of each primary study. This strategy provided the best available estimate of the proportion of females in each work environment. Several studies reported the proportion of females in both the sample and the organization as a whole, and the two proportions tended to be very similar. In addition, per our inclusion criteria, we did not include studies that (a) forced certain percentages of male or female respondents, or (b) intentionally oversampled male or female respondents. Thus, the tenets of critical mass theory and our careful screening of studies justified our use of study-level gender representation while examining gender token effects.

Next, we calculated *d* statistics (Cohen, 1988) that compared the job performance of male versus female employees. The numerator was the performance mean for males minus the performance mean for females. The denominator was the "pooled" (i.e., average) standard deviation of the two groups. For example, a *d* of −0.33 would demonstrate that females were rated higher than males by one third of a pooled standard deviation.

Next, we calculated the percentage of time the coders recorded the same values. The percentage of agreement for the male–female job performance *d*s was 88% for overall performance, 93% for task performance, and 99% for OCBs. The coders agreed about the percentage of females in each sample about 90% of the time, and agreed about the sample size for each sample about 92% of the time. The level of agreement was 100% for type of organization, 88% for number of organizations, 100% for the managerial status of jobs, 88% for the gender stereotypes associated with jobs, 91% for the

purpose of performance ratings, 100% for publication status, and 100% for the year the study was published.

We conducted psychometric meta-analysis to analyze the data (Hunter & Schmidt, 2004) using the Hunter and Schmidt Meta-Analysis Program (Schmidt & Le, 2004) to run random-effects models that weighted the study results by sample size. Furthermore, we used artifact distributions of internal consistency reliability estimates from our database to correct for measurement error because very few studies reported other types of reliability estimates (e.g., interrater reliability; Aytug et al., 2012). Median reliabilities were .90 for overall job performance, .88 for task performance, and .88 for OCBs. We did not correct for range restriction because we did not study applicant populations.

Sensitivity analyses. We used the specific sample removed method (Hunter & Schmidt, 2004; Kepes, McDaniel, Brannick, & Banks, 2013) to identify and remove potentially influential outliers from the meta-analytic results (Aguinis, Gottfredson, & Joo, 2013). Specifically, we used the sample-adjusted meta-analytic deviancy (SAMD) statistic (Beal, Corey, & Dunlap, 2002; Huffcutt & Arthur, 1995) to detect outliers to remove from the sensitivity analyses. The SAMD statistic identifies outliers with regard to effect size and sample size (Aguinis et al., 2013). First, we converted r values to z values using Fisher's formula for calculating Z (Beal et al., 2002). Then, we used the Multi-Purpose Meta-Analysis program (build 4.20.09; available at http://www.frankbosco.com/) to identify the outliers in our data using the t-distribution along which the SAMD is tested (i.e., SAMD $\geq |2.58|$). We report sensitivity subanalyses to report the findings obtained without studies that could unduly bias the inferences drawn from the study due to varying effect sizes (i.e., d values) and sample sizes (i.e., n) across studies.

Results

Male–Female Differences in Job Performance

Online Appendix A provides a list of studies that were included in the analyses, and Online Appendix B provides the main coding information for each study. Correlations between the performance variables are provided in Table 1. The results in Table 1 reveal that (a) few studies examined subjective and objective performance in tandem, and (b) the correlation between task performance and OCBs was large across studies ($\rho = .69$, $SD_\rho = .16$, $k = 45$, $N = 10{,}646$).

Table 1. Correlation Matrix for Performance Variables.

Variable	1	2	3
1. Task performance	—		
P	—		
*SD*ρ	—		
K	—		
N	—		
2. Organizational citizenship behaviors		—	
P	.69	—	
*SD*ρ	.16	—	
K	45	—	
N	10,646	—	
3. Objective performance			—
P	.31	—	—
*SD*ρ	.13	—	—
K	3	0	—
N	424	0	—

Note. ρ = the population estimate that corrects the zero-order bivariate correlation for measurement and sampling error across studies; *SDr* = standard deviation of the population correlation estimate across studies; *k* = number of studies included in the analysis; *N* = total sample size of all studies included in the analysis.

The *d* statistics that reflect male–female differences in job performance are reported in Table 2. The mean observed *d* for subjective measures of overall job performance was −0.09 ($k = 158$, $N = 101{,}071$), whereas the corrected *d* was −0.10. The results demonstrate that, on average, females outperformed males by one tenth of a standard deviation. The 80% credibility interval ranged from −0.28 to 0.08, which indicated that about 80% of the values in the estimated *d* distribution were between −0.28 and 0.08. The 95% confidence interval ranged from −0.12 to −0.08, which indicated that we can be about 95% certain that the true *d* value was between −0.12 and −0.08. Finally, about 8% of the variance was due to sampling error and unreliability in measures of performance. The results for the sensitivity analysis (i.e., analyses without outliers) were similar (mean $d = -0.10$, corrected $d = -0.11$, $k = 129$, $N = 95{,}071$).

Results for task performance also demonstrated a small effect that favored females (mean $d = -0.03$, corrected $d = -0.03$, $k = 82$, $N = 21{,}491$), as did the results for OCBs (mean $d = -0.04$, corrected $d = -0.04$, $k = 67$, $N = 14{,}597$) and objective performance (mean $d = -0.03$, corrected $d = -0.03$, $k = 19$, $N = 6{,}801$). In summary, the results for overall job performance, task

Table 2. Overall Meta-Analytic Results of the Gender–Job Performance Relationship for Subjective Performance Measures.

