ORIGINAL ARTICLE

The Woman's Nontraditional Sexuality Questionnaire-Short Form (WNSQ-SF): Development, Variance Composition, Reliability, Validity, and Measurement Invariance

Ronald Levant[1] · **Shana Pryor[1]** · **Kristin E. Silver[2]**

Abstract

The aim of the present study was to develop a psychometrically stronger version of the women's nontraditional sexuality questionnaire (WNSQ), one that could support a total scale score in addition to subscale scores. Using data from 519 college and community women, the variance composition of the WNSQ was assessed, from which the bifactor model showed the best model fit. This model had a general WNS factor and four group factors: casual sex, self-pleasuring, sexual interest, and sex-as-a-means-to-an-end. A trimmed model was developed based on updated guidelines for shortening composite measurement scales, and confirmed in a separate sample (N = 238). Next, the reliability of the WNSQ-SF was assessed using bifactor reliability and dimensionality diagnostic indices and found that the raw scores from the general factor and three out of the four group factors were reliable. Convergent and discriminant construct evidence of the validity of the subscales was found. Finally, strong and strict invariance across race/ethnicity and sexual orientation was demonstrated, meaning that members of both marginalized and dominant groups of women understand the scale scores in the same way, including the scale score points and zero points of the scales, and that the constructs assessed by the scale are measured with the same degree of precision.

Keywords Women's sexuality · Gender norms · Scale development · Structural equation modeling · Measurement invariance/equivalence

✉ Ronald Levant
  levant@uakron.edu

[1] Department of Psychology, University of Akron, 302 Buchtel Mall, Akron, OH 44325, USA

[2] VA Northeast Ohio Healthcare System, Cleveland, OH 44106, USA

Introduction

This study aimed to create a psychometrically stronger version of the only measure of overall nontraditional sexuality in women (i.e., sexual behaviors that are contrary to the traditional portrayal of women's sexuality), the woman's nontraditional sexuality questionnaire (WNSQ; [23]), which measures sexual behaviors and beliefs with 4 subscales: casual sex, self-pleasuring, sexual interest, and sex-as-a-means-to-an-end. While (as we will discuss shortly) there are other scales in the nomological network of the WNSQ, these are all much narrower than the WNSQ, which covers a broader range of women's nontraditional sexuality. Although the WNSQ has not been used much in research, that may be due to its present limitations.

For example, evidence was never provided for the use of a total scale score, which is important for investigators interested in an overall non-traditional orientation to sexuality, nor for its invariance by race and sexual orientation, which is important to assure that the scale is understood in the same way by different groups, and that therefore the groups' scores can be reliably compared. These two points are responsive to recent research that has pointed to the "hidden invalidity" of many psychological scales, with particular reference to structural validity—i.e., dimensionality (factor structure) and measurement equivalence/invariance [19]. It was thought that a psychometrically stronger version of the WNSQ might be of greater utility to researchers.

The WNSQ was developed using the theoretical framework of the gender role strain paradigm (GRSP), which posits that traditional gender ideologies define the prescriptive and proscriptive norms for gender role performance [26, 38, 39]. According to this paradigm, these ideologies guide childhood socialization as well as adult behavior. Gender role strain results regardless of the degree to which one adheres or does not to traditional gender norms.

Traditional femininity ideology (TFI) defines the traditional gender role for American girls and women. It consists of five norms: stereotypic images and activities, purity, caretaking, dependency/deference, and emotionality, which together mandate that women remain pure and docile, focusing more on the emotional components of a relationship rather than the sexual [22]. Traditional gender ideologies (such as TFI) reflect the values of the dominant group in the U.S.—namely White cisgender heterosexual Christian men—and are measured as individual's beliefs. Maintaining the status quo by adhering to these norms is often rewarded. Women who adhere to traditional norms of femininity are often regarded as more likable and benefit from more social acceptance [53]. By contrast, failure to abide by these norms may result in negative consequences.

The Purity norm for women's sexuality remains influential. For example, women who attend sexual novelty parties—a place to buy sex toys and other paraphernalia and to talk about sex—run the risk of being seen as insecure and less traditional [30]. Women who masturbate have reported feelings of shame and guilt when indulging in self-pleasure, and some may even feel that to masturbate and to give their sexual needs priority over the needs of others makes them

appear selfish [3]. There has long been a double standard for sexuality in society where women who engage in casual sex are viewed as 'slutty' whereas men who do so are seen as manly [11]. Many women fear that they would be considered promiscuous by having casual sex and thus keep those sexual experiences secret [14]. Several studies have demonstrated that casual sex is a reality for many women [2, 15, 48], and one that appears to be increasing relative to men [37], yet the idea of women being equally as sexual as men has not enjoyed general acceptance [32], perhaps because it conflicts with the purity norm. The male-centered nature of heterosexual sex may influence this hesitation to understand and accept the true nature of the sexual lives of heterosexual women. For example, pornography, which can serve as an enjoyable sexual outlet, has traditionally catered to men [40]. Additionally, men are often expected to relentlessly pursue sex whereas women are expected to hold more passive roles, leading many women in heterosexual relationships having unwanted sex or feeling pressured to have sex [13]. The sexual needs of men are often given priority over those of women and the idea of a sexually empowered woman is seen as more of a myth than a reality[51].

Much research in the area of women's sexual desire has reviewed problems associated with low sexual interest. However, research is emerging which highlights women with higher levels of sexual interest. Wentland et al. [57] reported that female participants with high sexual interests reported earlier onset of sexual intercourse, more frequent sexual activity, and higher numbers of both committed and casual partners than women with lower sexual interests. Other research indicates that women's sexuality can be used to gain resources, such as in the exotic dancing industry [9]. Exotic dancers capitalize on some men's need for sex by using their sexuality to earn money. Along similar lines, but in the area of leisure sports, the use of one's sexuality as a means to seek reward is also found in women's flat track roller derby in which derby girls create an image of being strong, aggressive, and sexy in order to gain fans [36]. In these two cases, women use their sexuality to compete for resources in a male-dominated world.

Recent scale development studies on aspects of women's sexuality create a nomological network which can be used to assess the evidence for the validity of the WNSQ subscales. To assess the evidence for the validity of the subscale tapping women's sexual interest, the Sex Drive Questionnaire [35] can be used. The Sex Is Power Scale [12] can provide a way to assess women's sexual empowerment, even tapping the extent to which sex is used as a means to an end (i.e., using sexuality to gain something of value or a desired response). Finally, the Sexual Assertiveness Scale [33] can allow the assessment of women's ability to take initiative for sex, as well as engaging in STD and pregnancy prevention, and the refusal of sex.

As the above review indicates, women's nontraditional sexuality has not been thoroughly explored. Current research on women's gender role and sexual behaviors indicates that women are increasingly violating traditional gender norms [27], and that differences between women and men in both attitudes toward sex and various sexual behaviors are decreasing [37]. These changes correspond to women's increased social and political freedom. A robust measure assessing overall nontraditional sexuality in women would thus be useful for future research on this topic. The purpose of the present study was to develop a psychometrically strong, short-form

version of a scale with increased utility to researchers. Trimming longer scales to remove weaker items can improve their psychometric properties [17]. Furthermore, shorter forms of scales are typically easier to use and reduce participant fatigue. Providing a useful tool to improve our understanding of women's sexuality may in turn support those who advocate for women's increased sexual freedoms and view women's sexuality as an important component of their overall humanity.

Development and Previous Examination of the WNSQ

The original WNSQ was published in 2012 [23] and consisted of 2 introductory questions regarding the sexual activity status of the participant. The questionnaire continued with 18 sexual practices and 3 belief items. Following recommended practices [59], an exploratory factor analysis (EFA) was followed by a confirmatory factor analysis (CFA) using a common factors model with a new sample, finding adequate fit for the four-factor model. Using Ferguson's [16] criteria to assess effect size for strength of association indices such as r, in which 0.2 is the recommended minimum effect size representing a practically significant effect. 0.5 represents a moderate effect and 0.8 represents a large effect, the WNSQ was found to positively correlate at a moderate level (total score, $r = 0.67$, $p < 0.01$; subscales ranged from $r = 0.32$ to 0.73, p's < 0.01) with a measure of non-relational sex used with both men and women, the Sociosexual Orientation Index (SOI; [49]). The WNSQ was also found to have a small negative correlation with the Purity subscale of the Femininity Ideology Scale (FIS, [22]; total score, $r = -0.42$, $p < 0.01$; subscales ranged from $r = -0.20$ to -0.41, p's from < 0.05 to < 0.01). These two sets of correlations provided convergent construct evidence of validity for the WNSQ. Finally, the WNSQ was found to be significantly, positively but weakly correlated (total score $r = 0.08$, $p < 0.05$; subscales ranged from $r = 0.00$ to 0.19, p's from n.s. to < 0.01) with the Health Protective Sexual Communication Scale (HPSCS; [4]), providing preliminary concurrent evidence for validity.

The Present Study

This study aims to improve the psychometric properties of the WNSQ. The *first objective* was to assess the variance composition of the WNSQ using a series of CFA's to estimate a set of dimensional models, to determine the extent to which general and group latent factors are caused by the items, which had not yet been done. In the common factors model, the items load only on the group latent factors corresponding to their subscales—that is the group latent factors cause their respective indicators. In the bifactor model, the items load both on their group factors (corresponding to their subscales) and on a general women's non-traditional sexuality factor (corresponding to the total scale score). In the hierarchical model the items load on their first order factors (corresponding to the subscales), which in turn load on the second order factor (corresponding to the total scale score). Finally, in the unidimensional model the items load only on the general women's non-traditional sexuality factor. Structural diagrams for all of these

except the unidimensional model are shown in Figs. 1–3. According to Rodriguez et al. [45], the parameters resulting from these analyses allow the determination of the degree to which variation in participants' responses to questions is attributable to the domain tapped by the subscale on which the question loads, or attributable to a general factor reflecting the total scale score. Namely, large ancillary bifactor indices for the general factor or very large second-order loadings would be evidence to support interpreting a total scale score. It should be noted that previous scale development studies of multidimensional scales developed using the Gender Role Strain Paradigm found them to be bifactor models. These include the Male Role Norms Inventory—Short Form (MRNI-SF; [24, 25] and the Femininity Ideology Scale—Short Form (FIS-SF,[27]. Therefore, we advance the following hypothesis:

Hypothesis 1 (H1) CFA will support the bifactor structure of the WNSQ over the common factors, hierarchical, and unidimensional models.

