Sex Roles (2025) 91:31
https://doi.org/10.1007/s11199-025-01580-8

ORIGINAL RESEARCH

# Measuring Intrapersonal Awareness of Nonverbal Behavior Associated With Gender: Development and Validation of the Gendered Mannerisms Scale

**Cheng Yu**[1] · **Charlotte Chucky Tate**[2]

Accepted: 5 February 2025


## Abstract
This study addresses the underexplored domain of self-reports of gendered nonverbal behaviors. We developed the Gendered Mannerisms Scale (GMS) to bridge self-perception and social perception of gendered nonverbal cues. Across three studies, the paper validates the GMS as a reliable self-report psychometric tool and determines which gendered nonverbal behaviors show consistent gender differences for cisgender heterosexual women and men in U.S. samples. The first two studies employed exploratory factor analysis with distinct samples from a U.S. data vendor platform to establish and replicate the GMS's initial factor structure, demonstrating the scale's robustness in capturing the self-reports of gendered nonverbal behaviors. The third study used a measurement invariance analysis with a community-based sample, further demonstrating the scale's validity and its applicability to more racially/ethnically diverse and younger samples (than Studies 1 and 2). Finally, the four factors of self-reportable nonverbal gendered behavior consistently identified by the GMS are: (a) dynamic movements and open posture, (b) swaying gait, (c) gesticulations, and (d) closed posture. Importantly, the first factor (dynamic) may actually be more influenced by extraversion than gender.

**Keywords** Gender · Nonverbal Behavior · Gender Differences · Gender Roles · Scale Development · Mannerisms

In the landscape of societal norms and behavioral expectations, gender emerges as a pivotal and salient social categorization. The societal understandings of gender (in any society) permeate various facets of life, from intrapersonal acts like self-categorization to interpersonal acts such as dress codes and adult influence on child toy preferences (Tate et al., 2014; Weinraub et al., 1984). One of the subtle ways that gender is associated with both intra- and interpersonal acts is in body language–the unspoken ways people behave, especially in the presence of others. Historically, research has spotlighted the power of nonverbal cues in conveying emotions and intentions in social communication. Facial expressions, for instance, are rich in emotional content (Ekman, 1993), while body cues can signal underlying needs or states (Lopez et al., 2017). In daily interactions, nonverbal behaviors, from vocal pitch to facial expressions, play a crucial role in communication and impression formation (Shrout & Fiske, 1981). Particularly in the context of gender, understanding how these cues are employed and interpreted has been a central research focus since the 1970 s (Higgins et al., 1977; Kozlowski & Cutting, 1977; Miller, 1970).

A notable area of exploration in the United States and Europe has been how individuals use nonverbal cues interpersonally to discern another's gender categorization. Vocal pitch, for example, stands out as an accepted distinguishing feature between genders categories of women and men (Pernet & Belin, 2012). Yet, while much attention has been given to interpersonal perceptions, there remains a gap in the literature on intrapersonal perceptions of gendered nonverbal behaviors. This self-perception of gendered nonverbal behaviors is crucial because research findings suggest that social actors themselves might not always be cognizant of the motivations behind their actions (Nisbett & Wilson, 1977). Indeed, this lack of self-understanding might imply a divergence between how individuals perceive and interpret their own nonverbal behaviors (intrapersonal dynamics) and how these behaviors are perceived and interpreted by others (interpersonal dynamics). By examining self-perceptions

✉ Cheng Yu
cheng.yu@utsa.edu

1 The University of Texas at San Antonio, San Antonio, USA

2 San Francisco State University, San Francisco, USA

Published online: 14 April 2025

of gendered nonverbal behavior, researchers can bridge the divide between how individuals understand their gender and how others perceive their gender.

In the current research, we developed a self-report measure to assess the nonverbal cues that have been most consistently associated with interpersonal perceptions of gender groups. The literature review encompassed two main areas: (1) interpersonal gender category recognition through walking motions, and (2) nonverbal communication dimensions, including sitting postures, talking gesticulations, and vocal pitch, that show differences by actor gender categories (namely, women and men). The insights gained from these areas of focus informed the development of items for the new measure, the Gendered Mannerisms Scale.

## Gender Recognition Through Walking Motions or Gait

Past research, mainly conducted in the United States, has examined interpersonal dynamics through which observers determine an individual's gender based on their walking patterns (Barclay et al., 1978; Cutting & Kozlowski, 1977; Kozlowski & Cutting, 1977). A meta-analysis conducted by Pollick et al. (2005) synthesized results from various experiments, indicating that observers were able to accurately identify gender from walking motions, achieving approximately 66% accuracy for side views and about 71% for frontal and oblique viewpoints. Considering that these experiments focused solely on distinguishing between women and men, with only two options available, a theoretical accuracy rate would be expected to hover around a 50–50 probability (assuming equal presentation of the two types of targets). The actual observed accuracy rate was approximately 66% and 71% and did not greatly exceed this theoretical baseline, suggesting that there are other aspects of the nuanced interplay between nonverbal cues and gender identity that remain unexplored and not fully understood.

Several variables might contribute to this accuracy rate. Among them, of course, are variations in research methodologies and the depth of information provided in different studies (Pollick et al., 2005). Additionally, as the understanding of gender identity has evolved beyond a binary framework, with modern perspectives recognizing a spectrum of gender identities that do not fit neatly into traditional male or female categories (Hyde et al., 2019; Richards et al., 2016; Tate et al., 2013, 2014), there is (and has always been) room for variability in walking patterns that might not align strictly with traditional "female" or "male" gaits—which are very often heterosexual expectations about behavior (cf., Dean & Tate, 2017). Individual actors will show variation based on their internalizations of gendered (and heterosexual) stereotypes, which means that the actor's walking behavior is not necessarily understood in the same way by the perceiver—showing the tension between the intrapersonal and the interpersonal understandings of gender stereotypes. This evolving understanding and the inherent diversity of gender expressions could contribute to the modest accuracy rate observed in gender recognition based on walking motions.

Another potential contributor to the accuracy rates observed in gait is the impact of (intrapersonal) personality on nonverbal behaviors (in addition to any interpersonal gender-based expectations). Personality traits, such as higher extraversion and conscientiousness, along with lower neuroticism, are associated with faster gait speed and more efficient, confident movement (Stephan et al., 2018). Consequently, observers might mistakenly interpret these personality-driven nonverbal cues as gender-specific, complicating the accuracy of gender recognition. This overlap underscores the importance of differentiating between the effects of gender and personality on nonverbal behaviors, especially when judgments of gender are likely to be influenced by a variety of factors. This interplay between psychological traits and nonverbal cues highlights the complexity in discerning gender and the potential for misattribution between personality traits and gendered behaviors—all of this in addition to the intrapersonal production and understanding of gendered behavior.

## Sitting, Talking, and Vocal Pitch From a Gendered Perspective

Other nonverbal cues related to gender have been studied in the literature, including sitting postures, conversational gestures (viz., gesticulations), and vocal pitch. Research indicates binary gender differences in these areas when participants are grouped into self-categorizations as women and men (or from researcher classifications of participants into these categories). For instance, men on average display expansive, open sitting postures, while women on average display more reserved, closed postures (Cashdan, 1998; Vrugt & Luyerink, 2000). In the context of a conversation, LaFrance et al. (2003) found that women in the United States use more expressive gestures and smile more than men. Additionally, Weston et al. (2015) observed that in the United Kingdom, women typically have higher-pitched voices. Pernet and Belin (2012) also demonstrated that participants in the United Kingdom could identify the gender of an individual based on voice pitch, particularly when other vocal characteristics were less distinct.

## The Gendered Mannerisms Scale

While past research has examined societal perceptions and interpersonal assessments of gender identity via gendered mannerisms as they are culturally associated with the two major gender categories, these studies do not capture an individual's self-understanding and awareness of enacting these specific gendered mannerisms. Consequently, the development of a self-report scale to assess individual awareness of enacting gendered mannerisms in everyday life is necessary to determine whether individuals are aware of these nonverbal behaviors. One of the only ways to provide an answer to this question is to allow individuals to report their perceptions regarding their nonverbal communication in the four domains that we have identified: (a) walking gait, (b) sitting postures, (c) gesticulation, and (d) vocal pitch. Moreover, in the studies presented below, we have focused exclusively on gendered mannerisms among sexual majority individuals. We did not make this decision lightly, and we believe that in many circumstances regarding questions of gender, there is no a priori reason to exclude sexual minority respondents. However, for the specific phenomenon of mannerisms, we believed that focusing on sexual majority respondents initially was allowable for two principled reasons.

The first reason was a pragmatic one and involved sample size. The number of sexual minority respondents combined across our three studies (and over all labels) was just over 100, which does not provide enough statistical power to describe meaningful or likely replicable patterns. We also did not want to reinforce the popular though suboptimal practice of combining different sexual orientation respondent groups into one large LGB category, which, for this specific phenomenon could mask subtleties in how lesbian and bisexual women, for example, report mannerisms. The second principled reason involved taking a holistic perspective on the role of nonverbal behaviors in social communication, which acknowledges that these behaviors can signal both gender and sexual identity simultaneously or separately. For example, gay men have been shown to exhibit more self-touch behaviors and adopt less stereotypically masculine postures in dyadic interactions compared to heterosexual men (Knöfler & Imhof, 2007). In this way, trying to analyze sexual majority and minority participants' data together would potentially mask nuances on mannerisms that covary by sexual orientation (e.g., gay and heterosexual) within the same gender identity (e.g., men).

## Overview of Current Research

The objectives of this research were to (a) develop a robust, multifaceted scale for assessing gendered nonverbal behaviors; (b) to investigate the initial factor structure of the GMS scores using exploratory factor analysis (EFA) in two separate samples; and (c) to offer preliminary evidence of the construct validity of these scores using measurement invariance (MI) on a third separate sample. All analyzed samples were from the United States because certain mannerisms, such as gesticulation frequencies, may vary across different cultures (Graziano & Gullberg, 2024). We also recognize that certain disabilities (e.g., mobility-based or sensory-based) may skew the results of the GMS's walking and talking questions. Consequently, data from these individuals will be excluded from our analysis.

Study 1 provided an initial test of the factor structure and psychometric properties of the GMS using EFA. Study 2 replicated Study 1, with a different sample to determine whether the initial factor structure is replicated and refine the items on the GMS as needed. Additionally, both Study 1 and Study 2 also established initial concurrent and divergent validity for the GMS by determining gender group differences on the extracted factors (concurrent) and by determining the association of personality traits with the extracted factors and further separating the gender differences from personality traits (divergent). In sum, Studies 1 and 2 test construct validity by examining the content, structure, and theoretical alignment of the GMS, ensuring reliability and identifying gender differences consistent with existing research. Study 3 employed a MI analysis via confirmatory factor analysis (CFA) on the GMS items (from Study 2), with a younger and more racially/ethnically diverse sample than Studies 1 and 2. Thus, Study 3 focuses on measurement invariance to confirm equivalence across gender and extends generalizability by validating the GMS in a more diverse demographic sample. This sequential approach, starting with two EFAs and then a MI, ensures a thorough validation process that is not likely to miss subtleties between the two EFA samples, before using MI in its true confirmatory role.