Analysis	Mean *d*	Corrected *d*	*k*	*N*	80% credibility interval	95% confidence interval	% artifacts
Overall subjective job performance	−0.09	−0.10	158	101,071	[−0.28, 0.08]	[−0.12, −0.08]	8
Sensitivity analysis	−0.10	−0.11	129	95,071	[−0.21, 0.00]	[−0.12, −0.09]	18
Task performance	−0.03	−0.03	82	21,491	[−0.36, 0.31]	[−0.08, 0.03]	6
Sensitivity analysis	−0.04	−0.04	76	19,925	[−0.29, 0.20]	[−0.09, 0.00]	11
Organizational citizenship behaviors	−0.04	−0.04	67	14,597	[−0.35, 0.27]	[−0.10, 0.02]	8
Sensitivity analysis	−0.05	−0.05	66	14,337	[−0.34, 0.23]	[−0.11, 0.00]	9
Objective performance	−0.03	−0.03	19	6,801	[−0.24, 0.18]	[−0.10, 0.05]	10
Sensitivity analysis	0.00	0.00	15	6,041	[−0.09, 0.10]	[−0.04, 0.05]	32

Note. Mean *d* is the observed *d* across studies. Corrected *d* is corrected for unreliability in job performance. *k* is the number of samples included in the analysis. *N* is the number of participants in the samples included in the analysis. % artifacts is the percentage of variance due to the artifacts of sampling error and unreliability. A negative *d* value indicates that women outperformed men.

Table 3. Meta-Analytic Results for Moderation Analyses.

Analysis	Mean *d*	Corrected *d*	*k*	*N*	80% credibility interval	95% confidence interval	% artifacts
Civilian versus military organization							
Civilian	−0.10	−0.10	152	96,991	[−0.29, 0.08]	[−0.13, −0.08]	8
Sensitivity analysis	−0.10	−0.11	124	91,679	[−0.21, 0.00]	[−0.12, −0.09]	18
Military	−0.05	−0.06	6	4,080	[−0.22, 0.11]	[−0.16, 0.05]	10
Sensitivity analysis	−0.05	−0.06	6	4,080	[−0.22, 0.11]	[−0.16, 0.05]	10
Single versus multiple organizations							
Single organization	−0.08	−0.09	96	33,495	[−0.35, 0.17]	[−0.13, −0.05]	7
Sensitivity analysis	−0.08	−0.09	84	31,928	[−0.28, 0.11]	[−0.12, −0.05]	11
Multiple organizations	−0.10	−0.11	62	67,576	[−0.24, 0.02]	[−0.13, −0.08]	9
Sensitivity analysis	−0.10	−0.11	56	66,546	[−0.20, −0.02]	[−0.13, −0.09]	17
Managerial jobs versus nonmanagerial jobs							
Managerial jobs	−0.09	−0.10	15	39,101	[−0.21, 0.01]	[−0.15, −0.06]	6
Sensitivity analysis	−0.09	−0.10	10	35,037	[−0.11, −0.08]	[−0.11, −0.09]	68
Nonmanagerial jobs	−0.09	−0.10	143	61,970	[−0.31, 0.12]	[−0.13, −0.07]	8
Sensitivity analysis	−0.10	−0.10	124	59,242	[−0.25, 0.05]	[−0.12, −0.08]	14
Stereotypes of men's work versus women's work							
Men's work	−0.08	−0.09	30	13,137	[−0.36, 0.19]	[−0.16, −0.01]	5
Sensitivity analysis	−0.10	−0.11	27	12,774	[−0.33, 0.12]	[−0.17, −0.04]	7
Women's work	−0.03	−0.03	11	2,021	[−0.32, 0.25]	[−0.17, 0.11]	11
Sensitivity analysis	−0.03	−0.03	11	2,021	[−0.32, 0.25]	[−0.17, 0.11]	11
Neutral or unclear	−0.10	−0.10	117	85,913	[−0.26, 0.06]	[−0.13, −0.08]	9
Sensitivity analysis	−0.10	−0.10	99	83,024	[−0.21, 0.00]	[−0.12, −0.09]	18
Administrative versus research purpose							
Administrative purpose	−0.10	−0.11	20	16,724	[−0.31, 0.10]	[−0.18, −0.04]	5

(continued)

Table 3. (continued)

Analysis	Mean *d*	Corrected *d*	*k*	*N*	80% credibility interval	95% confidence interval	% artifacts
Sensitivity analysis	−0.11	−0.11	18	16,473	[−0.28, 0.06]	[−0.18, −0.05]	6
Research purpose	−0.09	−0.10	123	78,983	[−0.26, 0.07]	[−0.12, −0.07]	10
Sensitivity analysis	−0.09	−0.10	101	74,988	[−0.20, 0.00]	[−0.12, −0.08]	21
Unclear purpose	−0.12	−0.13	15	5,364	[−0.44, 0.19]	[−0.26, 0.00]	5
Sensitivity analysis	−0.17	−0.18	10	4,859	[−0.31, −0.06]	[−0.25, −0.12]	19
Published versus unpublished research							
Published	−0.09	−0.09	140	82,058	[−0.28, 0.09]	[−0.12, −0.07]	9
Sensitivity analysis	−0.09	−0.10	118	77,358	[−0.21, 0.01]	[−0.12, −0.08]	19
Unpublished	−0.12	−0.13	18	19,013	[−0.30, 0.05]	[−0.19, −0.06]	5
Sensitivity analysis	−0.12	−0.13	10	17,665	[−0.21, −0.05]	[−0.17, −0.09]	13
Year of publication							
1960-1969	0.05	0.05	1	90	[0.05, 0.05]	[−0.15, 0.26]	—
Sensitivity analysis	0.05	0.05	1	90	[0.05, 0.05]	[−0.15, 0.26]	—
1970-1979	−0.26	−0.27	1	772	[−0.27, −0.27]	[−0.33, −0.20]	—
Sensitivity analysis	−0.26	−0.27	1	772	[−0.27, −0.27]	[−0.33, −0.20]	—
1980-1989	−0.25	−0.26	11	2,669	[−0.55, 0.02]	[−0.40, −0.13]	8
Sensitivity analysis	−0.24	−0.25	10	2,599	[−0.51, 0.01]	[−0.38, −0.12]	9
1990-1999	−0.10	−0.11	15	6,406	[−0.23, 0.02]	[−0.16, −0.05]	21
Sensitivity analysis	−0.11	−0.11	13	6,030	[−0.18, −0.05]	[−0.15, −0.08]	48
2000-2009	−0.09	−0.10	60	46,772	[−0.27, 0.07]	[−0.14, −0.07]	8
Sensitivity analysis	−0.11	−0.12	40	42,106	[−0.18, −0.05]	[−0.13, −0.10]	32
2010-2017	−0.06	−0.06	68	27,806	[−0.28, 0.16]	[−0.10, −0.02]	8
Sensitivity analysis	−0.07	−0.07	60	26,819	[−0.22, 0.08]	[−0.10, −0.04]	16