The *second objective* was to improve the psychometric properties of the WNSQ using procedures recommended by Goetz, et al. [17]. Longer versions of scales may include some items that may be poorer measures of the target factors, creating measurement error and thus misfit. Therefore, the development of a short form allows for the selection of items with higher factor loadings, thereby increasing model fit. Furthermore, shorter scales reduce participant fatigue, particularly when they are part of large battery, improving data quality. Hence our aim was to develop the WNSQ-Short Form (WNSQ-SF) and to confirm it in a new sample. We present the following hypothesis:

Hypothesis 2 (H2) The bifactor structure of the WNSQ will be retained in the WNSQ-SF with adequate fit and significant factor loadings for all subscales and will be the best fitting of all models tested.

Our *third objective* was to determine the extent to which the raw scores for the subscales and total scale score of the WNSQ-SF are reliable by calculating bifactor reliability and dimensionality diagnostic measures [10]. As there is no prior research on the WNSQ to use as a guide, this is an exploratory question without hypotheses.

Our *fourth objective* was to assess evidence for validity of the WNSQ-SF using a latent variable approach. Hypothesized effect size categories are provided and are based on previously reported correlations of each validity scale. We focused on specific convergent and discriminant construct evidence for the validity of the subscales, and put forth the following hypotheses:

Hypothesis 3 (H3) Convergent construct evidence for the validity of the Casual Sex subscale will be provided by a small significant positive correlation with a measure of sexual empowerment (Sex is Power Scale). Erchull and Liss [12] found a small, significant, and positive correlation between the Sex is Power Scale and

a scale measuring enjoyment of sexualized male attention. Finding enjoyment in men's attention is related to casual sex in that sexualized attention may lead to casual sex, providing the basis for our hypothesis that casual sex and the Sex Is Power Scale will be positively and significantly correlated, but because these connections are somewhat attenuated, we hypothesized a small effect size.

Hypothesis 4 (H4) Convergent construct evidence for the validity of the self-pleasuring subscale will be provided by significant moderate positive correlations with measures of (a) masturbation (Beliefs About Masturbation Questionnaire), and (b) sex drive (Sex Drive Questionnaire). The basis for hypothesizing the moderate link between self-pleasuring practices and beliefs about masturbation is that both scales tap a very similar phenomenon. With regard to H4b, Ostovich & Sabini [35] found evidence for the validity of the Sex Drive Questionnaire in positive, significant, and moderate correlations between it and the number of sexual partners per month and amount of intercourse in a month. Sex drive operationalized in terms of sex with a partner would likely be associated with self-pleasuring, providing the basis for our hypothesis that the Self-Pleasuring subscale and the Sex Drive Questionnaire will be positively and significantly correlated with a moderate effect size.

Hypothesis 5 (H5) Convergent construct evidence for the validity of the sexual interest subscale will be provided by significant moderate positive correlations with measures of (a) sex drive (Sex Drive Questionnaire), and (b) a measure of sexual assertiveness that reflects an ability to initiate sex (Initiation subscale of the Sexual Assertiveness Scale). With regard to H5a, the basis for hypothesizing the moderate link between our sexual interest subscale and the Sex Drive Questionnaire is that both scales tap a very similar phenomenon. With regard to H5b, we hypothesized a moderate positive correlation between the sexual interest subscale and the Initiation subscale of the Sex Drive Questionnaire because there is a moderate positive correlation between the Sex Drive Questionnaire and sexual intercourse (e.g., amount and number of partners per month; [35].

Hypothesis 6 (H6) Convergent construct evidence for the validity of the Sex as a Means to an End subscale will be provided by (a) a small-to-moderate significant positive correlation with a measure of sex for instrumental purposes (Sex is Power Scale) and by (b) a small-to-moderate significant negative correlation with a measure of sexual assertiveness that reflects an ability to refuse sex (Refusal subscale of the Sexual Assertiveness Scale), which is negative because the two scales are scored in the opposite direction. For H6a, Sex as a Means to an End assesses the inclination to use sex to achieve a desired outcome; the Sex is Power Scale was found to be moderately and significantly related to enjoyment in sexualization by men (as mentioned above; [12], which is empowering, providing the necessary foundation for using sex as a means to an end. For hypothesis 6b, as noted above, the Sexual Assertiveness scale has been found to be positively and mildly related to sexual initiation behaviors [33], and women who are likely to initiate sexual encounters are

also likely to refuse sexual encounters. Women who use sex as a means to an end will thus likely be willing to refuse sexual encounters.

Hypothesis 7 (H7) Discriminant construct evidence for the validity of the four WNSQ-SF subscales will be provided by their non-significant correlations with the validity scales that were not specified in hypotheses 3–6. For casual sex, these include measures of masturbation, sex drive, and the three measures of sexual assertiveness. For self-pleasuring, these include measures of sexual empowerment and the three measures of sexual assertiveness. For sexual interest these include measures of masturbation, sex drive, and two of the measures of sexual assertiveness (refuse sex and pregnancy—STD prevention). For sex as a means to an end these include measures of masturbation, sex drive, and two of the measures of sexual assertiveness (initiation and pregnancy—STD prevention).

The *fifth objective* was to assess the measurement invariance/equivalence (ME/I) of the WNSQ by race/ethnicity and sexual orientation. The assessment of ME/I focusses on the degree to which a scale is understood in the same way by different groups, including the meaning of the scale scores, the distance between scores, the zero points of the scale, and that the constructs assessed by the scale are measured with same degree of precision. This is necessary to ensure that the scale is free from construct bias. For example, without demonstrated scalar invariance (in which the unstandardized intercepts are constrained to be equal across the groups), one cannot compare mean scores across groups, because the intercepts "estimate the score on an indicator, given a true score of zero" on the latent factor [20], p. 398). We compared White women to women of color, and heterosexual women to sexual minority women. We were limited by low N's in the various groups of people of color and non-heterosexual identities, and thus could not test specific identities. We recognize the limitations both of aggregating these identities (i.e., bisexual women are different from lesbians) and of using a proxy variable like racial/ethnic identification [41]. We encourage investigators to examine specific identities using broader measures of racial/ethnic identity and sexual orientation in future studies. However, we were ultimately guided by Vera and Speight's [56] arguments for a social justice orientation in the social sciences, that understands and addresses the relative privilege and power of White and heterosexual people, and the systems of oppression that marginalize people of color and sexual minority people in the U. S. [1, 8]. Since there are no prior data on this topic, this is an exploratory question, without hypotheses.

Method

Participants

A total of 519 women participated in the first sample. Ages ranged from 18 to 68, with a mean of 26.46 (SD = 10.79, median = 21, mode = 19). A majority of

participants identified as White/European American (74%), Heterosexual (78%) and Christian (59.2%). Two thirds (67.2%) were in a romantic relationship, and 77.8% were sexually active.

A total of 238 women participated in the second sample. Ages ranged from 18 to 40, with a mean of 19.8 (SD = 2.31, median = 19, mode = 18). A majority of participants identified as White/European American (78.2%) and heterosexual (64.6%), and the largest group identified as Christian (48.3%). Almost half (47.9%) were in a romantic relationship, and 61.9% were sexually active. The full demographic details of both samples are in the Online Supplement Tables 1 and 2.

Procedure

The study was approved by the authors' university IRB. Students were recruited through the departmental research participation pool using the online SONA system and were offered extra credit for participation. Community members were recruited from internet sites such as Craigslist and were offered participation in a raffle for $50 gift cards in which 4 would win. Participants followed a link to Qualtrics, which hosted the survey. After providing informed consent, participants completed questionnaires, and were debriefed at the end of the survey. Upon completion of the survey, participants followed another link to a separate Qualtrics survey where they could confidentially enter their contact information for course credit or the raffle.

Measures

Demographic Questionnaire

The demographic questionnaire consisted of 10 items that asked questions about gender, age, race/ethnicity, religion, sexual orientation, preferred sexual partner, relationship status, highest degree completed, family/household income, and socioeconomic status.

Women's Nontraditional Sexuality Questionnaire (WNSQ)

The WNSQ [23] is a 23-item measure that is intended to tap women's nontraditional sexual behaviors (practices) and attitudes (beliefs). A definition of sex is first provided: "For all of the following questions, please consider the term "sex" to refer to any form of intimate physical contact involving more than kissing and hugging that is meant to express affection between you and another adult (of any sex)." This is followed by two yes or no questions inquiring whether the participant has had sex using that definition and whether she is currently sexually active. Questions 3–20 tap sexual behaviors using a 7-point Likert scale (1 = *never*, 7 = *frequently*) and questions 21–23 tap sexual attitudes and are measured on another 7-point Likert scale (1 = *strongly disagree*; 7 = *strongly agree*). In both case higher scores signify

a greater participation in or endorsement of nontraditional sexuality. The subscales for the WNSQ are (1) Sexual interest (SI; 4 items; e.g., "How often do you say what you want or need during sex?"), (2) Casual sex (CS; 7 items; e.g., "How often do you have sex with someone you just met?"), (3) Self-pleasuring (SP; 5 items; e.g., "How often do you masturbate?"), (4) Sex-as-a-Means-to-an-End (SME; 5 items; e.g., "How often have you had sex to end a fight?"). The WNSQ had Cronbach alphas of 0.84 for the full scale and from 0.67 to 0.82 for the subscales, indicating good reliability for all except for the Self-Pleasuring subscale. As discussed previously, convergent, and concurrent evidence for the validity of the WNSQ was provided by the scale developers. For the present study α's were SI $= 0.60$, CS $= 0.79$, SP $= 0.81$, SME $= 0.81$.