Of special note, to avoid androcentric bias in reporting on gender research (e.g., Hegarty & Buechel, 2006), we centered women participants as the focus of our comparative statements. Consequently, for all three studies, we present the gender difference analyses in the following consistent manner: positive signs (+) indicate women had higher scores than men and negative signs (-) indicate that women had lower scores than men. Thus, all *t*-values, effect size estimates (e.g., *d*s), and partial correlations presented in text and tables can be interpreted in the same quick manner to determine the direction of the gender difference. Finally, following statistical best practices, all *t*-tests were conducted with the Welch correction by default (see Delacre et al., 2017; Zimmerman, 2004) to account for even subtle inequality of variances between gender groups.

## Study 1: Initial GMS Factor Structure

Study 1 was conducted to establish the initial factor structure of the full GMS using EFA. We recruited a sample from an online panel service, Amazon Mechanical Turk (MTurk, hereafter) to determine the initial factor structure of the full set of GMS items (see Supplement A in the online supplement). This study served as the starting point for understanding the factor structure that was further refined in Study 2.

This study was conducted with two initial hypotheses: (a) the GMS will at least reveal three factors (e.g., walking, talking, and postures) that we expected to be consistent with the observational literatures and (b) women and men will self-report differences on the GMS consistent with the observational literature gender differences. In addition to the GMS, we included a personality scale (Soto & John, 2017) to determine divergent validity for the GMS factor derived. As argued above, including a self-report of personality dimensions makes a better distinction between the roles of gender and personality in nonverbal behaviors.

## Study 1 Method

### Participants

Data from 486 participants were collected via MTurk, with respondents required to reside in the United States. To ensure data integrity, duplicate IP addresses were excluded, and listwise deletion was performed on analyzed items to handle missing data ($n_{\text{excluded}} = 164$). It is important to note that missing data primarily stems from participants with disabilities, who completed a modified version of the initial GMS scale, as well as from other reasons, such as incomplete or inattentive responses. Data from 322 participants were analyzed ($M_{age} = 39.97$, $SD_{age} = 12.16$; 186 or 57.8% women, 136 or 42.2% men, all of whom were cisgender and heterosexual). Responses from individuals self-reporting disabilities in their arms, fingers, hands, or legs were not analyzed. The focus was on cisgender heterosexual respondents who are able-bodied, given their numerical majority in the United States and the empirical studies on gendered mannerisms to date. The majority ($n = 242$, 75.2%) identified as European American/White, followed by smaller percentages identifying as Asian American/Asian ($n = 22$, 6.8%), African American/Black ($n = 21$, 6.5%), Hispanic/Latino ($n = 16$, 5.0%), Native American ($n = 5$, 1.6%), Middle Eastern ($n = 1$, 0.3%), Pacific Islander ($n = 1$, 0.3%), and another unlisted identity ($n = 1$, 0.3%). Some participants chose multiple ethnicities or opted not to declare; thus, the percentages do not sum up to exactly 100%.

### Measures

#### The Gendered Mannerisms Scale (GMS)

The Gendered Mannerisms Scale (GMS) was developed to measure people's awareness of their enactment of common, daily gendered nonverbal behaviors. First, we developed scale items based on gendered observational literatures in gait, postures, gesticulation, and vocal pitch. For questions involving specific bodily movements, we created separate questions to reflect specific body parts of which we believed the participants to have the most awareness when behaving, that is arms, elbows, hands, fingers, feet, waist, and legs. Facial expressions (which are not always in conscious awareness of the actor) were excluded. Thus, the items presented on the GMS match the focus of the interpersonal mannerisms research and should be the dimensions of which any actor is most aware in daily life. Actions that were infrequent, specific to certain groups, potentially influenced by factors other than gender—such as cultural or religious beliefs—and possible cultural fads were also excluded. Thus, ensured the scale's focus remained on typical, repeatable gendered mannerisms.

Initially, the 35-items GMS focused on three key areas identified in the literature: (a) walking, (b) sitting posture, and (c) gesticulation. Additionally, it included nuanced mannerisms involving both large (e.g., shoulders, hips) and small (e.g., hands, fingers) body movements. The scale also included two items about vocal pitch. The GMS items were designed to assess these nonverbal behaviors from a first-person or intrapersonal perspective. Participants responded to each item using a frequency scale from 1 (*Not at all/Never*) to 5 (*Always*). Since the frequency response scale used for body parts does not apply to vocal pitch, we employed a separate response scale for vocal pitch. For vocal pitch, the response scale ranged from 0 (*Not at all*) to 4 (*Extremely*) and was used to measure both high pitch and low pitch independently. Because the vocal pitch items were on a different scale than the rest of the GMS items, the vocal pitch items were analyzed separately even while presented in the same instrument to participants.

Additionally, to be inclusive and accommodating, we included a disability inventory (see Supplement B in the online supplement) prior to seeing the GMS so that we could branch participants indicating any disability affecting gait, posture, gesticulation, or vocal pitch were directed to relevant GMS sections to accommodate their specific disability. Example GMS versions for individuals with lower-body disabilities and those who are deaf are provided in Supplements C and D in the online supplement.

#### Big Five Inventory 2.0 (BFI- 2)

This 60-item tool assesses five broad personality traits: open-mindedness, conscientiousness, extraversion, agreeableness, and negative emotionality (Soto & John, 2017). Responses are on a 1 (*Disagree strongly*)—5 (*Agree strongly*) scale, with items like "Tends to feel depressed" (negative emotionality item) or "Is outgoing" (extraversion item). Higher scores indicate stronger presence of the trait, and lower scores indicate the opposite of the trait being present (e.g., introversion is the opposite to extraversion). In our studies, the BFI- 2.0 demonstrated good reliability, with Cronbach's alphas for the five factors ranging from.85 to.93. The inclusion of BFI- 2.0 aimed to validate the GMS's divergent validity, as well as to determine whether personality traits influenced the self-reports of the nonverbal behaviors queried on the GMS.

#### The Two-Question Method of Assessing Gender Identity (2QAGI)

The 2QAGI (Tate et al., 2013) discerns cisgender and transgender spectrum self-categorizations. The first question inquiries about current gender identity with the options *woman*, *man*, *trans woman*, *trans man*, *nonbinary/genderqueer*, and *intersex*. The second question inquiries about gender assigned at birth (with three options: *female*, *male*, and *intersex*). Cisgender participants have similar answers for both questions (e.g., *woman*, *female*), while transgender spectrum participants answer differently to the two questions.

### Procedure

Participants viewed an implied consent agreement (because the study was delivered online), detailing the study and verifying if they were 18 years or older. Those not meeting the age criterion or declining consent were directed out of the survey. Eligible participants proceeded to answer (a) the 2QAGI for participant gender identity, (b) the disability inventory status that preceded the GMS (if participants identified as non-disabled, they completed the full 35-item GMS, see Supplement A in the online supplement; those identifying as disabled answered further questions about their disability and were directed to a GMS version tailored to their status, see Supplements C and D in the online supplement); (c) the BFI- 2, and (d) demographic questions (see Supplement E in the online supplement), including age, sexual orientation, political affiliation, ethnicity, U.S. state they live in, and educational level. The entire study was delivered on the Qualtrics survey deployment platform.

### Assumption Checking for the Exploratory Factor Analysis

Before we ran the EFA, we conducted three statistical tests to ensure that EFA was warranted on these data. We first conducted the KMO test of sampling adequacy and found that the 35-items were adequate for EFA, $KMO_{overall} = .91$ (which is above the minimum of.60), and indicated adequate sampling of the items for EFA. Additionally, the Bartlett's test suggested the correlation matrix structure is likely not perfectly orthogonal, $\chi^2(595) = 5561.61$, $p < .001$. Finally, we estimated the determinant from the correlation matrix for the 35 items and found the determinant to be.00000019 (not computationally zero), indicating that singularity of the correlation matrix was absent.

### Assumption Checking for the t-Tests

One of the core assumptions of the independent samples (aka two-sample) *t*-tests is the homogeneity (or equality) of variances between the samples. One test of this assumption is Levene's test (Levene, 1960) on each outcome variable across the two samples. The results for our four composites for the GMS indicated that one of the four factors violated this assumption in Study 1, suggesting that the variances between groups were not equal (see Table S1 in the online supplement), which biases the result of the Student version of the *t*-test (Welch, 1938), producing either Type I or Type II errors (cf. Delacre et al., 2017; Zimmerman, 2004). Rather than simply adjusting on the violations, following current best practices in applied statistics (see Delacre et al., 2017), we used Welch *t*-test for all four outcomes because this method provides more robust results even for slight inequalities in variance, thereby reducing the risk of Type I or Type II errors. By defaulting to the Welch *t*-test (and its degrees of freedom correction to the theoretical *t*-distribution used), we ensured that our analysis remained statistically sound in the presence of any variance disparities between the samples, enhancing the validity of our findings.

## Study 1 Results

### Exploratory Factor Analysis

In the investigation of the factor structure of the GMS, an EFA was employed using the psych package (Revelle, 2018) within the R statistical environment. Three methods were utilized to determine the number of factors to extract: Kaiser's first-generation jiffy (i.e., extracted eigenvalues over 1.00) suggested the extraction of eight factors, but with three factors being barely over the value 1.00 (see Figure S1 in the online supplement); Horn's parallel analysis indicated nine factors (see Figure S2 in the online supplement); and Cattell's scree plot pointed to four factors (see Figure S3 in the online supplement). Considering the proximity of Cattell's scree plot to the hypothesized model of at least three factors—gait, closed posture, and gesticulation—and that only five factors from the Kaiser jiffy were considerably over the value of 1.00, we decided to extract four factors initially.

For the 4-factor solution, we used both varimax and oblimin rotations. Both rotations revealed that 11 items did not meet the 0.45 loading standard that we chose a priori and were thus removed. The resulting GMS of Study 1 contained 24 items. Factor correlations ranged from $r=.25$ to $r=.39$ (see Table S2 in the online supplement), suggesting moderate independence between factors. However, based on the theoretical framework which posits that the constructs of self-perceptions of nonverbal behaviors are likely interrelated (Bem, 1967; Friedman, 2019), the oblimin rotation was selected for the final analysis. Details of the factor loadings, means, and standard deviations for the oblimin rotation are presented in Table 1 (with varimax rotation information available in Table S3 in the online supplement).

The first factor was named *Dynamic Movements and Posture* and included a total of 12 items, such as "How often do you shake your shoulders while you are sitting?" and "How often do you shake your legs while you are standing?" The second factor was named *Gesticulation* and comprised four items, such as "How often do you move your hands while you are talking?" and "How often do you 'talk with hands'?" The third factor was named *Swaying Gait,* and also consisted of four items, such as "How often do you sway your shoulders while you are walking?" and "How often do you sway your hands while you are walking?" The fourth and final factor was named *Closed Posture* and included four items, such as "How often do you cross your legs or ankle while you are sitting?" and "How often do you tap one of your feet while you are sitting?" The items within each factor showed excellent to acceptable reliability using the DeVellis standards (DeVellis, 2016): Dynamic Movements and Posture $\alpha=.90$; Gesticulation $\alpha=.86$; Swaying Gait $\alpha=.75$; and Closed Posture $\alpha=.71$. Additionally, descriptive statistics for each factor are presented in Tables S4 and S5 in the online supplement.