Note. Mean *d* is the observed *d* across studies. Corrected *d* is corrected for unreliability in job performance. *k* is the number of samples included in the analysis. *N* is the number of participants in the samples included in the analysis. % artifacts is the percentage of variance due to the artifacts of sampling error and unreliability. A negative *d* value indicates that women outperformed men.

performance, OCBs, and objective performance were all similar, and indicated small effects of gender on performance that favored females.

Next, we examined several different moderators that could affect the magnitude of our results. The results are presented in Table 3. First, we found that the results for civilian (mean $d = -0.10$, corrected $d = -0.10$, $k = 152$, $N = 96{,}991$) and military (mean $d = -0.05$, corrected $d = -0.06$, $k = 6$, $N = 4{,}080$) organizations were similar. Also, we found that the results for samples with a single organization (mean $d = -0.08$, corrected $d = -0.09$, $k = 96$, $N = 33{,}495$) were similar to the results for samples with multiple organizations (mean $d = -0.10$, corrected $d = -0.11$, $k = 62$, $N = 67{,}576$). Next, we found that the results for samples with managerial jobs (mean $d = -0.09$, corrected $d = -0.10$, $k = 15$, $N = 39{,}101$) were similar to the results for samples with nonmanagerial jobs (mean $d = -0.09$, corrected $d = -0.10$, $k = 143$, $N = 61{,}970$). Then, we found that the results for samples with work perceived as traditionally male (mean $d = -0.08$, corrected $d = -0.09$, $k = 30$, $N = 13{,}137$), work perceived as traditionally female (mean $d = -0.03$, corrected $d = -0.03$, $k = 11$, $N = 2{,}021$), and work with neutral or unclear (mean $d = -0.10$, corrected $d = -0.10$, $k = 117$, $N = 85{,}913$) gender stereotypes were similar.

Also, we found that the results for administrative (mean $d = -0.10$, corrected $d = -0.11$, $k = 20$, $N = 16{,}724$), research (mean $d = -0.09$, corrected $d = -0.10$, $k = 123$, $N = 78{,}983$), and unclear (mean $d = -0.12$, corrected $d = -0.13$, $k = 15$, $N = 5{,}364$) purposes were similar. Furthermore, we found similar results for published (mean $d = -0.09$, corrected $d = -0.09$, $k = 140$, $N = 82{,}058$) and unpublished (mean $d = -0.12$, corrected $d = -0.13$, $k = 18$, $N = 19{,}013$) studies. Finally, we found little evidence of temporal effects on our findings because the results for studies published in the 1990s (mean $d = -0.10$, corrected $d = -0.11$, $k = 15$, $N = 6{,}406$), 2000s (mean $d = -0.09$, corrected $d = -0.10$, $k = 60$, $N = 46{,}772$), and 2010s (mean $d = -0.06$, corrected $d = -.06$, $k = 68$, $N = 27{,}806$) were similar. Importantly, the 95% confidence intervals overlapped between subgroups for all of the moderator analyses reported above. Thus, the potential moderators we examined did not have a strong influence on the obtained results.

Gender Token Effect

The results in Table 4 provide tests of the three theoretical perspectives we presented earlier in this study: the linear token effect hypothesis (Hypothesis 1), the curvilinear token effect hypothesis (Hypothesis 2), and the no token effect hypothesis (Hypothesis 3). To test the competing hypotheses, we computed mean and corrected male–female *d*s for several gender proportion intervals noted in the tokenism literature, including 1% to 15% female (i.e., skewed

Table 4. Meta-Analytic Results of the Gender–Job Performance Relationship by Female Representation for Subjective Measures of Job Performance.

Analysis	Mean *d*	Corrected *d*	*k*	*N*	80% credibility interval	95% confidence interval	% artifacts
Overall subjective performance							
1%-15% females	−0.12	−0.13	15	13,266	[−0.35, 0.08]	[−0.22, −0.04]	4
Sensitivity analysis	−0.17	−0.17	8	11,750	[−0.17, −0.17]	[−0.19, −0.16]	100
16%-35% females	−0.10	−0.11	37	44,932	[−0.23, 0.01]	[−0.14, −0.08]	10
Sensitivity analysis	−0.10	−0.11	27	42,630	[−0.17, −0.05]	[−0.13, −0.09]	23
36%-50% females	−0.04	−0.04	29	12,970	[−0.31, 0.22]	[−0.12, 0.04]	6
Sensitivity analysis	−0.07	−0.08	24	11,814	[−0.27, 0.12]	[−0.14, −0.01]	9
51%-65% females	−0.03	−0.03	36	8,830	[−0.20, 0.14]	[−0.08, 0.02]	21
Sensitivity analysis	−0.03	−0.03	36	8,830	[−0.20, 0.14]	[−0.08, 0.02]	21
66%-85% females	−0.13	−0.14	35	19,969	[−0.30, 0.03]	[−0.18, −0.09]	10
Sensitivity analysis	−0.12	−0.12	29	18,947	[−0.24, −0.01]	[−0.16, −0.09]	17
86%-99% females	0.00	0.00	6	1,104	[−0.31, 0.31]	[−0.20, 0.20]	9
Sensitivity analysis	0.00	0.00	6	1,104	[−0.31, 0.31]	[−0.20, 0.20]	9
Task performance							
1%-15% females	0.14	0.16	9	3,748	[−0.35, 0.67]	[−0.10, 0.42]	2
Sensitivity analysis	0.14	0.16	9	3,748	[−0.35, 0.67]	[−0.10, 0.42]	2
16%-35% females	0.04	0.04	14	3,868	[−0.14, 0.22]	[−0.04, 0.12]	16
Sensitivity analysis	0.04	0.04	14	3,868	[−0.14, 0.22]	[−0.04, 0.12]	16
36%-50% females	−0.04	−0.05	14	4,094	[−0.30, 0.21]	[−0.16, 0.06]	9
Sensitivity analysis	−0.05	−0.05	12	3,792	[−0.22, 0.11]	[−0.13, 0.03]	18
51%-65% females	−0.08	−0.08	27	6,021	[−0.31, 0.15]	[−0.15, −0.01]	14
Sensitivity analysis	−0.06	−0.06	26	5,856	[−0.26, 0.13]	[−0.13, 0.00]	17
66%-85% females	−0.17	−0.18	15	3,112	[−0.50, 0.13]	[−0.31, −0.06]	8
Sensitivity analysis	−0.17	−0.18	15	3,112	[−0.50, 0.13]	[−0.31, −0.06]	8
86%-99% females	−0.07	−0.07	3	648	[−0.36, 0.21]	[−0.34, 0.19]	10
Sensitivity analysis	−0.07	−0.07	3	648	[−0.36, 0.21]	[−0.34, 0.19]	10

