

Algner, Mona; Fay, Doris; Lorenz, Timo

Article — Published Version

Validation and Measurement Invariance of the German and Spanish Gender Bias Scale for Women Leaders

Journal of Business and Psychology

Provided in Cooperation with:

Springer Nature

Suggested Citation: Algner, Mona; Fay, Doris; Lorenz, Timo (2024) : Validation and Measurement Invariance of the German and Spanish Gender Bias Scale for Women Leaders, Journal of Business and Psychology, ISSN 1573-353X, Springer US, New York, NY, Vol. 40, Iss. 1, pp. 155-178, <https://doi.org/10.1007/s10869-024-09946-7>

This Version is available at:

<https://hdl.handle.net/10419/317822>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



Validation and Measurement Invariance of the German and Spanish Gender Bias Scale for Women Leaders

Mona Algner^{1,2} · Doris Fay¹ · Timo Lorenz²

Accepted: 27 March 2024 / Published online: 2 May 2024
© The Author(s) 2024

Abstract

Gender bias still appears to contribute significantly to the gender disparity observed in upper leadership positions. The present study presents the translation, modification, and validation of the Gender Bias Scale for Women Leaders (GBSWL, Diehl et al., 2020) into German and Spanish. Using data from four samples of full-time employed women from Germany and Spain with and without leadership responsibility ($N=870$), we conducted confirmatory factor analyses to establish factorial validity, tested measurement invariance across the different job levels and countries, and tested construct validity. Our results indicate that the original factor structure does not hold in Germany and Spain. Therefore, utilizing the German leader sample as a construction sample, we modified the factor structure and validated the modified version using the remaining three samples. The modified version demonstrated good model fit, had metric measurement invariance across all samples, and resulted in a correlational pattern consistent with theory and the original study. Overall, results suggest that the German and Spanish versions of the GBSWL are reliable and valid instruments that scholars and practitioners can use to advance theory, research, and human resource practice in Germany and Spain.

Keywords Gender bias scale for women leaders · Confirmatory factor analysis · Measurement invariance analysis · Validation · HR management

The Role of Gender Bias for the Ongoing Gender Disparity in Leadership Positions

The *Global Gender Gap Report 2022* by the World Economic Forum (2022) reports that worldwide, only 33% of senior leadership positions are held by women. The report also estimates that, at the current pace, it will take 151 years to close the global gender gap regarding economic participation and opportunity. One of the main factors contributing to the ongoing gender disparity specifically in the upper

leadership echelons is gender bias (Hentschel et al., 2019; Jones et al., 2016; Kossek et al., 2017).

Early research on gender stereotypes traces these biases back to labor division by gender. The division of labor results in shared beliefs about how women and men *are* (descriptive gender norms) and *should be* (injunctive gender norms), which manifests in different social roles ascribed to women and men, respectively (i.e., homemaker vs. breadwinner; e.g., Eagly, 1987). The role congruity theory (RCT; Eagly & Karau, 2002) posits that the perceived incongruity between the (self-) ascribed female gender role and leadership roles leads to two forms of prejudice: (1) when considered as potential leaders, women are evaluated as less favorable than men; and (2) when exercising leadership behaviors, women experience backlash (i.e., are evaluated negatively when exhibiting typical leader behaviors). This is because male and leader stereotypes share agentic, that is assertive or dominant, qualities (e.g., “think manager, think male”; Schein, 1973) whereas *female* and *leader stereotypes* have less overlap. In contrast, women ought to be warm, sensitive, and helpful (i.e., present more communal qualities; Eagly, 1987). If women (in leadership positions)

Additional supplementary materials may be found here by searching on article title <https://osf.io/collections/jbp/discover>.

✉ Mona Algner
mona.algner@uni-potsdam.de

¹ Work and Organizational Psychology, University of Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany

² Department of Psychology, MSB Medical School Berlin, Berlin, Germany

exhibit agentic behaviors, they violate injunctive gender stereotypes and thus provoke backlash such as being punished, disliked, or rated less favorably (Eagly & Karau, 2002). Furthermore, women who act in accordance with their ascribed gender role (i.e., present more communal attributes) are perceived as not being fit for leadership roles. This is due to the discrepancy between the communal qualities ascribed to women and leadership prototypes, which are traditionally more masculine (Koenig et al., 2011).

Gender Bias as a Multifaceted Phenomenon

One mechanism through which the above-described phenomena unfold in the workplace is gender bias. Gender bias appears to be a multifaceted phenomenon: it manifests at different levels of analysis, with disadvantageous processes ranging from stereotypical societal views on gender to organizational structures and interpersonal processes (Eagly & Karau, 2002; Lyness & Grotto, 2018). At the organizational level, women can encounter gender bias in terms of male-dominated senior leadership, which influences organizational norms and culture (Diehl et al., 2020; Lyness & Grotto, 2018). From the perspective of Kanter's tokenism theory (Kanter, 1977), being a woman in a male-dominated organizational culture leads to the subjective experience of tokenism (i.e., increased salience of gender, social isolation, and visibility), which in turn negatively impacts organizational outcomes such as women's job satisfaction or turnover intention (King et al., 2010). At the individual level, women can face gender bias in the form of intrapersonal (e.g., self-limited career aspirations; Diehl et al., 2020) and interpersonal barriers (e.g., experiencing backlash when presenting agentic leadership behaviors; Hogue & Lord, 2007; Williams & Tiedens, 2016).

In addition to its manifestation at different levels of analysis, women experience gender bias in varying degrees of explicitness, ranging from traditional, overt forms of sexism (e.g., harassment) to more subtle, implicit gender biases. For example, men are more likely than women to emerge as leaders when engaging in agentic or communal task-oriented statements, suggesting that women evoke a different reaction than men to the same communicative behaviors (Schlamp et al., 2020).

The Gender Bias Scale for Women Leaders

Given the complexity of gender bias, research seeking to advance our understanding of its consequences requires measurement instruments that reflect its multifaceted nature. To specifically consider the above-named aspects, i.e., degree of explicitness and level of manifestation, Diehl et al. (2020) developed and validated the Gender Bias Scale for Women Leaders (GBSWL). The purpose of this instrument

is to assess both overt and subtle gender biases at multiple levels of analysis, integrating societal, organizational, and individual perspectives on gender bias.

The instrument consists of six constructs: *Male Privilege*, *Disproportionate Constraints*, *Insufficient Support*, *Devaluation*, *Hostility*, and *Acquiescence*. Each construct is represented by a hierarchical second-order factor model, comprising two to three first-order factors (see Fig. 1). To facilitate a better understanding of the model, we briefly introduce the first-order factors ordered by the second-order factors. Male privilege refers to a system of advantages in organizational environments that favors men based on their gender. The items of the corresponding first-order factors assess whether women are more likely to be promoted into high-risk positions (*glass cliff*), the extent to which leadership perceptions reflect masculine stereotypes (*male culture*), and the extent of formal and informal requirements placed on both spouses, even though only one is employed by the organization (*two-person career structure*). Disproportionate constraints refer to the constraints imposed on women's voices and choices and to the extent to which performance standards applied to women are different from those used for men. More specifically, the items assess the constraints placed on women, such as carefully managing one's communication to avoid being perceived as either too assertive or too submissive (*constrained communication*), choosing a career that is considered suitable for women rather than following one's interests (*constrained career choices*), and experiencing intense scrutiny of one's job performance (*unequal standards*). Insufficient support refers to the lack of support provided to women in navigating an organizational culture that predominantly favors men. Items in this construct assess whether women are excluded from (in-) formal organizational events (*exclusion*), and their access to organizational resources and support such as mentoring (*lack of mentoring*) and sponsorship (*lack of sponsorship*). Devaluation refers to the lack of recognition and appreciation of women's work contributions. It assesses devaluing organizational practices, such as interrupting women while they are speaking or refusing support for their ideas (*lack of acknowledgment*), as well as paying women less than men in equal positions (*salary inequality*). Hostility refers to hostile behaviors against women at work exercised by both men and women and to hindrances, such as being held back by other women (*self-group distancing*¹ or verbal abuse and sexual

¹ Following the suggestion of Derks and colleagues (e.g., Derks et al., 2016; Faniko et al., 2022), we use the term self-group distancing for the phenomenon previously described as "queen bee syndrome" to emphasize that the process of self-distancing from other women is not a natural tendency of women, but rather a "behavioral strategy prompted by (implicit) organizational definitions of success that are couched in masculine terms" (Faniko et al., 2021, p. 395).

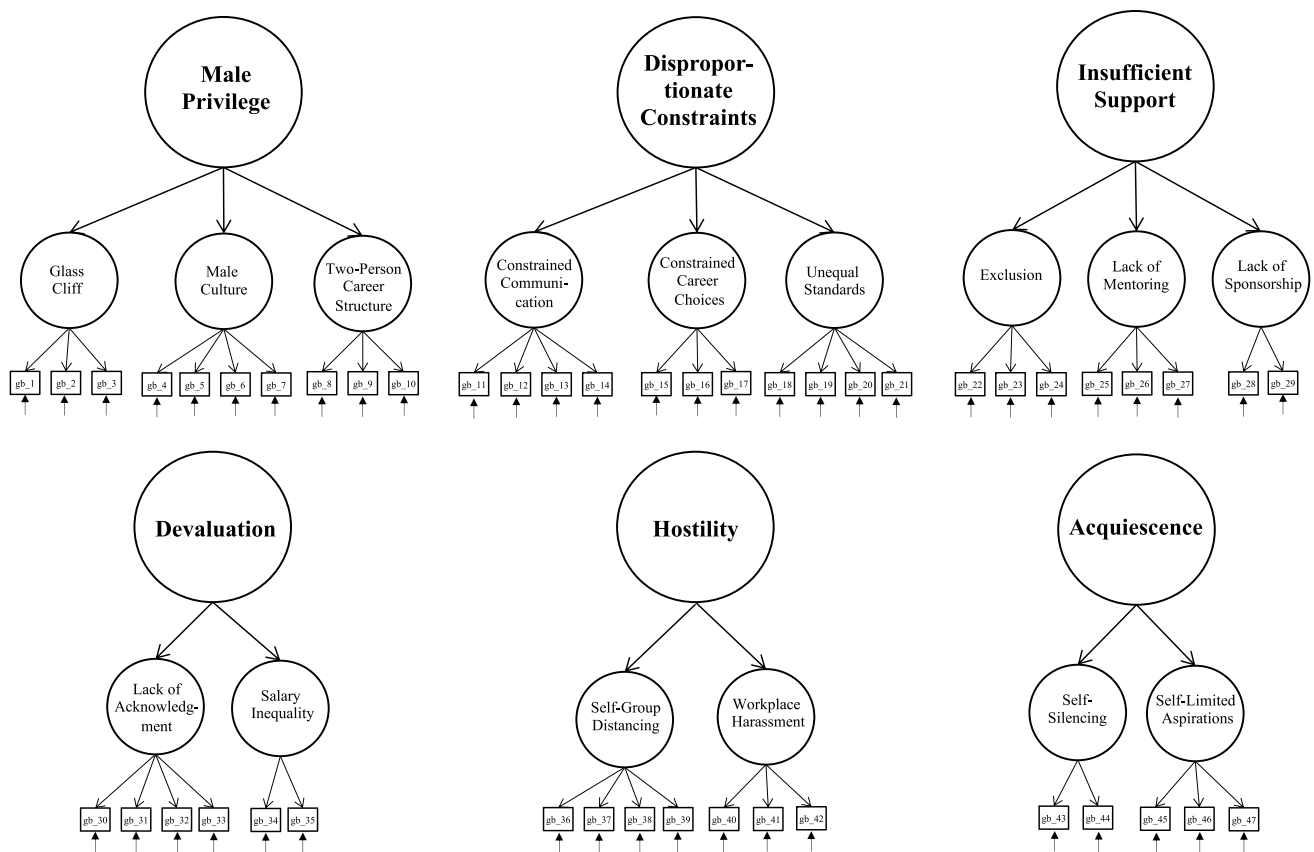


Fig. 1 Original factor structure of the Gender Bias Scale for Women Leaders. The first-order factor self-group distancing was previously termed “Queen Bee Syndrome”

harassment (*workplace harassment*). Finally, acquiescence refers to women’s seeming acceptance of their “place” in a male organizational environment. Items assess intraindividual barriers to women, such as not speaking up about women’s rights or challenges at work (*self-silencing*) and the psychological glass ceiling of limiting one’s career aspirations (*self-limited aspirations*).

Aim of this Study

Gender disparity exists in numerous countries (i.e., World Economic Forum, 2022), and even though progress in terms of equality has been made, disparity is still a tangible and measurable issue. In order to further advance the field, scholars called for new and improved measures that link implicit and explicit gender biases (Kossek et al., 2017). Furthermore, its (almost) ubiquitous existence resulted in calls for cross-cultural research (Diehl et al., 2020).

