Journal of Affective Disorders 278 (2021) 122–130

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Sexuality and gender invariance of the PHQ-9 and GAD-7: Implications for 16 identity groups

Nicholas C. Borgogna[a,*], Rachel E. Brenner[b], Ryon C. McDermott[a]

[a] The University of South Alabama, United States
[b] Colorado State University, United States

ARTICLE INFO

Keywords:
Minority stress
Measurement invariance
LGBTQ
Depression
Anxiety
Screeners

ABSTRACT

Background: The Patient Health Questionnaire-9 (PHQ-9) and the Generalized Anxiety Disorder-7 (GAD-7) are two commonly used screening tools for depression and anxiety, respectively. Despite the widespread use of these instruments, researchers have yet to examine whether scores may differ as a function of gender identity or sexuality orientation. **Method**: Using data from the 2018 and 2019 National Healthy minds study (N = 46,672), the present study tested each instrument for measurement invariance across 16 gender and sexual minority groups. **Results**: Multigroup structural equation modeling revealed that several sexual minority groups evidenced questionable fit indices for both measures. Gay men, questioning women, and queer men had unacceptable fit indices for the PHQ-9. Only cisgender heterosexual men and women evidenced residual invariance on the PHQ-9. All gender minority groups evidenced significantly higher factor loadings for item 9 (the self-harm indicator) for the PHQ-9. Most identity groups evidenced scalar or partial scalar invariance for the GAD-7; however, no groups evidenced residual invariance on the GAD-7. **Limitations**: Findings may not generalize to non-college student populations. **Conclusions**: Researchers should weight means when conducting between group comparisons for groups that failed scalar invariance. Gender and sexual minorities may have inflated scores using the PHQ-9 and GAD-7 cut-offs.

1. Introduction

A strong body of research indicates sexual and gender minority (LGBTQ+) individuals disproportionately suffer from psychopathology compared to their cisgender and heterosexual counter parts (Borgogna et al., 2018; Ross et al., 2018; Sutter and Perrin, 2016). Despite improvements in mental health awareness, prevention, and interventions for LGBTQ+ individuals (Holman and Oswald, 2016; Martin et al., 2018; Pryor et al., 2017), continued work is necessary. Particularly, psychometric comparisons of popular mental health measures represent an area in need of further research. Most mental health outcome measures have been validated using primarily cisgender/heterosexual samples. Even if LGBTQ+ individuals participate in validation studies, they are often removed or are combined with other LGBTQ+ participants when examined. This practice, though at times necessary for statistical modeling, promulgates an empirical assumption that all LGBTQ+ people respond to mental health measures in the same manner as cisgender/heterosexual participants. In turn, any potential measurement bias from this practice could lead to inaccurate representations of the true mental health discrepancies across LGBTQ+ populations. The Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001) and the Generalized Anxiety Disorder-7 (GAD–7; Spitzer et al., 2006) are arguably the most popular mental health measures in both research and practice. By extension, psychometric validations of the factor structures of the PHQ-9 and GAD-7 are necessary across LGBTQ+ populations to ensure comparisons using these measures are not distorted. However, researchers have yet to conduct large scale measurement invariance examinations on the PHQ-9 and GAD-7 across specific LGBTQ+ groups. We addressed this problem by conducting a large-scale measurement invariance study of the PHQ-9 and GAD-7 across 16 sexual and gender identity samples.

1.1. LGBTQ+ minority stress

Meyer's minority stress model (2003, 2015) is the premier theoretical framework for understanding LGBTQ+ mental health disparities. Meyer's model extends social stress theory (c.f., Burke, 1991) to LGBTQ+ individuals. Specifically, sexual and gender minority groups experience additive stress from having a marginalized identity. Discrimination experiences (including perceived discrimination;

* Corresponding author.
E-mail address: nicholascborgogna@gmail.com (N.C. Borgogna).

https://doi.org/10.1016/j.jad.2020.09.069
Received 14 February 2020; Received in revised form 29 May 2020; Accepted 11 September 2020
Available online 15 September 2020

Borgogna and McDermott, 2019) such as microaggressions and discriminatory policies function as an external stressors in this framework. Internalized heterosexism and homonegativity function as internalized stressors. Empirical findings align with the minority stress model in predicting mental and physical health problems for LGBTQ+ individuals (Lefevor et al., 2019; Meyer and Frost, 2013; Parra et al., 2016).

Sexual and gender minority individuals also experience nuanced forms of minority stress unique to specific LGBTQ+ groups. For instance, transgender individuals experience gender-related forms of minority stress (e.g., misgendering) that are not necessarily experienced by sexual minority individuals (McLemore, 2018; Velez et al., 2017). Similarly, gender and sexual minority stressors differ *within* gender minority groups. Indeed, genderqueer individuals experience unique stressors associated with identifying outside of the gender binary (Hyde et al., 2019; Matsuno and Budge, 2017; Richards et al., 2016). Similarly, non-monosexual individuals (e.g., bisexual, pansexual) experience minority stressors associated with being a minority within the sexual minority community (e.g., biphobia; Alarie and Gaudet, 2013; Ghabrial and Ross, 2018; La Roi et al., 2019; Lambe et al., 2017).