(continued)

Table 4. (continued)

Analysis	Mean *d*	Corrected *d*	*k*	*N*	80% credibility interval	95% confidence interval	% artifacts
Organizational citizenship behaviors							
1%-15% females	0.25	0.27	2	661	[0.21, 0.34]	[0.17, 0.38]	53
Sensitivity analysis	0.25	0.27	2	661	[0.21, 0.34]	[0.17, 0.38]	53
16%-35% females	−0.07	−0.07	9	3,109	[−0.35, 0.20]	[−0.22, 0.07]	7
Sensitivity analysis	−0.07	−0.07	9	3,109	[−0.35, 0.20]	[−0.22, 0.07]	7
36%-50% females	−0.01	−0.01	13	2,506	[−0.27, 0.24]	[−0.13, 0.10]	14
Sensitivity analysis	−0.05	−0.05	12	2,362	[−0.24, 0.15]	[−0.14, 0.05]	21
51%-65% females	−0.02	−0.02	21	4,276	[−0.34, 0.30]	[−0.13, 0.09]	8
Sensitivity analysis	−0.06	−0.07	20	4,016	[−0.32, 0.18]	[−0.16, 0.02]	13
66%-85% females	−0.12	−0.13	19	3,397	[−0.44, 0.18]	[−0.24, −0.02]	10
Sensitivity analysis	−0.12	−0.13	19	3,397	[−0.44, 0.18]	[−0.24, −0.02]	10
86%-99% females	0.03	0.04	3	648	[−0.27, 0.35]	[−0.25, 0.33]	8
Sensitivity analysis	0.03	0.04	3	648	[−0.27, 0.35]	[−0.25, 0.33]	8
Objective performance							
1%-15% females	−0.05	−0.05	4	807	[−0.37, 0.27]	[−0.31, 0.20]	7
Sensitivity analysis	−0.05	−0.05	4	807	[−0.37, 0.27]	[−0.31, 0.20]	7
16%-35% females	−0.17	−0.17	5	582	[−0.45, 0.11]	[−0.38, 0.03]	15
Sensitivity analysis	−0.12	−0.12	4	551	[−0.16, −0.08]	[−0.21, −0.03]	90
36%-50% females	0.02	0.02	2	4,080	[0.02, 0.02]	[−0.01, 0.05]	100
Sensitivity analysis	0.02	0.02	2	4,080	[0.02, 0.02]	[−0.01, 0.05]	100
51%-65% females	−0.39	−0.39	3	400	[−0.57, −0.21]	[−0.57, −0.21]	22
Sensitivity analysis	−0.39	−0.39	3	400	[−0.57, −0.21]	[−0.57, −0.21]	22
66%-85% females	0.02	0.02	5	932	[−0.17, 0.21]	[−0.13, 0.16]	19
Sensitivity analysis	0.02	0.02	5	932	[−0.17, 0.21]	[−0.13, 0.16]	19
86%-99% females	—	—	0	0	—	—	—
Sensitivity analysis	—	—	0	0	—	—	—

Note. Mean *d* is the observed *d* across studies. Corrected *d* is corrected for unreliability in job performance. *k* is the number of samples included in the analysis. *N* is the number of participants in the samples included in the analysis. % artifacts is the percentage of variance due to the artifacts of sampling error and unreliability. A negative *d* value indicates that women outperformed men.

settings for females), 16% to 35% female (i.e., tilted settings for females), 36% to 50% female, 51% to 65% female, 66% to 85% female (tilted settings for males), and 86% to 99% female (skewed settings for males) categories.

The corrected *d*s for the main overall subjective job performance analyses ranged from −0.14 to 0.00. The effects were consistently small and negative across categories. Also, all 80% credibility intervals included 0.00. Interestingly, the results demonstrated that females in skewed (i.e., 1%-15% females; mean $d = -0.12$, corrected $d = -0.13$, $k = 15$, $N = 13{,}266$) and tilted (i.e., 16%-35% females; mean $d = -0.10$, corrected $d = -0.11$, $k = 37$, $N = 44{,}932$) conditions were rated more favorably than males. The small and negative effect sizes indicate that females tend to outperform males in the two conditions in which females are thought to be most disadvantaged.

The corrected *d*s for the main task performance analyses ranged from −0.18 to 0.16. Next, the corrected *d*s for the main OCB analyses ranged from −0.13 to 0.27. Finally, the corrected *d*s for the main objective performance analyses ranged from −0.39 to 0.02. Similar to the results for overall subjective job performance, there was generally no markedly different pattern of results across gender proportionality levels, and most 80% credibility intervals included 0.00. The main exception to the overall trend in findings was the small performance advantage males had relative to females for skewed conditions for task performance (mean $d = 0.14$, corrected $d = 0.16$, $k = 9$, $N = 3{,}748$). A similar effect was found for OCBs (mean $d = 0.25$, corrected $d = 0.27$, $k = 2$, $N = 611$), though we caution that this effect was based on only two primary studies.