However, research striving to understand commonalities and differences in the biases that female leaders encounter across the globe needs assessment tools that are culturally invariant, because only then meaningful comparisons between cultures can be made. Having the goal of

measurement invariance in mind, the primary aim of this study is therefore to translate and validate an instrument that assesses the extent of gender bias faced by female leaders. To this end, we drew on the GBSWL by Diehl et al. (2020), which was originally developed in the English language and in the North American context and translated and tested it in two European countries: Germany and Spain. We specifically chose these countries based on, first, their similar patterns of culture values and practices as reported in the Global Leadership and Organizational Behavior Effectiveness (GLOBE) study (House et al., 2004), second, based on the assumption that gender bias may have particularly strong negative impacts in those countries (Triana et al., 2019), and third, because there is a widespread societal and academic interest in this topic.

First, rigorous quantitative cross-cultural research requires instruments that are invariant across cultures, i.e., those that measure the same construct in different cultures and languages. The likelihood of achieving measurement invariance appears to be higher when focusing on cultures that are fairly similar. Therefore, we chose Germany and Spain, two cultures with many similarities. Across the different cultural dimensions, Germany and Spain exhibit

similarities in their societal culture values and practices. Most relevant to the present study is their similarity with regard to *gender egalitarianism*, both in terms of their societal culture *value* and their societal culture *practice* (Hanges & Gupta, 2004). Given their cultural similarities, we assume that we do not need to establish a differential measure but hope that the GBSWL can serve as an instrument that will be applicable in a similar fashion in both Germany and Spain.

Second, it is meaningful to examine gender bias where its impact is likely to have particularly strong negative consequences. This appears to be especially the case in cultures exhibiting a larger discrepancy between culture values and practices. In their meta-analysis, Triana et al. (2019) extended relative deprivation theory to show that perceived gender discrimination is more strongly (negatively) associated with employee outcomes (e.g., anxiety and job performance) in countries with high gender egalitarianism. They argued that individuals in countries with high gender egalitarianism have a lower threshold for tolerating gender inequity; thus, being confronted with gender discrimination has a greater impact on them because of a larger relative sense of deprivation (Triana et al., 2019).

This finding makes Germany and Spain promising countries to study gender bias. Germany and Spain rank in the upper 12% on the Global Gender Gap Index (i.e., having higher gender equality), but a closer look at the sub-indices indicates that both countries exhibit mediocre gender equality in terms of economic participation and opportunity (Germany, 75th rank; Spain, 64th rank). Thus, while on a global scale, Germany and Spain are relatively gender equal (e.g., culture values, women's health, and educational attainment), there remains a disparity when it comes to women in leadership positions in politics and industry (World Economic Forum, 2022). This disparity suggests that despite the presence of gender egalitarian values, there are underlying factors that contribute to the persistence of gender inequality in these areas. Since the cultural values of gender egalitarianism in Germany and Spain are high but actual gender equality is only mediocre, it seems plausible that women in these countries experience a greater sense of deprivation.

A third reason for focusing on Germany and Spain is the substantial interest in the topic, both within society and academia. Notably, both countries have adopted Feminist Foreign Policies (Zilla, 2022), further emphasizing their commitment to promoting gender equality. In 2022, the European Union approved a directive to promote gender diversity in leadership roles across its member states (European Commission, 2022). Germany had already implemented a gender quota for board directors, and Spain is currently exploring similar initiatives, underlining a broader trend toward gender equality in corporate governance across Europe.

Similarly, a literature search in the Web of Science revealed 1763 articles on this topic from these two countries.

The number of articles increased after 2017, with 1463 of the 1763 articles being published after that year. This increase coincides with the popularization of the #MeToo movement, which gained significant traction after a post by the American actress Alyssa Milano encouraging women to reply “Me too” if they had been sexually harassed or assaulted went viral (Milano, 2017). The #MeToo movement was originally started by Tarana Burke in 2006 (Me too. Movement, 2023), but it gained widespread attention and participation after Alyssa Milano's post. The social media movement has sparked discussions and actions worldwide, leading to increased awareness of issues related to gender inequality and sexual harassment. This development seems to be reflected in an increased academic interest, as evidenced by the fact that most studies on this topic have been published after 2017.

We believe that it is vital to provide psychometrically sound instruments that may help uncover the reasons why gender inequality persists, even in the face of gender egalitarian values. By focusing on Germany and Spain, we hope to provide an instrument that in the future will help to shed light on the factors that contribute to gender inequality in two of the largest economies in the European Union (Eurostat, 2023). Our research provides a tool that will be useful in developing a better understanding of the underlying causes of gender inequality, which in turn can aid in developing effective strategies to promote gender equality in these and other countries. Thus, we translate the original GBSWL developed by Diehl et al. (2020) from English to German (GBSWL-G) and Spanish (GBSWL-S), test the original factor structure, and adjust it to two European leadership samples.

The second purpose of this study is to test its usefulness for female non-leaders. Whereas Diehl et al. (2020) originally developed an instrument for female leaders, we propose that developing a measure that captures bias experienced by female *non*-leaders is also a timely task. Women are still less likely to emerge as leaders (see Badura et al., 2018 for a meta-analysis). One reason could be that women without leadership responsibility also experience gender bias. These biases may function as barriers to pursuing and obtaining leadership positions, which could explain the “leaky” leadership pipeline (i.e., at each stage of their career women leave talent pipelines at a higher rate than men, resulting in their underrepresentation at higher levels). Following a recent call for research by Shen and Joseph (2021) to better understand the under-emergence of women in leadership positions, the second purpose of our study is thus to test whether the proposed (or modified) scale is appropriate for German and Spanish women who *do not* have any leadership responsibility. This is important because it would allow researchers to use the GBSWL to explore potential factors that keep women from reaching leadership positions. We therefore also investigate the factor structure

in two non-leadership samples. Finally, we provide evidence of the construct validity of the instrument. Following Diehl et al. (2020), we test the relation between several gender biases and outcomes relevant to organizations, that is, turnover intention and job satisfaction.

Research question 1: Does the GBSWL have a similar factor structure in Germany (GBSWL-G) and Spain (GBSWL-S) as in the original study?

Research question 2: Can results from the GBSWL-G and GBSWL-S be compared in cross-national research, in that both scales are measurement equivalent?

Research question 3: Does the factor structure of the GBSWL-G and GBSWL-S also hold for non-leaders?

Construct Validity of the Gender Bias Scale for Women Leaders

To establish external validity, we examine the relation of different gender bias factors with two relevant organizational outcome variables: job satisfaction and turnover intention. Job satisfaction has been associated with several beneficial organizational outcomes, such as increased performance, higher organizational effectiveness, and less withdrawal behavior, such as absenteeism or turnover intention (Judge et al., 2017, 2020). We investigate whether experiencing gender bias affects women's job satisfaction. Furthermore, we explore the relation of gender bias with turnover intention, another widely researched topic in management and organizational science (see Hom et al., 2017). Gallup estimates that voluntary turnover costs organizations approximately half to two times the employee's annual salary (McFeely & Wigert, 2019). Hence, fostering a better understanding of the drivers of turnover intention is in most organizations' interests.

Theoretical Background and Hypotheses

We use the conservation of resources theory (COR; Hobfoll, 1989; Hobfoll et al., 2018) and person-environment fit theory (e.g., Edwards et al., 1998) as our overarching frameworks to explain the relations between gender bias and turnover intention on the one hand, and job satisfaction on the other.

Conservation of Resources

In its basic tenet, COR theory posits that individuals strive to obtain, retain, foster, and protect resources they deem valuable (Hobfoll, 1989; Hobfoll et al., 2018). Hobfoll (2002) broadly defines resources as those entities that are either valued in their own right (e.g., self-esteem, health, or inner peace) or are instrumental in acquiring valuable outcomes (e.g., money,

social support). According to COR theory, a stress reaction is elicited when resources are threatened by potential loss, when they are actually lost, or when fostering resources is not possible despite considerable effort (Hobfoll et al., 2018).

We argue that experiencing workplace gender bias acts as a barrier to women's career advancement, thus preventing them from fostering resources. More specifically, the gender bias facets of male culture (i.e., organizationally cultivated masculine leadership prototypes and male gatekeeping), unequal standards (i.e., holding women to higher performance standards), and self-group distancing (i.e., women in leadership positions fail to support or deliberately block job opportunities for junior women) reflect an imbalance between the gain of resources and invested effort.

Furthermore, gender bias threatens women's well-being (a centrally valued resource; Hobfoll et al., 2018) and motivates them to leave situations in which it occurs (i.e., the workplace or organization). We argue that workplace harassment (i.e., intimidating behavior, verbal abuse, or sexual harassment) reflects either a threat to or an actual loss of (psychological) safety and well-being. Following the assumptions of the COR theory, women should be motivated to remove themselves from situations that threaten the conservation or further accumulation of resources, and as a result, should intend to leave the organization.

Empirical studies provide support for the assumption that perceived gender inequality (e.g., organizational practices and procedures that favor men) increases women's turnover intention (King et al., 2010). In line with research indicating that women leaders are often perceived as less effective (Koch et al., 2015) and that women frequently face the need to receive higher ratings on performance evaluations to secure promotions (Lyness & Heilman, 2006), these are factors likely to contribute to women's turnover intention. With regard to self-group distancing and workplace harassment, previous research suggests that both hostile work behaviors increase women's intention to leave (Derks et al., 2016; Diehl et al., 2020; Jones et al., 2016; Raver & Nishii, 2010). Thus, we hypothesize as follows:

Hypothesis 1: (a) Male culture, (b) unequal standards, (c) workplace harassment, and (d) self-group distancing are positively correlated with turnover intention.

Person–Environment-Fit Models

Person–environment fit models propose that a (perceived) fit between an individual and their work environment (e.g., job, organization, team, or supervisor) positively impacts job attitudes, such as job satisfaction (Kristof-Brown et al., 2005). Within this research domain, different forms of fit exist: demands-abilities fit, and needs-supplies fit. Demands-abilities fit describes how well an individual's abilities (e.g.,

skills, training, or aptitude) match the demands (e.g., qualitative and quantitative requirements of the job and group and organizational norms) of a specific job (Edwards et al., 1998). Needs–supplies fit entails the congruence between an individual's needs (e.g., biological and psychological requirements, values, and motives to achieve desired outcomes) and the supplies (e.g., salary, social involvement, and opportunities to achieve desired outcomes) that an organization provides (Edwards et al., 1998). We argue that gender bias decreases women's job satisfaction because of both *perceived* demands–abilities misfit and needs–supplies misfit. More specifically, we assume that the gender bias facets male culture and unequal standards reduce job satisfaction because of women's underrepresentation in the upper leadership echelon (i.e., male culture), and having to work harder to obtain the same acknowledgment as men (i.e., unequal standards) emphasizes the incongruence between the social role of women and leadership positions. These experiences could lead to a perceived misfit, which, in turn, decreases job satisfaction. Furthermore, we presume that the gender bias facets of lack of mentoring, lack of acknowledgment, workplace harassment, and self-group distancing reflect a form of needs-supplies misfit, in that they (a) impede opportunities to achieve career advancement (i.e., lack of mentoring, lack of acknowledgment and self-group distancing) and (b) pose a threat to the physical and psychological safety of women (i.e., workplace harassment).

Empirical evidence supports our assumptions. For example, King et al. (2010) show that perceived inequality in the workplace (e.g., male culture or unequal standards) is associated with lower job satisfaction. In contrast, being recognized at work and receiving mentoring increases job satisfaction (Fowler et al., 2021; Pfister et al., 2020); thus, lack of acknowledgment and lack of mentoring are likely to reduce job satisfaction. Furthermore, meta-analytic evidence suggests that workplace gender discrimination in general is associated with low job satisfaction (Jones et al., 2016). In particular, women who experience hostility from other women (e.g., self-group distancing) report a negative affective reaction that reduces their job satisfaction (Gabriel et al., 2018). Hence, we hypothesize as follows:

Hypothesis 2: (a) Male culture, (b) unequal standards, (c) lack of mentoring, (d) lack of acknowledgment, (e) workplace harassment, and (f) self-group distancing are negatively correlated with job satisfaction

Method

Participants and Procedure

Using an online panel provider (Talk online panel), we recruited a total sample of $N=926$ German and Spanish

women with and without leadership responsibility (Germany, leadership responsibility, $n_1=252$; Germany, no leadership responsibility, $n_2=212$; Spain, leadership responsibility, $n_3=252$; Spain, no leadership responsibility, $n_4=210$). Online panel data has been associated with questionable data quality and threats to validity resulting from, e.g., inattentiveness or careless responding (Aguinis et al., 2021; Meade & Craig, 2012). However, meta-analytic evidence showed that online panel data and conventionally sourced data produced similar results with regard to psychometric properties and criterion validity, suggesting that, with appropriate caution, online panel data are a suitable source for applied psychological research (Walter et al., 2019). Therefore, we followed best-practice recommendations in the planning, implementation, and data-cleaning stages (Aguinis et al., 2021; Curran, 2016; Meade & Craig, 2012; Walter et al., 2019).