Given the unique nature of LGBTQ+ identities (Scheer et al., 2019), it is possible that the mental health presentations are also unique. That is, the depression associated with LGBTQ+ minority stress is categorically different from the depression associated with a divorce or poor diet. Researchers have highlighted this problem in conjunction with the classification of psychological disorders more broadly (e.g., Hayes et al., 1996; Jablensky, 2016; Westen, 2012). Though a solution to the syndromal classification of mental illness is beyond the scope of the current report, researchers have developed sophisticated techniques to examine whether measures that are designed to assess latent constructs, such as depression, are equivalent between groups in terms of their meaning, scale, and degree of precision. By extension, researchers can test whether commonly used instruments that were validated in primarily cisgender/heterosexual samples measure the same constructs in LGBTQ+ samples. While many researchers have conducted such psychometric evaluations on the PHQ-9 and GAD-7 (e.g., Teymoori et al., 2019), there is a notably lack of comparisons across sexual/gender diverse samples.

1.2. Measurement invariance

Measurement invariance testing is the primary statistical technique for such explorations (Meade et al., 2008; Vandenberg and Lance, 2000). Measurement invariance is generally tested in four forms/steps.

1. *Configural Invariance*: At the most basic level, responses to an instrument will demonstrate the same factor structure. For example, a unidimensional depression measure in cisgender/heterosexuals should evidence a unidimensional pattern in LGBTQ+ individuals.
2. *Metric Invariance:* Assuming configural invariance is achieved, researchers can test whether the instrument items have the same relationship to the underlying latent construct across groups (i.e., equivalent factor loadings; Xu and Tracey, 2017). For example, metric invariance would suggest that the depression construct, as measured by the PHQ–9, is equivalent for cisgender/heterosexuals and LGBTQ+ individuals.
3. *Scalar Invariance:* Assuming metric invariance is met, researchers can then test the equivalence of intercepts between groups. Support for scalar invariance suggests equivalent means between groups. In other words, a mean score of 15 on the PHQ-9 represents the same level of depression for cisgender/heterosexuals as LGBTQ+ individuals.
4. *Residual Invariance:* Finally, if scalar invariance is supported, researchers can test the equivalence of errors. This is the most rigorous form of invariance. Support for residual invariance would suggest that the instrument measures the same underlying construct between groups with an equivalent degree of precision. If residual invariance is achieved, an instrument may be assumed to be entirely equivalent between groups in terms of its measurement function.

Understanding the measurement invariance of the PHQ-9 and GAD-7 between LGBTQ+ and cisgender/heterosexual individuals is critical for researchers and clinicians, as these instruments were validated using primarily cisgender/heterosexual samples. Furthermore, it is important to understand how these instruments function across differing LGBTQ+ groups (i.e., Is the PHQ-9 measuring depression equally between gay men and bisexual women?). Preliminary research on highly specified scales suggests some degree of equivalence among these groups. For instance, all eight subscales of the Lesbian, Gay, and Bisexual Identity Scale (Mohr and Kendra, 2011) have demonstrated scalar invariance between gay men and lesbian women (Niepel et al., 2019). Similarly, the Spiritual Values/Religion subscale of Self-Description Questionnaire III have demonstrated scalar invariance across heterosexual and non-heterosexual groups (Ong et al., 2019). Conversely, the Everyday Discrimination Scale (Williams et al., 1997) has demonstrated partial metric invariance between cisgender and non-cisgender participants (as well as between specific non-cisgender groups), whereas the Discrimination-Related Vigilance Scale (Clark et al., 2006) has demonstrated metric invariance between cisgender and non-cisgender groups broadly, but only partial metric invariance across specific gender identities (Bauerband et al., 2019). These findings suggest there may be important similarities or differences across sexual/gender minority groups, which could influence the way they respond to specific instruments.

To date, the PHQ-9 and GAD-7 lack measurement invariance examinations across sexual and gender minority groups. This is a substantial gap in the literature given the popularity of the PHQ-9 and GAD-7. To address this limitation, we conducted a large measurement invariance study across a sexual orientation and gender diverse dataset of college students. Given there is limited LGBTQ+ measurement invariance research across these instruments, our analyses were highly exploratory. However, consistent with the minority stress model (Meyer, 2003) and findings suggesting LGBTQ+ mental health problems are associated with factors that are not present in traditional mental health problems for cisgender/heterosexual individuals (e.g., Bauerband et al., 2019), we broadly hypothesized that some degree of measurement non-invariance would be present for both measures across every LGBTQ+ group compared to cisgender/heterosexual participants.

2. Method

2.1. Participants/Procedure

We analyzed data from the 2017–2018 and 2018–2019 Healthy Minds Study's (HMS) to explore the measurement invariance of multiple sexual and gender identities on the PHQ-9 and GAD-7. The HMS is an annual survey examining various mental health variables across college students from 60 different universities in the United States and Canada. We initially sampled $N = 62{,}025$ participants from the 2018–2019 dataset; however, because substantially unequal sample sizes between groups can bias measurement invariance analyses (Yoon and Lai, 2018), we randomly sampled participants in each identity group to establish equal or near equal groups, which involved occasional supplementation from the 2017–2018 dataset (specifics are noted in the results section and Tables 1 and 2). Prior to organizing identity groups, we cleaned the data by removing suspicious write-in demographic responses (e.g., apache helicopter), as well as participants who completed less than 80% of each measure. Demographically, the mean age was 23.05 ($SD = 6.80$). Racial distribution was 72% White, 13.6% Asian, 7.6% Black, 10.1% Latinx, 2.2% Middle Eastern, 1.5%

Table 1
Fit indices from individual group CFA's on the PHQ-9.