Sex is Power Scale (SIPS)

The SIPS [12] is a 12-item measure that assesses heterosexual women's beliefs that sexuality provides power over men. High scores indicate greater endorsement that sexuality provides power over men. The SIPS has two subscales of which we used only the Self-Sex Is Power Scale (S-SIPS,7 items, e.g., "I can get what I want using my feminine wiles.") The items are measured on a 6-point Likert scale (1 = *disagree strongly*; 6 = *agree strongly*). The S-SIPS demonstrated a Cronbach alpha of 0.89, indicating good reliability. The S-SIPS was found to be significantly correlated ($r = 0.43$, $p < 0.001$) with the body evaluation subscale of the Interpersonal Sexual Objectification Scale (ISOS; [21] demonstrating convergent evidence for validity. The S-SIPS failed to correlate with shame ($r = 0.10$, $p > 0.05$), a subscale of the Objectified Body Consciousness Scale (OBCS; [31] which provided discriminant evidence for validity. For the present study the α was 0.92.

Sexual Assertiveness Scale (SAS)

The SAS [33] is an 18-item measure created to assess sexual assertiveness in women with higher scores reflecting greater assertiveness. The SAS is made up of 3 subscales, each with 6 items: (1) Initiation (e.g., "I begin sex with my partner if I want to"),(2) Refusal (e.g., "I refuse to have sex if I don't want to, even if my partner insists"); and (3) Pregnancy—STD Prevention (e.g., "I refuse to have sex if my partner refuses to use a condom or latex barrier"). The 18 questions are measured on a 5-point Likert scale (0 = *disagree* strongly; 4 = *agree* strongly). Internal consistency was good with Cronbach alphas of 0.77 for Initiation, 0.71 for Refusal, and 0.83 for pregnancy–STD prevention. The SAS demonstrated construct evidence of validity by moderately correlating with a single item measure of general assertiveness ($r = 0.58$, $p < 0.001$). The SAS demonstrated moderate test–retest reliability with correlation values ranging from 0.59 to 0.77 over a 6-month and 1-year time period, respectively. For the present study the α's were 0.69 for Initiation, 0.75 for Refusal, and 0.85 for pregnancy–STD Prevention.

Sex Drive Questionnaire (SDQ)

The SDQ [35] is a 4-item measure intended to assess sex drive, with higher scores indicating higher levels of sex drive. Different Likert-type scales are used. Question 1 ("How often do you experience sexual desire?") is measured on a 7-point scale (1 = *never*, 7 = *several times a day*). Questions 2 ("How often do you orgasm in the average month?") and 3 ("How many times do you masturbate in the average month?") are measured on a 6-point scale (1 = *never*; 6 = *several times a day*). Finally, question 4 ("How would you compare your level of sex drive with that of the average person of your gender and age?" is measured on a 7-point scale (1 = *very much lower*; 7 = *very much higher*). The SDQ was found to be correlated with another known measure of sex drive, the Sexual Desire Inventory (SDI; [50] thereby demonstrating convergent construct evidence for validity. The SDQ has demonstrated good reliability with Cronbach alphas being 0.79 for men and 0.83 for women. Test–retest reliability was found over a 6- to 8-week period for both men and women ($r = 0.91$, $r = 0.90$, respectively). For the present study the α was 0.82.

Beliefs About Masturbation Questionnaire

Given that little is known about women's attitudes about masturbation, we introduced three questions asking women about their beliefs surrounding masturbation. All questions were measured using a 5-point Likert scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*). Question 1 ("It is okay for me to meet my own sexual needs through masturbation"), question 2 ("I believe masturbating can be an exciting experience"), and reverse scored question 3 ("I believe masturbation is wrong"). For the present study α was 0.86.

Data Analytic Procedures

Sample Size Considerations

We used MacCallum et al. ([29] Table 4) RMSEA-based method for estimating sample size in structural equation modeling (SEM). For the WNSQ, the smallest number of parameters in the analyses conducted is 67, and the recommended sample size for a test of not-close fit is 224. For the WNSQ-SF, the smallest number of parameters in the analyses conducted is 48, and the recommended sample size for a test of not-close fit is 279. For the ME/I multi-group analyses, the smallest number of parameters in the analyses conducted is 70, and the recommended sample size for a test of not-close fit is 219. Our sample of 519 is more than adequate for these analyses.

Overview

Several measurement models of the WNSQ were estimated and compared via CFA: common factors, bifactor, hierarchical, and unidimensional models, as described

above. Next, using the best fitting of these models, the scale was trimmed by deleting weak items, following the updated guidelines developed by Goetz et al. [17] for shortening composite measurement scales (CMS), based on their literature review of 91 scale-shortening projects conducted from 1995 to 2009. Thus, item selections were based on the strength of factor loadings as well as evaluation of the individual items to ensure that the content reflected each of the original subscales while avoiding having too similar items and preserving content validity. We planned a priori to generate a 3 items-per-subscale version of the WNSQ-SF because construction of latent variables in SEM requires use of at least 3 manifest variables to indicate a latent factor without causing local identification problems [20]. Finally, following Goetz et al., we confirmed the trimmed model in a new sample.

Although there is some controversy regarding the practice of parceling [28], we chose this method for the unidimensional validity scales or subscales because of the greater reliability afforded when dealing with unidimensional scales [20, 28]. We thus followed the recommendations of Russell et al., [46] for the validity hypotheses and created three to four item parcels from the manifest variables for each validity scale that had six or more observed items, which included the Sex Is Power Scale and the three subscales of the Sexual Assertiveness Scale. For the Masturbation and Sex Drive Questionnaires the observed items were used to assess the latent factors. Item parcels were created by performing a principle axis exploratory factor analysis with a one-factor solution for the items comprising the scale. Iterative assignment of items into one of the two parcels was done to ensure that parcel loadings were balanced [46]. Evidence for validity was assessed using SEM, which has the advantages of analyzing the composition of the variance into general and group factors and controlling for multiple sources of measurement error that may otherwise attenuate validity estimates [20]. Specifically, validity was assessed using correlational analysis to estimate the strength, significance, and valence of the associations of WNSQ-SF latent factors with the latent factors of the validity scales. Finally, the WNSQ-SF was then used as a basis for specifying multi-group models testing measurement invariance. Testing for configural, metric and scalar invariance was performed using the new M*plus* Invariance Shortcut Code [34], whereas that for residuals invariance required the writing of syntax.

Statistical Analyses

The descriptive statistics were calculated using SPSS 26. Mplus v.8 [34] was utilized to conduct the single and multiple group CFA's and validity analyses with latent variables. The scaled chi-square goodness-of-fit test was used to assess the overall fit of all CFA models. However, alternative fit indices were also utilized (Kahn, 2006) because the goodness-of-fit statistic has been found to be overly sensitive to inconsequential sources of model misfit when sample sizes are large [7]. Based on the recommendations of Kline [20], alternative indices included: Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), for which values of ≥ 0.90 indicate reasonable fit, and values of ≥ 0.95 indicate good fit; Root Mean Square Error of Approximation (RMSEA), where values between 0.05 and 0.08 suggest reasonable fit and a value of 0.05 or lower is considered good fit; and standardized root mean square residual

(SRMR), for which values of less than 0.10 are considered acceptable and values of 0.05 or lower suggest good fit.

The fits of relevant CFA models were compared using a scaled chi-square difference tests, which was adjusted for the use of the maximum likelihood estimation with robust standard errors (MLR) on the recommendations of Satorra and Bentler [47], to accommodate for the non-normality in the samples. However, similar to the chi-square goodness-of-fit test, the scaled chi-square difference test ($\Delta\chi2$) is affected by large sample sizes [6, 7]. Since the $\Delta\chi2$ is expected to be statistically significant in samples larger than 300 [20], we utilized a ΔCFI with a cut-off score of < 0.01 [5–7].

Results

Information on data cleaning, missing data, outliers, and normality for both samples is in Online Supplement 1 and 2.

Objective 1: Assessment of Variance Composition of the WNSQ

To accomplish this objective, single-group CFA models were fit to the data set of 519 participants. A common factors model was first tested, shown in Fig. 1, which provides the factor loadings. The resulting chi-square goodness of fit statistic was statistically significant as is usually the case in large samples, χ^2 (183) $=$ 535.10; $p < 0.001$. Some of the remaining indices were within the guidelines described earlier and others were not, providing mixed evidence on the fit of this model: CFI $= 0.872$; TLI $= 0.854$; RMSEA $= 0.061$ (90% CI $= 0.055, 0.067$); SRMR $= 0.059$. All but one of the standardized factor loadings (loading on the DSI factor) were significant at the $p < 0.01$ level and ranged from 0.37 to 0.87.
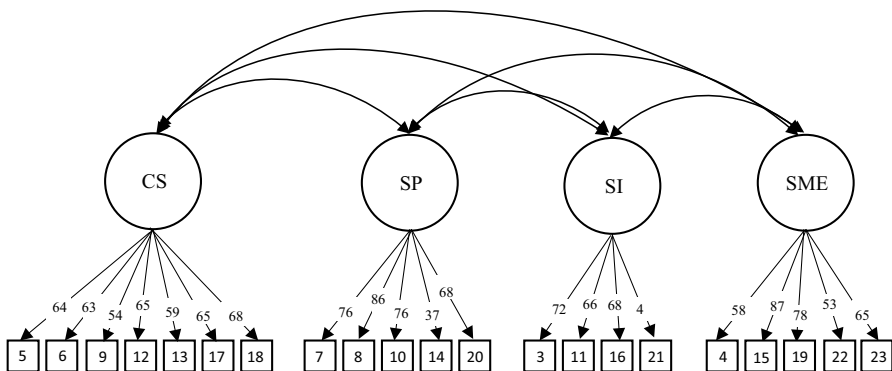