## Gender Differences on Factor Composites as Concurrent Validity

We conducted Welch two-sample *t*-tests on each of the four factors of the GMS (see Table 2), all of which were relatively uncorrelated (all *r*s $\leq .40$). Men ($M=2.37$, $SD=0.81$) scored significantly higher on average than women ($M=2.07$, $SD=0.61$) in Dynamic Movements and Posture, $t(239.16)=-3.72$, $p<.001$, $d=-0.44$. Conversely, women

**Table 1** GMS Exploratory Factor Analysis Factor Loadings (Study 1, Oblimin)

| GMS Item | Factor Loading | | | |
|---|---|---|---|---|
| | PA1 | PA2 | PA3 | PA4 |
| Factor 1: Dynamic Movements and Posture | | | | |
| 20. When sitting, how often do you move/shake your shoulders while you are sitting? | **.81** | .01 | .06 | -.12 |
| 12. When standing, how often do you shake your legs while you are standing? | **.75** | -.08 | .03 | .11 |
| 10. When standing, how often do you move/shake your shoulders while you are standing? | **.76** | -.05 | .13 | -.07 |
| 9. When standing, how often do you tap one of your feet while you are standing? | **.62** | -.11 | .09 | .28 |
| 11. When standing, how often do you rest your hand on your face while you are standing? | **.69** | .05 | -.08 | -.03 |
| 16. When sitting, how often do you have your hands on your waist and hips while you are sitting? | **.64** | .03 | .10 | -.03 |
| 25. When sitting, how often do you hold your elbows while you are sitting? | **.61** | .12 | .01 | .08 |
| 8. When standing, how often do you cross your legs or ankle while you are standing? | **.55** | -.01 | -.04 | .27 |
| 15. When sitting, how often do you move your waist and hips while you are sitting? | **.62** | .02 | .06 | -.13 |
| 14. When standing, how often do you hold your elbows while you are standing? | **.55** | .16 | -.10 | .12 |
| 13. When standing, how often do you move your arms while you are standing? | **.50** | .16 | .08 | .04 |
| 27. When sitting, how often do you move your arms while you are sitting? | **.47** | .21 | -.02 | -.04 |
| Factor 2: Gesticulation | | | | |
| 31. When talking, how often do you move your hands while you are talking? | -.10 | **.84** | .02 | .05 |
| 30. When talking, how often do you move your arms while you are talking | .06 | **.78** | .04 | -.03 |
| 29. When talking, how often do you "talk with your hands"? | -.09 | **.71** | .03 | .11 |
| 32. When talking, how often do you move your fingers while you are talking? | .24 | **.72** | .00 | -.09 |
| Factor 3: Closed Posture | | | | |
| 18. When sitting, how often do you cross your legs or ankle while you are sitting? | -.18 | .10 | .12 | **.62** |
| 26. When sitting, how often do you cross your legs over the knee while you are sitting? | .02 | .09 | .11 | **.45** |
| Factor 4: Swaying Gait | | | | |
| 3. When walking, how often do you sway your shoulders while you are walking? | .27 | -.02 | **.68** | -.08 |
| 2. When walking, how often do you sway your hips while you are walking? | -.02 | .05 | **.67** | .08 |
| 4. When walking, how often do you sway your hands while you are walking? | .02 | .04 | **.61** | -.01 |
| 1. When walking, how often do you sway your arms while you are walking? | -.19 | .19 | **.51** | .15 |

($M$ = 3.24, $SD$ = 0.75) scored higher on average than men ($M$ = 2.96, $SD$ = 0.84) in Gesticulation, $t(306.34) = 3.06$, $p = .002$, $d = 0.34$; and women scored higher ($M$ = 3.41, $SD$ = 0.72) on average than men ($M$ = 2.96, $SD$ = 0.91) on Closed Posture, $t(249.3) = 4.82$, $p < .001$, $d = 0.56$. No significant differences were observed for Swaying Gait between women ($M$ = 2.80, $SD$ = 0.77) and men ($M$ = 2.76, $SD$ = 0.83), $t(277.52) = 0.42$, $p = .670$, $d = 0.05$.

## Gender Differences on the GMS Factors as Divergent Validity With Personality Traits

To determine divergent validity of the gender differences with personality traits on self-reported GMS factors, we conducted several partial correlation analyses (see Table 3; for the full correlation matrix please view Table 4). The partial correlations (*pr*s) isolate the gender difference on each GMS factor to remove all five personality traits (i.e., openness, conscientiousness, extraversion, agreeableness, and emotional stability [measured by the BFI- 2]) from the focal correlation between the gender difference and the GMS factor. This procedure establishes whether the gender difference is uniquely present (by mathematically removing the correlations of all five personality traits from the relationship). Further, the partial correlations also allow for the unique correlation of any single personality trait any GMS factor, removing the remaining personality traits and the gender difference. This procedure is therefore the best test of divergent validity—answering the question: Is the gender difference on any GMS factor attributable to gender or is it confounded with personality? Significant partial correlations were observed between gender and Dynamic Movements and Posture, $pr(315) = -.20$, $p < .001$, with women

**Table 2** Results of Welch Two-Samples t-Tests on the Four Composites From the GMS (Studies 1 & 2)

| Session | | Cis Heterosexual Women | | | Cis Heterosexual Men | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *M* | *SD* | *N* | *M* | *SD* | *n* | *t* | *p* | *d* |
| Study 1 | Dynamic Movements† | 2.07 | 0.61 | 186 | 2.37 | 0.81 | 136 | − 3.72 | <.001 | − 0.44 |
| | Gesticulation | 3.24 | 0.84 | 186 | 2.96 | 0.75 | 136 | 3.06 | .002 | 0.34 |
| | Closed Posture | 3.41 | 0.72 | 186 | 2.96 | 0.91 | 136 | 4.82 | <.001 | 0.56 |
| | Swaying Gait | 2.80 | 0.77 | 186 | 2.76 | 0.83 | 136 | 0.42 | .670 | 0.048 |
| Study 2 | Dynamic Movements† | 2.03 | 0.60 | 193 | 1.91 | 0.53 | 150 | 1.95 | .052 | 0.21 |
| | Gesticulation | 3.24 | 0.80 | 193 | 2.78 | 0.86 | 150 | 5.03 | <.001 | 0.55 |
| | Closed Posture | 3.53 | 0.71 | 193 | 2.93 | 0.88 | 150 | 6.8 | <.001 | 0.76 |
| | Swaying Gait | 2.73 | 0.81 | 193 | 2.46 | 0.79 | 150 | 3.04 | .003 | 0.33 |

† Dynamic Movements = Dynamic Movements and Dynamic Posture. The *d* represents Cohen's *d* effect size for two independent samples

**Table 3** Results of Partial Correlations on Cisgender Heterosexual Group Gender Differences While Removing the Big Five Personality From Each of the Composite Differences (Studies 1 & 2)

| Session | | Gender Difference | O | C | E | A | N |
|---|---|---|---|---|---|---|---|
| Study 1 | Dynamic Movements† | -.20** | .05 | -.04 | .26** | -.04 | .12* |
| | Gesticulation | .17* | -.04 | .07 | .03 | .09 | .07 |
| | Closed Posture | .28** | -.02 | .08 | -.03 | .0 | .14* |
| | Swaying Gait | .04 | -.02 | .05 | .13* | .03 | .03 |
| Study 2 | Dynamic Movements† | .10 | .08 | .02 | .12* | .08 | .08 |
| | Gesticulation | .28** | -.11* | .90 | .18** | .07 | -.03 |
| | Closed Posture | .37* | -.09 | .04 | .02 | .01 | .10 |
| | Swaying Gait | .15* | .11 | .01 | .09 | -.00 | .04 |

* = $p < .05$, ** = $p < .001$. † = Dynamic Movements and Dynamic Posture. For the gender difference variable, women were coded as 0 and men were coded as 1. (Thus, negative values indicate that women had a higher average than men, while positive values indicate that men had higher average than women.) O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Negative Emotionality. Each letter-named column for personality traits aggregates the effect of that trait across all participants, and removing the gender difference (and the other four traits) from the partial correlation

**Table 4** Correlation Matrix Between GMS Factors and BFI-2.0 Factors for Studies 1 and 2

| Session | Dynamic Movement | Gesticulation | Closed Posture (2 items) | Closed Posture (4 items)[a] | Swaying Gait |
|---|---|---|---|---|---|
| Study 1 ($n = 322$) | | | | | |
| O | .29** | .12* | .11* | .20** | .15* |
| C | .21** | .21** | .22** | .28** | .19** |
| E | .39** | .16* | .11* | .18* | .23** |
| A | .23** | .22** | .20** | .28** | .19** |
| N | .29** | .20** | .24** | .26** | .18* |
| Study 2 ($n = 343$) | | | | | |
| O | .31** | .08 | .00 | – | .21* |
| C | .28** | .17* | .06 | – | .11* |
| E | .37** | .23** | .04 | – | .19* |
| A | .32** | .19* | .06 | – | .13* |
| N | .27** | .10 | .10 | – | .12* |

* $= p < .05$, ** $= p < .001$. O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Negative Emotionality

[a]In the exploratory factor analysis for Study 1, 4 items were suggested to comprise the Closed Posture factor. However, in the second study, two of these items failed to load properly. As a result, we decided to proceed with the two-item version of the Closed Posture factor as the definitive measure. Consequently, we performed correlation analyses for both the two-item and four-item versions to compare their relationships with other variables in Study 1

reporting lower scores than men, controlling for all personality traits; on Gesticulation, $pr(315) = .17$, $p = .002$, with women reporting higher scores than men, controlling for all personality traits; and on Closed Posture, $pr(315) = .28$, $p < .001$, with women reporting higher scores than men, controlling for all personality traits. However, no significant partial correlation was found between gender and Swaying Gait, $pr(315) = .04$, $p = .508$ showing no gender difference when controlling for all personality traits.

Additionally, two personality dimensions–Extraversion and Negative Emotionality–were significantly correlated with certain GMS factors when controlling for the remaining personality traits and the gender difference. Specifically, extraversion showed a significant partial correlation, controlling for the remaining personality traits and the gender difference, with Dynamic Movements and Posture, $pr(315) = .26$, $p < .001$, such that higher extraversion was associated with more dynamic movements and posture; and with Swaying Gait, $pr(315) = .13$, $p = .018$, such that higher extraversion was associated with more swaying gait. Negative Emotionality showed a significant partial correlation with Dynamic Movements and Posture, controlling for the remaining personality traits and the gender difference, $pr(315) = .122$, $p = .030$, such that higher emotional stability was associated with more dynamic movements and posture; and with Closed Posture, $pr(315) = .14$, $p = .013$, such that higher emotional stability was associated with more closed posture.