Group	*n*	X^2	*df*	CFI	TLI	RMSEA	90% CI RMSEA	SRMR	α
Heterosexual Men	226	47.17**	27	.970	.960	.057	[.028, 0.084]	.035	.894
Heterosexual Women	226	97.792***	27	.912	.882	.108	[.085, 0.131]	.052	.900
Gay Men	226	93.462***	27	.880	.840	.104	[.082, 0.128]	.065	.861
Lesbian Women	226	90.553***	27	.920	.893	.102	[.079, 0.126]	.051	.900
Bisexual Men	226	94.572***	27	.909	.879	.105	[.083, 0.129]	.050	.892
Bisexual Women	226	90.367***	27	.922	.895	.102	[.079, 0.125]	.048	.887
Questioning Men	226	80.453***	27	.913	.884	.094	[.070, 0.118]	.053	.873
Questioning Women	226	107.787***	27	.881	.841	.115	[.093, 0.138]	.059	.866
Queer Men	226†	121.697***	27	.861	.814	.125	[.103, 0.148]	.064	.866
Queer Women	226	64.980***	27	.946	.928	.079	[.055, 0.104]	.043	.885
Cisgender Men	240	48.215**	27	.967	.956	.057	[.029, 0.083]	.037	.889
Cisgender Women	240	95.449***	27	.911	.881	.103	[.081, 0.126]	.053	.888
Transgender Men	240†	108.673***	27	.923	.897	.112	[.091, 0.135]	.043	.918
Transgender Women	157†	59.293***	27	.945	.927	.087	[.057, 0.117]	.041	.900
Genderqueer-M	240†	113.479***	27	.892	.856	.116	[.094, 0.138]	.055	.888
Genderqueer-F	240	110.870***	27	.901	.868	.114	[.092, 0.136]	.053	.887

Note: X^2 = scaled chi-square, df = degrees of freedom, CFI = Comparative Fit Index, TLI = Tucker Lewis Index,.
RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, α = Internal consistency.
*$p < .05$, **$p < .01$, ***$p < .001$, †=Supplemented with 2017–2018 HMS.

Native American or Alaskan Native, 0.6% Hawaiian Native or Pacific Islander, and 1.5% write-in identity. HMS allows participants to identify with more than one race (i.e., percentages do not add up to 100%).

Participants specified sex assigned at birth (male, female) and gender identity (male, female, trans male / trans man, trans female / trans woman, genderqueer / gender non-conforming). Gender terms (i.e. "men," "women,") will be used in place of sex terms "male," "female" for the remainder of this paper. Participants were also allowed to write-in a gender identity. Based on these responses, those retained after the initial data cleaning procedures were organized into the following gender groups: cisgender men, cisgender women, transmen, transwomen, genderqueer-M (male assigned at birth) and genderqueer-F (female assigned at birth). Similar to race, participants could select multiple sexual orientations from a list of heterosexual, gay, lesbian, bisexual, questioning, queer, and a write-in category. Participants were first organized by those who only selected one sexual orientation. Responses with multiple selections were then reviewed for inclusion. For example, those who selected heterosexual, gay, and bisexual were included in the bisexual group. Similarly, if questioning was indicated, they were categorized in the "questioning" group. However, those who selected every category were removed for potential suspicious responding. Write-in participants with valid responses such as pansexual or pansexual/queer were included in the "queer" group. Although some participants also indicated an asexual identity, they could not be included due to insufficient power (see supplementary Table 1 for raw gender and sexual orientation demographics).

2.2. Measures

The Patient Health Questionnaire-9. The PHQ-9 (Kroenke et al., 2001) is a nine-item screener for major depressive disorder. Individuals rates items on a 4-point Likert-scale from 0 (*not at all*) to 3 (*nearly every day*). The PHQ-9 has demonstrated evidence of validity through associations with anxiety, depressive cognitions, and discrimination (Burns et al., 2012; Timmins et al., 2017), and has demonstrated evidence of internal consistency among sexual minority college students (α = 0.86; Grant et al., 2014), transgender and non-conforming adults (α = 0.91; Timmins et al., 2017), and heterosexual individuals (α = 0.89; Kroenke et al., 2001). We provide the current study alpha's in Table 1.

The Generalized Anxiety Disorder-7. The GAD-7 (Spitzer et al., 2006) is a seven item screener designed to assess for the presence of generalized anxiety disorder. Individuals rate the frequency of seven anxiety symptoms on a 4-point Likert-scale from 0 (*not at all*) to 3 (*nearly every day*). The GAD-7 has demonstrated evidence of validity through

Table 2
Fit indices from individual group CFA's on the GAD-7.

Sample	*n*	X^2	*df*	CFI	TLI	RMSEA	90% CI RMSEA	SRMR	α
Heterosexual Men	222	43.643**	14	.965	.947	.098	[.066, 0.131]	.031	.938
Heterosexual Women	222	55.430***	14	.934	.901	.115	[.085, 0.148]	.044	.910
Gay Men	222	46.577**	14	.959	938	.102	[.071, 0.136]	.042	.912
Lesbian Women	222	71.964***	14	.920	.880	.137	[.106, 0.169]	.064	.903
Bisexual Men	222	30.345**	14	.975	.963	.073	[.037, 0.108]	.033	.897
Bisexual Women	222	77.255***	14	.925	.888	.143	[.113, 0.175]	.060	.912
Questioning Men	222	37.431**	14	.960	.941	.087	[.054, 0.121]	.043	.892
Questioning Women	222	32. 436**	14	.967	.950	.077	[.042, 0.112]	.042	.879
Queer Men	222†	30.957**	14	.975	.962	.074	[.038, 0.109]	.031	.898
Queer Women	222	38.302**	14	.962	.943	.088	[.056, 0.123]	.043	.898
Cisgender Men	240	68.257***	14	.923	.884	.127	[.098, 0.158]	.043	.910
Cisgender Women	240	29.006**	14	.984	.975	.067	[.031, 0.101]	.023	.927
Transgender Men	240†	48.092***	14	.948	.922	.101	[.070, 0.133]	.049	.891
Transgender Women	155†	37.080**	14	.956	.933	.103	[.063, 0.144]	.044	.905
Genderqueer-M	240†	52.114**	14	.949	.923	.107	[.077, 0.138]	.034	.907
Genderqueer-F	240	35.085**	14	.970	.954	.079	[.047, 0.112]	.036	.886