We invited participants based on self-identified attributes (e.g., gender, leadership responsibility, and native speakers) and screened out those who did not meet our inclusion criteria. Respondents were paid at least a minimum wage, with compensation ranging from 3.45€ (employees without leadership responsibility) to 8.81€ (employees with leadership responsibility) for 15 min of survey time. The aim was to calibrate the compensation in a way that motivated, but not over-motivated respondents (Walter et al., 2019). Furthermore, we tracked participants' anonymous identification numbers to ensure that each respondent would participate only once. Following Aguinis et al. (2021) and Meade and Craig (2012), we implemented two instructed-response items to check for inattentiveness. Participants were screened out if they failed to correctly answer any one of the instructed-response items. They could not restart the survey and were thus not included in the analyses.

In the data-cleaning stage, we applied several methods to detect careless and inattentive responding (i.e., response time, long-string analysis, and Mahalanobis distance). We first examined the time each respondent needed to complete the survey and, following Huang et al.'s (2012) suggestion, removed participants with an average response time of < 2 s per item, since at this speed a thorough completion of the survey seemed unlikely (removed participants from each subsample: Germany: leadership responsibility, $n=24$; no leadership responsibility, $n=14$; Spain: leadership responsibility, $n=10$; no leadership responsibility, $n=8$). We then applied long-string analysis to check for the longest sequential string of the same response. Finally, we computed the Mahalanobis distance to flag potential outliers for deeper examination (Curran, 2016). Neither analysis revealed any further suspicious cases. In total, we removed 6% of the participants. The remaining total sample size was $N=870$, with the following subsample sizes: Germany: leadership responsibility, $n_1=228$; no leadership responsibility, $n_2=198$;

Spain: leadership responsibility, $n_3 = 242$; no leadership responsibility, $n_4 = 202$). We used a forced-choice answer format; hence, no data were missing.

Sample

Demographic Variables

Demographic variables for all samples are presented in Table 1.

Measures

Gender Bias

Gender bias was assessed using the Gender Bias Scale for Women Leaders (Diehl et al., 2020). For the present study, the original English questionnaire was translated into German and Spanish, using a back-translation procedure (Guillemin et al., 1993). Participants rated 47 items such as “My job performance has been scrutinized more closely than that of my male colleagues.” or “The behavior of my male coworkers has sometimes made me feel uncomfortable.” The items had either a response format assessing participants’ agreement ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) or a response format that assessed the frequency with which they had experienced the described event (1 = *never* to 5 = *always*).

Turnover Intention

We assessed women’s intention to leave in the same way as Diehl et al. (2020) using a single item with a dichotomous answering format. We asked participants whether they intend to leave their position within the next 12 months (yes/no).

Job Satisfaction

Like Diehl et al. (2020), we assessed job satisfaction with an adapted version of the Job Satisfaction Subscale of the Michigan Organizational Assessment Questionnaire (Camman et al., 1979). On a rating scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), the participants rated their job satisfaction with three items (e.g., “In general, I am satisfied with my job.”).

Data Analysis

We first tested the original factor structure (Diehl et al., 2020). Following Diehl et al. (2020), we tested each of the

six second-order factor models separately. If the original GBSWL model fit the data reasonably well, we proceeded with the analysis. In the case of a non-acceptable model fit, we analyzed local fit to obtain information on how to improve the measurement model. After having established a well-fitting model, we continued the analyses by testing measurement invariance across samples. If modifications had to be made, to simplify the procedure, we used the German sample with leadership responsibility as the construction sample and subsequently validated the solution in the remaining samples.

Model Specification and Evaluation of Model Fit

Diehl et al. (2020) conceptualized gender bias as a model with 15 latent first-order factors (e.g., glass cliff), which then loaded on six second-order factors (e.g., male privilege). In their study, the authors did not report any constraints imposed on their model, which is necessary for model identification. In our study, the model was identified by fixing the first unstandardized first- and second-order loadings of an indicator or latent variable, respectively, to 1 and factor means to 0. All cross-loadings were constrained to zero.

We estimated the measurement models using confirmatory factor analyses (CFA), run with the R package ‘lavaan’ (version 0.6–9; Rosseel, 2012). Previous research suggested that maximum likelihood estimators performed slightly better than diagonally weighted least squares when the sample size is small (i.e., $N = 200$) or—as is the case with the GBSWL—when five or more answering categories are used (Li, 2016). To account for non-normal distribution, we used a robust maximum likelihood estimator (MLR). Model fit was evaluated using a Satorra–Bentler-scaled χ^2 -test (Satorra & Bentler, 2001) as well as criteria proposed by Hu and Bentler (1999). They comprised a standardized root mean square residual ($\text{SRMR} \leq 0.08$ in combination with at least one of the following fit indices: a root-mean-square error of approximation ($\text{RMSEA} \leq 0.06$, a lower bound of the 90% confidence interval (CI) of the $\text{RMSEA} \leq 0.06$, a comparative fit index ($\text{CFI} \geq 0.95$, or a Tucker–Lewis index ($\text{TLI} \geq 0.95$).

Evaluation of Measurement Invariance

To investigate measurement invariance between German- and Spanish-speaking samples as well as between samples with and without leadership responsibility, we ran multi-group CFA using the R packages ‘lavaan’ (version 0.6–9; Rosseel, 2012) and ‘semTools’ (version 0.5–6; Jorgensen et al., 2022). We analyzed measurement invariance between (a) leaders from both countries, (b) non-leaders from both countries, (c) leaders and (d) non-leaders in each country,

Table 1 Demographic variables for all samples

Variable	Germany		Spain	
	Leadership responsibility	No leadership responsibility	Leadership responsibility	No leadership responsibility
<i>M (SD)</i>				
Age ^a	42.73 (10.50)	46.14 (12.03)	38.33 (10.27)	42.91 (10.73)
Working experience ^a	17.26 (11.25)	20.51 (12.93)	13.45 (8.95)	16.20 (10.08)
Weekly hours	39.40 (6.92)	36.97 (5.13)	39.57 (5.98)	38.07 (4.56)
Tenure at current organization ^a	10.68 (8.07)	9.33 (9.25)	8.60 (7.89)	8.83 (9.02)
Absolute (relative) frequency				
Education				
Master's degree or higher	90 (39.47%)	28 (14.15%)	106 (43.80%)	71 (35.15%)
Bachelor's degree	34 (14.91%)	15 (7.58%)	35 (14.46%)	45 (22.27%)
Apprenticeship	79 (34.65%)	98 (49.49%)	15 (6.20%)	24 (11.88%)
Secondary education, high school, or a different degree	24 (10.53%)	56 (28.28%)	48 (19.48%)	46 (22.77%)
Did not graduate	1 (0.35%)	1 (0.51%)	38 (15.70%)	16 (7.92%)
Permanently employed	190 (83.00%)	165 (81.68%)	174 (71.9%)	151 (74.75%)
Sectors ^b	Social, health, and education (<i>n</i> = 56, 24.56%); commercial services, commerce, distribution, and tourism (<i>n</i> = 61, 30.08%); social, health, and education (<i>n</i> = 48, 21.05%); construction, architecture, and engineering (<i>n</i> = 44, 19.30%)	Commercial services, commerce, distribution, and tourism (<i>n</i> = 61, 30.08%); social, health, and education (<i>n</i> = 49, 24.75%); accounting, law, and administration (<i>n</i> = 28, 14.15%)	Commercial services, commerce, distribution, and tourism (<i>n</i> = 62, 25.62%); accounting, law, and administration (<i>n</i> = 59, 24.38%); social, health, and education (<i>n</i> = 56, 23.14%)	Commercial services, commerce, distribution, and tourism (<i>n</i> = 53, 26.24%); social, health, and education (<i>n</i> = 59, 29.21%); accounting, law, and administration (<i>n</i> = 37, 18.32%)

N = 870; Germany (leaders) *n* = 228, Germany (non-leaders) *n* = 198; Spain (leaders) *n* = 242, Spain (non-leaders) *n* = 202

^aResults in years

^bSectors refer to the three sectors most participants worked in

and (e) between all groups. To determine which level of measurement invariance was achieved, we calculated the chi-square difference test, as well as changes in CFI and RMSEA at different levels of increasingly constrained model parameters. According to Putnick and Bornstein (2016) and Chen (2007), a difference of <0.01 for CFI and <0.015 for RMSEA indicated an acceptable relative fit. Corresponding to the four hierarchical levels of measurement invariance (Meredith, 1993), the four models we tested were (1) configural, (2) metric (or weak factorial), (3) strong (or scalar factorial), and (4) strict invariance.

Configural invariance means that the factorial structure of the measurement model is the same in different groups, reflecting the construct's theoretical consistency (Flake & Luong, 2021). The next level of measurement invariance, metric invariance, assumes equality of the factorial structure and adds an equality constraint to the factor loadings of the indicators. This means that the linear relation between each indicator and its respective latent factor is equal across groups. In addition to the equality of the factor structure and factor loadings, strong invariance assumes equality of item intercepts across groups. Achieving strong invariance allows researchers to compare observed or latent scores without running the risk of biased results (Chen, 2007; Flake & Luong, 2021). The fourth and strictest level of invariance, strict invariance, is achieved when the factorial structure, factor loadings, intercepts, and residual variances are equal across groups. Strict invariance implies that the unexplained variance for each item is equal across groups, indicating identical measurement at the item level of the construct with equal reliability (Flake & Luong, 2021).

Evaluation of Reliability and External Validity

We used McDonald's ω_i and Cronbach's α as the estimates of the total reliability of a test. We chose to report McDonald's ω_i as an estimate of the total reliability of a test in addition to Cronbach's α since the latter is only an appropriate measure of reliability under rather strict assumptions (e.g., tau-equivalent models, independence of error terms; see McNeish, 2018; Revelle & Condon, 2019). These assumptions are often violated, and many methodologists have argued against using Cronbach's α since it only assesses a fraction of the reliability concept (i.e., internal consistency) and underestimates population reliability (Cortina et al., 2020; McNeish, 2018). However, we still present Cronbach's α to facilitate comparisons with results from Diehl et al. (2020). We computed Pearson's correlation coefficients with relevant organizational outcome measures (i.e., job satisfaction and turnover intention), to evaluate the convergent and divergent validity of the GBSWL-G and GBSWL-S.

Correlations are evaluated as follows: >0.1 , small; >0.3 , moderate; and >0.5 , strong.

Results

Owing to space constraints, we only report the results of the German leadership sample in the manuscript. Please refer to the Supplementary Material for the results of the remaining three samples.

Model Fit and Latent Structure

We aimed to replicate the findings of Diehl et al., (2020; see Fig. 1). We followed the steps described by Diehl et al., (2020) and compared each of the six proposed second-order factor models with two alternative models. First, we estimated a general factor model in which all items loaded onto a single factor. We then estimated a first-order factor model in which all items loaded onto their respective first-order factors (e.g., glass cliff), and these factors could covary. Finally, we tested the a priori theorized second-order factor model, in which all items loaded onto their respective first-order factor, and each first-order factor loaded onto its second-order factor (e.g., male privilege).

We first fitted the CFA models for the German sample with leadership responsibility. The results of the CFA are presented in Table 2. With reference to the fit criteria by Hu and Bentler (1999), model fit in the different constructs ranged from poor to acceptable. Overall, the one-factor model failed to exhibit an acceptable model fit. The data did not support any model for the construct male privilege and disproportionate constraints. Both the first- and second-order factor models were supported for the constructs devaluation, hostility, and acquiescence.

However, a closer inspection of the models indicated several Heywood cases (i.e., negative residual variance or negative latent variable variance). For example, item 35 of the first-order factor salary inequality (e.g., "I have made less money than men who have held my position prior to me.") displayed a negative residual variance in the first- and second-order factor model. Results also indicated a negative latent variable variance of the first-order factor self-group distancing, as well as of the second-order factor acquiescence. The second-order factor model of insufficient support did not converge. Furthermore, the first-order factor exclusion presented an item with negative residual variance (e.g., item 24, "I have been excluded from leadership events (e.g., off-sites, retreats) because of my gender."). Considering these findings, we decided to modify the original factor structure using the German sample with leadership responsibility as a construction sample and subsequently cross-validated the solution in the remaining samples.

Model Specification of a Modified Model

Altogether, three of the originally established second-order factor models had an unacceptable model fit in the present data (Table 2). To address this, we made the following changes. We excluded three first-order factors (i.e., two-person career structure, exclusion, and salary inequality). We dissolved the second-order factor devaluation, which included the first-order factor salary inequality, and rearranged the first-order factor lack of acknowledgment into the second-order factor male privilege. Furthermore, we excluded three items from different factors (see Fig. 2 for a visualization of the modified model). The modifications are based on the following theoretical considerations and empirical observations.

The second-order factor male privilege presented an unacceptable model fit. Thus, we removed the factor two-person career structure, which included items such as “My organization vets spouses/partners of senior leaders as part

of the hiring process.” In the German leader sample, this factor exhibited a weak loading onto its second-order factor, male privilege (standardized loading = 0.42), as compared to the other first-order factors (glass cliff, standardized loading = 0.87; male culture, standardized loading = 0.71). Given that incorporating spouses into hiring or other human resource (HR) practices is uncommon in Germany, for cultural reasons, this factor may hold a different meaning in our sample than in the original North American context.