Fig. 1 Structural equation model of the common factors model. *Note* Diagram uses standardized loadings
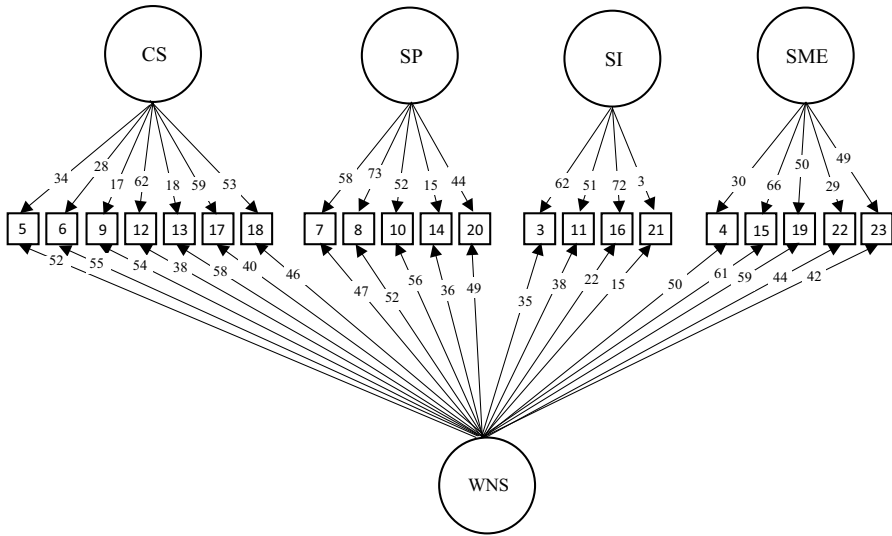
Fig. 2 Structural equation model of the bifactor model. *Note* Diagram uses standardized loadings

Next, a bifactor model was investigated as displayed in Fig. 2. The chi-square goodness of fit statistic was statistically significant, χ^2 (168) $=$ 505.37; $p < 0.001$. Some of the remaining indices were within the guidelines described earlier and others were not, providing mixed evidence on the fit of this model: CFI $= 0.878$; TLI $= 0.847$; RMSEA $= 0.062$ (90% CI $= 0.056$, 0.068), SRMR $= 0.054$. All but four of the standardized factor loadings loading on both the group factors (range $= 0.28$—0.83) and the general factor (range $= 0.15$—0.61) were significant at the $p < 0.01$ level. Of the non-significant factor loadings two loaded on ICS, and one each on SP and DSI. Comparing the common factors with the bifactor model using the scaled chi square difference test, the bifactor model significantly improved fit: $\Delta\chi^2$ (15) $= 36.48$, $p < 0.01$. Furthermore, the CFI of the bifactor model was 0.006 *larger* than that of the common factors model. Thus, the bifactor model showed an improvement in fit when compared to the common factors model.

Next estimated was the hierarchical factor model, displayed in Fig. 3. The chi-square goodness of fit statistic was statistically significant, χ^2 (185) $=$ 555.67; $p < 0.001$. Some of the remaining indices were within the guidelines described earlier and other were not, providing mixed evidence on the fit of this model: CFI $= 0.866$; TLI $= 0.847$; RMSEA $= 0.062$ (90% CI $= 0.056$, 0.068), SRMR $= 0.063$. All but one (loading on DSI) of the standardized factor loadings for both the lower-order and higher order factors were significant at the $p < 0.001$ level. The more constrained hierarchical model showed a decrement in fit by the chi square difference test ($\Delta\chi^2$ (2) $=$ 30.45; $p < 0.001$). However, the Δ CFI was 0.006, which is less than the 0.01 cutoff. Thus, comparison of the fit of hierarchical model with the common factors model was equivocal. Comparing the bifactor and hierarchical models on the basis of ΔCFI, the bifactor model had a CFI that was 0.006 larger.
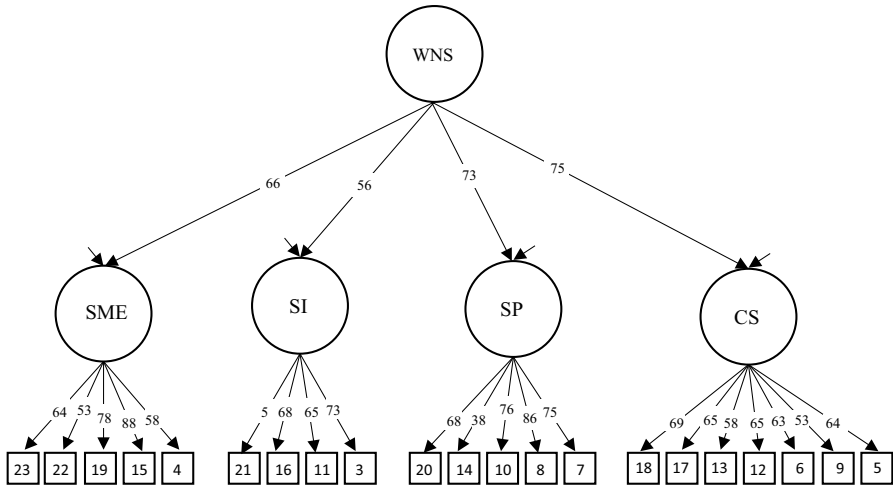
Fig. 3 Structural equation model of the Hierarchical model. *Note* Diagram uses standardized loadings

Finally, a unidimensional model was assessed. The chi-square goodness of fit statistic for unidimensional model was statistically significant, $\chi2$ (189) = 1352.96; $p < 0.001$. None of the remaining indices were within the guidelines described earlier, suggesting poor model fit, CFI = 0.578; TLI = 0.531; RMSEA = 0.109 (90% CI = 0.104, 0.114); SRMR = 0.096. Furthermore, four of the standardized factor loadings were not significant; those that were ranged from 0.30 to 0.81. Considering the fit criteria, the unidimensional model for the WNSQ-SF does not seem plausible.

To summarize these results, of the set of single-group models that we tested, the bifactor model fit better than the common factors, hierarchical and unidimensional models, supporting hypothesis H1, although there is definitely room for improvement, both in terms of the fit with the data, and the non-significant factor loadings.

Objective 2: Creating and Confirming the WNSQ Short Form (WNSQ-SF)

Using the bifactor model of the WNSQ, the model was trimmed to create the WNSQ-SF using the Goetz et al. [17] guidelines. The first two authors first sorted the items on each subscale into the various facets of the construct they were measuring. For example, the 5 Self-Pleasuring items broke down into 2 facets: self-stimulation and purchasing. The aim was then to select the highest loading items that loaded at least 0.40 from each facet to ensure coverage of the construct and to eliminate overlapping items. The items and factor loadings for the WNSQ-SF are shown in Table 1. The chi-square goodness of fit statistic was statistically significant as is typical, $\chi2$ (42) = 75.52; $p < 0.001$. The remaining indices were within the guidelines described earlier, providing evidence of the good fit of this model: CFI = 0.977; TLI = 0.964; RMSEA = 0.039 (90% CI = 0.024, 0.053); SRMR = 0.037. All standardized factor loadings were significant at the $p < 0.01$ level and ranged from 0.20 to 0.74 for the group factors and 0.32 to 0.62 for the general factor.

Table 1 Standardized factor loadings from confirmatory factor analysis of the WNSQ-SF items as a Bifactor model

Scale and Item	General	CS	SP	SI	SME
Casual Sex (CS)					
12. How often do you have sex outside of an exclusive relationship?	.40	.65			
17. Do you ever have sex with a friend with whom you are not interested in dating (so-called "friends with benefits")?	.42	.58			
18. How often do you have sex with someone you just met?	.50	.47			
Self-pleasuring (SP)					
7. How often do you masturbate?	.59		.60		
10. How often do you purchase sex toys?	.62		.20		
20. How often do you watch pornography alone?	.56		.51		
Sexual interest (SI)					
3. Given the chance, how often would you choose to have sex?	.42			.58	
11. How often do you say what you want or need during sex?	.50			.41	
16. How often do you fantasize about having sex with your current partner?	.35			.66	
Sex as a means to an end (SME)					
15. How often do you use sex to get something you want?	.50				.74
19. How often have you had sex to get someone to do something for you?	.50				.60
23. I would not use sex to get something I wanted (R)	.32				.56

Item numbers refer to the number of the item in the original WNSQ. R=reverse-scored item. Standardized factor loadings are reported. All factor loadings significant at *p*<.001. General=General Women's Non-traditional Sexuality Factor

As a check on the WNSQ-SF we estimated common factors and unidimensional models to first, compare the bifactor model with the common factors model, and second to run bifactor reliability and dimensionality diagnostic measures. The WNSQ-SF bifactor model fit better than the common factors model, $\Delta\chi2$ (6)=46.76, $p<0.01$. Furthermore, the CFI of the bifactor model was 0.023 *larger* than that of the common factors model. Finally, we confirmed the WNSQ in the second sample: $\chi2$ (42)=91.03; $p<0.001$; CFI=0.929; TLI=0.888; RMSEA=0.070 (90% CI=0.050, 0.090); SRMR=0.049. All standardized factor loadings were significant at the $p<0.01$ level and ranged from 0.28 to 0.94 for the group factors and 0.21 to 0.67 for the general factor. Thus, hypothesis H2 was supported.