Overall, the above results provide support for the no token effect hypothesis (Hypothesis 3), rather than the linear token effect hypothesis (Hypothesis 1) or the curvilinear token effect hypothesis (Hypothesis 2). Specifically, the negative *d*s in skewed and tilted conditions and the similar *d*s across conditions all demonstrate support for the no token effect hypothesis.

Discussion

Theory and research in various literatures suggest that being a minority (e.g., female) in a mostly majority environment can negatively affect minorities' performance. This meta-analysis provides a large-scale test of the effects of gender tokenism on job performance. Results reveal that females consistently score higher than males on measures of overall job performance, task performance, OCBs, and objective measures of performance, regardless of gender representation in organizations. Importantly, the results hold for key conditions where tokenism has been theorized to occur (i.e., skewed and tilted conditions). Thus, the results challenge the commonly held belief that female employees demonstrate lower job performance than males when

females represent a smaller proportion of organizations' workforces. Instead, corrected mean differences between male and female employees were in the range of −0.03 to −0.10 across types of performances (see Table 2), and were generally similar across various features of study designs (i.e., moderators; see Table 3) and gender representations (see Table 4).

Implications for Theory and Practice

Our results have important implications for theory and practice in areas such as gender differences and gender discrimination. First, we did not find support for a linear gender token effect (e.g., Kanter, 1977a, 1977b) or a curvilinear gender token effect (e.g., Laws, 1975). The results are important because we tested linear, curvilinear, and no token effect theoretical perspectives using data from a variety of organizations, jobs, and measures of job performance. As such, tokenism does not appear to affect the job performance of female employees, as has been theorized and widely purported. Thus, our results suggest that certain aspects of theories that predict or imply token effects (e.g., critical mass theory) should be revisited.

However, the results certainly do not mean that scholars should discontinue researching gender or other minority group differences in organizational settings. We focused on measures of job performance as an important criterion, but gender differences could be evident in other types of personnel decisions, such as reward decisions (Joshi, Son, & Roh, 2015) and promotion opportunities (Hoobler, Masterson, Nkomo, & Michel, 2016). Also, it is possible that token effects may be more prevalent for other demographic characteristics (e.g., race/ethnicity) than they are for gender.

Our meta-analytic findings also have implications for practitioners (Le, Oh, Shaffer, & Schmidt, 2007): First, the results reiterate that it is important for managers to have as much information as possible when assessing subordinates, so that managers can rely more on relevant job-performance-related information than on stereotypes and biases. Prior research has noted that job-performance-related information is important for employment interviews (Goldberg & Cohen, 2004; McCarthy et al., 2010), and we reiterate its importance for performance appraisals. Working to ensure job-performance-relevant information is available may be particularly important in settings where managers may not have direct contact with their workers on a regular basis (e.g., police work per Landy, Farr, Saal, & Freytag, 1976; see also Cascio & Aguinis, 2011). Furthermore, our findings demonstrate that managers can expect male and female employees to demonstrate similar levels of performance for most jobs. Such expectations may, in and of themselves, begin to counter the potential influence of gender-based stereotypes.

Although our results demonstrate that females tend to demonstrate similar levels of job performance to males, females may be disadvantaged on other outcomes. For example, research has found that females are often paid less than males for doing similar work, and females are underrepresented in managerial positions (e.g., Lyness & Schrader, 2006; Ostroff & Atwater, 2003). Furthermore, Joshi, Son, and Roh (2015) recently found meta-analytic evidence that differences in rewards for males and females (d = .56) are much larger (and favor males) than the differences in job performance between males and females that we found, such that males are rewarded disproportionately more than females. Stereotypes may also impede females' ability to attain promotions (Olsen, Parsons, Martins, & Ivanaj, 2016; Schwab, Werbel, Hofmann, & Henriques, 2015). Thus, organizational leaders should be mindful of gender-related stereotypes and biases when making these and other personnel decisions. Ultimately, awareness of gender-related stereotypes and biases can improve gender relations in organizations, and help organizations avoid claims of gender discrimination (Leslie & Gelfand, 2008).

Limitations

Like all studies, our study has limitations that readers should consider when drawing conclusions from the results (Brutus, Aguinis, & Wassmer, 2013). First, we did not have access to true scores of job performance. Thus, we could not directly test for gender bias within the measures of job performance. However, a similar pattern of results for subjective and objective performance measures may alleviate potential concerns about this issue.

Second, we relied on study-level percentages of gender representation to serve as a proxy for organizational-level percentages of gender representation. Although we did not find information in the studies that indicated that the samples had markedly different gender representations than respondents' organizations, we could not determine the level of consistency between study- and organizational-level gender representations. This issue may be particularly relevant to studies that included participants from multiple organizations.

Third, although the overall results are based on a very large number of studies (k = 220) and participants (N = 119,528), results for some subanalyses are based on modest numbers of primary studies. As a result, we recommend caution while interpreting some of our ancillary findings.

Fourth, we made corrections for unreliability based on internal consistency estimates. Such corrections are likely conservative because they do not account for rater error, which has been found to be large in subjective measures of performance (e.g., Viswesvaran, Ones, & Schmidt, 1996). The limitations noted above are common for meta-analyses, in that the results are limited by the data

and information available in the underlying primary studies (Hunter & Schmidt, 2004). Finally, the high correlation between task performance and OCBs reported in Table 1 ($\rho = .69$, $SD_\rho = .16$, $k = 45$, $N = 10{,}646$) suggests that a latent factor (e.g., overall subjective performance or overall organizational contribution) may account for the similar findings we obtained for task performance and OCBs. Thus, we recommend that inferences from our study account for the high correlation between these variables in our study.