Because the second-order factor insufficient support did not converge in its original factor structure, we inspected the individual items to identify the potential causes of misfit. One problem might have been the simultaneous use of negatively and positively worded items (van Sonderen et al., 2013). Specifically, the first-order factors lack of mentoring and lack of sponsorship comprised exclusively reverse-coded items, whereas the first-order factor exclusion consisted of a mix of items—one reverse-worded item and two positively worded items. Integrating reverse-worded items into

Table 2 Results of the confirmatory factor analyses^a in the German leadership sample (replication of the original model)

Model	χ^2	<i>df</i>	<i>p</i>	SRMR	RMSEA [90% CI]	CFI	TLI	AIC	BIC	Acceptable model fit ^b
Male privilege										
One-factor model	507.44	35	< .001	0.148	0.262 [0.242; 0.283]	0.52	0.38	6931.87	7034.75	No
First-order factor model	170.62	32	< .001	0.112	0.148 [0.126; 0.170]	0.86	0.81	6543.47	6656.64	No
Second-order factor model	170.62	32	< .001	0.112	0.148 [0.126; 0.170]	0.86	0.81	6543.47	6656.64	No
Disproportionate constraints										
One-factor model	253.04	44	< .001	0.104	0.157 [0.138; 0.176]	0.70	0.63	7848.59	7961.76	No
First-order factor model	117.10	41	< .001	0.072	0.097 [0.077; 0.118]	0.89	0.86	7691.30	7814.80	No
Second-order factor model	117.10	41	< .001	0.072	0.097 [0.077; 0.118]	0.89	0.86	7691.30	7814.80	No
Insufficient support										
One-factor model	107.34	20	< .001	0.094	0.143 [0.117; 0.170]	0.70	0.57	5665.46	5747.76	No
First-order factor model ^c	25.19	17	.091	0.047	0.047 [0.000; 0.083]	0.97	0.96	5582.93	5675.52	Yes
Second-order factor model	NC									
Devaluation										
One-factor model	140.68	9	< .001	0.115	0.264 [0.226; 0.303]	0.74	0.57	4017.51	4079.23	No
First-order factor model ^c	13.53	8	.095	0.041	0.056 [0.000; 0.106]	0.99	0.98	3881.12	3946.28	Yes
Second-order factor model ^c	11.84	7	.106	0.041	0.060 [0.000; 0.117]	0.99	0.98	3883.12	3951.70	Yes
Hostility										
One-factor model	44.43	14	< .001	0.060	0.114 [0.077; 0.153]	0.92	0.88	4699.13	4771.15	No
First-order factor model	28.07	13	.009	0.047	0.080 [0.039; 0.121]	0.96	0.94	4675.83	4751.27	Yes
Second-order factor model ^d	25.91	12	.011	0.047	0.084 [0.038; 0.128]	0.96	0.94	4677.83	4756.70	Yes
Acquiescence										
One-factor model	33.13	5	< .001	0.093	0.176 [0.122; 0.235]	0.79	0.58	3555.26	3606.70	No
First-order factor model	11.37	4	.023	0.054	0.098 [0.033; 0.168]	0.95	0.87	3529.32	3584.19	Yes
Second-order factor model ^d	8.53	3	.036	0.054	0.114 [0.026; 0.207]	0.95	0.83	3531.32	3589.62	Yes

Sample consisted of *N* = 228 German women with leadership responsibility

AIC Akaike's information criterion, BIC Bayesian information criterion, NC non-convergence

^aCFA using MLR estimator and calculated with Satorra–Bentler-adjusted χ^2

^bModel fit evaluated according to the criteria of Hu and Bentler (1999)

^cHeywood case (negative residual variance)

^dHeywood case (negative latent variable variance)

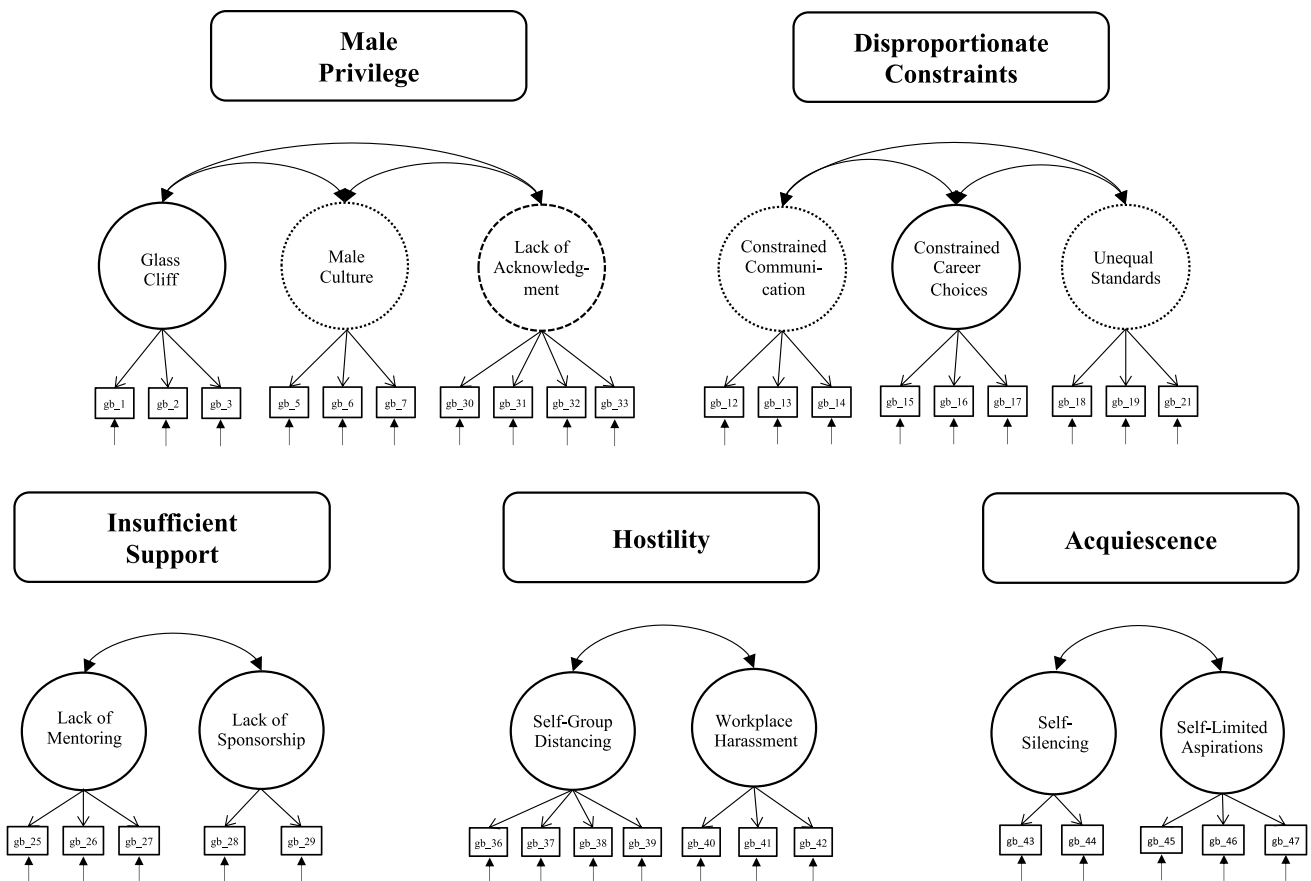


Fig. 2 Modified factor structure of the Gender Bias Scale for Women Leaders. Dotted lines indicate an item was removed from this factor; dashed lines indicate the factor was moved from the former second-

order factor devaluation; the factor self-group distancing was previously termed “Queen Bee Syndrome”

questionnaires is commonly used to reduce response bias and acquiescence (Swain et al., 2008). However, scholars have argued the use of reverse-worded items could confuse participants or increase the frequency of inattentive response behaviors (van Sonderen et al., 2013). Furthermore, on a conceptual level, it is possible that positively and negatively worded items differ in meaning. For instance, feeling welcome when attending social events (item 22) refers conceptually to inclusion, while having male colleagues who socialize without the respondent (item 23) points to ostracism on a non-work-related occasion. Finally, being excluded from leadership events because of one’s gender (item 24) is an active act of exclusion at work. The opposite of exclusionary coworker behavior is not necessarily inclusive coworker behavior; hence, the items in this factor might not conceptually relate to exclusion which could explain the non-convergence. As a result, we removed the first-order factor exclusion.

In the second-order factor devaluation, item 35 of the first-order factor salary inequality (e.g., “I have made less money than men who have held my position prior to me.”)

exhibited a negative residual variance, resulting in an inadmissible solution. Constraining the residual variance (e.g., to 0, or imposing equality constraints), as is sometimes done, is only a proper way of solving a Heywood case when theoretical considerations suggest it (Chen et al., 2001; Kline, 2011). Another conceptual issue with this factor was the question of whether salary inequality is a latent construct or, in fact, a manifestation of the existence of gender bias. Furthermore, from a cultural perspective, it is uncommon for Germans to discuss their salary with others, so it is fairly unlikely that respondents have knowledge about their predecessor’s salary. Thus, participants’ responses are likely to be based on speculation about salary inequality and may result in answers that reflect personality tendencies rather than facts. Consequently, we decided to dissolve the second-order factor devaluation by eliminating the first-order factor salary inequality. The remaining factor lack of acknowledgment was integrated into the second-order factor male privilege. Lack of acknowledgment comprised items such as “At work, I am interrupted by men when I am speaking.”, which could also reasonably represent the concept of male privilege.

Next, we dropped three items. Based on modification indices, we identified item 4 of the first-order factor male culture (“In my organization, there is pressure to conform to gender stereotypes.”) as problematic and excluded it from the model. Modification indices of the second-order factor disproportionate constraints suggested that item 11 of the first-order factor constrained communication (“I am mindful of my communication approach when exercising authority at work.”) and item 20 of the first-order factor unequal standards (“As a woman I am expected to be nurturing at work.”) are likely to be the sources of the local misfit.

The modification resulted in a factor structure with five second-order factors (i.e., male privilege, disproportionate constraints, insufficient support, hostility, and acquiescence), each consisting of either two or three first-order factors. However, hierarchical CFA models are only identified when there are at least three first-order factors (Kline, 2011) unless additional constraints are introduced (e.g., equality constraints on factor loadings). Instead of introducing such constraints to the second-order factor models, we proceeded with model testing using the first-order factor models. Specifically, we modeled gender bias as five first-order factor models with correlating factors. The composition of the five first-order factor models was based on the original structure (Diehl et al., 2020) and theoretical considerations, resulting in models that represent the topical domains of male privilege, disproportionate constraints, insufficient support, hostility, and acquiescence (Fig. 2).

We propose that modeling five distinct first-order correlated factor models is the most suitable approach to capture a phenomenon as complex and heterogeneous as gender bias. Second-order factor models run the risk of oversimplifying the complexity of the phenomenon and also limit the ability to comprehensively assess the multifaceted nature of gender bias. Our modeling approach also acknowledges the bandwidth-fidelity dilemma, which highlights the trade-off between the breadth and precision of a measurement instrument (Ones & Viswesvaran, 1996). In the context of gender bias research, it is important to consider the varying predictive impact of the different factors of gender bias. Depending on the criterion under investigation, certain factors may exhibit a greater predictive power than others. For example, experiencing sexual harassment at work (vs. self-group distancing) could have a stronger (weaker) impact on workplace safety. This underscores the theoretical significance of refraining from second-order factor models or even a general factor model of gender bias. Implementing such models would limit the ability to assess the nuanced ways in which different facets of gender bias could affect different outcomes.

The final factor models were always identified by fixing the first unstandardized loading of an indicator to 1 and factor means to 0. All correlations between factors were freed and all cross-loadings were constrained to zero.

Model Fit and Latent Structure of the Modified Model in the Construction Sample

The model fits of the modified models estimated in the construction sample (i.e., German women with leadership responsibility) are presented in Table 3. The data indicated a good model fit for all first-order factor models, i.e., CFI range = [0.95–0.98], TLI range = [0.87–0.97], RMSEA range = [0.05–0.10], and SRMR range = [0.03–0.05]. All items loaded statistically significantly on their respective factor (for factor loadings and standard errors, as well as results of the CFAs in all samples, please refer to the Supplementary Material). Several items showed a relatively low standardized factor loading (i.e., < 0.40). Especially, item 27 (factor lack of mentoring; “I have had to learn how to lead on my own.”) exhibited a standardized factor loading of 0.19. Like Diehl et al. (2020), we retained this item to ensure content validity.

Model Fit and Latent Structure in Three Separate Validation Samples

Since the purpose of this study was to investigate whether the factor structure of the GBSWL would hold in different languages and across leaders and non-leaders, we sought to validate our solution by applying it to three separate samples. This approach also reduced the potential risk of capitalizing on chance when making adaptations based on modification indices (Flora & Flake, 2017).

The model fit for all first-order factor models was good in the validation samples (i.e., Spanish leaders, German and Spanish non-leaders), except for those in the domains disproportionate constraints in the German non-leader sample and acquiescence in the Spanish non-leader sample (CFI range = [0.94–1.00], TLI range = [0.90–1.02], RMSEA range = [0.00–0.10] and SRMR range = [0.01–0.06]).