## Gender Differences on Swaying Gaits Items as Concurrent Validity

We employed a multivariate analysis of variance (MANOVA) to assess gender differences in self-reported frequency of swaying hips and shoulders while walking, given their substantial correlation ($r_{\text{overall}} = .53$). Women ($M = 2.87$, $SE = 0.06$) on average reported significantly more frequent hip swaying compared to men ($M = 2.10$, $SE = 0.07$), $F_{\text{stepdown}}(1, 319) = 34.33$, $p < .001$, $\hat{d} = 0.81$. Conversely, women ($M_{adj} = 2.15$, $SE = 0.08$) on average indicated significantly less frequent shoulder swaying than men ($M_{\text{adj}} = 2.50$, $SE = 0.09$), when controlling for hip swaying, $F_{\text{stepdown}}(1, 318) = 7.67$, $p = .006$, $\hat{d} = -0.31$ (Table 5).

## Gender Differences on Vocal Pitch Items as Concurrent Validity

Two separate Welch two-sample $t$-tests were chosen because the high-pitch voice and low-pitch voice items were relatively uncorrelated ($r_{\text{overall}} = -.25$). Women ($M = 2.69$, $SD = 0.92$) on average reported significantly more high-pitched voices compared to men ($M = 2.28$, $SD = 0.96$), $t(320.72) = 4.17$, $p < .001$, $d = 0.44$. Conversely, women ($M = 2.56$, $SD = 0.91$) reported significantly less low-pitched voices than men ($M = 3.23$, $SD = 0.87$), $t(341.41) = -7.30$, $p < .001$, $d = -0.75$ (Table 6).

**Table 5** Results of Multivariate Analysis of Variance on Cisgender Heterosexual Group Differences in Response to the Walk—Sway Hips/Shoulders Questions (Studies 1 & 2)

| | | Respondent Group | | | | |
|---|---|---|---|---|---|---|
| | | Cis Heterosexual Women | Cis Heterosexual Men | $F$ | $df_w$ | $\hat{d}$ |
| Session | | $N$= 186 | $N$= 136 | | | |
| Study 1 | Sway hips | 2.87 (0.06) | 2.10 (0.07) | 34.33** | 319 | 0.81 |
| | Sway shoulders | 2.15 (0.08) | 2.50 (0.09) | 7.67* | 318 | − 0.31 |
| Study 2 | Sway hips | 2.62 (0.05) | 1.76 (0.057) | 63.65** | 340 | 1.20 |
| | Sway shoulders | 2.10 (0.08) | 2.13 (0.09) | 0.06 | 339 | − 0.03 |

* $=p<.01$, ** $=p<.001$. The $\hat{d}$ represents the Cohen's $d$ effect size but using the variability of the whole ANOVA per stepdown. Negative $\hat{d}$ values indicate that women had a higher average than men, while positive values indicate that men had higher average than women

## Study 1 Discussion

The first study aimed to clarify the structure of the GMS and to investigate gender differences in self-reported nonverbal behaviors. The EFA results provided a preliminary four-factor structure for the GMS that included the expected domains of walking, sitting, and gesticulation. These initial findings were in line with prior research, which has reported distinct gendered patterns in nonverbal behaviors (Johnson & Tassinary, 2005; Kozlowski & Cutting, 1977). The congruence between the GMS structure and the anticipated domains offers early convergence of individual's intrapersonal perceptions of gendered mannerisms with others' social perceptions of gendered mannerisms.

In addition, from the intrapersonal domains identified, we found participant gender group differences consistent with past literature, with men reporting greater awareness of enacting gendered dynamic movements and posture, and women reporting greater awareness of enacting gendered gesticulations and closed posture. These self-reported differences paralleled those observed in prior studies, reinforcing the GMS's capacity to capture the conscious aspect of these nonverbal behaviors. Notably, when personality factors were controlled, all gender differences remained, indicating that personality was not masking this effect. What is more we did observe that some nonverbal behaviors were correlated with personality traits, when removing the gender difference on the behaviors. Thus, personality may have an independent contribution to some nonverbal behaviors—even while a gender difference can be observed too. Vocal pitch results further supported prior interpersonal gender-group differences (e.g., Pernet & Belin, 2012), with women reporting higher frequencies of high-pitched voices, and men lower frequencies of low-pitched voices. Collectively, these results provided early evidence supporting the construct validity of the GMS.

## Study 2: Replication of GMS Factor Structure

Acknowledging that Study 1 findings represent preliminary evidence, we conducted an exact replication study (Study 2) to assess the robustness of the Study 1 EFA structure. We therefore provided the same 35-item GMS to another sample on MTurk. We conducted a second EFA to determine to ensure the reliability of the GMS structure before subjecting the scale to a MI in Study 3.

**Table 6** Results of Welch Two-Samples t-Tests on Cisgender Heterosexual Group Differences in Response to the Voice—High/Low Pitched Questions (Studies 1 & 2)

| | | Group | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Cis Heterosexual Women | | | Cis Heterosexual Men | | | | | |
| Session | | $M$ | $SD$ | $n$ | $M$ | $SD$ | $n$ | $t$ | $p$ | $d$ |
| Study 1 | High pitched | 2.69 | 0.92 | 228 | 2.28 | 0.96 | 155 | − 4.17 | <.001 | − 0.44 |
| | Low pitched | 2.56 | 0.91 | 228 | 3.23 | 0.87 | 155 | 7.29 | <.001 | 0.75 |
| Study 2 | High pitched | 2.67 | 0.92 | 241 | 2.26 | 0.90 | 199 | − 4.71 | <.001 | − 0.45 |
| | Low pitched | 2.64 | 0.94 | 241 | 3.23 | 0.84 | 199 | 6.88 | <.001 | 0.65 |

$d$= Cohen's $d$ effect size for two independent samples. Negative $d$ values indicate that women had a higher average than men, while positive values indicate that men had higher average than women

## Study 2 Method

### Participants, Measures, and Procedure

In Study 2, we replicated the method followed in Study 1, employing MTurk for data collection from a new set of U.S. residents ($N$= 754), ensuring a distinct sample by verifying that we had unique IP addresses between the samples. The demographic focus on cisgender heterosexual individuals without disabilities remained, with 343 participants ($M_{age}$= 41.77, $SD_{age}$= 12.68; $n$= 193 or 56.3% women, $n$= 149 or 43.4% men) included for analysis. The ethnic breakdown was predominantly European American/White ($n$= 264, 76.9%), with Asian American/Asian ($n$= 26, 7.6%), African American/Black ($n$= 18, 5.2%), Hispanic/Latino ($n$= 17, 5.0%). A smaller proportion of participants identified as Middle Eastern ($n$= 2, 0.6%), Native American ($n$= 1, 0.3%), or with an unlisted identity ($n$= 1, 0.3%). No participants identified as Pacific Islander. Since participants could select more than one ethnicity, the total percentages do not add up to 100%. Consistent with Study 1, the same measures and procedures were employed, as noted below.

### Measures

The same measures from Study 1 were used in Study 2: the 35-item Gender Mannerisms Scale (GMS), the 60-item Big Five Inventory 2.0 (BFI- 2), and the Two-Question Method of Assessing Gender Identity (2QAGI).

### Assumption Checking for the Exploratory Factor Analysis

Assumptions for Study 2 mirrored those of Study 1. The KMO test yielded $KMO_{overall}$ =.91, confirming sample adequacy of the items for EFA. The Bartlett's test, $\chi^2(595) = 6528.266$, $p$ <.001, indicated the correlation matrix was not entirely orthogonal. The determinant for the 35 items was.0000000868 (not computationally zero), suggesting absence of singularity in the correlation matrix. Thus, EFA was deemed appropriate.

### Assumption Check for the t-tests

As done in Study 1, in Study 2, we performed Levene's test (Levene, 1960) to determine whether the assumption of homogeneity of variances was met for the two-sample $t$-tests that we planned to conduct. The results showed that one of the four GMS composites violated this assumption (see Table S1 in the online supplement). Yet, as with Study 1, we used the recommended Welch $t$-test for all composites to account for any inequality of variances, detectable at $p$ <.05 or not (see Delacre et al., 2017; Zimmerman, 2004). Once again, using the Welch correction by default provides a robust interpretation (minimizing Type I and Type II errors), and helps ensure the reliability of the results for any level of variance inequality.

## Study 2 Results

### Exploratory Factor Analysis

Based on the Study 1 analysis, we tested a 4-factor solution with both varimax and oblimin rotations on the Study 2 sample to see whether the 4-factor solution could be recovered for this sample with the same item locations on factors spontaneously emerging. Using the oblimin rotation, in Study 2 we found the intercorrelations between the pairs of factors to be between $r$=.152 and $r$=.485, again showing a fair amount of independence.

Compared to Study 1, Study 2 showed the same four factors, but one factor had a different number of items loading on it. The first factor was Dynamic Movements and Posture (same 12 items from Study 1, α =.90), Gesticulation (same four items from Study 1, α =.90), and Swaying Gait (same four items from Study 1, α =.77). For the fourth factor (Closed Posture), there were only two items that loaded above 0.45 in both Study 1 and 2 (see Tables 1 and 7). To create consistency across the samples, we retained these two items for the Closed Posture in Study 1 (two items, $r$=.45) and Study 2 (two items, $r$=.40). For more details about loadings, means, and standard deviations see Table 7 and Table S6 in the online supplement. Descriptive statistics for each factor can also be found in Table 8 and Table S4 in the online supplement.

### Gender Differences on Factor Composites as Concurrent Validity

Similar to Study 1, four Welch $t$-tests were conducted for Study 2 (Table 2), with all factors being relatively uncorrelated (all $r$s ≤.477). Replicating Study 1, for the Gesticulation factor, women ($M$= 3.24, $SD$= 0.80), on average, scored significantly higher than men ($M$= 2.78, $SD$= 0.86), $t(308.51) = 5.03$, $p$ <.001, $d$= 0.55; for the Closed Posture factor, women ($M$= 3.53, $SD$= 0.71), on average, scored significantly higher than men ($M$= 2.93, $SD$= 0.88), $t(281.82) = 6.81$, $p$ <.001, $d$= 0.76; and, for the Swaying Gait factor, women ($M$= 2.73, $SD$= 0.81) scored significantly higher than men ($M$= 2.46, $SD$= 0.79), $t(324.14) = 3.04$, $p$ =.003, $d$= 0.33. However, no significant differences found between women ($M$= 2.03, $SD$= 0.60) and men ($M$= 1.91, $SD$=