Agenda for Future Research

Below, we provide an actionable agenda for future research that describes opportunities to make immediate and incremental contributions to theory and research based on our findings (Carlson & Ji, 2011). One interesting area for future research involves job performance raters' cognitive processes that drive implicit judgments of performance for token groups (Biernat, 2003; Biernat & Kobrynowicz, 1997; Biernat, Manis, & Nelson, 1991). For example, males and females might demonstrate the same level of performance on job tasks that require physical ability, but females might be judged to be more capable because they compare more favorably with other females than do the males with other males. Also, future research could examine how individuating information (i.e., trait-based individuating features such as behaviors, personality, social roles, and traits; Pratto & Bargh, 1991) affects raters of job performance. For example, individuating information may attenuate or amplify tokenism effects on job performance. Thus, we encourage future primary studies and meta-analyses to explore this important area of inquiry.

Also, it is possible that raters make conscious efforts to control information not related to performance when they evaluate job performance (e.g., Hilton & Fein, 1989). For example, people may simultaneously incorporate job-related individuating and categorical information (e.g., group membership) as they attempt to mitigate biases and stereotypes that jointly influence evaluations of job applicants and/or current employees (e.g., Kunda & Thagard, 1996). Indeed, there is evidence that individuals can reduce stereotype processes (Hilton & Fein, 1989). Thus, future research could explore the extent to which managers consider nonperformance information when evaluating their subordinates' job performance, as well as how this information affects managers' gender stereotypes and biases.

Next, we encourage future research to examine performance management systems as a potential boundary condition of the relationship between gender proportionality and job performance. For example, future research could examine how criteria for reward decisions (e.g., promotions and compensation) contribute to tokenism effects. Organizations that develop effective performance

management systems generally report stronger financial performance than other organizations (Cascio, 2006), so performance management systems and the various components within them are important areas for future research.

Another important area for future research involves the cognitive processes and behaviors of performance *ratees*. Although we did not find support for linear or curvilinear gender token effects, issues related to assessments of male and female workers deserve further investigation. For example, females may perceive greater performance pressures and fewer networking opportunities than their male counterparts. If so, how do females react to such conditions? Do females work extra hours, develop goal attainment strategies, increase persistence of effort, or search for mentors (e.g., King et al., 2010)? Specifically, how do the cognitive processes of female workers manifest in behavior that contributes to their performance relative to males?

In addition, future research could investigate the possible role of organizational factors that contribute to gender differences in job performance. For example, an "up or out" organizational climate (e.g., some accounting and law firms) could trigger feelings of competitiveness and have negative effects that result in decreasing proportions of minority groups (Milkman & McGinn, 2010). Perhaps organizational efforts to promote diversity and diversity climate (Van Dijk, Van Engen, & Van Knippenberg, 2012) could decrease gender differences in competitive organizational climates. We encourage future research to examine whether and how prodiversity initiatives and climates affect performance differently at the individual, team, and organizational levels (Joshi & Roh, 2009).

Conclusion

We conducted a large-scale meta-analytic test of the gender token effect on job performance. This study's results challenge the widely held belief that females' performance is lower than their male counterparts' performance when females comprise a small proportion of organizations. Instead, our findings demonstrate a consistent, small effect of gender on job performance across measures of overall subjective job performance, task performance, OCBs, and objective performance that favors female employees instead of male employees. The results were relatively consistent across various levels of gender representation and study design features. Yet, critical questions remain. Thus, we encourage future research to examine the nature, magnitude, and consistency of gender token effects on organizational outcomes.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplementary material

Supplementary material is available for this article online.

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*, 270-301.

Aytug, Z. G., Rothstein, H. R., Zhou, W., & Kern, M. C. (2012). Revealed or concealed? Transparency of procedures, decisions, and judgment calls in meta-analyses. *Organizational Research Methods*, *15*, 103-133.

Baron, J. N. (1984). Organizational perspectives on stratification. *Annual Review of Sociology*, *10*, 37-69.

Beal, D. J., Corey, D. M., & Dunlap, W. P. (2002). On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in the meta-analysis of correlations. *Journal of Applied Psychology*, *87*, 583-589.

Berger, J., & Webster, M. (2006). Expectations, status, and behavior. In P. Burke (Ed.), *Contemporary social psychological theories* (pp. 268-300). Stanford, CA: Stanford University Press.

Biernat, M. (2003). Toward a broader view of social stereotyping. *American Psychologist*, *58*, 1019-1027.

Biernat, M., & Kobrynowicz, D. (1997). Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology*, *72*, 544-557.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*, 485-499.

Bobbitt-Zeher, D. (2011). Gender discrimination at work: Connecting gender stereotypes, institutional policies, and gender composition of the workplace. *Gender & Society*, *25*, 764-786.

Bowen, C. C., Swim, J. K., & Jacobs, R. R. (2000). Evaluating gender biases on actual job performance of real people: A meta-analysis. *Journal of Applied Social Psychology*, *30*, 2194-2215.

Brutus, S., Aguinis, H., & Wassmer, U. (2013). Self-reported limitations and future directions in scholarly reports: Analysis and recommendations. *Journal of Management*, *39*, 48-75.

Carlson, K. D., & Ji, F. X. (2011). Citing and building on meta-analytic findings: A review and recommendations. *Organizational Research Methods*, *14*, 696-717.

Cascio, W. F. (2006). *Managing human resources: Productivity, quality of work life, profits*. Boston, MA: McGraw-Hill.

Cascio, W. F., & Aguinis, H. (2011). *Applied psychology in human resource management* (7th ed.). Boston, MA: Prentice Hall.

Cleveland, J. N., Vescio, T. K., & Barnes-Farrell, J. L. (2005). Gender discrimination in organizations. In R. Dipboye & A. Colella (Eds.), *Discrimination at work: The psychological and organizational bases* (pp. 149-176). Mahwah, NJ: Lawrence Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Courtright, S. H., McCormick, B. W., Postlethwaite, B. E., Reeves, C. J., & Mount, M. K. (2013). A meta-analysis of sex differences in physical ability: Revised estimates and strategies for reducing differences in selection contexts. *Journal of Applied Psychology*, *98*, 623-641.