Most items loaded statistically significantly on their respective factor, except item 27 (factor lack of mentoring, “I have had to learn how to lead on my own.”), which exhibited statistically non-significant loadings in all three validation samples. Item 33 (“My efforts at creating harmony at work are noticed.”) of the factor lack of acknowledgment and none of the items of the factors self-silencing and self-limited aspirations showed a statistically significant factor loading in the Spanish non-leader sample, the latter indicating that these factors might be problematic. Additionally, in all three validation samples, items 33 (see above) and 39 (factor self-group distancing, “High-level women in my organization help other women succeed.”) displayed relatively low standardized factor loadings (i.e., < 0.40). However, following Diehl et al. (2020), we retained the items to ensure sufficient content validity.

Table 3 Results of the confirmatory factor analyses^a of the modified first-order factor structures in the German leadership sample

Model	χ^2	<i>N</i>	<i>df</i>	<i>p</i>	SRMR	RMSEA [90% CI]	CFI	TLI	Acceptable model fit ^b
Male privilege	47.95	228	32	.035	0.052	0.050 [0.014; 0.078]	0.98	0.97	Yes
Disproportionate constraints	45.77	228	24	.005	0.054	0.068 [0.037; 0.098]	0.96	0.95	Yes
Insufficient support	9.40	228	4	.052	0.033	0.077 [NA; 0.146]	0.98	0.94	Yes
Hostility	28.07	228	13	.009	0.047	0.080 [0.039; 0.121]	0.96	0.94	Yes
Acquiescence	11.37	228	4	.023	0.054	0.098 [0.033; 0.168]	0.95	0.87	Yes

^aCFA using MLR estimator and calculated with Satorra–Bentler-adjusted χ^2

^bModel fit evaluated according to criteria of Hu and Bentler (1999)

Measurement Invariance

For the results of the measurement invariance analysis between leaders from both countries refer to Table 4. Results of all other analyses of measurement invariance are presented in more detail in the Supplementary Material.

German and Spanish Women With Leadership Responsibility

All first-order factor models exhibited at least metric invariance, indicating equivalence of the factor structure in both samples. Factor models in the domains male privilege, disproportionate constraints, and acquiescence achieved metric invariance with good model fit (CFI range = [0.97–0.98], RMSEA range = [0.05–0.07], SRMR range = [0.04–0.06]). These results suggest equivalence of both the factor structure of the measurement models and of the item loadings in both groups. The measurement model of the first-order factor model acquiescence failed to converge at higher levels of measurement invariance (i.e., strong, and strict). Factor models in the domains insufficient support and hostility achieved strict equivalence with good model fit (CFI = range [0.98–1.00], RMSEA range = [0.02–0.05], SRMR range = [0.04–0.05]), indicating equivalence of the factor structure, item loadings, intercepts, and residuals in both leadership samples. These results suggest the appropriateness of comparing correlations in all first-order factor models and comparing means in the domains insufficient support and hostility between samples of German and Spanish women with leadership responsibility.

German and Spanish Women Without Leadership Responsibility

Except for acquiescence, all first-order factor models reached at least configural invariance with an acceptable model fit. A closer analysis of the first-order factor model acquiescence revealed Heywood cases in either

the German sample only (when testing for configural or metric invariance) or in both samples (when testing for strong or strict invariance), preventing the interpretation of the results. The first-order factor model hostility reached configural invariance, while male privilege exhibited metric invariance with an acceptable model fit (CFI = 0.96, RMSEA = 0.06, SRMR = 0.06). Disproportionate constraints fulfilled strong invariance (CFI = 0.94, RMSEA = 0.07, SRMR = 0.06) and insufficient support reached strict invariance with a good model fit (CFI = 0.98, RMSEA = 0.07, SRMR = 0.06).

This means that when applied to non-leaders in Germany and Spain, scholars can compare correlations between the samples in the case of male privilege, disproportionate constraints, and insufficient support. Furthermore, the latter two allow for mean comparisons.

German Women With and Without Leadership Responsibility

In the German samples, all first-order factor models fulfilled configural measurement invariance with acceptable model fit (CFI range = [0.95–0.99], RMSEA range = [0.06–0.09], SRMR range = [0.03–0.06]). However, the configural measurement model of acquiescence resulted in a Heywood case in the German non-leader sample. The first-order factor models male privilege and hostility fulfilled metric invariance (CFI range = [0.95–0.96], RMSEA range = [0.06–0.08], SRMR range = [0.06–0.06]). Restraining the intercepts resulted in a statistically significantly worse model fit in the first-order factor models disproportionate constraints and insufficient support. Thus, only configural invariance was obtained. Nonetheless, in the case of insufficient support, despite failing to fulfill metric invariance, model fit was good (CFI = 0.98, RMSEA = 0.07, SRMR = 0.06) and the differences in model fit were only marginal (Δ CFI = −0.013, Δ RMSEA = 0.017). The results suggest that in samples of German women with and without leadership responsibility, scholars could compare correlations of male privilege and hostility and potentially insufficient support.

Table 4 Results of measurement invariance analyses for the modified first-order factor structures in the leadership samples

Model	$\chi^2(df)$	Scaled χ^2	CFI	TLI	RMSEA[90% CI]	SRMR	AIC	BIC	Modelcomp	$\Delta\chi^2(\Delta df)^a$	p^b	ΔCFI	$\Delta RMSEA$	$\Delta SRMR$	Decision
Male privilege															
Configural	105.62 (64)	89.33*	0.984	0.978	0.045 [0.019; 0.065]	0.042	13,702.14	13,976.22							Accept
Metric	120.15 (71)	102.75*	0.980	0.975	0.047 [0.025; 0.066]	0.053	13,702.67	13,947.68	Configural	13.84 (7)	.054	−0.004	0.003	0.011	Accept
Strong	156.86 (78)	136.22***	0.964	0.959	0.060 [0.043; 0.077]	0.059	13,725.38	13,941.32	Metric	37.84 (7)	<.001	−0.016	0.013	0.005	Reject
Strict	195.81 (88)	163.47***	0.951	0.951	0.066 [0.050; 0.082]	0.061	13,744.33	13,918.74	Strong	24.97 (10)	.005	−0.013	0.006	0.002	Reject
Disproportionate constraints															
Configural	104.48(48)	85.10**	0.968	0.953	0.064 [0.041; 0.085]	0.050	12,999.59	13,248.76							Accept
Metric	108.98 (54)	89.86**	0.970	0.960	0.059 [0.036; 0.079]	0.052	12,992.09	13,216.34	Configural	4.12 (6)	.661	0.001	−0.005	0.001	Accept
Strong	139.32 (60)	111.44***	0.958	0.949	0.066 [0.046; 0.085]	0.056	13,003.69	13,203.02	Metric	24.05 (6)	<.001	−0.012	0.007	0.004	Reject
Strict	175.19 (69)	139.32***	0.941	0.938	0.073 [0.055; 0.090]	0.066	13,022.23	13,184.18	Strong	26.57 (9)	.002	−0.017	0.007	0.010	Reject
Insufficient support															
Configural	13.16 (8)	11.94	0.992	0.979	0.048 [0.000; 0.101]	0.028	7474.46	7607.35							Accept
Metric	14.67 (11)	12.34	0.997	0.994	0.025 [0.000; 0.081]	0.033	7469.97	7590.40	Configural	1.06 (3)	.786	0.005	−0.023	0.005	Accept
Strong	17.57 (14)	15.13	0.997	0.996	0.020 [0.000; 0.072]	0.036	7466.87	7574.84	Metric	2.73 (3)	.434	0.001	−0.005	0.003	Accept
Strict	25.17 (19)	21.26	0.995	0.995	0.024 [0.000; 0.069]	0.044	7464.47	7551.68	Strong	6.09 (5)	.298	−0.003	0.005	0.008	Accept
Hostility															
Configural	60.92 (26)	49.86**	0.975	0.959	0.069 [0.039; 0.098]	0.043	9964.09	10,146.81							Accept
Metric	65.46 (31)	53.65**	0.976	0.968	0.062 [0.032; 0.089]	0.046	9958.63	10,120.58	Configural	3.75 (5)	.586	0.001	−0.007	0.003	Accept
Strong	77.08 (36)	64.81**	0.970	0.965	0.064 [0.038; 0.088]	0.051	9960.25	10,101.45	Metric	11.63 (5)	.040	−0.006	0.002	0.005	Accept
Strict	81.14 (43)	64.97*	0.976	0.977	0.052 [0.023; 0.077]	0.053	9950.31	10,062.44	Strong	2.61 (7)	.919	0.006	−0.011	0.002	Accept
Acquiescence															
Configural	15.41 (8)	14.80	0.980	0.951	0.061 [0.000; 0.110]	0.033	7407.78	7540.66							Accept
Metric	22.234 (11)	22.49*	0.968	0.943	0.066 [0.025; 0.105]	0.044	7408.60	7529.03	Configural	8.06 (3)	.045	−0.012	0.005	0.011	Accept
Strong	NC														
Strict	NC														

N = 470; Germany, n = 228; Spain, n = 242

NC non-convergence

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$ ^aChi-square difference test refers to standard (rather than robust) test statistics and was Satorra–Bentler-scaled (see Satorra and Bentler (2001))^bp-value corresponds to the chi-square difference test

Spanish Women With and Without Leadership Responsibility

In the Spanish samples, except for acquiescence, all first-order factor models fulfilled metric measurement invariance with good model fit (CFI range = [0.97–1.00], RMSEA range = [0.00–0.06], SRMR range = [0.03–0.06]). Restraining factor loadings in the measurement model of acquiescence resulted in a statistically significantly worse model fit and a Heywood case. The first-order factor model disproportionate constraints fulfilled strict measurement invariance with good model fit (CFI = 0.97, RMSEA = 0.05, SRMR = 0.05). These results indicate that in samples of Spanish women with and without leadership responsibility, scholars can compare correlations of all first-order factor models except acquiescence, as well as compare means in the first-order factor model disproportionate constraints.

German and Spanish Women With and Without Leadership Responsibility

When all four groups were compared regarding their level of measurement invariance, results suggested that every first-order factor model fulfilled metric invariance with good model fit (CFI range = [0.96–0.99], RMSEA range = [0.04–0.07], SRMR range = [0.05–0.06]). Again, the CFA of the configural measurement model of acquiescence resulted in a Heywood case in the German non-leader sample. This indicates that in cross-national research in Germany and Spain on women with and without leadership responsibility scholars could compare correlations across samples when being cautious about interpreting the first-order factor model acquiescence.

Descriptive Statistics, Correlations, and Reliability

Descriptive statistics, McDonald's ω , Cronbach's α , and correlations for the German leadership sample are presented in Table 5 (for all other samples, please refer to the Supplementary Material). In all samples, absolute skewness values for the GBSWL were mostly below one, indicating that the distribution did not substantially deviate from a normal distribution. Exceptions were the first-order factor models workplace harassment and glass cliff in the German non-leader sample (skew = 1.29 and 1.06, respectively), and workplace harassment in the German and Spanish leader samples (skew = 1.11 and 1.02, respectively). However, all absolute skewness values were below the threshold of 2, at which estimation problems previously arose (Curran & West, 1996). In general, job satisfaction was high across all samples ($M_{\text{range}} = [4.91; 5.68]$), with slightly lower job satisfaction in the non-leader samples. According to the

Welch test, this difference was statistically significant in both Germany ($t(371.14) = 4.46$, $p < 0.001$, 95% CI [0.37; 0.97], $d = 0.44$) and Spain ($t(376.56) = 4.16$, $p < 0.001$, 95% CI [0.32; 0.90], $d = 0.40$). Women's intentions to leave their current organization ranged from 14 to 24%.

To assess the reliability of the GBSWL, we computed McDonald's ω and Cronbach's α . Overall, values of all twelve first-order gender bias factors were satisfactory in all samples with coefficients ranging from $\omega = 0.70$ to 0.90 and $\alpha = 0.69$ to 0.90, except for self-silencing ($\omega = 0.67$ and $\alpha = 0.67$) and self-limited aspirations ($\omega = 0.56$ and $\alpha = 0.56$) in the Spanish non-leader sample. Some first-order factors exhibited unsatisfactory values across all samples. For example, McDonald's ω (Cronbach's α) for the first-order factors lack of mentoring and self-limited aspirations ranged from $\omega = 0.48$ to 0.60 ($\alpha = 0.38$ to 0.48), and $\omega = 0.51$ to 0.68 ($\alpha = 0.50$ to 0.68), respectively.

Construct Validity

We tested Hypotheses 1 and 2 with bivariate correlations. Of the altogether 40 hypothesized correlations between specific gender bias factors and job satisfaction and turnover intention, almost all, i.e., 39, were statistically significant and in the proposed direction (Table 6). The only non-significant relation emerged between workplace harassment and turnover intention. Thus, the general picture of correlations provides evidence for the construct validity of the measures.