**Table 7** GMS Exploratory Factor Analysis Factor Loadings (Study 2, Oblimin)

| GMS Item | Factor Loading | | | |
|---|---|---|---|---|
| | PA1 | PA2 | PA3 | PA4 |
| Factor 1: Dynamic Movements and Posture | | | | |
| 20. When sitting, how often do you move/shake your shoulders while you are sitting? | **.78** | .03 | .01 | -.16 |
| 12. When standing, how often do you shake your legs while you are standing? | **.74** | -.04 | .06 | .01 |
| 10. When standing, how often do you move/shake your shoulders while you are standing? | **.80** | .00 | .09 | -.12 |
| 9. When standing, how often do you tap one of your feet while you are standing? | **.59** | -.06 | .09 | .11 |
| 11. When standing, how often do you rest your hand on your face while you are standing? | **.68** | .03 | .02 | .03 |
| 16. When sitting, how often do you have your hands on your waist and hips while you are sitting? | **.65** | -.05 | .10 | -.04 |
| 25. When sitting, how often do you hold your elbows while you are sitting? | **.49** | -.02 | .02 | .31 |
| 8. When standing, how often do you cross your legs or ankle while you are standing? | **.44** | .02 | .06 | .22 |
| 5. When sitting, how often do you move your waist and hips while you are sitting? | **.57** | .03 | .11 | .02 |
| 14. When standing, how often do you hold your elbows while you are standing? | **.52** | .02 | .02 | .26 |
| 13. When standing, how often do you move your arms while you are standing? | **.53** | .15 | -.01 | .18 |
| 27. When sitting, how often do you move your arms while you are sitting? | **.44** | .06 | .00 | .22 |
| Factor 2: Gesticulation | | | | |
| 31. When talking, how often do you move your hands while you are talking? | -.05 | **.92** | -.01 | -.02 |
| 30. When talking, how often do you move your arms while you are talking | .10 | **.83** | -.04 | -.03 |
| 29. When talking, how often do you "talk with your hands"? | -.11 | **.84** | .04 | .03 |
| 32. When talking, how often do you move your fingers while you are talking? | .10 | **.72** | .05 | -.02 |
| Factor 3: Closed Posture | | | | |
| 26. When sitting, how often do you cross your legs or ankle while you are sitting? | -.16 | .09 | .10 | **.53** |
| 18. When sitting, how often do you cross your legs over the knee while you are sitting? | -.11 | .04 | .12 | **.50** |
| Factor 4: Swaying Gait | | | | |
| 3. When walking, how often do you sway your shoulders while you are walking? | .21 | -.03 | **.74** | -.17 |
| 2. When walking, how often do you sway your hips while you are walking? | .00 | .00 | **.72** | .14 |
| 4. When walking, how often do you sway your hands while you are walking? | .07 | .10 | **.58** | .02 |
| 1. When walking, how often do you sway your arms while you are walking? | -.17 | .10 | **.55** | .12 |

0.53) in Dynamic Movements and Posture $t(336.2) = 1.95$, $p = .052$, $d = 0.21$. This finding did not replicate Study 1.

## Gender Differences on the GMS Factors as Divergent Validity with Personality Traits

In Study 2, we again conducted partial correlations as tests of divergent validity to isolate whether any gender differences on the GMS factors could be uniquely attributed to gender and not confounded with personality traits. The partial correlation results appear in Table 3 (and for correlation matrix please view Table 4). For the Gesticulation factor, there was a significant gender difference when controlling for all personality traits, $pr(336) = .28$, $p < .001$, such that women still had significantly higher scores than men, replicating Study 1; for Closed Posture, there was a significant gender difference controlling for all personality traits, $pr(336) = .37$, $p < .001$, such that women still had significantly higher scores than men, replicating Study 1. Unlike Study 1, the Swaying Gait factor showed a significant partial correlation for the gender difference controlling for all personality traits, $pr(336) = -.15$, $p = .005$, with women reporting significantly lower swaying gait than men. However, in Study 2, for the Dynamic Movements and Posture, the gender difference was not statistically significant when controlling for all personality traits, $pr(336) = .10$, $p = .072$.

In terms of specific personality factors, replicating Study 1, Extraversion maintained a significant relationship with Dynamic Movements and Posture, controlling for the other personality traits and the gender difference, $pr(336) = .123$, $p = .024$, such that higher extraversion was associated with more dynamic movements and posture. However, extraversion did not show significant partial correlations with swaying gait in Study 2 (see Table 3; $p = .113$). Additionally, negative emotionality did not show significant partial correlations with any mannerisms in Study 2 (all $ps > .073$). Finally, for Study 2, there were new personality correlates with GMS factors. Specifically, for the Gesticulation factor, extraversion showed a significantly correlation, controlling for the other personality traits and the gender difference, $pr(336) = .18$, $p < .001$, such that higher extraversion was associated with more gesticulation; as did Open-Mindedness, $pr(336)$

**Table 8** Descriptive Statistics for Dynamic Movement, Gesticulation, Closed Posture, and Sway Gait by Gender and U.S. Ethnicity (Study 2)

| F | Ethnicity | Gender | *n* | *M* | *SD* | Median | Skewness | Kurtosis | *SE* |
|---|---|---|---|---|---|---|---|---|---|
| D | Asian | Women | 15 | 2.27 | 0.92 | 2.00 | 1.57 | 2.40 | .24 |
| | | Men | 11 | 1.65 | 0.66 | 1.50 | 0.36 | − 1.71 | .20 |
| | Black | Women | 10 | 1.74 | 0.49 | 1.54 | 0.41 | − 1.39 | .15 |
| | | Men | 8 | 2.02 | 0.60 | 2.00 | 0.33 | − 1.46 | .21 |
| | Latinx | Women | 8 | 1.96 | 0.83 | 1.88 | 0.14 | − 1.86 | .29 |
| | | Men | 9 | 2.19 | 1.05 | 2.08 | 0.59 | − 1.17 | .35 |
| | White | Women | 150 | 2.05 | 0.65 | 1.92 | 1.05 | 1.04 | .05 |
| | | Men | 114 | 1.96 | 0.61 | 1.83 | 1.31 | 2.07 | .06 |
| G | Asian | Women | 15 | 3.05 | 0.77 | 3.00 | 0.81 | 0.55 | .20 |
| | | Men | 11 | 2.50 | 0.75 | 2.50 | − 0.44 | − 0.50 | .23 |
| | Black | Women | 10 | 3.28 | 1.03 | 3.38 | 0.12 | − 1.53 | .33 |
| | | Men | 8 | 2.69 | 0.84 | 2.75 | 0.11 | − 1.55 | .30 |
| | Latinx | Women | 8 | 2.59 | 1.13 | 3.00 | − 0.27 | − 1.39 | .40 |
| | | Men | 9 | 2.97 | 1.07 | 3.00 | − 0.47 | − 1.31 | .36 |
| | White | Women | 150 | 3.26 | 0.76 | 3.12 | − 0.08 | 0.09 | .06 |
| | | Men | 114 | 2.78 | 0.88 | 2.75 | 0.12 | − 0.51 | .08 |
| C | Asian | Women | 15 | 3.60 | 0.74 | 3.50 | 0.56 | − 0.76 | .19 |
| | | Men | 11 | 2.59 | 0.70 | 2.50 | 0.30 | − 0.80 | .21 |
| | Black | Women | 10 | 3.50 | 0.75 | 3.50 | 0.00 | − 1.62 | .24 |
| | | Men | 8 | 1.81 | 0.46 | 2.00 | − 0.32 | − 1.06 | .16 |
| | Latinx | Women | 8 | 3.50 | 0.80 | 3.75 | − 0.55 | − 1.07 | .28 |
| | | Men | 9 | 2.33 | 0.97 | 2.50 | 0.05 | − 1.17 | .32 |
| | White | Women | 150 | 3.56 | 0.70 | 3.50 | − 0.27 | − 0.17 | .06 |
| | | Men | 114 | 3.09 | 0.82 | 3.00 | − 0.59 | 0.00 | .08 |
| S | Asian | Women | 15 | 3.27 | 1.07 | 2.75 | 0.69 | − 1.23 | .28 |
| | | Men | 11 | 2.09 | 0.64 | 2.25 | − 0.30 | − 1.42 | .19 |
| | Black | Women | 10 | 2.38 | 0.62 | 2.50 | − 0.97 | − 0.16 | .19 |
| | | Men | 8 | 2.56 | 0.58 | 2.50 | 0.40 | − 1.55 | .20 |
| | Latinx | Women | 8 | 2.62 | 0.78 | 2.75 | − 0.31 | − 1.13 | .28 |
| | | Men | 9 | 2.33 | 0.89 | 2.25 | 0.34 | − 0.94 | .30 |
| | White | Women | 150 | 2.70 | 0.82 | 2.75 | 0.17 | 0.05 | .07 |
| | | Men | 114 | 2.54 | 0.85 | 2.50 | 0.51 | 0.10 | .08 |

F =Factor; D= Dynamic Movements and Posture; G= Gesticulation; C= Closed Posture; S= Swaying Gait. The U.S. Ethnicities with fewer than seven women or men were excluded from this table

$=-.11$, $p=.037$, such that higher openness correlated with lower gesticulation.

## Gender Differences on Swaying Gaits Items as Concurrent Validity

To replicate the effects observed in Study 1, a MANOVA was conducted on the swaying gait items in Study 2 (the swaying gait items showed a correlation of $r_{\text{overall}}=.54$). Replicating Study 1 (Table 5), women ($M= 2.62$, $SE= 0.050$) on average reported more frequent hip swaying than men ($M= 1.76$, $SE= 0.057$), $F(1, 340) = 63.65$, $p<.001$, $\hat{d} = 1.23$. However, there were no significant differences in shoulder swaying between gender-groups (women: $M_{\text{adj}} = 2.10$, $SE= 0.08$ and men: $M_{\text{adj}} = 2.13$, $SE= 0.09$), when controlling for hip swaying, $F(1, 339) = .06$, $p=.809$, $\hat{d} = -0.028$.

## Gender Differences on Vocal Pitch Items as Concurrent Validity

Replicating the findings of Study 1 (Table 6), women reported significantly greater high-pitched voices ($M= 2.67$, $SD= 0.92$) than men ($M= 2.26$, $SD= 0.90$), $t(425.08) = 4.71$, $p<.001$, $d= 0.45$. Similarly, men reported significantly greater low-pitched voices ($M= 3.23$, $SD= 0.84$) compared to women ($M= 2.64$, $SD= 0.94$), $t(435.34) = -6.88$, $p<.001$, $d= -0.65$.

## Study 2 Discussion

Study 2 replicated the methodology of Study 1 and replicated the GMS's reliability with a new participant sample. This finding suggests that the GMS taps aspects of nonverbal behavior that generalize beyond the initial sample. However, as noted in the results section for the factor analysis, we did drop items on the Closed Posture factor that did not consistently load above our standard between Studies 1 and 2. In this way, we refined the item content of the GMS by using two separate samples to mitigate the undue influence of having only one sample of participants respond to the initial items. In any case, the consistent replication of factor structures and gender differences in nonverbal behaviors between the two studies lends further support to the GMS's validity and underscores the reason to conduct two EFAs before a MI. Additionally, we identified that the personality trait of extraversion is likely the unique associate of the variance in dynamic movements and posture given that it correlated with this dimension in both Study 1 and Study 2, controlling for both the gender difference and the other personality traits. Consequently, while an important factor for self-reported mannerisms, Dynamic Movements and Posture is more associated with extraversion than gender. Given that other personality dimensions showed inconsistent correlations with GMS factors across Studies 1 and 2, researchers should continue to examine the unique associations of personality traits with the GMS factors to provide a nuanced picture.