Note: X^2 = scaled chi-square, df = degrees of freedom, CFI = Comparative Fit Index, TLI = Tucker Lewis Index,.
RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual.
*$p < .05$, **$p < .01$, ***$p < .001$, †=Supplemented with 2017–2018 HMS.

associations with depression, well-being, self-esteem, and discrimination (Timmins et al., 2017; Woodford et al., 2014), and has demonstrated evidence of internal consistency among sexual minority college students (α = 0.90; Woodford et al., 2014), transgender and nonconforming adults (α = 0.91; Timmins et al., 2017), and heterosexual individuals (α = 0.92; Spitzer et al., 2006). We provide the current study alpha's in Table 2.

2.3. Analytic plan

We conducted two sets of invariance analyses for each measure. The first set of analyses tested invariance across sexual identity groups nested within cisgender identities: heterosexual men, heterosexual women, gay men, gay women, bisexual men, bisexual women, questioning men, questioning women, queer men, and queer women. Our second set of analyses examined measurement invariance across gender identity groups: cisgender men, cisgender women, transmen, transwomen, genderqueer-M, and genderqueer-F.

To maintain equal or near equal group sample sizes, we conducted multiple descriptive analyses for each measure and randomly reduced the sample size of each group to be equal to the smallest "baseline" group for each set of measurement invariance analyses on both measures. We followed guidelines to determine what constituted enough power for our baseline group. This involved reviewing multiple "rules of thumb" for powering measurement invariance analyses (Chen, 2007; Kline, 2016; Kyriazos, 2018; Wang and Wang, 2012). Given both measures are primarily used as one-factor instruments (e.g., Teymoori et al., 2019) and contain less than 10 indicators, we attempted to maintain baseline samples of $n > 200$ per group. However, we allowed some samples of less than 200 (though we did not reduce the other groups), specifically for the transwomen, genderqueer-M, questioning men, and queer men groups respectively. For groups below 100, we supplemented the samples with participants from the 2017–2018 HMS (which underwent an identical data cleaning procedure). Because measurement invariance analyses may become untrustworthy in small samples, we did not conduct any analyses with groups of $n < 100$ (unless supplemented).

Primary analyses were conducted using Mplus version 8 (Muthén and Muthén, 2016). Prior to conducting our multi-group invariance analyses, we conducted individual confirmatory factor analyses (CFA's) for each group to ensure sound factor structure. For evaluating model-fit across models, we used the following indices and cutoffs (Kline, 2016): Comparative Fit Index (CFI), Tucker-Lewis index (TLI; ≥.90 indicate acceptable fit and ≥ 0.95 indicate preferable fit for CFI and TLI), the root-mean-square error of approximation (RMSEA) with 90% confidence intervals (CIs; low values of 0.06 or less and high values ≤ 0.10 indicate a good fit), and the standardized root-mean-square residual (SRMR; ≤ 0.08 indicates a good fit). The chi-square test statistic was also reported (non-significance indicates perfect fit); however, this was interpreted with caution, given its sensitivity to sample size. All SEM analyses utilized a maximum likelihood estimator with robust standard errors, as well as full information maximum likelihood estimation (FIML) to address any missing inputs.

Any group that evidenced unacceptable model fit (CFI below 0.90 and/or SRMR above 0.08) was not included in the invariance analyses. Because both measures were validated on primarily cisgender heterosexual men and women samples, we used cisgender men, cisgender women, heterosexual men, and heterosexual women as our referent groups. For each analysis, we used whichever group between cisgender/heterosexual men and women evidenced the best initial fit as determined by the size of the chi-square statistic.

Once an acceptable CFA structure was identified for each group, we tested configural, metric, scalar, and residual invariance for each measure across our specified sexual and gender identity groups. Conducting each invariance test was predicated upon successful identification of the former (e.g., configural invariance had to be demonstrated before metric invariance was examined). To evaluate invariance in each model, we initially conducted a scaled chi-square difference test. A non-significant chi-square difference provides support for invariance (Kline, 2016). However, the chi-square difference test is extremely sensitive to large sample sizes. Thus, in the event of a significant chi-square difference test we utilized alternative measurement invariance tests. First, we examined the change in the CFI, where, ΔCFI ≤ 0.01 suggests invariance (Cheung and Rensvold, 2002). We also calculated the bias-corrected bootstrapped CIs of the difference between groups on parameters of interest (e.g., factor loadings, intercepts, or residuals; Cheung and Lau, 2012) to identify particular items that may be invariant/non-invariant. A confidence interval containing zero for the between-groups difference on a particular unstandardized coefficient suggests invariance (Cheung and Lau, 2012). If more than half the items were found to be invariant, the measure was considered "partially invariant" and further measurement invariance analyses proceeded with the invariant items.