Objective 3: Calculate WNSQ-SF Bifactor Reliability and Dimensionality Diagnostic Measures

Bifactor ancillary diagnostic measures were calculated for the WNSQ-SF and are displayed in Table 2. In estimating the bifactor model, all group factors were specified as orthogonal to the general factor and to each other by fixing their intercorrelations to zero, which allowed us to obtain estimates of each ancillary bifactor

Table 2 Explained common variance and model-based dimensionality and reliability estimates for the WNSQ-SF

	General	CS	SP	SI	SME
ECV	.42	.63	.39	.63	.67
Omega	.88	.76	.79	.74	.82
Omega H	.66	–	–	–	–
Omega H S	–	.48	.28	.46	.55
Relative Omega	.75	.63	.35	.63	.68

ECV = explained common variance. Omega = A model-based estimate of internal reliability of the multidimensional composite. Omega H (Hierarchical) = Percentage of variance in raw total scores that can be attributed to the individual differences on the general factor. Omega HS (Hierarchical Subscale) = Percentage of reliable variance of a subscale score after partitioning out variance attributed to the general factor. Relative Omega = Omega Hierarchical divided by Omega

measure uncontaminated by any shared variance [42, 45]. The values for the ECV (proportion of explained common variance) indicate that WNSQ-SF general factor accounted for 42.2% of the common variance of all the items. The values for the group factors are only relative to the items loading on that factor. Ordered from largest to smallest, SME, CS, SI, and SP accounted for 67.2%, 62.9%, 62.9% and 38.8% of the variance of the items loading on their group factors, respectively. These results indicate that for all group factors except SP, the majority of the variance taps the group factors, whereas for SP the majority of the variance re-measures the general factor.

In addition, Table 2 summarizes the Omega coefficients for the WNSQ-SF bifactor model. Omega (ω) is a factor analytic model-based estimate of the internal reliability of the multidimensional composite. For the general factor all items are taken into account, whereas for the group factors only those items that load on a factor are considered in the calculation. The ω values range from 0.74 to 0.88, indicating reliability of all factors. Omega Hierarchical (ωH) reflects the percentage of variance in raw total scores that can be attributed to the general factor, whereas Omega hierarchical subscale (ωHS) reflects the percentage of reliable variance of a subscale score after removing variance attributed to the general factor. Definitive guidelines for evaluating ωH and ωHS do not yet exist; however, Reise et al. [43] indicated that "tentatively, we can propose that a minimum would be greater than 0.50, and values closer to 0.75 would be much preferred" (p. 137). From this we can see that the general factor at 0.66 and SME at 0.55 meet the lower of these criteria, although CS (0.48) and SI (0.46) come close. SP is once again on the low end at 0.28.

Relative Omega (ωH/ω) is Omega Hierarchical and Omega Hierarchical Subscale divided by Omega. For the general factor this represents the percentage of reliable variance in the multidimensional composite that is due to the general factor; for group factors, it represents the percentage of reliable variance in the subscale composite that is independent of the general factor. The relative ω for the general factor was 0.75, indicating that 75% of the reliable variance in the WNSQ-SF total score was due to the general factor. Thus, model-based reliability estimates support

the use of the raw WNSQ-SF total score to represent the general women's nontraditional sexuality factor. Ordered from largest to smallest, SME, CS, SI and SP had 68%, 63%, 63% and 35% of the variance of the items loading on their group factors that was independent of the general factor, tentatively supporting the use of the raw scores as measures of those subscale constructs, although the value for SP is a bit low, indicating caution when using raw scores for that subscale.

Finally, using the Reise et al. [44] criteria, our Percentage of Uncontaminated Correlations (PUC) value was 0.82, failing to meet the criterion < 0.80, our Omega Hierarchical was 0.66, failing to meet the criterion of > 0.70, and our general factor ECV value was 0.42, missing the criterion of > 0.60. Hence, modeling the WNSQ-SF as a unidimensional instrument would likely lead to significant measurement parameter bias (i.e., biased item factor loadings). This is also supported by the poor fit statistics for the unidimensional model: $\chi 2$ (54) = 700.68; $p < 0.001$; CFI = 0.562; TLI = 0.465; RMSEA = 0.152 (90% CI = 0.142, 0.162); SRMR = 0.110. Finally, there is relative parameter bias—the difference between an item's loading in the unidimensional solution and its general factor loading in the bifactor (i.e., the truer model), divided by the general factor loading in the bifactor model. "According to Muthén, Kaplan, and Hollis (1987), [average] parameter bias less than 10–15% is acceptable and poses no serious concern." [45], p. 145). Our average relative parameter bias is slightly above this at 18.0%.

In summary, the bifactor indices generally support the bifactor structure of the WNSQ-SF allowing the use of raw total scale and subscale scores to measure the general and group factors, although caution is indicated in using raw scores to tap the SP group factor.

Objective 4: Assessment of Evidence for Validity

Using the bifactor model for the WNSQ-SF, we estimated the correlations of the WNSQ-SF's general and group latent factors with the latent validity factors. With the exception of TLI, the CFA of the overall measurement model (i.e., the WNSQ-SF and the validity scales together) produced reasonable fit to the data: $\chi 2$ (267) = 716.10, $p < 0.001$, CFI = 0.921, TLI = 0.896, RMSEA = 0.057 (90% CI = 0.052, 0.062), SRMR = 0.049. For the WNSQ-SF, all of the standardized loadings on the general and group factors were significant and ranged from 0.22 to 0.70. For the validity variables, whether measured by parcels or items, all had significant standardized loadings on their respective factors, and ranged from 0.64 to 0.98. However, a warning was issued that the latent variable covariance matrix was non-positive definite for the Sex Drive Questionnaire latent variable. We sought to resolve this problem by re-running the analysis using the common factors model, which fit less well: $\chi 2$ (279) = 772.02, $p < 0.001$, CFI = 0.914, TLI = 0.891, RMSEA = 0.058 (90% CI = 0.053, 0.063), SRMR = 0.050. But even more significant, the common factors model inflated the correlations between the latent factors of the WNSQ-SF subscales and the validity variables by remeasuring the general factor in the subscales. We had earlier pointed out that one advantage of assessing validity using a latent variable approach is the ability to separate out variation due to

Table 3 Standardized correlation coefficients of the WNSQ-SF subscales with the validity variables

WNSQ-SF Factors and Validity Variables

Scale	CS	SP	SI	SME	WNS	MAST	SDQ	SIPS	INI	REF
MAST	− .05	**.47**	.09	− .06	.42**	-	-	-	-	-
SDQ	*.18**	*.49**	*.26**	*.17**	*.38**	*.46**	-	-	-	-
SIPS	.08	**.66**	**.60**	.14**	.52**	.37**	.38**	-	-	-
INI	.36**	.65**	.65**	.33**	.67**	.15**	.29**	.15**	-	-
REF	**.22**	− .03	.12	**.56**	.46**	*.11**	.34**	*.06*	.12	-
STD	.49**	.29**	.27**	.68**	.59**	.27**	.25**	− .20**	.14**	.35**
	− .26	− .20	**.21**	− .15	.52**	*.15**	− .09	− .19**	− .07	.35**
	.10	*.22**	*.34**	*.10*	*.25**	− .02	− .06	-.10	− .02	
	.01	− .06	.08	**− .32**	− .12	− .01	− .10	− .13**		
	− .11**	− .13**	*.00*	− .32**	− .20**	− .03	− .09*			
	.13	.21**	.01	.01	− .28**	− .05				
	− .09*	− .12**	− .12**	− .20**	− .18**					

CS casual sex, *SP* self-pleasuring, *SI* sexual interest, *SME* sex as a means to an end, *WNS* women's general nontraditional sexuality; *MAST* Masturbation; *SDQ* Sex Drive Questionnaire; *SIPS* Sex Is Power Scale; *INI* initiation; *REF* Refusal; *STD* pregnancy—STD prevention; coefficients based on latent factors (regular font) are in the upper row and correlations based on raw scores (italic font) are in the lower row. Boldface correlations were used for hypothesis testing

*p < .05, **p < .01

the group factors from that due to the general factor. To illustrate, using the bifactor model, only 10 of 30 correlations between latent group factors and validity variables were significant, whereas using the common factors model 19 of 24 correlations were significant, reflecting the remeasurement of the general factor. Hence, we reported the bifactor results and express caution about relying on the findings involving the Sex Drive Questionnaire variable.

Table 3 displays the correlations of each of the WNSQ latent factors with each of the validity latent factors, and for comparison's sake, the raw score correlations as well. Hypotheses H3—H6 concerned convergent construct evidence for validity. Hypothesis H3 was supported; evidence for the validity of the Casual Sex subscale was found by a small significant positive correlation with a measure of sexual empowerment ($r = 0.22$, $p < 0.001$). Hypothesis H4 was supported; evidence for the validity of the Self-Pleasuring subscale was provided by moderate significant positive correlations with measures of (a) masturbation ($r = 0.47$, $p < 0.001$) and (b) sex drive ($r = 0.66$, $p < 0.001$). Hypothesis H5 was partially supported; evidence for the validity of the sexual interest subscale was provided by (a) a moderate significant and positive correlation with a measure of sex drive ($r = 0.60$, $p < 0.001$) but not by a (b) moderate significant and positive correlation with a measure of initiating sex ($r = 0.21$, $p = 0.22$). Finally, hypothesis H6 was supported; evidence for the validity of the Sex as a Means to an End subscale was provided by a small-to-moderate significant positive correlation with (a) a measure of sexual empowerment ($r = 0.56$, $p < 0.001$), and (b) by a small-to-moderate significant negative correlation with a measure of sexual assertiveness that reflects an ability to refuse sex ($r = − 0.32$, $p < 0.001$).

With regard to discriminant validity, hypothesis H7 was fully supported. Casual Sex was not significantly correlated with measures of masturbation, sex drive, and the three measures of sexual assertiveness. Self-Pleasuring was not significantly correlated with measures of sexual empowerment and three measures of sexual assertiveness. Sexual interest was not significantly correlated with measures of masturbation, sex drive, and two of the measures of sexual assertiveness (refuse sex and pregnancy—STD prevention). Sex as a Means to an End was not significantly correlated with measures of masturbation, sex drive, and two of the measures of sexual assertiveness (initiative and pregnancy—STD prevention).