In addition to testing the relations proposed in the hypotheses, we also compared the four patterns of correlations of the present study with the findings presented by Diehl et al. (2020). This comparison provided preliminary insights into the consistency of our findings with prior research and with the proposed conceptual framework by Diehl et al. (2020). Overall, with some minor exceptions, the results of the present study are very similar to the findings by Diehl et al. (2020). The correlations with the outcomes were small to moderate i.e., the gender bias facets correlated negatively with job satisfaction and positively with turnover intention (Table 6). One noteworthy difference from the findings of Diehl et al. (2020) is the correlation between constrained career choices and job satisfaction in the sample of German leaders, which unexpectedly showed a positive correlation ($r = 0.16$, $p < 0.05$).

The only relevant structural change we made to the original model was to remove the first-order factor two-person career structure from the second-order factor model of male privilege and to replace it with lack of acknowledgment, a first-order factor previously categorized under the now dissolved second-order factor devaluation. Given the altered configuration of the first-order factor model for male privilege, we deemed it important to assess the relations among the individual first-order factors to ensure coherence

Table 5 Descriptive statistics, correlations, McDonald's Omega, and Cronbach's alpha in the German leadership sample

	Descriptives				Correlations													
	M	SD	SKW	KRT	GC	MC	CC	CCC	US	LM	LS	LA	SGD	WH	SES	SLA	JS	TI
GC	2.08	1.07	0.85	−0.02	0.81 (0.81)													
MC	2.34	1.08	0.57	−0.45	0.41	0.77 (0.74)												
CC	2.68	0.95	0.44	−0.10	0.25	0.36	0.69 (0.68)											
CCC	2.17	1.04	0.75	−0.15	0.20	0.28	0.29	0.75 (0.74)										
US	2.46	1.22	0.40	−0.79	0.33	0.56	0.42	0.44	0.86 (0.85)									
LM	3.31	0.89	0.07	−0.27	0.01	0.06	0.00	−0.34	−0.00	0.58 (0.46)								
LS	2.66	1.15	0.35	−0.66	0.08	0.01	−0.05	−0.28	−0.15	0.39	0.81 (0.81)							
LA	2.24	0.77	0.54	0.10	0.43	0.52	0.23	0.05	0.49	0.13	0.04	0.70 (0.64)						
SGD	2.52	0.98	0.57	−0.44	0.37	0.42	0.29	0.07	0.46	0.13	0.06	0.50	0.77 (0.73)					
WH	1.77	0.88	1.11	0.64	0.41	0.46	0.29	0.12	0.48	0.09	−0.02	0.50	0.55	0.70 (0.70)				
SES	2.42	1.08	0.58	−0.33	0.04	−0.10	−0.06	−0.22	−0.20	0.14	0.25	0.00	−0.07	−0.15	0.78 (0.78)			
SLA	2.13	0.86	0.55	−0.08	0.34	0.30	0.32	0.35	0.35	−0.13	−0.17	0.38	0.36	0.36	−0.20	0.51 (0.50)		
JS	5.58	1.35	−1.06	1.12	−0.26	−0.30	−0.08	0.16	−0.25	−0.24	−0.16	−0.39	−0.40	−0.35	−0.14	−0.11	0.93 (0.93)	
TI					0.08	0.19	0.16	0.01	0.28	0.12	−0.09	0.04	0.17	0.14	−0.04	0.11	−0.37	
Age	42.73	10.50	0.22	−0.91	−0.24	−0.03	−0.14	−0.21	−0.13	0.24	0.03	−0.04	0.01	−0.07	0.04	−0.09	−0.09	−0.06

Diagonal displays McDonald's omega (Cronbach's alpha); $N = 228$. All correlations $> |0.13|$ are significant at $p < .05$; correlations $> |0.17|$ are significant at $p < .01$; correlations $> |0.22|$ are significant at $p < .001$

SKW skewness, KRT kurtosis, GC glass cliff, MC male culture, CC constrained communication, CCC constrained career choices, US unequal standards, LM lack of mentoring, LS lack of sponsorship, LA lack of acknowledgment, SGD self-group distancing, WH workplace harassment, SES self-silencing, SLA self-limited aspirations, JS job satisfaction, TI turnover intention (coded, 0 no, 1 yes)

Table 6 Correlations of the gender bias facets with job satisfaction and turnover intention

Gender bias	Job satisfaction				Turnover intention			
	Leaders		Non-leaders		Leaders		Non-leaders	
	German	Spanish	German	Spanish	German	Spanish	German	Spanish
GC	−0.26***	−0.36***	−0.46***	−0.48***	0.08	0.21**	0.28***	0.27***
MC	−0.30***	−0.50***	−0.24***	−0.39***	0.19*	0.27***	0.21**	0.23***
CC	−0.08	−0.07	−0.23***	−0.13	0.16*	0.11	0.17*	0.10
CCC	0.16*	−0.20**	−0.06	−0.22**	0.01	0.09	−0.00	0.09
US	−0.25***	−0.49***	−0.38***	−0.46***	0.28***	0.27***	0.31***	0.40***
LM	−0.24***	−0.22***	−0.22**	−0.27***	0.12	0.13*	0.20**	0.22**
LS	−0.16*	−0.25***	−0.03	−0.21**	−0.09	0.09	0.07	0.01
LA	−0.39***	−0.49***	−0.42***	−0.41***	0.04	0.20**	0.24***	0.31***
SGD	−0.40***	−0.39***	−0.40***	−0.39***	0.17**	0.13*	0.23***	0.23***
WH	−0.35***	−0.44***	−0.40***	−0.42***	0.14	0.16*	0.16*	0.22**
SES	−0.14*	0.12	−0.05	−0.09	−0.04	−0.05	0.04	−0.01
SLA	−0.11	−0.49***	−0.28***	−0.17*	0.11	0.25***	0.15	0.12

$N_{\text{German leaders}}$ 228, $N_{\text{Spanish leaders}}$ 242, $N_{\text{German non-leaders}}$ 198, $N_{\text{Spanish non-leaders}}$ 202; turnover intention (coded, 0 no, 1 yes)

GC glass cliff, MC male culture, CC constrained communication, CCC constrained career choices, US unequal standards, LM lack of mentoring, LS lack of sponsorship, LA lack of acknowledgment, SGD self-group distancing, WH workplace harassment, SES self-silencing, SLA self-limited aspirations

* $p < .05$; ** $p < .01$; *** $p < .001$

within this factor model. Therefore, we assessed the latent correlations of the factors in the newly constructed domain male privilege, i.e., between the first-order factors lack of acknowledgment and glass cliff, and male culture, respectively. We found a moderate to high correlation between lack of acknowledgment and glass cliff, as well as between lack of acknowledgment and male culture.

Discussion

This study presents the translation and validation of the Gender Bias Scale for Women Leaders (Diehl et al., 2020) in German and Spanish samples of women with ($n = 470$) and without ($n = 400$) leadership responsibility. The purpose of this study was twofold. First, we aimed to develop an instrument to assess the gender biases experienced by women leaders in two European countries. We translated the GBSWL (Diehl et al., 2020), which was developed and validated for the North American context, into German (GBSWL-G) and Spanish (GBSWL-S). We thereby answered a recent call to present improved measures that link implicit and explicit gender biases in the workplace (Kossek et al., 2017). Our second goal was to test whether the factor structure also holds in samples of women without leadership responsibility. This provides future research seeking to uncover what prevents women from entering leadership positions with a suitable instrument. This is especially relevant against the backdrop of reports on the broken rung

phenomenon. According to the current Women in the Workplace report by McKinsey and Company (2023), the primary challenge women encounter is not reaching top management positions (i.e., glass ceiling), but rather their initial promotion to a management position (i.e., broken rung).

The modified version of the GBSWL can be used to assess the degree of gender bias experienced by women with and without leadership responsibility in Germany and Spain. Because of its measurement invariance across samples, it is suitable for cross-cultural research, which strives to compare structural relations between gender bias and outcome variables.

Factorial Validity and Measurement Invariance

Confirmatory factor analyses of the original factor structure revealed an unsatisfactory model fit in all samples. Based on theoretical and empirical considerations, we modified the original higher-order factor models into correlated first-order factor models. Specifically, we reorganized the original six second-order factors into five correlated first-order factor models retaining most of the original gender bias domains. In doing so, we were able to construct measurement models suitable for German and Spanish women leaders and non-leaders. Only the first-order factor model acquiescence failed to exhibit an acceptable model fit in the Spanish non-leader sample and yielded a Heywood case in the German non-leader sample. A potential explanation might be that factor models with only two indicators per factor can be

problematic (Kenny et al., 1998). In particular, when sample sizes are small, the probability of improper solutions (i.e., Heywood cases) increases (Marsh et al., 1998).

To facilitate cross-national research and enable scholars to compare results between leaders and non-leaders, we tested the GBSWL-G and GBSWL-S for measurement invariance. Analyses comparing both leader samples indicated metrics (e.g., male privilege and disproportionate constraints) and strict invariance (e.g., insufficient support and hostility) for the respective domains. Regarding acquiescence, the results suggested configural measurement invariance; however, the model fit of the model with equality constraints on the factor loadings was still very good (CFI = 0.97, RMSEA = 0.07, SRMR = 0.04). Therefore, we are cautiously optimistic that using the GBSWL-G and GBSWL-S for correlational analyses between German and Spanish leaders is appropriate.

In the German samples, the first-order factor model insufficient support only fulfilled the criteria for configural measurement invariance; however, the metric measurement model still exhibited good model fit (CFI = 0.98, RMSEA = 0.07, SRMR = 0.06). Scholars can still use this first-order factor model but should interpret the results with a grain of salt or even better establish measurement invariance in their own samples.

When comparing all samples, measurement invariance analyses indicated that every first-order factor model fulfilled metric invariance with good model fit (CFI range = [0.96–0.99], RMSEA range = [0.04–0.07], SRMR range = [0.04–0.06]). This suggests that the linear relation between the indicators and their respective latent factors is equal across groups. Thus, future research can use the GBSWL-G and GBSWL-S in order to compare relations of gender bias and other constructs of interest between Spanish and German individuals with and without leadership responsibility.

Construct Validity

Of the 40 correlations between specific gender bias factors and job satisfaction and turnover intention (Hypotheses 1 and 2), 39 are statistically significant and in the proposed direction. Furthermore, taking the remaining correlations, i.e., the ones not implied in the hypotheses, into consideration, the correlational patterns are theoretically plausible and similar to those reported by Diehl et al. (2020). Some observations are worthy to be discussed. With few exceptions, gender bias, in general, is associated with lower job satisfaction across all samples. An exception to this is the small positive correlation between constrained career choices and job satisfaction in the German leadership sample. This factor assesses the degree to which women experience societal constraints on their educational and career choices. In line

with RCT (Eagly & Karau, 2002), a possible explanation for the unexpected association might be that women leaders in occupations deemed suitable for women violate gender norms to a lesser degree, thus having to deal with fewer conflicts and being more satisfied.

Exploring specifically the relation between turnover intention and the gender bias factors *not* covered in Hypothesis 2 reveals that many correlations are not statistically significant (e.g., constrained career choices, self-silencing, and lack of sponsorship). One explanation could be that women develop an inurement to gender discrimination (Raver & Nishii, 2010), similar to adapting to a repeated stimulus, which consequently elicits a reduced reaction (see psychological adaptation theory, Helson, 1947). For example, women who have been confronted with gender discrimination in the past might have adapted to this stimulus. Further experience of gender bias might produce only minor reactions. Additionally, many of the gender bias facets in the GBSWL are subtle forms of discrimination (e.g., self-silencing), whose impact could be below the threshold of what is considered a shock in the unfolding model of voluntary turnover (e.g., Holtom et al., 2005; Lee & Mitchell, 1994). These shocks are jarring events that drive turnover. However, if women evaluate the gender bias experience as not sufficiently severe, it is possible that neither turnover intention nor actual turnover behavior is affected, which could explain the non-significant correlations.

These findings should be interpreted in light of the fact that we assessed turnover intentions using a single-item measure. Single-item measures can be problematic, especially when used to capture complex psychological constructs (e.g., emotion). In the case of a multifaceted construct, using a single-item measure would not be a convincing choice. However, as Allen et al. (2022) noted, since the late 1990s several authors have challenged the notion that all single-item measures are unsuitable to capture psychological outcomes. In this study, the goal was to keep the research design as similar to Diehl et al. (2020) as possible. Following the authors we used a single-item measure which corresponds to classic operationalizations of turnover intention as a dichotomous construct (i.e., we asked participants whether they intend to leave their position within the next 12 months).

The factor self-silencing exhibited a complex correlational pattern. For example, it correlated positively with the two gender bias factors lack of mentoring and lack of sponsorship, but negatively with most others (e.g., glass cliff, unequal standards, workplace harassment). One possible explanation could be methodological in that the factor consists of two reverse-scored items. This could make it difficult for participants to process the items, and thus represent a method effect. On the other hand, the gender

bias factors that correlated negatively with self-silencing represented relatively explicit forms of gender bias (e.g., workplace harassment and unequal standards). It is possible that experiencing these overt forms of sexism might elicit a verbal response insofar as women will speak up if they witness gender inequality.