3. Results

3.1. Participant selection

PHQ-9. After data cleaning, most groups had *n*'s > 200, with exception to queer men ($n = 180$), transmen ($n = 156$), transwomen ($n = 61$), and genderqueer-M ($n = 149$) participants; thus, we supplemented these groups with responses from the 2017–2018 HMS. Specifically, we randomly selected 46 queer men, 254 transmen, 96 transwomen, and 91 genderqueer-M individuals from the 2017–2018 dataset to meet the minimum baseline criteria for each group. Upon including these additional participants, the supplemented sample sizes for queer men ($n = 226$) and genderqueer-M ($n = 240$) met our baseline group criteria. Transwomen were the only group below the threshold ($n = 157$); however, they were retained and analyzed because the sample still exceeded 100. With the exception of transwomen, we then randomly selected 226 participants from the remaining sexual orientation groups and 240 from the gender groups in the 2018–2019 dataset to establish equal sample sizes.

GAD-7. After data cleaning, $N = 44{,}360$ had acceptable responses. Queer men ($n = 135$), transmen ($n = 126$), transwomen ($n = 47$), and genderqueer-M ($n = 104$) groups all fell below our minimum *n* of 200. Therefore, we supplemented these groups with 87 queer men, 279 transmen, 108 transwomen, and 149 genderqueer-M individuals from the 2017 to 2018 HMS. With supplementation, queer men ($n = 222$) and genderqueer-M ($n = 240$) groups met our baseline criteria; thus, to establish equal sample sizes we randomly reduced every other sexuality group to $n = 222$ and gender group to $n = 240$. The transwomen group was the only group below our threshold at $n = 155$; however, they were retained and analyzed because the sample was $n > 100$.

3.2. Descriptive statistics

Descriptive analyses indicated a slight positive skew was evidenced in the sample of heterosexual women (skew = 1.06) and the GAD–7 (modest skews evidenced among heterosexual men, skew = 1.07, and cisgender men, skew = 1.09). Across measures and groups, no univariate outliers ($z \geq 3.29$) were identified. Missing items were rare (< 0.59% of cases) and addressed using FIML Means and standard deviations by group are available on Table 3. Notably, heterosexual/cisgender participants demonstrated the lowest means on both instruments, followed by gay men/lesbian women, who in turn were followed by bisexual, questioning, and queer participants. Non-cisgender participants demonstrated higher means, with genderqueer-F participants having the highest scores on both instruments.

Table 3
Means and standard deviations.

Sexual Orientation	Het Men	Het Women	Gay Men	Les Women	Bi Men	Bi Women	Qt Men	Qt Women	Qr Men	Qr Women
PHQ-9										
Mean (*SD*)	7.13 (6.08)	7.73 (6.00)	8.17 (5.61)	9.80 (6.65)	10.16 (6.61)	11.94 (6.57)	10.05 (6.07)	11.54 (6.15)	10.00 (5.98)	11.24 (6.37)
GAD-7										
Mean (*SD*)	5.33 (5.70)	7.18 (5.39)	7.49 (5.48)	7.90 (5.49)	7.78 (5.48)	10.15 (6.10)	7.67 (5.33)	9.56 (5.08)	8.10 (5.47)	9.89 (5.71)
Gender	C. Men	C. Women	T. Men	T. Women	GQ-M	GQ-F				
PHQ-9										
Mean (*SD*)	6.99 (5.91)	7.89 (5.83)	13.13 (7.51)	10.88 (7.22)	12.86 (6.93)	14.26 (6.78)				
GAD-7										
Mean (*SD*)	5.51 (5.15)	7.34 (5.78)	10.29 (5.71)	8.63 (5.93)	10.08 (5.95)	10.85 (5.53)				

Note: For sexual orientation: Het = Heterosexual, Les = Lesbian, Bi = Bisexual, Qt = Questioning, Qr = Queer. For gender, *C* = Cisgender, *T* = Transgender, GQ-*M* = Genderqueer Male assigned at birth, GQ-*F* = Genderqueer Female assigned at birth.

3.3. Measurement invariance testing

Sexual Orientation PHQ-9. Table 1 displays the fit indices from the initial CFA's across groups. Only heterosexual men and queer women evidenced good fit across all indices. Because heterosexual women, lesbian women, bisexual men, bisexual women, and questioning men all evidenced acceptable CFI's and SRMR's, we included them in the initial configural invariance model. Because gay men, questioning women, and queer men demonstrated unacceptable CFI's, we excluded them from further analyses.[1]

For readability, we present the fit indices of all invariances tests in Table 4 and results in Table 5. The initial configural model demonstrated acceptable fit, indicating the same basic structure across groups. The metric model also demonstrated acceptable fit. The scaled-chi square difference test indicated the metric model did not significantly degrade fit from the configural, suggesting the PHQ-9 measures the same underlying construct across most sexual orientations. We then tested scalar invariance by constraining the intercepts to be equal across groups. However, the scalar model demonstrated poor fit and the scaled chi-square difference test indicated a significant degradation from the metric model. Additionally, the change in CFI exceeded 0.01. Thus, we calculated the 99% bias-corrected bootstrapped CIs (1000 samples) of the item-intercept difference between our groups (see Table 4). Results indicated that the intercepts between heterosexual men and heterosexual women were fully scalar invariant. However, no sexual minority groups demonstrated any degree of scalar invariance when compared to heterosexual men or heterosexual women. Notably, some of the intercepts were invariant or partially invariant between sexual minority groups. Because the scalar model evidenced poor fit, we did not test a full residual invariance model. However, because the PHQ-9 was validated on a primarily heterosexual sample, we conducted an additional post-hoc bootstrap analyses on the difference between residuals for heterosexual men and heterosexual women; results failed to indicate any significant differences. Thus, the PHQ-9 evidenced full measurement invariance only between heterosexual men and heterosexual women.