To summarize these results, out of a total of 24 tests, 23 fully supported the hypotheses, providing convergent and discriminant evidence for the validity of the casual sex, self-pleasuring, sexual interest, and sex as a means to an end scales. As mentioned above, the findings involving the Sex Drive Questionnaire should be interpreted with caution. This impacts the convergent evidence for validity of the self-pleasuring and sexual interest subscales.

Objective 5: Measurement Equivalence/Invariance

Assessment of Measurement Invariance of the WNSQ-SF by Race/Ethnicity

Multi-group CFA's were estimated to assess the configural, metric, scalar, and residuals invariance of the WNSQ-SF responses across race/ethnicity, estimating a series of nested models with the White women (WW) and women of color (WoC) participants treated as separate subsamples. The dimensional structure for all models was bifactor. The $\chi 2$ was statistically significant for all models, while the CFI, TLI, RMSEA, and SRMR, were often at acceptable levels. We examined models with increasing cross-group equality constraints to test for measurement invariance (c.f. [20] using $\Delta\chi 2$ and ΔCFI, as discussed above.

We first assessed the least parsimonious model, namely *configural* invariance, to ascertain whether the same pattern of indicators loading on factors held across the racial/ethnic groups. This model showed reasonable fit to the data, supporting configural invariance: Total sample: $\chi 2$ (84) = 147.96; $p < 0.001$; [for each group— WW $\chi 2$ (42) = 76.32; $p = 0.001$, WoC $\chi 2$ (42) = 71.64; $p = 0.003$]; CFI = 0.960; TLI = 0.937; RMSEA = 0.055 (90% CI = 0.040, 0.069); SRMR = 0.045. All but one (10, loading on SP) of the standardized factor loadings for both groups were significant. Ranges for the significant factor loadings on the group factors were WW 0.39–0.78, WoC 0.37–0.74. Ranges for the general factors were WW 0.35–0.65, WoC 0.32–0.64. We conclude that configural invariance was largely demonstrated in this data set for the five latent factors.

We constrained factor loadings to be equal for the two groups to test for *metric* invariance and found adequate fit: Total sample: $\chi 2$ (103) = 165.12; $p < 0.001$; [for each group—WW $\chi 2 = 75.61$, WoC $\chi 2 = 89.51$]; CFI = 0.961; TLI = 0.950; RMSEA = 0.049 (90% CI = 0.034, 0.062); SRMR = 0.055. All standardized factor loadings for the 2 groups were significant. Ranges for the significant factor loadings

on the group factors were WW 0.21–0.78, WoC 0.22–0.70. Ranges for the general factors were WW 0.35–0.61, WoC 0.32–0.60. Using the scaled chi square difference test, the more parsimonious metric invariance model did not degrade fit, $\Delta\chi^2$ (19) $=$ 22.44; $p = 0.264$. Furthermore, the ΔCFI was 0.001, with that of the metric model larger. Hence, the evidence supports the full metric invariance by race-ethnicity of the WNSQ-SF.

We next estimated a model of *scalar* invariance by constraining the item intercepts for the latent factors to equality across the 2 groups. This model fit well also: Total sample: χ^2 (110) = 173.47; $p < 0.001$; [for each group—WW $\chi^2 = 81.28$, WoC $\chi^2 = 92.19$]; CFI = 0.960; TLI = 0.952; RMSEA = 0.047 (90% CI = 0.033, 0.061); SRMR = 0.056. All but one (10, loading on SP) of the standardized factor loadings for the 2 groups were significant. Ranges for the significant factor loadings on the group factors were WW 0.39–0.79, WoC 0.44–0.79. Ranges for the general factors were WW 0.36–0.66, WoC 0.29–0.62. The scalar model did not degrade fit when compared with the metric model: $\Delta\chi^2$ (7) $=$ 7.60, $p = 0.369$, and ΔCFI = 0.001, below the cutoff of 0.01, indicating that full scalar invariance held.

Establishing scalar invariance, i.e., strong invariance [20], is a precondition for continuing to a test of strict factorial invariance [18, 55], thus, we proceeded to test residuals invariance. This model also fit adequately: Total sample: χ^2 (93 = 167.45; $p < 0.001$; [for each group—WW $\chi^2 = 83.19$, WoC $\chi^2 = 84.26$]; CFI = 0.964; TLI = 0.948; RMSEA = 0.567 (90% CI = 0.042, 0.069); SRMR = 0.047. All but one (for item 10, loading on SP) of the standardized factor loadings for White women were significant; whereas for women of color there were 4 non-significant factor loadings (10 & 7, both loading on SP and 15 & 23, both loading on the general factor. Ranges for the significant factor loadings on the group factors were WW 0.39–0.78, WoC 0.32–0.74. Ranges for the general factors were WW 0.35–0.65, WoC 0.34–0.66. The residuals model did not degrade fit when compared with the metric model: $\Delta\chi^2$ (17) $=$ 26.81, $p = 0.061$, and ΔCFI = 0.004 with that of the residuals model larger, indicating that residuals and therefore strict invariance held. Furthermore, bootstrapped confidence intervals indicated that, for every single item loading on both the group factors and the general factor, the comparison between WW and WoC was invariant. To summarize these results, evidence was found for strong and strict invariance by race/ethnicity.

Assessment of Measurement Invariance of the WNSQ-SF by Sexual Orientation

Finally, we assessed the invariance of the WNSQ-SF responses across sexual orientation, estimating a series of nested models with the heterosexual women (HW) and sexual minority women (SMW) participants treated as separate subsamples. Again, the dimensional structure for all models was bifactor; the χ^2 was statistically significant for all models, while the CFI, TLI, RMSEA, and SRMR, were at acceptable levels.

We first assessed *configural* invariance, finding good fit to the data: total sample: χ^2 (84) = 127.04; $p < 0.001$; for each group—HW χ^2 (42) = 82.27; $p < 0.001$, SMW χ^2 (42) = 44.76; $p = 0.357$]; CFI = 0.972; TLI = 0.955; RMSEA = 0.045 (90%

CI = 0.028, 0.060); SRMR = 0.046. All but one (10, loading on Self-Pleasuring) of the standardized factor loadings for the HW group were significant; all but three (7, 10, & 20, all loading on SP) for the SMW group were significant. Ranges for the significant factor loadings on the group factors were: HW 0.21–0.74, SMW 0.28–0.91. Ranges for the general factors were: HW 0.32–0.60, SMW 0.29–0.66. Thus, configural invariance by sexual orientation was largely evidenced.

Next, we tested for *metric* invariance, finding good fit: Total sample: $\chi 2$ (103) = 158.37; $p < 0.001$; [for each group—HW $\chi 2 = 88.77$, SMW $\chi 2 = 69.60$]; CFI = 0.964; TLI = 0.953; RMSEA = 0.046 (90% CI = 0.031, 0.059); SRMR = 0.055. All but one (10, loading on SP) of the standardized factor loadings for the HW group was significant, and all for the SMW group were significant. Ranges for the significant factor loadings on the group factors were: HW 0.41–0.72, SMW 0.21–0.79. Ranges for the general factors were: HW 0.33–0.61, SMW 0.33–0.59. Using the scaled chi square difference test, this more parsimonious metric invariance model did degrade fit, $\Delta\chi^2$ (19) = 33,50; $p = 0.021$. However, the ΔCFI was only 0.008 smaller, missing the criterion of 0.01, suggesting that the more constrained metric model did not worsen fit. Hence, we conclude that full metric invariance is largely supported by these data, but some unknown sources of inequality in factor loadings remain.

We next estimated *scalar* invariance, which also fit well: Total sample: $\chi 2$ (110) = 171.06; $p < 0.001$; [for each group—HW $\chi 2 = 92.28$, SMW $\chi 2 = 78.78$]; CFI = 0.960; TLI = 0.952; RMSEA = 0.046 (90% CI = 0.032, 0.060); SRMR = 0.054. All but one (10, loading on SP) of the standardized factor loadings for the 2 groups were significant. Ranges for the significant factor loadings on the group factors were: HW 0.42–0.75, SMW 0.36–0.80. Ranges for the general factors were: HW 0.33–0.59, SMW 0.33–0.59. The scalar model did not degrade fit when compared with the metric model: $\Delta\chi^2$ (7) = 13.26, $p = 0.066$, and ΔCFI = 0.004, below the cutoff of 0.01; thus, scalar invariance held.

Hence, we proceeded to test residuals invariance. This model also fit adequately: Total sample: $\chi 2$ (93) = 184.01; $p < 0.001$; [for each group—HW $\chi 2 = 98.59$, SMW $\chi 2 = 85.42$]; CFI = 0.954; TLI = 0.935; RMSEA = 0.062 (90% CI = 0.048, 0.075); SRMR = 0.063. All but one (10, loading on SP) of the standardized factor loadings for the HW group were significant; whereas for SMW, there were 2 non-significant factor loadings (7 & 10) both loading on SP. Ranges for the group factors were: HW 0.44–0.78, SMW 0.32–0.86. Ranges for the significant factor loadings on the general factors were: HW 0.36–0.60, SMW 0.20–0.61. The residuals model did not degrade fit when compared with the metric model: $\Delta\chi^2$ (17) = 5.11, $p = 0.997$, and ΔCFI = − 0.006, missing the cutoff of 0.01, indicating that residuals and therefore strict invariance held. Furthermore, bootstrapped confidence intervals indicated that, for every single item loading on both the group factors and the general factor, the comparison between HW and SMW was invariant. To summarize these results, evidence was found for strong and strict invariance by sexual orientation.