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology*, *65*, 221-249.

Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior*, *56*, 225-248.

Dovido, J. F., & Hebl, M. R. (2005). Discrimination at the level of the individual: Cognitive and affective factors. In R. Dipboye & A. Colella (Eds.), *Discrimination at work: The psychological and organizational bases* (pp. 11-36). Mahwah, NJ: Lawrence Erlbaum.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, *117*, 125-145.

Fogarty, T. J., Parker, L. M., & Robinson, T. (1998). Where the rubber meets the road: Performance evaluation and gender in large public accounting organizations. *Women in Management Review*, *13*, 299-231.

Frink, D. D., Robinson, R. K., Reithel, B., Arthur, M. M., Ammeter, A. P., Ferris, G. R., . . . Morrisette, H. S. (2003). Gender demography and organization performance: A two-study investigation with convergence. *Group & Organization Management*, *28*, 127-147.

Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology*, *55*, 178-186.

Goldberg, C., & Cohen, D. J. (2004). Walking the walk and talking the talk: Gender differences in the impact of interviewing skills on applicant assessments. *Group & Organization Management*, *29*, 369-384.

Gustafson, J. L. (2008). Tokenism in policing: An empirical test of Kanter's hypothesis. *Journal of Criminal Justice*, *36*, 1-10.

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 5, pp. 269-298). Greenwich, CT: JAI Press.

Heilman, M. E., & Eagly, A. H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology*, *1*, 393-398.

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology*, *92*, 81-92.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, *89*, 416-427.

Helmick, T. A. (1987). *The effect of gender, job classification and self-esteem upon job performance* (Unpublished master's thesis). Central Missouri State University, Warrensburg.

Hilton, J. L., & Fein, S. (1989). The role of typical diagnosticity in stereotype-based judgments. *Journal of Personality and Social Psychology*, *57*, 201-211.

Hoobler, J. M., Masterson, C. R., Nkomo, S. M., & Michel, E. J. (2016). The business case for women leaders: Meta-analysis, research critique, and path forward. *Journal of Management*, *44*, 2473-2499. doi:10.1177/0149206316628643

Huffcutt, A. I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, *80*, 327-334.

Huffman, M. L., Cohen, P. N., & Pearlman, J. (2010). Engendering change: Organizational dynamics and workplace gender desegregation, 1975-2005. *Administrative Science Quarterly*, *55*, 255-277.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting for error and bias in research findings* (2nd ed.). Newbury Park, CA: SAGE.

Jawahar, I. M., & Williams, C. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, *50*, 905-992.

Joecks, J., Pull, K., & Vetter, K. (2013). Gender diversity in the boardroom and firm performance: What exactly constitutes a "critical mass?" *Journal of Business Ethics*, *118*, 61-72.

Joshi, A., Liao, H., & Jackson, S. E. (2006). Cross-level effects of workplace diversity on sales performance and pay. *Academy of Management Journal*, *49*, 459-481.

Joshi, A., Neely, B., Emrich, C., Griffiths, D., & George, G. (2015). Gender research in AMJ: An overview of five decades of empirical research and calls to action. *Academy of Management Journal*, *58*, 1459-1475.

Joshi, A., & Roh, H. (2009). The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal*, *52*, 599-627.

Joshi, A., Son, J., & Roh, H. (2015). When can women close the gap? A meta-analytic test of sex differences in performance and rewards. *Academy of Management Journal*, *58*, 1516-1545.

Kanter, R. M. (1977a). *Men and women of the corporation*. New York, NY: Basic Books.

Kanter, R. M. (1977b). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology*, *82*, 965-999.

Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, *15*, 624-662.

Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS (the meta-analytic reporting standards). *Journal of Business and Psychology*, *28*, 123-143.

King, E. B., Hebl, M. R., George, J. M., & Matusik, S. F. (2010). Understanding tokenism: Antecedents and consequences of a psychological climate of gender inequity. *Journal of Management*, *36*, 482-451.

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, *100*, 128-161.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*, 284-308.

Landy, F. J. (2009). Performance ratings: Then and now. In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 227-248). New York, NY: Routledge.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, *87*, 72-107.

Landy, F. J., Farr, J. L., Saal, F. E., & Freytag, W. R. (1976). Behaviorally anchored scales for rating the performance of police officers. *Journal of Applied Psychology*, *61*, 750-758.

Laws, J. L. (1975). The psychology of tokenism: An analysis. *Sex Roles*, *1*, 51-67.

Le, H., Oh, I. S., Shaffer, J., & Schmidt, F. (2007). Implications of methodological advances for the practice of personnel selection: How practitioners benefit from meta-analysis. *Academy of Management Perspectives*, *21*(3), 6-15.

Leslie, L. M., & Gelfand, M. J. (2008). The who and when of internal gender discrimination claims: An interactional model. *Organizational Behavior and Human Decision Processes*, *107*, 123-114.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, *5*, 821-831.

Lyness, K. S., & Schrader, C. A. (2006). Moving ahead or just moving? An examination of gender differences in senior corporate management appointments. *Group & Organization Management*, *31*, 651-676.

McCarthy, J. M., Van Iddekinge, C. H., & Campion, M. A. (2010). Are highly structured job interviews resistant to demographic similarity effects? *Personnel Psychology*, *63*, 325-359.

McDaniel, M. A., Rothstein, H. R., & Whetzel, D. L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology*, *59*, 927-953.

McKay, P. F., & McDaniel, M. A. (2006). A reexamination of Black-White mean differences in work performance: More data, more moderators. *Journal of Applied Psychology*, *91*, 538-554.

Milkman, K. L., & McGinn, K. L. (2010, August). *Will I stay or will I go: Cooperative and competitive effects of relational demography on turnover*. Paper presented at the annual meeting of the Academy of Management, Montreal, Québec, Canada.