Gender PHQ-9. Single group CFA's indicated the PHQ-9 only demonstrated acceptable fit across all indices for cisgender men and transgender women (see Table 1). Cisgender women, transgender men, and genderqueer-F participants demonstrated some acceptable fit indices (CFI's above 0.90, SRMR's below 0.08) and were included in the measurement invariance analyses. Genderqueer-M participants were removed from further analyses for poor fit (CFI below 0.90). Since cisgender men demonstrated the best fit, they were entered as the referent.

The initial configural model demonstrated acceptable fit (Factor loadings across groups for both instruments are available in supplementary file 2), indicating the same basic structure across groups (see Table 4). We then tested metric invariance by constraining the factor loadings to be equal. The metric invariance model demonstrated marginally acceptable fit. The scaled-chi-square test indicated the metric model significantly degraded from the configural model. Moreover, the change in CFI exceeded 0.01. We employed the bootstrap procedure to identify which factor loadings were non-invariant. Results indicated full metric invariance between cisgender men and cisgender women. All other groups demonstrated partial metric invariance with cisgender men and cisgender women. The only item to significantly differ was item 9 ("*Thoughts that you would be better off dead, or of hurting yourself*"); transmen, transwomen, and genderqueer-F participants all demonstrated significantly higher factor loadings compared to cisgender men and women; there were no significant within-group differences across non-cisgender groups.[2]

We then tested scalar invariance by constraining all intercepts except item 9 to be equal between groups (item 9 intercepts were constrained between cisgender men and women). The scalar invariance model demonstrated poor fit. The scaled-chi-square test indicated significant degradation from the metric model. The change in CFI also exceeded 0.01. Results from the bootstrap procedure indicated cisgender men and women were fully scalar invariant, with varying degrees of invariance/partial invariance between non-cisgender groups (see Table 4). Because the scalar model evidenced poor fit, we did not test residual variance between all groups. However, we employed the bootstrap method to examine differences in the residuals between cisgender men and women. Results indicated full residual invariance on the PHQ-9 between cisgender men and women.

Sexual Orientation GAD-7. The GAD-7 individual CFA's across groups generally evidenced acceptable fit (see Table 2). Because heterosexual men evidenced better fit than heterosexual women, they were selected as the referent; though notably the gay men, bisexual men, questioning men and women, and queer men and women groups evidenced better fit than one or both of the heterosexual groups. The initial configural model demonstrated acceptable fit, indicating the same basic structure across identity groups (see Table 4). The metric model also demonstrated acceptable fit. The scaled-chi square difference test indicated it did not significantly degrade from the configural. We then tested scalar invariance; however, the scalar model demonstrated poor fit. Furthermore, the scaled chi-square difference test indicated a significant degradation from the metric model and the change in CFI exceeded 0.01. We then calculated the 99% bias-corrected bootstrapped CIs of between-group intercept differences. The intercepts between heterosexual men and all other groups, including heterosexual women, were non-invariant. However, gay men, lesbian women, questioning men, and

[1] A configural model with these groups included evidenced unacceptable fit (contact first author for specific coefficients).

[2] We tested a post-hoc analysis with the genderqueer-M participants. The same pattern was evident with the factor loading for item 9 being significantly higher than cisgender men and women, but not significantly different from transgender men, transgender women, and genderqueer-F participants.

Table 4
Results and fit indices from the multi-group invariance analyses.

Measure	Model	X^2	*df*	CFI	TLI	RMSEA	90% CI RMSEA	SRMR	ΔX^2	Δdf	ΔCFI
PHQ-9 (S)	Configural	559.98***	189	.927	.903	.093	[.084, 0.102]	.048	—	—	—
	Metric	631.35***	237	.923	.918	.086	[.078, 0.094]	.067	63.01	48	.004
	Scalar	834.72***	291	.894	.908	.091	[.084, 0.098]	.108	212.19***	54	.029
PHQ-9 (G)	Configural	414.99***	135	.928	.904	.096	[.086, 0.107]	.046	—	—	—
	Metric	525.78***	167	.908	.901	.098	[.089, 0.108]	.088	112.11***	32	.020
	Partial Scalar	674.90***	197	.877	.888	.104	[.096, 0.113]	.176	158.95***	30	.031
GAD-7 (S)	Configural	460.79***	140	.954	.931	.102	[.091, 0.112]	.045	—	—	—
	Metric	544.30***	194	.950	.945	.090	[.081, 0.099]	.059	61.41	54	.004
	Scalar	768.51***	257	.926	.940	.095	[.087, 0.102]	.113	231.32***	63	.024
	Scalar ^{P}	332.32***	122	.938	.947	.088	[.077, 0.100]	.066	—	—	—
	Residual ^{P}	383.76***	150	.931	.952	.084	[073, 0.094]	.065	52.61**	28	.007
GAD-7 (G)	Configural	274.60***	84	.955	.933	.100	[.087, 0.114]	.039	—	—	—
	Metric	341.95***	114	.946	.941	.094	[.083, 0.106]	.063	57.50**	30	.009
	Scalar	550.07***	147	.905	.918	.110	[.100, 0.120]	.140	219.67***	33	.036

Note: X^2 = scaled chi-square, df = degrees of freedom, CFI = Comparative Fit Index, TLI = Tucker Lewis Index,.
RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual. The indications of (S) and (G) represent analyses across Sexual Orientation and Gender groups respectively. ^{P} = Separate model only including heterosexual women, gay men, lesbian women, questioning men, and queer men with heterosexual women entered as the referent.
$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 5
Invariance Levels Across Sexual Orientation and Gender Identity groups on the PHQ-9.