In sum, given the reliability of the GMS and the consistent evidence of construct validity established in the earlier studies, the next phase, Study 3, involved a new sample and a different recruitment method to conduct a MI. This step was necessary for a more thorough examination of the scale's factor structure. In Study 3, the MI affirmed the structures identified through the EFAs of the previous studies and further supported the GMS structure with a sample that differed in terms of age and racial/ethnic diversity from the previous samples. This refinement process involved assessing each item's alignment with its intended factor, enhancing the scale's precision and effectiveness for future research.

## Study 3: Measurement Invariance of the GMS Factor Structure

Building upon the findings from two previous studies, Study 3 examined the generalizability of the GMS to a community-based sample, which was both younger on average and less U.S. geographically diverse than the MTurk samples from Studies 1 and 2, but simultaneously more racially/ethnically diverse. We could have conducted separate confirmatory factor analysis (CFAs) for different gender groups in the community-based sample to see if the structure fits the EFA structure found from Studies 1 and 2, indicating excellent model fit across both gender groups. However, it is not easy to determine the mean differences from separate CFAs alone. Consequently, a measurement invariance (MI) analysis was utilized to test the fit of the 4-factor solution across three types of invariances: configural, metric, and scalar. While it is possible to conduct MI using either CFA or item-response theory (IRT), we chose CFA because of its popularity in psychology literature (cf. Putnick & Bornstein, 2016). The Study 3 sample consisted of friends and acquaintances of students enrolled in a human sexual behavior course at a San Francisco State University (SFSU) (on the West Coast of the United States). Based on the student permanent residences at the SFSU, the majority of the sample participants were from the Los Angeles and San Francisco metro areas in California, with fewer from the Pacific Northwest and the Southwest of the United States. This recruitment strategy was designed to capture a broader and more varied set of respondents in terms of U.S. race/ethnicity, gender role expectations, and to more reflective of community norms (than is likely happening on a data vendor like MTurk). Importantly, international students enrolled in the courses were able to recruit international participants and did; however, we analyzed only the participants from the United States for this report.

## Study 3 Method

### Participants, Measures, and Procedure

In Study 3, the participants were not directly recruited by the researchers; instead, participants were reached through a two-step process. Students enrolled in two separate sections of human sexual behavior courses at SFSU served as intermediaries, inviting their friends, family, and acquaintances to participate. We initially collected a sample of 1,032 participants. To be consistent with Studies 1 and 2, we only analyzed cisgender heterosexual women and men who are able-bodied from the United States to determine whether these patterns generalized to the community-based sample of the same gender identity and sexual orientation demographics. For Study 3, the GMS was the sole measure analyzed. Nonetheless, like Studies 1 and 2 other measures were included, specifically the BFI- 2.

The analyzed participants from the United States only had an average age of 24.39 years ($SD$ = 8.14 years). In terms of U.S. ethnicity categories, the majority predominantly identified as Hispanic/Latino ($n$ = 91, 31.70%), then European American/White ($n$ = 71, 23.99%), then Asian American/Asian ($n$ = 61, 20.61%). Smaller percentages identified as African American/Black ($n$ = 19, 6.42%), Native American ($n$ = 6, 2.15%), Pacific Islander ($n$ = 7, 2.40%), and Other ($n$ = 12, 4.38%). None of the participants selected Middle Eastern. Consistent with prior studies, the analysis focused on cisgender heterosexual women ($n$ = 164, 55.40% of the sample) and men ($n$ = 132, 44.60% of the sample). This approach was taken to ensure comparability of findings across all three studies, thus providing a consistent basis for assessing the GMS's applicability in diverse populations.

### Measures

The same measures from Study 1 and Study 2 were used in Study 3: the 35-item Gender Mannerisms Scale (GMS), the 60-item Big Five Inventory 2.0 (BFI- 2), and the Two-Question Method of Assessing Gender Identity (2QAGI).

### Data Preparation

One of the advantages of using MI for groups that have known differences (such as the gender groups differing on the GMS scores; see Studies 1 and 2) is that MI allows researchers to confirm that score differences reflect true variations in the phenomenon understudy rather than disparities in response interpretations (cf., Luong & Flake, 2023) by allowing us to examine configural and metric invariances across the gender groups. We would not be able to achieve this goal using separate CFAs per gender group.

MI maintains the psychometric rigor required by CFA's statistical constraints, including multivariate normality. Accordingly, like Studies 1 and 2, we used the unweighted least squares (ULS) estimator was selected for CFA for its suitability with both ordinal and ordinal-like data and with non-normal distributions, as noted by Li (2016). While our response options can be treated as Likert-like, since the phrasing indicates increasing frequency, participants may treat the response scale more like an ordinal one. The ULS estimator is particularly adept at handling the ordinal-like response scales present and non-normality in this study to more reliable estimates of factor loadings and inter-factor correlations as compared to the default maximum likelihood estimator (Li, 2016).

### Assumption Checking for the Measurement Invariance

In preparation for MI, the dataset underwent outlier treatment through winsorization, per gender group (MI is sensitive to outliers, especially affecting multivariate normality, see Luong & Flake, 2023). Subsequently, Mardia's test indicated significant deviations from multivariate normality ($p$ <.001). However, the overall KMO measure yielded high values (KMO =.87), suggesting sampling adequacy of the items for factor analysis. Bartlett's test of perfect orthogonality indicated that inter-item correlations were likely not perfectly orthogonal in both samples ($p$s <.001), justifying the factorability of the correlation matrices. Additionally, the determinant of the correlation matrix was not computationally zero for either gender group, allowing for matrix algebra necessary for MI. The existence of non-normality is why we chose the ULS estimator instead of the default maximum likelihood estimator (see also Li, 2016). Having met the necessary assumptions of MI, we proceeded with this analysis.

## Study 3 Results

We conducted a MI analysis using the lavaan package (Rosseel, 2012) in R, applying the ULS estimator and the nonlinear minimization by bounded least squares method. For descriptive statistics of each factor, as well as their intercorrelations, please consult Table 9 and Tables S2 and S4 in the online supplement. Our sample consisted of 164 women and 132 men (with complete cases for all GMS items and all BFI- 2 items), and we analyzed configural, metric, and scalar invariances to evaluate the consistency of the GMS across gender groups, with the model statistics presented in Table 10. We present the results for each type of invariance using the change values for $\chi^2$ statistics and the common fit indices of root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI), following suggested best practices (Putnick & Bornstein, 2016).

### Configural Invariance of the GMS

The initial model, testing configural invariance, the model comprised 144 parameters and demonstrated a good fit with a $\chi^2_{overall.configural}$ = 352.05 for 406 degrees of freedom. The fit for women was $\chi^2_{women.configural}$ = 196.45 and for men $\chi^2_{men.configural}$ = 155.61, showing a strong base model performance. Fit indices were exceptionally good, with $RMSEA_{configural}$ = 0.00, $CFI_{configural}$ = 1.00, and $TLI_{configural}$ = 1.01, indicating an excellent model structure without overfitting. High factor loadings (>.62) across the latent variables of (a) dynamic movement, (b) gesticulation, (c) swaying gait, and (d) closed posture underscored the robustness of the model structure across genders (Table 11).

**Table 9** Descriptive Statistics for Dynamic Movement, Gesticulation, Closed Posture, and Sway Gait by Gender and U.S. Ethnicity (Study 3)

| F | Ethnicity | Gender | *n* | *M* | *SD* | Median | Skewness | Kurtosis | *SE* |
|---|---|---|---|---|---|---|---|---|---|
| D | Asian | Women | 23 | 2.17 | 0.60 | 2.08 | 0.33 | − 1.07 | .13 |
| | | Men | 30 | 2.26 | 0.77 | 2.08 | 1.23 | 1.09 | .14 |
| | Black | Women | 10 | 2.46 | 0.80 | 2.67 | − 0.34 | − 1.52 | .25 |
| | | Men | 6 | 2.39 | 0.53 | 2.21 | 0.37 | − 1.93 | .21 |
| | Latinx | Women | 49 | 2.27 | 0.59 | 2.25 | − 0.07 | − 0.70 | .08 |
| | | Men | 32 | 1.99 | 0.58 | 1.92 | 0.50 | 0.18 | .10 |
| | White | Women | 38 | 2.26 | 0.59 | 2.38 | − 0.29 | − 0.50 | .15 |
| | | Men | 26 | 2.34 | 0.74 | 2.21 | 0.92 | 0.22 | .19 |
| G | Asian | Women | 23 | 3.43 | 0.96 | 3.50 | 0.01 | − 0.87 | .20 |
| | | Men | 30 | 3.12 | 0.92 | 3.00 | 0.14 | − 0.83 | .17 |
| | Black | Women | 10 | 3.70 | 0.65 | 3.62 | 0.66 | − 0.90 | .21 |
| | | Men | 6 | 3.17 | 0.74 | 3.25 | − 0.36 | − 1.58 | .30 |
| | Latinx | Women | 49 | 3.41 | 0.81 | 3.50 | 0.38 | − 0.79 | .12 |
| | | Men | 32 | 2.95 | 0.99 | 3.00 | 0.03 | − 0.21 | .17 |
| | White | Women | 38 | 3.43 | 0.71 | 3.75 | − 1.14 | 1.76 | .11 |
| | | Men | 26 | 3.21 | 0.96 | 3.00 | 0.36 | − 0.67 | .19 |
| C | Asian | Women | 23 | 3.72 | 0.80 | 4.00 | − 0.24 | − 0.89 | .17 |
| | | Men | 30 | 3.03 | 0.97 | 3.00 | − 0.58 | − 0.44 | .18 |
| | Black | Women | 10 | 3.75 | 0.89 | 3.75 | 0.16 | − 1.63 | .28 |
| | | Men | 6 | 2.67 | 1.03 | 3.00 | − 0.37 | − 1.37 | .42 |
| | Latinx | Women | 49 | 3.76 | 0.88 | 4.00 | − 0.96 | 1.04 | .13 |
| | | Men | 32 | 2.38 | 1.00 | 2.50 | 0.38 | − 0.26 | .18 |
| | White | Women | 38 | 3.86 | 0.69 | 4.00 | − 0.34 | − 0.17 | .11 |
| | | Men | 26 | 3.50 | 1.07 | 4.00 | − 0.45 | − 0.90 | .21 |
| S | Asian | Women | 23 | 2.75 | 0.74 | 2.75 | 0.84 | 1.75 | .15 |
| | | Men | 30 | 2.42 | 0.75 | 2.25 | 0.50 | 0.08 | .14 |
| | Black | Women | 10 | 2.95 | 1.03 | 2.62 | 0.90 | − 0.65 | .32 |
| | | Men | 6 | 2.33 | 0.74 | 2.62 | − 0.82 | − 1.05 | .30 |
| | Latinx | Women | 49 | 2.80 | 0.87 | 2.75 | 0.29 | 0.53 | .12 |
| | | Men | 32 | 1.80 | 0.68 | 2.00 | 0.61 | − 0.28 | .12 |
| | White | Women | 38 | 2.78 | 0.55 | 3.00 | − 0.62 | − 0.59 | .09 |
| | | Men | 26 | 2.62 | 1.02 | 2.75 | 0.33 | − 0.68 | .20 |