Mischel, W., & Shoda, Y. (2010). The situated person. In B. Mesquita, L. Barrett, & E. Smith (Eds.), *The mind in context* (pp. 149-173). New York, NY: The Guilford Press.

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika*, *8*, 161-168.

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, *10*, 71-83.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, *79*, 475-480.

Nieva, V. F., & Gutek, B. A. (1980). Sex effects on evaluation. *Academy of Management Review*, *5*, 267-276.

Olian, J. D., Schwab, D. P., & Haberfield, Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations: A meta-analysis of experimental studies. *Organizational Behavior and Human Decision Processes*, *41*, 180-195.

Olsen, J. E., Parsons, C. K., Martins, L. L., & Ivanaj, V. (2016). Gender diversity programs, perceived potential for advancement, and organizational attractiveness: An empirical examination of women in the United States and France. *Group & Organization Management*, *41*, 271-309.

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.

Ostroff, C., & Atwater, L. E. (2003). Does whom you work with matter? Effects of referent group gender and age composition on managers' compensation. *Journal of Applied Psychology*, *88*, 725-774.

Ott, E. M. (1989). Effects of the male-female ratio at work. *Psychology of Women Quarterly*, *13*, 41-57.

Pfeffer, J., & Davis-Blake, A. (1987). The effect of the proportion of women on salaries: The case of college administrators. *Administrative Science Quarterly*, *32*, 1-24.

Pratto, F., & Bargh, J. A. (1991). Stereotyping based on apparently individuating information: Trait and global components of sex stereotypes under attention overload. *Journal of Experimental Social Psychology*, *27*, 26-47.

Reskin, B. F. (2000). The proximate causes of employment discrimination. *Contemporary Sociology*, *29*, 319-328.

Reskin, B. F., & Hartmann, H. I. (1986). *Women's work, men's work*. Washington, DC: The National Academies Press.

Reskin, B. F., McBrier, D. B., & Kmec, J. A. (1999). The determinants and consequences of workplace sex and race composition. *Annual Review of Sociology*, *25*, 335-361.

Riordan, C. M., Schaffer, B. S., & Stewart, M. M. (2005). Relational demography within groups: Through the lens of discrimination. In R. Dipboye & A. Colella (Eds.), *Discrimination at work: The psychological and organizational bases* (pp. 37-62). Mahwah, NJ: Lawrence Erlbaum.

Roth, P. L., Purvis, K. L., & Bobko, P. (2012). A meta-analysis of gender group differences for measures of job performance in field studies. *Journal of Management*, *38*, 719-739.

Sackett, P. R., Dubois, C. L. Z., & Noe, A. W. (1991). Tokenism in performance evaluation: The effects of work group representation on male-female and white-black differences in performance ratings. *Journal of Applied Psychology*, *76*, 263-267.

Schmidt, F. L., & Le, H. (2004). Software for the Hunter-Schmidt meta-analysis methods. Iowa City: Department of Management and Organizations, University of Iowa.

Schwab, A., Werbel, J. D., Hofmann, H., & Henriques, P. L. (2016). Managerial gender diversity and firm performance: An integration of different theoretical perspectives. *Group & Organization Management*, *41*, 5-31.

Shore, L. M., & Thornton, G. C. (1986). Effects of gender on self- and supervisory ratings. *Academy of Management Journal*, *29*, 115-129.

Shoss, M. K., Eisenberger, R., Restubog, S. L. D., & Zagenczyk, T. J. (2013). Blaming the organization for abusive supervision: The roles of perceived organizational support and supervisor's organizational embodiment. *Journal of Applied Psychology*, *98*, 158-168.

Stone, T. H., Foster, J., Webster, B., Harrison, J., & Jawahar, I. M. (2016). Gender differences in supervisors' multidimensional performance ratings: Large sample evidence. *Human Performance*, *29*, 428-446.

Taylor, E. K., & Wherry, R. J. (1951). A study of leniency in two rating systems. *Personnel Psychology*, *4*, 39-47.

Thomas, K. M., & Chrobot-Mason, D. (2005). Group-level explanations of workplace discrimination. In R. Dipboye & A. Colella (Eds.), *Discrimination at work: The psychological and organizational bases* (pp. 63-88). Mahwah, NJ: Lawrence Erlbaum.

Tsui, A. S., & Gutek, B. A. (1984). A role set analysis of gender differences in performance, affective relationships, and career success of industrial middle managers. *Academy of Management Journal*, *27*, 619-635.

Van Dijk, H., Van Engen, M. L., & Van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, *119*, 38-53.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, *81*, 557-574.

Wood, J. (2008). Methodology for dealing with duplicate effects in a meta-analysis. *Organizational Research Methods*, *11*, 79-95.

Wood, W., & Karten, S. J. (1986). Sex difference in interaction style as a product of perceived differences in competence. *Journal of Personality and Social Psychology*, *50*, 341-347.

Yoder, J. D., Adams, J., & Prince, H. T. (1983). The price of a token. *Journal of Political and Military Sociology*, *11*, 325-337.

Associate Editor: Orlando Richard
Submitted Date: November 20, 2016
Revised Submission Date: August 11, 2017
Acceptance Date: August 14, 2017

Author Biographies

Jeremy D. Mackey is an assistant professor of Management in the Raymond J. Harbert College of Business at Auburn University. He received his PhD from Florida State University. His current research interests include abusive supervision, interpersonal mistreatment, and meta-analysis.

Philip L. Roth is a professor of Management in the College of Business at Clemson University. He received his PhD from the University of Houston. His current research interests include organizational staffing, gender issues in organizations, the impact of social media in organizations, and meta-analysis.

Chad H. Van Iddekinge is the Bank of America professor of Management in the College of Business at Florida State University. He received his PhD from Clemson University. His current research interests include employee selection, the measurement and prediction of job performance, measure development and validation, and meta-analysis.

Lynn A. McFarland is an assistant professor of Management in the Darla Moore School of Business at the University of South Carolina. She received her PhD from Michigan State University. Her current research interests include staffing, social influence in organizations, and workplace diversity.