Group	1	2	3	4	5	6	7	8	9	10
1. Heterosexual Men	—-									
2. Heterosexual Women	RI	—-								
3. Gay Men	—	—	—-							
4. Lesbian Women	MI	MI	—-	—-						
5. Bisexual Men	MI	MI	—-	SI-P	—-					
6. Bisexual Women	MI	MI	—-	—-	—-	—-				
7. Questioning Men	MI	MI	—-	SI-P	SI	SI	—-			
8. Questioning Women	—-	—-	—-	—-	—-	—-	—-	—-		
9. Queer Men	—-	—-	—-	—-	—-	—-	—-	—-	—-	
10. Queer Women	MI	MI	—-	SI-P	SI-P	—-	—-	SI-P	—-	—-
Group	1	2	3	4	5	6				
1. Cisgender Men	—-									
2. Cisgender Women	RI	—-								
3. Transmen	MI-P	MI-P	—-							
4. Transwomen	MI-P	MI-P	SI	—-						
5. Genderqueer-M	—-	—-	—-	—-	—-					
6. Genderqueer-F	MI-P	MI-P	SI	MI	—-	—-				

Note: MI = Metric Invariant, SI = Scalar Invariant, RI = Residual Invariant, *P* = Partial. Measurement invariance analyses were not undertaken for Gay Men, Questioning Women, Queer Men, and Genderqueer-M participants due to poor CFA fit.

Table 6
Invariance Levels Across Sexual Orientation and Gender Identity groups on the GAD-7.

Group	1	2	3	4	5	6	7	8	9	10
1. Heterosexual Men	—-									
2. Heterosexual Women	MI	—-								
3. Gay Men	MI	SI	—							
4. Lesbian Women	MI	SI	SI	—-						
5. Bisexual Men	MI	SI-P	SI	SI	—-					
6. Bisexual Women	MI	MI	MI	MI	MI	—-				
7. Questioning Men	MI	SI	SI	SI	SI	MI	—-			
8. Questioning Women	MI	MI	MI	SI-P	MI	SI	MI	—-		
9. Queer Men	MI	SI	SI	SI	SI	MI	SI	SI-P	—-	
10. Queer Women	MI	MI	MI	MI	MI	SI	MI	SI	SI-P	—-
Group	1	2	3	4	5	6				
1. Cisgender Women	—-									
2. Cisgender Men	MI	—-								
3. Transmen	MI	MI	—-							
4. Transwomen	SI-P	MI	SI	—-						
5. Genderqueer-M	**MI-P**	**MI-P**	**SI-P**	SI	—-					
6. Genderqueer-F	MI	MI	SI	SI-P	SI	—-				

Note: MI = Metric Invariant, SI = Scalar Invariant, RI = Residual Invariant, *P* = Partial.

queer men evidenced full scalar invariance with heterosexual women (see Table 6). Thus, we constructed an alternate scalar model (scalar ^{P}) with women entered as the referent and proceeded to test residual invariance with other invariant groups. Both the scalar ^{P} and residual ^{P} models evidenced good fit. The scaled chi-square indicated a marginal degradation from the scalar ^{P} model, further the change in CFI was modest. However, the bootstrap procedure suggested significant residual differences across groups. Thus, we determined the GAD-7 to be scalar invariant across heterosexual women, gay men, lesbian women, questioning men, and queer men.

Gender GAD-7. Single group CFA's indicated acceptable fit across groups (see Table 2). Unlike our previous analyses, cisgender women evidenced better fit than cisgender men, who displayed the worst fit. Thus, cisgender women were entered as the referent. The configural model demonstrated acceptable fit, indicating the same basic structure across gender groups (see Table 4). The metric invariance model also demonstrated acceptable fit. The scaled-chi-square test indicated the metric model significantly degraded from the configural model. However, the change in CFI was modest. We employed the bootstrap procedure to identify which factor loadings were non-invariant. Results indicated full metric invariance across groups, except for item 7 for genderqueer-M individuals being non-invariant with cisgender men and women, and transmen (item 6 was also non-invariant between genderqueer-M individuals and transmen). Thus, we concluded the GAD-7 to be measuring the same underlying anxiety construct across groups, with exception to items 6 and 7 for genderqueer-M participants.

We proceeded to test scalar invariance by constraining the intercepts to be equal (parameters were not constrained for item 7 for genderqueer-M participants). The scalar invariance model demonstrated poor fit. The scaled-chi-square test and ΔCFI indicated significant degradation from the metric model. We employed the bootstrap procedure to identify which intercepts were non-invariant. Results indicated cisgender women were partially scalar invariant with transwomen. No groups were scalar invariant with cisgender men. Various degrees of scalar invariance/partial invariance were observed across non-cisgender groups (see Table 6). Because the scalar model evidenced poor fit, and no groups were fully scalar invariant with the referent, we

did not test a residual invariance model.

4. Discussion

Our results suggest the measurement accuracy of the PHQ-9 and GAD-7 in capturing depression and anxiety across LGBTQ+ groups is complex. Results were consistent with our broad hypothesis that some degree of measurement non-invariance would be present for both instruments across groups. Only cisgender/heterosexual men and women evidenced full residual invariance on the PHQ-9. Notably, we had to exclude gay men, questioning women, queer men, and genderqueer-M individuals from measurement invariance analyses on the PHQ-9 due to fit problems with the PHQ-9 among these groups. For the GAD-7, no LGBTQ+ group evidenced scalar invariance with heterosexuals, though almost every group (with exception to genderqueer-M participants) evidenced metric invariance. Importantly, both instruments demonstrated questionable TLI's and RMSEA's in several groups despite having acceptable CFI's and SRMR's.