Table 4 Raw score means, standard deviations, and alpha coefficients of the WNSQ-SF and validity variables

Scale	M	SD	α
WNS	2.70	0.92	.87
CS	1.72	0.90	.79
SP	2.37	1.28	.81
SI	3.92	1.15	.60
SME	2.24	1.04	.81
SDQ	.00	0.81	.82
SIPS	2.45	1.28	.92
SAS	3.02	0.61	.81
INI	3.03	0.74	.69
REF	3.13	0.91	.75
STD	2.88	1.11	.85
MAST	4.12	0.93	.86

WNS Women's general nontraditional sexuality- short form, *CS* casual sex, *SP* self pleasuring, *SI* sexual interest, *SME* sex as a means to an end; *SDQ* Sex Drive Questionnaire; *SIPS* Sex Is Power Scale; *SAS* Sexual Assertiveness Scale; *INI* Initiation; *REF* Refusal; *STD* Pregnancy—*STD* Prevention; *MAST* Masturbation

Descriptive Statistics

Table 4 shows the means, SD's and alpha coefficients for the WNSQ-SF and the validity variables. In addition, we report the means and SD's data separately by race/ethnicity and sexual orientation in the Online Supplement Tables 3 and 4.

Discussion

The purpose of the present study was to develop and assess the psychometric properties of the WNSQ-SF using a large (N=519) and somewhat diverse sample (24.2% non-White, 21.4% non-heterosexual), and to confirm these results using a second sample (N=238). This study extends prior research on the WNSQ [23] and provides an abbreviated questionnaire that offers improved psychometric properties. The WNSQ-SF measures women's nontraditional sexuality through four subscales (casual sex, self-pleasuring, sexual interest, and sex as a means to an end) and a general women's non-traditional sexuality factor, corresponding to the total scale score.

The present study determined first that the bifactor model provided the best fit for the WNSQ compared with other dimensional models (i.e., common factors, hierarchical, unidimensional), although there was room for improvement in terms of both the fit statistics and non-significant factor loadings. The bifactor model partitions the variance in each item-level response between its respective group factor, the general women's non-traditional sexuality factor, and error. This is important because subscales may remeasure the general factor to varying degrees.

This study found that the WNSQ-SF also best fit the data as a bifactor model in comparison to the other dimensional models, and that its fit statistics were greatly

improved over those of the WNSQ. In addition, we calculated bifactor reliability and dimensionality diagnostic measures for the WNSQ-SF, finding that the omega coefficients suggest good reliability for the instrument. We also found that the total raw score can be used to represent women's general nontraditional sexuality as can the raw subscale scores for casual sex, sexual interest, and sex as a means to an end. That is because in all group factors except for self-pleasuring (SP), the majority of the variance taps the group factors, whereas for SP the majority of the variance remeasures the general factor. We will return to the SP subscale when we discuss the measurement invariance results. Finally, using these indices we found that it would not be advisable to use the WNSQ-SF as a unidimensional model.

We also examined construct (convergent and discriminant) evidence for the validity of the subscales, using the latent variable approach, which has the advantage of separating out variation due to the group factors from that due to the general factor. Using measures of sex drive, sexual empowerment, masturbation, and sexual assertiveness as validity scales, we found that all 23 of 24 tests fully supported the hypotheses, providing convergent and discriminant evidence for the validity of the casual sex, self-pleasuring, sexual interest, and sex as a Means to an End scales. However, as noted above, the findings involving the Sex Drive Questionnaire used as a validity variable should be interpreted with caution, which impacts the convergent evidence for validity of the Self-Pleasuring and sexual interest subscales. Future research should investigate other evidence for validity including concurrent evidence, which for the WNSQ was only weakly supported [23].

Finally, we examined measurement invariance between groups based on race/ethnicity (comparing White to non-White participants) and sexual orientation (comparing heterosexual to non-heterosexual participants), finding in both sets of analyses evidence for strong and strict invariance. This means that members of both dominant and marginalized groups based on race/ethnicity and sexual orientation understand the scale scores in the same way, including the difference between the scale score points, and the zero points of the scales, which allows the comparisons of means between these groups. Because residuals invariance was found, it also means that the scale measures the constructs with the same degree of precision between groups.

Further, we did not consider differences in women's nontraditional sexuality outside of the gender binary; hence, future research should consider assessing this scale with the transgender/gender non-conforming population. Finally, it was observed that through these two sets of ME/I analyses that one or more of the items loading on the SP factor were nonsignificant. These results in combination with those from the bifactor reliability and dimensionality diagnostic measures suggest that the SP subscale should be used cautiously. On the other hand, we found evidence using a latent variable approach (that separates out the influence of the general from the group factors) supporting the validity of the SP scale, and validity is typically considered more important than reliability.

The present findings should be interpreted cautiously with respect to several key limitations. Some limitations result from several aspects of the sample. First, not all participants were either in a relationship or were sexually active, which is a limitation of the present study. Women not in a relationship or sexually inactive might not have viewed the question the same as sexually active women and

women in a relationship. However, it is difficult to know what kind of difference that might have made in the results. Next, although the sample was large and drawn from both community and college sources, it was still a convenience sample, and participants self-selected. Third, most of our sample consisted of participants who were predominantly White and Christian and a significant portion was young and currently in college. Hence, the present results are likely not generalizable to the overall population. However, it is difficult to know what kind of difference that might have made in the results.

Fourth, although we made a case in the Introduction for comparing White women to women of Color, and heterosexual women to sexual minority women, aggregating minoritized identities is a considerable limitation, and therefore the results must be interpreted cautiously. There might be differences in understanding the scale between different minoritized racial and ethnic groups, or between different sexual orientation groups, that would be missed in the aggregated analyses that we conducted. For example, Asian-American women might understand the scale score, the difference between the scores, and the zero-point in very different way from African American women, and such differences would be masked in our approach. There is also substantial diversity between members within any particular cultural group. Further, we did not consider differences in women's nontraditional sexuality outside of the gender binary; hence, future research should consider assessing this scale with the transgender/gender non-conforming population. Additional research is needed using more sophisticated sampling procedures to address these limitations by gathering a truly representative sample of the United States population in terms of race, ethnicity, sexual orientation and gender identity, all of whom are in a relationship and are sexually active. Finally, it would have been informative to determine if there were differences between the community and college samples; unfortunately, we did not collect data on whether participants were college students or a non-student community member, and thus could not run the analyses, which is another limitation.

Second, by consulting the recommendation of Tabachnick and Fidell [52] to require a loading of 0.32 to load, we found that there are two low factor loadings on the WNSQ-SF: item 10, loading on the Self-Pleasuring subscale, and item 23, loading on the General Women's Non-Traditional Sexuality scale. Although these are limitations, they are partially mitigated by the alpha coefficients for Self-Pleasuring of 0.81, and for the General scale, of 0.87.

Finally, the study relies on self-report data which introduces the possibility of socially desirable responding (SDR). SDR was not measured in our study; however, a recent article demonstrated that SDR is not always a problem [54]. In addition, the data are cross-sectional. Furthermore, the alpha coefficients for two scales were > 0.70: sexual interest (0.60), a subscale of the WNSQ-SF, and Initiation, a validity variable. The former is partially mitigated by its Omega value of 0.74. We also did not test for test–retest reliability; this would be important to achieve in future research.

There are implications of the present study for practitioners and for college student personnel administrators. The WNSQ-SF would be an appropriate and quick resource in assessing for women's engagement in nontraditional sexuality. There

could be value in examining whether nontraditional sexuality is a contributing factor to stress or conflict for women who subscribe to traditional femininity ideology. Additionally, the WNSQ-SF is short enough to minimize participant fatigue, allowing for a more accurate examination of women's nontraditional sexuality in larger batteries. With regard to college administrators, some evidence links specific traditionally feminine beliefs to behaviors directly associated with sexual assault risk and indicates that sexual assertiveness is a protective factor [58]. The sex as a means to an end subscale could be used to assess sexual assertiveness.

Conclusions

In conclusion, the present study provides an abbreviated version of a measure with much improved psychometric properties, confirms that the WNSQ-SF is best modeled as bifactor structure, provides convergent and discriminant construct evidence for the validity of the subscales, and provides evidence of the measure's strict invariance between two racial/ethnic groups and two groups based on sexual orientation for the five latent factors. As such, the current refinement demonstrates significant psychometric strengths, and we encourage its adoption for investigating women's non-traditional sexuality and its correlates. Finally, the present study demonstrates the advantages of using SEM and latent variables in assessing psychometric properties of scales, especially regarding dimensionality, validity, and measurement equivalence.

References

1. Alexander, M. (2012). *The New Jim Crow: Mass incarceration in the age of colorblindness* (2nd ed.). New Press.
2. Beres, M. A., & Farvid, P. (2010). Sexual ethics and young women's accounts of heterosexual casual sex. *Sexualities, 13*, 377–393. https://doi.org/10.1177/1363460709363136
3. Bowman, C. P. (2014). Women's masturbation: Experiences of sexual empowerment in a primarily sex-positive sample. *Psychology of Women Quarterly, 38*, 363–378. https://doi.org/10.1177/0361684313514855
4. Catania, J. A. (1998). Health protective sexual communication scale. In J. Nageotte (Ed.), *Sexual Risk* (pp. 544–547). Sage Publications.
5. 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
6. Cheung, G. W., & Lau, R. S. (2012). A direct comparison approach for testing measurement invariance. *Organizational Research Methods, 15*, 167–198. https://doi.org/10.1177/1094428111421987
7. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.
8. Connell, R. W., & Messerschmidt, J. W. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society, 19*, 829–859. https://doi.org/10.1177/0891243205278639
9. Deshotels, T. H., Tinney, M., & Forsyth, C. J. (2012). McSexy: Exotic dancing and institutional power. *Deviant Behavior, 33*, 140–148. https://doi.org/10.1080/01639625.2011.573370