Theoretical Implications

The GBSWL aims to cover two perspectives—a rather broad conceptualization of gender bias via its domains (e.g., disproportionate constraints) and a more nuanced view of different facets of biases women face at work (the first-order factors, e.g., unequal standards). Whereas the original GBSWL was developed and tested for women in leadership positions, the modified versions GBSWL-G and GBSWL-S were also tested and found valid for women without leadership responsibility. The applicability of the instrument for two distinct job levels makes it suitable for research that seeks to contribute to our understanding of what prevents women from (a) entering leadership positions in the first place and (b) advancing into the upper echelons of leadership. The former is especially important since reaching the initial management position (i.e., broken rung) proves to be one of the biggest challenges for women's career advancement (McKinsey & Company, 2023). Thus, by providing a scale to assess the gender biases faced by women in non-leadership positions, we enable scholars and organizations alike to gain insights into the broken-rung phenomenon.

Furthermore, women at different levels of the organizational hierarchy may encounter different barriers that they have to overcome—another area of research deserving attention. For example, junior women might experience self-group distancing more strongly than women leaders at a higher hierarchical level. On the other hand, and in accordance with RCT (Eagly & Karau, 2002), when women hold upper-level leadership positions, the perceived lack of fit between stereotypical attributes and job requirements becomes more pronounced. This may contribute to stricter performance standards and less favorable performance ratings for women in line management positions compared to women in staff jobs or men in both line and staff jobs (Lyness & Heilman, 2006). Thus, using the GBSWL to develop a more nuanced insight into the different manifestations of gender bias at different hierarchical levels could help advance gender equality.

Our study also provides evidence that gender bias in Germany and Spain is to some extent conceptually different from that in the United States. In particular, the facets two-person career structure, which includes spousal vetting in the hiring process, and salary inequality are not directly applicable from the North American to the German and Spanish contexts. First, jobs in Europe are typically not structured in

such a way that a fully available, unpaid support partner is required (e.g., a spouse hosting events). Moreover, spousal vetting in the hiring process is unusual, highlighting the cultural differences in workplace expectations between the U.S. and Germany and Spain, respectively.

While Diehl et al. (2020) intentionally recruited a sample of women working in diverse industries, this gender bias facet was initially identified through interviews with leaders in faith-based organizations (e.g., religious nonprofit institutions based on evangelicalism) and higher education leaders (see Diehl & Dzubinski, 2016). Particularly noteworthy are the experiences of women leaders in religious organizations, who often operate in conservative environments that may emphasize traditional gender roles even more than other industries. Diehl and Dzubinski (2016) found that the negative aspects of a two-person career structure were more prevalent in religious organizations, which could suggest that factors contributing to gender bias may vary across different professional contexts. It is important to recognize that the original factor structure may not fully capture the experiences of leaders in different industries or cultural contexts. This raises the question of whether a two-person career structure is indeed a universal challenge that women leaders face or whether it is contingent on environmental factors such as the conservativeness of the workplace or industry-specific expectations.

Second, in light of cultural norms, open discussions about salaries are rare in Germany,² making it unlikely for employees to be aware of their predecessors' salaries. As a result, participants' responses may be influenced by speculation regarding salary inequality, potentially reflecting individual personality traits rather than factual information.

To accommodate cultural differences and consider theoretical aspects, we dropped these first-order factors. Rather than opting for a literal translation of the questionnaire, our research highlights the importance of culturally informed translation and adaptation when applying questionnaires to different languages and contexts. Furthermore, we removed the factor of exclusion due to a lack of conceptual clarity rather than cultural differences. This factor comprised a mix of inclusionary behavior and exclusion at work, as well as from non-work-related social gatherings. While we

² In 2017, Germany introduced a law on pay transparency, granting employees the right to access individual salary information. The directive requires employers with over 200 employees to explain, upon request, how employees are paid. Furthermore, at the European level, a similar directive on pay transparency was adopted in 2023 by the European Union. However, a recent assessment of the directive's implementation in Germany revealed that only 4% of employees utilized their individual right, and employers seldom disclose this information voluntarily (Brändle et al., 2023). The need for such legal measures highlights the rarity of salary discussions in Germany, as indicated by the limited use of the individual right to access salary.

do acknowledge that being excluded from informal gatherings could have negative implications for women's career advancement, we do not believe that the items in their original conceptualization reflect this notion in the most optimal way.

In its modified version, we have successfully tailored the GBSWL for use in (at least) two European countries, offering a valuable tool for studies with a cross-cultural perspective on gender bias. Given the substantial volume of research emerging from Germany and Spain, particularly in the wake of the #MeToo movement, the modified GBSWL comes at an opportune time to further support researchers in these countries as they seek a more nuanced understanding of gender bias at work.

Practical Implications

Our results suggest that gender bias is generally associated with lower job satisfaction. Although the correlations in our study were small to moderate, lowered job satisfaction due to gender bias should be a concern for organizations and HR professionals. Meta-analytical evidence on job satisfaction previously indicated associations with performance ($\rho = 0.30$; Judge et al., 2001) and turnover ($\rho = -0.19$; Griffith et al., 2000). Even a moderately low job satisfaction can result in severe financial repercussions for organizations. Thus, organizations should be interested in reducing gender bias in the workplace.

The GBSWL is also of interest to organizations who want to develop gender equity programs. A recent meta-analysis showed that training programs are most effective when based on needs analysis (Lacerenza et al., 2017). HR professionals can use the GBSWL to analyze the extent of specific gender bias facets. This knowledge can then be used to tailor interventions specifically to the needs of the organization. Rather than providing generic recommendations or implementing training targeting gender bias in general, HR professionals can create interventions based on a needs analysis.

Limitations and Future Directions

Our study has some limitations. For example, some of our study variables exhibited low McDonald's ω and Cronbach's α values (e.g., lack of mentoring and self-limited aspirations). In some cases, this could be caused by the reverse-scored items. Another explanation might be that some of the factors represent a diverse set of gender bias experiences, as is the case with self-limited aspirations. This factor contains items related to intrapersonal restraint due to a lack of confidence ("I have turned down a promotion because I felt unqualified."), but also regarding matters external to the

woman ("My personal obligations have prevented me from pursuing opportunities for advancement at work.").

Furthermore, several items had low (i.e., < 0.40), or statistically non-significant factor loadings (e.g., self-limited aspirations, self-silencing). Diehl et al. (2020) constructed the original instrument to cover a broad range of gender biases. In their study, they too found some low factor loadings, but kept those items to "ensure adequate content validity" (Diehl et al., 2020, p. 272). From this perspective, our results are consistent with the original study. However, the results should be interpreted considering this limitation.

The construct validation of the German and Spanish versions has only been conducted in Germany and Spain. Thus, before applying the instrument to other German-speaking (e.g., Austria, specific regions of Switzerland, and northern Italy) or Spanish-speaking (e.g., Mexico, Argentina, Columbia) samples, further tests on measurement invariance and validity are advisable, because nuances in terms of language might affect the meaning of items and thus the validity of the instrument.

We also want to recognize the temporal context of our study. We collected the data in November 2021 during the height of the COVID-19 pandemic, a time of uncertainty in several areas of life (e.g., health, job security). To contain the pandemic, employers took measures that reduced direct and personal contact with colleagues and supervisors massively. At the same time, household responsibilities increased, especially for women with children (Collins et al., 2021). Dealing with these immediate concerns and uncertainties might be more important than forming the intention to leave a stable job. It is an open question whether these factors affected the strength of the relation between gender bias, turnover intention, and job satisfaction. Therefore, it appears worthwhile to look into the structural relations in more stable times.

Future Directions

In this study, we present a factorial valid and measurement invariant instrument to assess gender bias in Germany and Spain. We further show theoretically plausible correlations with meaningful organizationally relevant outcome variables. Future research should explore convergent and discriminant validity.

Furthermore, it would be interesting to examine boundary conditions in terms of exacerbating or buffering factors. For example, we did not hypothesize but found that in the German samples being promoted into high-risk, precarious roles (i.e., glass cliff) correlated more strongly negatively with job satisfaction for women without leadership responsibilities. According to the *think crisis, think female* paradigm (Ryan et al., 2011), stereotypical female leadership qualities

(e.g., being understanding or showing concern for others) are valued in times of crisis, especially when crisis management involves managing people. It is possible that women in leadership positions already have experience in managing crises through past appointments. Having proved oneself in crisis situations and the resulting self-efficacy may mitigate the impact of being assigned a high-risk role.

We examined the association of gender bias with job satisfaction and turnover intention. Several other outcome variables may also be of interest. For example, internalizing a leader role identity results in behaving more leader-like and seeking opportunities to practice leadership behaviors, which then strengthens the leader identity (Day & Sin, 2011) and increases the probability of emerging as a leader (Kwok et al., 2018). However, this process may be more challenging for women than men. Mayo et al. (2012), for example, suggest that women tend to respond more strongly to (negative) feedback regarding their leadership competencies by aligning their self-evaluations with others' views of them. It seems plausible that experiencing gender bias interferes with women's identity work (i.e., constructing and internalizing a leader role identity), thereby reducing their probability of emerging as leaders.

In our study, we did not apply an intersectional perspective on gender bias. However, having intersecting marginalized identities (e.g., ethnicity, sexual orientation) likely interacts with women's career advancement. For example, the 2023 Women in the Workplace report (McKinsey & Company, 2023) indicates that women of color encounter greater difficulty in initially attaining management positions compared to other women and men. While we have demonstrated the applicability of the GBSWL for women without leadership responsibility, we encourage researchers to examine its suitability for research that takes an intersectional approach.

Conclusion

To achieve gender equality, it is imperative to understand the barriers that prevent women from entering and succeeding in leadership positions. As originally intended by Diehl et al. (2020), scholars and practitioners in Germany and Spain can use the GBSWL-G and GBSWL-S to assess the broad concept of gender bias at different levels of analysis. It is feasible to use the entire instrument to obtain an overview of the degree to which women experience bias in their daily work life or to use specific first-order factor models to identify particular areas in which organizations exhibit a possible gender bias and tailor interventions specifically to address them. Thus, the modified GBSWL offers a comprehensive tool for analyzing barriers likely to affect women's career advancement.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10869-024-09946-7>.

Author Contribution All authors contributed to the study conception and design. Material preparation and analysis were performed by Mona Algnier. The first draft of the manuscript was written by Mona Algnier and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management*, 47(4), 823–837. <https://doi.org/10.1177/0149206320969787>
- Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science. *European Journal of Psychological Assessment*, 38(1), 1–5. <https://doi.org/10.1027/1015-5759/a000699>
- Badura, K. L., Grijalva, E., Newman, D. A., Yan, T. T., & Jeon, G. (2018). Gender and leadership emergence: A meta-analysis and explanatory model. *Personnel Psychology*, 71(3), 335–367. <https://doi.org/10.1111/peps.12266>
- Brändle, T., Kalweit, R., Koch, A., Scheu, T., Schlachter-Voll, M., & Wiemann, J. S. (2023). Zweiter Bericht der Bundesregierung zur Wirksamkeit des Gesetzes zur Förderung der Entgelttransparenz zwischen Frauen und Männern sowie zum Stand der Umsetzung des Entgeltgleichheitsgebots in Betrieben mit weniger als 200 Beschäftigten. Bundesministerium für Familie, Senioren, Frauen und Jugend. Retrieved March 3, 2024, from <https://www.bmfsfj.de/resource/blob/229484/7c1deeee794163514c0221663af940d2/zweiter-bericht-entgelttransparenzgesetz-data.pdf>
- Camman, C., Fichman, M., Jenkins, D., & Klesh, J. (1979). The Michigan organizational assessment questionnaire. (Unpublished manuscript).
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Chen, F., Bollen, K., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research*, 29(4), 468–508. <https://doi.org/10.1177/0049124101029004003>