F = Factor; D = Dynamic Movements and Posture; G = Gesticulation; C = Closed Posture; S = Swaying Gait. To keep parallel structure with Table 8, we only listed the values for Asian, Black, Latinx, and White respondents

**Table 10** Measurement Invariance Results Across Gender Groups

| Model | $\chi^2$ | *df* | $\Delta\chi^2$ | $\Delta df$ | *p*-value | RMSEA | CFI | TLI | SRMR |
|---|---|---|---|---|---|---|---|---|---|
| Configural Invariance | 352.05 | 406 | – | – | – | .00 | 1.00 | 1.01 | .07 |
| Metric Invariance | 407.72 | 424 | 55.67 | 18 | <.001 | .00 | 1.00 | 1.00 | .08 |
| Scalar Invariance | 513.74 | 442 | 106.02 | 18 | <.001 | .04 | .90 | .99 | .08 |

In the first row of the table, $\chi^2$ stands for the Chi-square statistic, df represents the degrees of freedom, $\Delta\chi^2$ indicates the change in the Chi-square statistic, and $\Delta df$ denotes the change in degrees of freedom. The RMSEA, CFI, TLI, and SRMR are all overall statistics, not differences between models. Metric invariance is the difference between configural and metric invariances, and similarly, scalar invariance is the difference between metric and scalar invariances

## Metric Invariance of the GMS

Adding 18 equality constraints to the factor loadings, the metric invariance model exhibited a $\chi^2_{\text{overall.metric}} = 407.72$ over 424 degrees of freedom. The fit indices for this model were $\text{RMSEA}_{\text{metric}} = 0.00$, $\text{CFI}_{\text{metric}} = 1.00$, and $\text{TLI}_{\text{metric}} = 1.00$, suggesting minimal impact from the constraints and

**Table 11** GMS Confirmatory Factor Analysis Factor Loadings and Fit Indices (Study 3)

| GMS Item | Factor Loading | |
|---|---|---|
| | Women | Men |
| Factor 1: Dynamic Movements and Posture | | |
| 8. When standing, how often do you cross your legs or ankle while you are standing? | 1.00 (fixed) | 1.00 (fixed) |
| 12. When standing, how often do you shake your legs while you are standing? | 1.38 | 1.13 |
| 10. When standing, how often do you move/shake your shoulders while you are standing? | 1.51 | 0.95 |
| 9. When standing, how often do you tap one of your feet while you are standing? | 1.41 | 0.73 |
| 11. When standing, how often do you rest your hand on your face while you are standing? | 1.18 | 0.96 |
| 16. When sitting, how often do you have your hands on your waist and hips while you are sitting? | 1.15 | 0.62 |
| 25. When sitting, how often do you hold your elbows while you are sitting? | 1.05 | 0.91 |
| 20. When sitting, how often do you move/shake your shoulders while you are sitting? | 1.27 | 0.88 |
| 5. When sitting, how often do you move your waist and hips while you are sitting? | 1.21 | 0.75 |
| 14. When standing, how often do you hold your elbows while you are standing? | 0.94 | 0.89 |
| 13. When standing, how often do you move your arms while you are standing? | 1.34 | 0.95 |
| 27. When sitting, how often do you move your arms while you are sitting? | 1.26 | 0.81 |
| Factor 2: Gesticulation | | |
| 29. When talking, how often do you “talk with your hands”? | 1.00 (fixed) | 1.00 (fixed) |
| 30. When talking, how often do you move your arms while you are talking | 1.45 | 0.9 |
| 31. When talking, how often do you move your hands while you are talking? | 1.14 | 0.90 |
| 32. When talking, how often do you move your fingers while you are talking? | 1.44 | 1.26 |
| Factor 3: Closed Posture | | |
| 26. When sitting, how often do you cross your legs or ankle while you are sitting? | 1.00 (fixed) | 1.00 (fixed) |
| 18. When sitting, how often do you cross your legs over the knee while you are sitting? | 1.26 | 1.09 |
| Factor 4: Swaying Gait | | |
| 1. When walking, how often do you sway your arms while you are walking? | 1.00 (fixed) | 1.00 (fixed) |
| 2. When walking, how often do you sway your hips while you are walking? | 1.85 | 1.15 |
| 3. When walking, how often do you sway your shoulders while you are walking? | 2.22 | 1.56 |
| 4. When walking, how often do you sway your hands while you are walking? | 2.59 | 1.66 |
| Fit Indices | | |
| Comparative Fit Index | .96 | .99 |
| Tucker-Lewis Index | .96 | .99 |
| Root Mean Square Error of Approximation | .05 | .03 |
| Standardized Root Mean Square Residual | .08 | .07 |

maintaining high comparability across genders. There was a significant chi-squared difference ($\Delta\chi^2 = 55.67$, $df = 18$, $p < .001$), indicating changes with metric constraints. Nonetheless, the $p$-value for the $\Delta\chi^2$ statistic is known to be influenced by increasing sample size (e.g., Chen, 2007; Putnick & Bornstein, 2016). Accordingly, we used the difference in the fit indices as a more stable indication of invariance (see Cheung & Rensvold, 2002). The differences in fit indices were minimal with $\Delta RMSEA_{metric} = 0.00$, $\Delta CFI_{metric} = 0.00$, and $\Delta TLI_{metric} = 0.01$ units. The overall model fits for metric invariance still showed excellent fits using existing MI standards, $RMSEA_{metric} = .00$, $CFI_{metric} = 1.00$, $TLI_{metric} = 1.00$, $SRMR_{metric} = .08$. Thus, these results demonstrate excellent metric invariance across the gender groups (see Table 10).

## Scalar Invariance of the GMS

Adding 40 equality constraints on item intercepts and factor loadings, the scalar model displayed a $\chi^2_{overall.scalar} = 513.74$ with 442 degrees of freedom. The comparison between the metric and scalar models showed a substantial chi-squared increase ($\Delta\chi^2 = 106.02$, $df = 18$, $p < .001$), reflecting the rigorous constraints of scalar invariance. Specifically, the fit indices slightly deteriorated compared to the metric model (with $\Delta RMSEA_{scalar} = -.04$, $\Delta SRMR_{scalar} = -.00$, $\Delta CFI_{scalar} = .01$, and $\Delta TLI_{scalar} = .02$), but remained within acceptable ranges as overall model fits: $RMSEA_{scalar} = 0.04$, $CFI_{scalar} = 0.99$, and $TLI_{scalar} = 0.99$, effectively demonstrating the model's capability to maintain equivalence across groups using existing MI standards (Table 10).

## Study 3 Discussion

Overall, the MI analysis conducted across configural, metric, and scalar models has robustly demonstrated the scale's consistency across gender groups. The analysis showed that the model structure remained strong, and the constraints introduced at each level of invariance—while affecting the fit to varying degrees—did not compromise the overall utility of the scale. Specifically, the scale's ability to measure latent constructs consistently between women and men was affirmed, indicating its reliability for valid group comparisons. These results ensure that any observed differences in scale scores between genders are likely reflective of true differences in the constructs measured, rather than artifacts of measurement bias, thus supporting the scale's applicability and effectiveness in diverse demographic settings. This validation supports the scale's broad applicability, making it a reliable tool for diverse demographic settings.

## General Discussion

These three studies collectively provided compelling evidence that the GMS is a valid and reliable psychometric instrument that captures salient aspects of intrapersonal, declarative self-knowledge regarding mannerisms that have been previously studied as interpersonally gendered. The validity and reliability of the GMS had been rigorously evaluated through EFA and MI across multiple studies. Specifically, the use of two EFAs before a MI allowed us to refine the scale itself, ensuring that the items included were both representative and indicative of the underlying constructs they aimed to measure. Of note, the second EFA showed that at least two items identified by the first EFA were not replicable. This outcome would have been masked by using only one EFA and then a MI. What is more, the two EFA results consistently supported the 4-factor multidimensionality of the scale, aligning with theoretical expectations of gendered nonverbal behaviors in general and providing a further breakdown of the dynamic movement dimension for intrapersonal understanding. Subsequent MI (via CFA) provided robust support for the scale's structure, demonstrating good model fit on a more diverse sample in terms of age band and U.S. ethnicity. This level of validation spoke to the scale's reliability, indicating that the GMS could be confidently used to measure gendered nonverbal cues with consistency over time.

The existence of the GMS can now provide theoretical perspectives on and empirical connections between the dynamics between self-perception and social perception of gendered nonverbal behavior—finally connecting the intrapersonal to the interpersonal. These connections are possible because the scale's development was rooted in interpersonal nonverbal behavior.

One noteworthy possible connection is that the middling accuracy rates in gender recognition through walking motions, as highlighted by Pollick et al. (2005) can now attempt to reconcile the interpersonal perception dynamics with a potential reliance on the perceivers intrapersonal insights about how their own mannerisms connect to their gender self-categorization. The consistent relationship between extraversion and dynamic movement and posture (in Studies 1 and 2) underscores the role of personality traits in shaping gendered nonverbal expression, at least in terms of self-perception, illustrating how individual differences like extraversion are interwoven with gender categorizations to influence self-perceived nonverbal behaviors. Additionally, because extraversion emerged as a consistent correlate of dynamic movement and posture, studies can also combine the thin-slicing perceiver methodology (e.g., Kolar et al., 1996) of inferring personality traits with the existing gender-based categorizations to get a finer-grain understanding of how perceivers use multiple information sources to make their judgments about nonverbal behavior.

### Limitations and Future Directions

A salient limitation of the GMS and all self-report measures is that the GMS responses do not sufficiently separate actuarial behavior from the potential biases of societal gender norms and the internalization of gender roles, either consciously or unconsciously. This could mean that some of the differences in scores between gender groups might reflect biases toward social stereotypes rather than actual behaviors. Without an objective evaluation of behaviors, it is currently challenging to distinguish between accurate self-reports and those influenced by stereotypes. However, new self-report tools like the adult felt-pressure to conform to gender stereotypes (Chessin et al., 2019) offer ways to address these issues at the level of self-reporting alone. Further studies should examine how accurate the self-reports for the GMS factors are vis-à-vis actual manifest behavior when recorded and coded.

Another important limitation is that the GMS in this study has been validated primarily within the U.S. context for cisgender, heterosexual populations, thereby overlooking sexual minorities. To initially address this limitation, we conducted supplementary EFA using oblimin rotation, incorporating both sexual minorities and majorities in Studies 1 and 2 (see Table S7 in the online supplement). The factor analysis revealed the same factor structure and item composition as observed in Studies 1 and 2 with heterosexual participants, with differences only in item loadings. Nonetheless, the initial addressal needs to be (and will be) coupled with future research focusing on recruiting large samples of cisgender queer individuals and, separately, heterosexual and queer transgender individuals to verify whether the GMS accurately captures gendered mannerisms across diverse sexual orientations and gender identities. Thus, our investigation can be considered as the first step in a larger, multi-step research process.