10. Dueber, D. M. (2016, November). Bifactor Indices Calculator: A Microsoft Excel-based tool to cal-
     culate various indices relevant to bifactor CFA models. http://sites.education.uky.edu/apslab/resou
     rces.
11. Endendijk, J. J., van Baar, A. L., & Deković, M. (2020). He is a stud, she is a slut! A meta-analysis
     on the continued existence of sexual double standards. *Personality & Social Psychology Review, 24*,
     163–190.
12. Erchull, M., & Liss, M. (2013). Exploring the concept of perceived female sexual empowerment:
     Development and validation of the Sex is Power Scale. *Gender Issues.* https://doi.org/10.1007/
     s12147-013-9114-6
13. Fahs, B., Swank, E., & Shamb, A. (2020). "I just go with it": Negotiating sexual desire discrepan-
     cies for women in partnered relationships. *Sex Roles, 83*, 226–239.
14. Farvid, P., & Braun, V. (2018). "You worry, 'cause you want to give a reasonable account of your-
     self'": Gender, identity management, and the discursive positioning of "risk" in men's and women's
     talk about heterosexual casual sex. *Archives of Sexual Behavior, 47*, 1405–1421. https://doi.org/10.
     1007/s10508-017-1124-0
15. Farvid, P., Braun, V., & Rowney, C. (2017). 'No girl wants to be called a slut!': Women, hetero-
     sexual casual sex and the sexual double standard. *Journal of Gender Studies, 26*, 544–560. https://
     doi.org/10.1080/09589236.2016.1150818
16. Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional
     Psychology: Research and Practice, 40*, 532–538.
17. Goetz, C., Coste, J., Lemetayer, F., Rat, A., Montel, S., Recchia, S., Debouverie, M., Pouchot, J.,
     Spitz, E., & Guillemin, F. (2013). Item reduction based on rigorous methodological guidelines is
     necessary to maintain validity when shortening composite measurement scales. *Journal of Clinical
     Epidemiology, 66*, 710–718. https://doi.org/10.1016/j.jclinepi.2012.12.015
18. Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance
     in aging research. *Experimental Aging Research, 18*, 117–144. https://doi.org/10.1080/0361073920
     8253916
19. Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social
     and personality psychology. *Advances in Methods and Practices in Psychological Science.* https://
     doi.org/10.1177/2515245919882903
20. Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford.
21. Kozee, H. B., Tylka, T. L., Augustus-Horvath, C. L., & Denchik, A. (2007). Development and psy-
     chometric evaluation of the Interpersonal Sexual Objectification Scale. *Psychology of Women Quar-
     terly, 31*, 176–189. https://doi.org/10.1111/j.1471-6402.2007.00351.x
22. Levant, R. F., Richmond, K., Cook, S., House, A., & Aupont, M. (2007). The femininity ideology
     scale: Factor structure, reliability, validity, and social contextual variation. *Sex Roles, 57*, 373–383.
23. Levant, R. F., Rankin, T. J., Hall, R. J., Smalley, K. B., David, K., & Williams, C. (2012). The meas-
     urement of nontraditional sexuality in women. *Archives of Sexual Behavior, 41*, 283–295.
24. Levant, R. F., Hall, R. J., & Rankin, T. J. (2013). Male Role Norms Inventory-Short Form (MRNI-
     SF): Development, confirmatory factor analytic investigation of structure, and measurement invari-
     ance across gender. *Journal of Counseling Psychology, 60*, 228–238. https://doi.org/10.1037/a0031
     545
25. Levant, R. F., Hall, R. J., Weigold, I. K., & McCurdy, E. R. (2015). Construct distinctiveness and
     variance composition of multidimensional instruments: Three short-form masculinity measures.
     Journal of Counseling Psychology, 62, 488–502. https://doi.org/10.1037/cou0000092
26. Levant, R. F., & Richmond, K. (2016). The gender role strain paradigm and masculinity ideologies.
     In Y. J. Wong & S. R. Wester (Eds.), *APA Handbook on Men and Masculinities* (pp. 23–49). Ameri-
     can Psychological Association.
27. Levant, R. F., Alto, K. M., McKelvey, D. K., Richmond, K., & McDermott, R. C. (2017). Variance
     composition, measurement invariance by gender, and validity of the Femininity Ideology Scale-
     Short Form. *Journal of Counseling Psychology, 64*, 708–723.
28. Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel:
     Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151–173. https://doi.
     org/10.1207/S15328007SEM0902_1
29. MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of
     sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149. https://doi.org/
     10.1037/1082-989X.1.2.130

30. Marks, M. J., & Wosick, K. (2017). Exploring college men's and women's attitudes about women's sexuality and pleasure via their perceptions of female novelty party attendees. *Sex Roles: A Journal of Research, 77*, 550–561. https://doi.org/10.1007/s11199-017-0737-z

31. McKinley, N. M., & Hyde, J. S. (1996). The objectified body consciousness scale: Development and validation. *Psychology of Women Quarterly, 20*, 181–215. https://doi.org/10.1111/j.1471-6402.1996.tb00467.x

32. Milnes, K. (2004). What lies between romance and sexual equality? A narrative study of young women's sexual experiences. *Sexualities, Evolution and Gender, 6*(2–3), 151–170. https://doi.org/10.1080/14616660412331325169

33. Morokoff, P. J., Quina, K., Harlow, L. L., Whitmire, L., Grimley, D. M., Gibson, P. R., & Burkholder, G. J. (1997). Sexual Assertiveness Scale (SAS) for women: Development and validation. *Journal of Personality and Social Psychology, 73*, 790–804. https://doi.org/10.1037/0022-3514.73.4.790

34. Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén. Myers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: Design and interpretation* (2nd Ed). Thousand Oaks, CA: Sage

35. Ostovich, J. M., & Sabini, J. (2004). How are sociosexuality, sex drive, and lifetime number of sexual partners related? *Personality and Social Psychology Bulletin, 30*(1255), 1266. https://doi.org/10.1177/0146167204264754

36. Parry, D. C. (2016). "Skankalicious": Erotic capital in women's flat track roller derby. *Leisure Sciences, 38*, 295–314. https://doi.org/10.1080/01490400.2015.1113149

37. Petersen, J. L., & Hyde, J. S. (2010). A meta-analytic review of research on gender differences in sexuality, 1993–2007. *Psychological Bulletin, 136*(1), 21–38. https://doi.org/10.1037/a0017504.supp(Supplemental)

38. Pleck, J. H. (1981). *The myth of masculinity*. MIT Press.

39. Pleck, J. H. (1995). The gender role strain paradigm: An update. In R. F. Levant & W. S. Pollack (Eds.), *A new psychology of men* (pp. 11–32). Basic Books.

40. Price, J., Patterson, R., Regnerus, M., & Walley, J. (2016). How much more XXX is generation X consuming? Evidence of changing attitudes and behaviors related to pornography since 1973. *Journal of Sex Research, 53*, 12–20. https://doi.org/10.1080/00224499.2014.1003773

41. Priem, R. L., Lyon, D. W., & Dess, G. G. (1999). Inherent limitations of demographic proxies in top management team heterogeneity research. *Journal of Management, 25*, 935–953. https://doi.org/10.1177/014920639902500607

42. Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behaviorual Research, 47*, 667–696. https://doi.org/10.1080/00273171.2012.715555

43. Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

44. Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling a bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5–26. https://doi.org/10.1177/0013164412449831

45. Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. https://doi.org/10.1037/met0000045

46. Russell, D. W., Kahn, J. H., Spoth, R., & Altmaier, E. M. (1998). Analyzing data from experimental studies: A latent variable structural equation modeling approach. *Journal of Counseling Psychology, 45*, 18–29. https://doi.org/10.1037/0022-0167.45.1.18

47. Satorra, A., & Bentler, P. M. (1999). A scaled difference chi-square test statistic for moment structure analysis. #260, *UCLA Statistics Series*.

48. Sevi, B., Aral, T., & Eskenazi, T. (2018). Exploring the hook-up app: Low sexual disgust and high sociosexuality predict motivation to use Tinder for casual sex. *Personality and Individual Differences, 133*, 17–20. https://doi.org/10.1016/j.paid.2017.04.053

49. Simpson, J. A., & Gangestad, S. (1991). Individual differences in sociosexuality: Evidence for convergent and discriminant validity. *Journal of Personality and Social Psychology, 60*, 870–883.

50. Spector, I. P., Carey, M. P., & Steinberg, L. (1996). The Sexual Desire Inventory: Development, factor structure, and evidence of reliability. *Journal of Sex & Marital Therapy, 22*, 175–190. https://doi.org/10.1080/00926239608414655

51. Strager, S. (2003). What men watch when they watch pornography. *Sexuality & Culture: An Interdisciplinary Quarterly, 7*, 50–61. https://doi.org/10.1007/s12119-003-1007-5

52. Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (Fifth Editon)*. Pearson.

53. Takiff, H. A., Sanchez, D. T., & Stewart, T. L. (2001). What's in a name? The status implications of students' terms of address for male and female professors. *Psychology of Women Quarterly, 25*, 134–144. https://doi.org/10.1111/14716402.00015

54. Tracey, T. J. G. (2016). A note on socially desirable responding. *Journal of Counseling Psychology, 63*, 224–232. https://doi.org/10.1037/cou0000135

55. Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70. https://doi.org/10.1177/109442810031002

56. Vera, E. M., & Speight, S. L. (2003). Multicultural competence, social justice, and counseling psychology: Expanding our roles. *The Counseling Psychologist, 31*, 253–271. https://doi.org/10.1177/0011000003031003001

57. Wentland, J. J., Herold, E. S., Desmarais, S., & Milhausen, R. R. (2009). Differentiating highly sexual women from less sexual women. *Canadian Journal of Human Sexuality*, *18*, 169–182. Retrieved from   http://search.ebscohost.com.ezproxy.uakron.edu:2048/login.aspx?direct=true&db=psyh&AN=2010-01507-002&site=ehost-live

58. Wigderson, S., & Katz, J. (2015). Feminine ideology and sexual assault: Are more traditional college women at greater risk? *Violence Against Women, 21*(5), 616–631. https://doi.org/10.1177/1077801215573333

59. Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*, 806–838. https://doi.org/10.1177/0011000006288127