- Collins, C., Landivar, L. C., Ruppanner, L., & Scarborough, W. J. (2021). COVID-19 and the gender gap in work hours. *Gender, Work and Organization*, 28(S1), 101–112. <https://doi.org/10.1111/gwao.12506>
- Cortina, J., Sheng, Z., Keener, S., Keeler, K., Grubb, L., Schmitt, N., Tonidandel, S., Summerville, K., Heggstad, E., & Banks, G. (2020). From alpha to omega and beyond! A look at the past, present, and (possible) future of psychometric soundness in the Journal of Applied Psychology. *Journal of Applied Psychology*, 105. <https://doi.org/10.1037/apl0000815>
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4–19. <https://doi.org/10.1016/j.jesp.2015.07.006>
- Curran, P. J., & West, S. G. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16–29. <https://doi.org/10.1037/1082-989X.1.1.16>
- Day, D. V., & Sin, H.-P. (2011). Longitudinal tests of an integrative model of leader development: Charting and understanding developmental trajectories. *The Leadership Quarterly*, 22(3), 545–560. <https://doi.org/10.1016/j.leaqua.2011.04.011>
- Derks, B., Van Laar, C., & Ellemers, N. (2016). The queen bee phenomenon: Why women leaders distance themselves from junior women. *The Leadership Quarterly*, 27(3), 456–469. <https://doi.org/10.1016/j.leaqua.2015.12.007>
- Diehl, A. B., & Dzubinski, L. M. (2016). Making the invisible visible: A cross-sector analysis of gender-based leadership barriers. *Human Resource Development Quarterly*, 27(2), 181–206. <https://doi.org/10.1002/hrdq.21248>
- Diehl, A. B., Stephenson, A. L., Dzubinski, L. M., & Wang, D. C. (2020). Measuring the invisible: Development and validation of the gender bias scale for women leaders. *Human Resource Development Quarterly*, 31(3), 249–280. <https://doi.org/10.1002/hrdq.21389>
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573–598. <https://doi.org/10.1037/0033-295X.109.3.573>
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation* (1st ed.). Psychology Press. <https://doi.org/10.4324/9780203781906>
- Edwards, J. R., Caplan, R. D., & Harrison, R. V. (1998). Person-environment fit theory: Conceptual foundations, empirical evidence, and directions for future research. In C. L. Cooper (Ed.), *Theories of organizational stress* (pp. 28–67). Oxford University Press.
- European Commission. (2022). Proposal for a directive of the European Parliament and of the Council on improving the gender balance among non-executive directors of companies listed on stock exchanges and related measures. Retrieved November 1, 2023 from <https://eur-lex.europa.eu/legal-content/EN-DE/TEXT/?uri=CELEX:52012PC0614>
- Eurostat. (2023). Gross domestic product at market prices. Eurostat data browser. Retrieved September 9, 2023 from <https://ec.europa.eu/eurostat/databrowser/view/TEC00001/bookmark/table?lang=en&bookmarkId=e3030c9f-8b66-48ae-b1be-43199d1060eb>
- Faniko, K., Ellemers, N., & Derks, B. (2021). The queen bee phenomenon in academia 15 years after: Does it still exist, and if so, why? *British Journal of Social Psychology*, 60(2), e12408. <https://doi.org/10.1111/bjso.12408>
- Faniko, K., Ellemers, N., & Derks, B. (2022). Lack of ambition or lack of support? Diverging career experiences of men and women explain the persistence of gender bias. *Journal of Applied Social Psychology*, 52(9), 851–864. <https://doi.org/10.1111/jasp.12883>
- Flake, J. K., & Luong, R. (2021). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/qr32u>
- Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science*, 49(2), 78–88. <https://doi.org/10.1037/cbs0000069>
- Fowler, J. L., Fowler, D. S., & O’Gorman, J. G. (2021). Worth the investment? An examination of the organisational outcomes of a formal structured mentoring program. *Asia Pacific Journal of Human Resources*, 59(1), 109–131. <https://doi.org/10.1111/1744-7941.12252>
- Gabriel, A. S., Butts, M. M., Yuan, Z., Rosen, R. L., & Sliter, M. T. (2018). Further understanding incivility in the workplace: The effects of gender, agency, and communion. *Journal of Applied Psychology*, 103(4), 362–382. <https://doi.org/10.1037/apl0000289>
- Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26(3), 463–488. [https://doi.org/10.1016/S0149-2063\(00\)00043-X](https://doi.org/10.1016/S0149-2063(00)00043-X)
- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology*, 46(12), 1417–1432. [https://doi.org/10.1016/0895-4356\(93\)90142-N](https://doi.org/10.1016/0895-4356(93)90142-N)
- Hanges, P. J., & Gupta, V. (2004). Regional and climate clustering of societal cultures. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Culture, leadership and organizations. The GLOBE study of 62 societies* (pp. 178–218). Sage Publications.
- Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *The American Journal of Psychology*, 60(1), 1–29. <https://doi.org/10.2307/1417326>
- Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The multiple dimensions of gender stereotypes: A current look at men’s and women’s characterizations of others and themselves. *Frontiers in psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.00011>
- Hobfoll, S. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, 44, 513–524. <https://doi.org/10.1037/0003-066X.44.3.513>
- Hobfoll, S. (2002). Social and psychological resources and adaptation. *Review of General Psychology*, 6, 307–324. <https://doi.org/10.1037/1089-2680.6.4.307>
- Hobfoll, S., Halbesleben, J., Neveu, J.-P., & Westman, M. (2018). Conservation of resources in the organizational context: The reality of resources and their consequences. *Annual Review of Organizational Psychology and Organizational Behavior*, 5(1), 103–128. <https://doi.org/10.1146/annurev-orgpsych-032117-104640>
- Hogue, M., & Lord, R. G. (2007). A multilevel, complexity theory approach to understanding gender bias in leadership. *The Leadership Quarterly*, 18(4), 370–390. <https://doi.org/10.1016/j.leaqua.2007.04.006>
- Holtom, B. C., Mitchell, T. R., Lee, T. W., & Inderrieden, E. J. (2005). Shocks as causes of turnover: What they are and how organizations can manage them. *Human Resource Management*, 44(3), 337–352. <https://doi.org/10.1002/hrm.20074>
- Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, 102(3), 530–545. <https://doi.org/10.1037/apl0000103>
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership and organizations. The GLOBE study of 62 societies*. Sage Publications.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99–114. <https://doi.org/10.1007/s10869-011-9231-8>
- Jones, K. P., Peddie, C. I., Gilrane, V. L., King, E. B., & Gray, A. L. (2016). Not so subtle: A meta-analytic investigation of the correlates of subtle and overt discrimination. *Journal of Management*, 42(6), 1588–1613. <https://doi.org/10.1177/0149206313506466>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). semTools: Useful tools for structural equation modeling (R package 0.5–6). [Computer software]. Retrieved June 6, 2022, from <https://cran.r-project.org/web/packages/semTools/citation.html>
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127(3), 376–407. <https://doi.org/10.1037/0033-2909.127.3.376>
- Judge, T. A., Weiss, H. M., Kammeyer-Mueller, J. D., & Hulin, C. L. (2017). Job attitudes, job satisfaction, and job affect: A century of continuity and of change. *Journal of Applied Psychology*, 102(3), 356–374. <https://doi.org/10.1037/apl0000181>
- Judge, T. A., Zhang, S. (Carrie), & Glerum, D. R. (2020). Job satisfaction. In V. I. Sessa & N. A. Bowling (Eds.), *Essentials of job attitudes and other workplace psychological constructs* (pp. 207–241). Routledge.
- Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *The American Journal of Sociology*, 82(5), 965–990.
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). McGraw-Hill.
- 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(2), 482–510. <https://doi.org/10.1177/0149206308328508>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.
- Koch, A., D’Mello, S., & Sackett, P. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, 100(1), 128–161. <https://doi.org/10.1037/a0036734>
- Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616–642. <https://doi.org/10.1037/a0023557>
- Kossek, E. E., Su, R., & Wu, L. (2017). “Opting out” or “pushed out”? Integrating perspectives on women’s career equality for gender inclusion and interventions. *Journal of Management*, 43(1), 228–254. <https://doi.org/10.1177/0149206316671582>
- Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals’ fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. *Personnel Psychology*, 58(2), 281–342. <https://doi.org/10.1111/j.1744-6570.2005.00672.x>
- Kwok, N., Hanig, S., Brown, D. J., & Shen, W. (2018). How leader role identity influences the process of leader emergence: A social network analysis. *The Leadership Quarterly*, 29(6), 648–662. <https://doi.org/10.1016/j.leaqua.2018.04.003>
- Lacerenza, C. N., Reyes, D. L., Marlow, S. L., Joseph, D. L., & Salas, E. (2017). Leadership training design, delivery, and implementation: A meta-analysis. *Journal of Applied Psychology*, 102(12), 1686–1718. <https://doi.org/10.1037/apl0000241>
- Lee, T. W., & Mitchell, T. R. (1994). An alternative approach: The unfolding model of voluntary employee turnover. *The Academy of Management Review*, 19(1), 51–89. <https://doi.org/10.2307/258835>
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Lyness, K. S., & Grotto, A. R. (2018). Women and leadership in the United States: Are we closing the gender gap? *Annual Review of Organizational Psychology and Organizational Behavior*, 5(1), 227–265. <https://doi.org/10.1146/annurev-orgpsych-032117-104739>
- Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91(4), 777–785. <https://doi.org/10.1037/0021-9010.91.4.777>
- Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33(2), 181–220. https://doi.org/10.1207/s15327906mbr3302_1
- Mayo, M., Kakarika, M., Pastor, J. C., & Brutus, S. (2012). Aligning or inflating your leadership self-image? A longitudinal study of responses to peer feedback in MBA teams. *Academy of Management Learning and Education*, 11(4), 631–652. <https://doi.org/10.5465/amle.2010.0069>
- McFeely, S., & Wigert, B. (2019). This fixable problem costs U.S. businesses \$1 trillion. Gallup. Retrieved September 21, 2022, from <https://www.gallup.com/workplace/247391/fixable-problem-costs-businesses-trillion.aspx>
- McKinsey & Company. (2023). Women in the workplace 2023. McKinsey & Company. Retrieved October 11, 2023, from <https://www.mckinsey.com/featured-insights/diversity-and-inclusion/women-in-the-workplace?cid=other-eml-dre-mip-mck&hlkid=d623bb07a3c4494395f48004e3afe625&hctky=14741475&hdpid=b929a855-cc08-488d-80de-fbc29a07c441>
- McNeish, D. (2018). Thanks coefficient alpha, we’ll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>
- Me too. Movement. (2023). Get To Know Us | History & inception. Me too. Movement. Retrieved September 20, 2023 from <https://metoomvmt.org/get-to-know-us/history-inception/>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
- Milano, A. (2017). If you’ve been sexually harassed or assaulted, write ‘me too’ as a reply to this tweet. [Tweet]. *Twitter*. Retrieved November 11, 2020, from https://twitter.com/alyssa_milano/status/919659438700670976?lang=de
- Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, 17(6), 609–626. [https://doi.org/10.1002/\(SICI\)1099-1379\(199611\)17:6%3c609::AID-JOB1828%3e3.0.CO;2-K](https://doi.org/10.1002/(SICI)1099-1379(199611)17:6%3c609::AID-JOB1828%3e3.0.CO;2-K)
- Pfister, I. B., Jacobshagen, N., Kälin, W., & Semmer, N. K. (2020). How does appreciation lead to higher job satisfaction? *Journal of Managerial Psychology*, 35(6), 465–479. <https://doi.org/10.1108/JMP-12-2018-0555>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Raver, J. L., & Nishii, L. H. (2010). Once, twice, or three times as harmful? Ethnic harassment, gender harassment, and

- generalized workplace harassment. *Journal of Applied Psychology*, 95(2), 236–254. <https://doi.org/10.1037/a0018377>
- Revelle, W., & Condon, D. M. (2019). Reliability from α to ω : A tutorial. *Psychological Assessment*, 31(12), 1395–1411. <https://doi.org/10.1037/pas0000754>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Ryan, M. K., Haslam, S. A., Hersby, M. D., & Bongiorno, R. (2011). Think crisis—think female: The glass cliff and contextual variation in the think manager—think male stereotype. *Journal of Applied Psychology*, 96(3), 470–484. <https://doi.org/10.1037/a0022133>
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507–514. <https://doi.org/10.1007/BF02296192>
- Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*, 57(2), 95–100. <https://doi.org/10.1037/h0037128>
- Schlamp, S., Gerpott, F., & Voelpel, S. (2020). Same talk, different reaction? Communication, emergent leadership, and gender. *Journal of Managerial Psychology*, 36. <https://doi.org/10.1108/JMP-01-2019-0062>
- Shen, W., & Joseph, D. L. (2021). Gender and leadership: A criterion-focused review and research agenda. *Human Resource Management Review*, 31(2), 100765. <https://doi.org/10.1016/j.hrmr.2020.100765>
- Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, 45(1), 116–131. <https://doi.org/10.1509/jmkr.45.1.116>
- Trianadel, M. C., Jayasinghe, M., Pieper, J. R., Delgado, D. M., & Li, M. (2019). Perceived workplace gender discrimination and employee consequences: A meta-analysis and complementary studies considering country context. *Journal of Management*, 45(6), 2419–2447. <https://doi.org/10.1177/0149206318776772>
- van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS One*, 8(7), e68967. <https://doi.org/10.1371/journal.pone.0068967>
- Walter, S. L., Seibert, S. E., Goering, D., & O'Boyle, E. H. (2019). A tale of two sample sources: Do results from online panel data and conventional data converge? *Journal of Business and Psychology*, 34(4), 425–452. <https://doi.org/10.1007/s10869-018-9552-y>
- Williams, M. J., & Tiedens, L. Z. (2016). The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin*, 142(2), 165–197. <https://doi.org/10.1037/bul0000039>
- World Economic Forum. (2022). Global gender gap report 2022. World economic forum. Retrieved July 18, 2022, from https://www3.weforum.org/docs/WEF_GGGR_2022.pdf
- Zilla, C. (2022). Feminist foreign policy—concepts, core components and controversies (SWP Comment). German Institute for International and Security Affairs. <https://doi.org/10.18449/2022C48>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.