Additionally, future studies should examine the GMS responses and factor structure with more inclusive demographics within the United States as well as in different national cultures. For extending the GMS to other national cultures, international researchers are encouraged to adapt the GMS for their cultural contexts through measurement invariance analysis (via CFA or IRT) and tweak the items for their specific national cultural context if necessary. This international expansion will enhance our understanding of nonverbal gender expressions across cultures and identify both universal (or near-universal) dynamics in addition to culturally specific mannerisms. Gender research writ large should do a better job of identifying culture-specific and culture-general dynamics.

Moreover, the GMS can now be used in the U.S. context at least to explore other influences on how individuals self-report their nonverbal behaviors specific to gendered mannerisms, including factors such as a person's household religion (e.g., Calabro, 2013), childhood experiences (e.g., Daines et al., 2021), self-presentation on social media (Jackson & Luchner, 2018), social desirability (e.g., Reynolds, 1982), and gender norm conformity concerns (Mahalik et al., 2003, 2005).

Another application of the GMS worth noting in this journal is its potential use in therapy and counseling settings. Given the GMS's analysis focus on cis heterosexual women and men, clinicians and counselors should be cautious at present in applying these initial findings to queer and trans clients. Nonetheless, to the extent that clinicians rely on tools such as the Bem Sex Role Inventory (Bem, 1974) to ascertain adherence to traditional (U.S.) femininity, traditional (U.S.) masculinity, and so-called psychological androgyny (the combination of femininity and masculinity), the GMS may be one tool to remind and/or illustrate to clients that trait-endorsement and mannerisms may or may not be intimately connected. What is more, given that some mannerisms—namely, dynamic movement—are also correlated with personality traits, clients can see their own behavioral repertoire in a broader context of how gender and personality converge and separate. When the GMS is correlated with actuarial behavior, this understanding will become vital to concepts like behavioral gender fluidity (as opposed to a core nonbinary sense of self; cf. Tate et al., 2020), behavioral heterosexual gender non-conformity, in addition to other ways that heterosexual and queer people and cis people and trans people challenge or move with traditional heteronormative gender norms. All of these insights will eventually be available to therapy and counseling contexts, helping pave the way for more nuanced and supportive counseling practices that promote well-being among gender traditional and gender diverse populations.

## Conclusion

This study presents the GMS as a promising tool for quantifying self-reports of gendered nonverbal behaviors, filling a gap between self-perception (intrapersonal) and social perception (interpersonal). Validated through rigorous statistical analyses, the GMS offers a nuanced understanding of gender expression in everyday life—albeit for heterosexual cis women and men at the moment. Nonetheless, the GMS's development marks a key advancement in gender studies in psychological science and related behavioral sciences, providing a foundation for further research to explore the rich nexus of intrapersonal and interpersonal communication for all genders and sexual orientations in all countries.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s11199-025-01580-8.

**Funding** Beyond the aforementioned compensation to participants, there were no external funding sources or financial support received for this research. No individual or organization outside of the research team had any role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript. There are no personal, professional, or financial relationships that could potentially be construed as a conflict of interest in relation to this research.

**Data Availability** The datasets used during the current study are available from the corresponding author upon reasonable request.

## Declarations

**Research Involving Human Participants** This study involved human participants who responded to online surveys. The study was non-invasive, and participants were only required to provide their responses through an online platform Qualtrics.

**Competing interests** The authors declare no competing interests.

## References

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics, 23*(2), 145–152. https://doi.org/10.3758/BF03208295

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*(3), 183–200. https://doi.org/10.1037/h0024835

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*(2), 122–162. https://doi.org/10.1037/t00748-000

Calabro, D. M. (2013). Gestures of praise: Lifting and spreading the hands in biblical prayer. In J. R. Chadwick, M. J. Grey, & D. R. Seely (Eds.), *Ascending the mountain of the Lord: Temple, praise, and worship in the Old Testament* (pp. 105–121). Religious Studies Center, Brigham Young University; Deseret Book.

Cashdan, E. (1998). Smiles, speech, and body posture: How women and men display sociometric status and power. *Journal of Nonverbal Behavior, 22*(4), 209–228. https://doi.org/10.1023/A:1022967721884

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504. https://doi.org/10.1080/10705510701301834

Chessin, K., Lazo, A., Strait, M., & Tate, C. C. (2019). Facilitating and inhibiting factors for self-reports of same-gender attraction in cisgender heterosexual-identifying women and men. *The Journal of Sex Research, 56*(9), 1155–1167. https://doi.org/10.1080/00224499.2019.1632252

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9*(5), 353–356. https://doi.org/10.3758/BF03337021

Daines, C. L., Hansen, D., Novilla, M. L. B., & Crandall, A. (2021). Effects of positive and negative childhood experiences on adult family health. *BMC Public Health, 21*, 651. https://doi.org/10.1186/s12889-021-10732-w

Dean, M. L., & Tate, C. C. (2017). Extending the legacy of Sandra Bem: Psychological androgyny as a touchstone conceptual advance for the study of gender in psychological science. *Sex Roles, 76*(11–12), 643–654. https://doi.org/10.1007/s11199-016-0713-z

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology, 30*, 92–101. https://doi.org/10.5334/irsp.82

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage Publications.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384. https://doi.org/10.1037/0003-066X.48.4.384

Friedman, H. S. (2019). Introduction to the special issue on theory in nonverbal communication. *Journal of Nonverbal Behavior, 43*(2), 107–109. https://doi.org/10.1007/s10919-019-00299-x

Graziano, M., & Gullberg, M. (2024). Providing evidence for a well-worn stereotype: Italians and Swedes do gesture differently. *Frontiers in Communication, 9*, 1314120. https://doi.org/10.3389/fcomm.2024.1314120

Hegarty, P., & Buechel, C. (2006). Androcentric reporting of gender differences in APA journals: 1965–2004. *Review of General Psychology, 10*, 377–389. https://doi.org/10.1037/1089-2680.10.4.377

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*(2), 141–154. https://doi.org/10.1016/S0022-1031(77)80007-3

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist, 74*(2), 171–193. https://doi.org/10.1037/amp0000307

Jackson, C. A., & Luchner, A. F. (2018). Self-presentation mediates the relationship between self-criticism and emotional response to Instagram feedback. *Personality and Individual Differences, 133*, 1–6. https://doi.org/10.1016/j.paid.2017.04.052

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving sex directly and indirectly: Meaning in motion and morphology. *Psychological Science, 16*(11), 890–897. https://doi.org/10.1111/j.1467-9280.2005.01633.x

Knöfler, T., & Imhof, M. (2007). Does sexual orientation have an impact on nonverbal behavior in interpersonal communication? *Journal of Nonverbal Behavior, 31*, 189–204. https://doi.org/10.1007/s10919-007-0032-8

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality, 64*(2), 311–337. https://doi.org/10.1111/j.1467-6494.1996.tb00513.x

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics, 21*, 575–580. https://doi.org/10.3758/BF03198740

LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: A meta-analysis of sex differences in smiling. *Psychological Bulletin, 129*(2), 305–334. https://doi.org/10.1037/0033-2909.129.2.305

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods, 21*(3), 369–387. https://doi.org/10.1037/met0000093

Lopez, L. D., Reschke, P. J., Knothe, J. M., & Walle, E. A. (2017). Postural communication of emotion: Perception of distinct poses of five discrete emotions. *Frontiers in Psychology, 8*, 710. https://doi.org/10.3389/fpsyg.2017.00710

Luong, R., & Flake, J. K. (2023). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods, 28*(4), 905–924. https://doi.org/10.1037/met0000441

Mahalik, J. R., Locke, B. D., Ludlow, L. H., Diemer, M. A., Scott, R. P., Gottfried, M., & Freitas, G. (2003). Development of the Conformity to Masculine Norms Inventory. *Psychology of Men & Masculinity, 4*(1), 3–25. https://doi.org/10.1037/1524-9220.4.1.3

Mahalik, J. R., Morray, E. B., Coonerty-Femiano, A., Ludlow, L. H., Slattery, S. M., & Smiler, A. (2005). Development of the Conformity to Feminine Norms Inventory. *Sex Roles, 52*, 417–435. https://doi.org/10.1007/s11199-005-3709-7

Miller, A. G. (1970). Role of physical attractiveness in impression formation. *Psychonomic Science, 19*(4), 241–243. https://doi.org/10.3758/BF03328797

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology, 3*, 23. https://doi.org/10.3389/fpsyg.2012.00023

Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1247–1265. https://doi.org/10.1037/0096-1523.31.6.1247

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychology research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Revelle, W. (2018). psych: Procedures for personality and psychological research. Northwestern University, Evanston, Illinois. R package version 1.8.4. https://CRAN.R-project.org/package=psych

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe‐Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38*(1), 119-125.https://doi.org/10.1002/

1097-4679(198201)38:1%3C119::AID-JCLP2270380118%3E3.0.CO;2-I
Richards, C., Bouman, W. P., Seal, L., Barker, M. J., Nieder, T. O., & T'Sjoen, G. (2016). Non-binary or genderqueer genders. *International Review of Psychiatry, 28*(1), 95–102. https://doi.org/10.3109/09540261.2015.1106446
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
Shrout, P. E., & Fiske, D. W. (1981). Nonverbal behaviors and social evaluation. *Journal of Personality, 49*(2), 115–128. https://doi.org/10.1111/j.1467-6494.1981.tb00732.x
Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and accessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. https://doi.org/10.1037/pspp0000096
Stephan, Y., Sutin, A., Bovier-Lapierre, G., & Terracciano, A. (2018). Personality and walking speed across adulthood. *Social Psychological and Personality Science, 9*, 773–780. https://doi.org/10.1177/1948550617725152
Tate, C. C., Ledbetter, J. N., & Youssef, C. P. (2013). A two-question method for assessing gender categories in the social and medical sciences. *Journal of Sex Research, 50*(8), 767–776. https://doi.org/10.1080/00224499.2012.690110
Tate, C. C., Youssef, C. P., & Bettergarcia, J. N. (2014). Integrating the study of transgender spectrum and cisgender experiences of self-categorization from a personality perspective. *Review of General Psychology, 18*(4), 302–312. https://doi.org/10.1037/gpr0000019
Tate, C. C., Ben-Hagai, E., & Crosby, F. J. (2020). *Undoing the gender binary*. Cambridge University Press.
Vrugt, A., & Luyerink, M. (2000). The contribution of bodily posture to gender stereotypical impressions. *Social Behavior and Personality: An International Journal, 28*(1), 91–103. https://doi.org/10.2224/sbp.2000.28.1.91
Weinraub, M., Clemens, L. P., Sockloff, A., Ethridge, T., Gracely, E., & Myers, B. (1984). The development of sex role stereotypes in the third year: Relationships to gender labeling, gender identity, sex-types toy preference, and family characteristics. *Child Development, 55*, 1493–1503. https://doi.org/10.2307/1130019
Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29*(3/4), 350–362. https://doi.org/10.1093/biomet/29.3-4.350
Weston, P. S. J., Hunter, M. D., Sokhi, D. S., Wilkinson, I. D., & Woodruff, P. W. R. (2015). Discrimination of voice gender in the human auditory cortex. *NeuroImage, 105*, 208–214. https://doi.org/10.1016/j.neuroimage.2014.10.056
Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology, 57*, 173–181. https://doi.org/10.1348/000711004849222

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