4.1. Research implications

When constructing studies that involve comparisons on the PHQ-9 and GAD-7 between cisgender/heterosexual and LGBTQ+ individuals, we recommend that researchers weight items before conducting mean comparisons between any groups that failed to achieve scalar invariance. This is necessary as those from LGBTQ+ backgrounds responded differently to the items than cisgender/heterosexual individuals (c.f., Cheung and Rensvold, 2000). Indeed, unweighted totals may be distorted as a result of the differential response bias across these instruments. Although the depression and anxiety constructs may hold the same meaning, there may be unique aspects to the experiences of gender and sexual minority respondents that impact the way they respond to certain items.

Researchers should also be mindful when conducting correlational studies between groups that failed metric invariance. One of the most striking findings was that metric invariance failed between cisgender and non-cisgender groups on the functioning of item nine of the PHQ-9. The factor loading for this item was significantly higher for every gender minority group compared to cisgender men and women. This means the “self-harm” construct attributed to item nine is different for gender minority individuals. Furthermore, this item was more strongly related to the underlying depression construct for gender minority individuals compared to cisgender participants (per the significantly stronger factor loadings). This is consistent with the large body of evidence indicating non-cisgender individuals disproportionality suffer from suicide ideation/self-harm (Kuper et al., 2018; Tucker, 2019), and signals that while suicide ideation is likely associated with depression for most people, it is *more strongly* associated with depression, as measured by the PHQ-9, for non-cisgender individuals. The differential item functioning of item nine could also be the result of stigma/response bias. Many cisgender individuals may feel reluctant to score item nine or score it highly, whereas non-cisgender individuals may be more open to scoring it due to openness about suicidality within the non-cisgender community (Green et al., 2015; McDermott et al., 2018). Future research using item response theory to examine the PHQ-9 in sexual/gender diverse samples may be useful in understanding the non-invariance.

It is concerning that the PHQ-9 demonstrated poor/questionable factor structure in several LGBTQ+ groups. This was the first study to validate the factor structure in many of these groups. While this is not necessarily conclusive evidence, it is consistent with the notion that depressive symptoms can function differently as the result of identity-related factors. Consistent with the minority stress model (Meyer, 2003), LGBTQ+ unique stressors could lead to identity-unique mental health presentations. Future researchers should replicate the current findings and extend them by conducting sexual/gender identity diverse measurement invariance analyses on other popular measures of depression.

4.2. Limitations

While we took a heavily nuanced and rigorous approach to our analyses, some limitations are still present. Notably, our college student sample. The use of a large nationally representative college student sample had several advantages, as it allowed us to gather sufficient samples across our groups. Furthermore, college is often a time of identity exploration. Many participants may have felt comfortable disclosing their sexual/gender minority statuses on a college survey compared to assessments in other contexts. However, our results may not generalize to other populations. Moreover, our samples were primarily white. Additional research should examine these instruments among community samples, while also exploring intersectional aspects to the PHQ-9 and GAD-7 (e.g., studies that examine how race and sexual orientation moderate measurement function).

4.3. Clinical implications

The scalar invariance results may be most relevant to clinicians, given the use of PHQ-9 and GAD-7 cutoff scores in applied settings with assumptions of a unidimensional structure. Although scalar invariance was not supported with either measure among sexual and gender minorities, use of the PHQ-9 and GAD-7 is still permissible in clinical settings with this population given that LGBTQ+ given support for full or partial metric invariance in most instances. These differences suggest that LGBTQ+ individuals may be more likely to receive a false positive score on these screening measures, which is preferred to false negatives. In other words, LGBTQ+ individuals may experience overestimation of depression and anxiety, and overestimating the prevalence of depression, anxiety, and suicidal thoughts can be addressed and clarified through following up with clients, whereas underestimation could lead these concerns going undetected. That said, possible inflation of such scores also highlights the importance of not replying on scores alone. Moreover, LGBTQ+ individuals tend to experience worse depression, anxiety, and suicidality, largely due to discrimination and stigma (Ross et al., 2018; Sutter and Perrin, 2016). Thus, these results may also reflect genuine group differences. Proper follow up will help clinicians ensure depression, anxiety, and suicidality are properly addressed and addressed.

Despite general support for the use of the PHQ-9 and GAD-7, there are notable exceptions in using the PHQ-9. Namely, the PHQ-9 did not demonstrate a sound factor structure among gay men, questioning women, queer men, and genderqueer-M individuals, and thus invariance on the PHQ-9 was not examined for these groups. Clinicians might consider alternate measures screening tools for depression. Additional testing is needed to identity whether certain items are permissible in clinical settings, as well as how commonly used cut-offs function for these groups. Until then, clinicians might exercise caution in their use of the PHQ-9 and GAD-7 as a screening tool with such individuals.

Author contributions

Nicholas C. Borgogna is a Ph.D. candidate in the Combined-Integrated Clinical and Counseling Psychology program at the University of South Alabama. Ryon C. McDermott is an associate professor in the Counseling and Instructional Sciences Department at the University of South Alabama. Rachel E. Brenner is now an Assistant Professor in the Psychology Department at Colorado State University. Nicholas C. Borgogna is the corresponding author: nicholascborgogna@gmail.com. The authors would also like to acknowledge Damion Whittington for his help configuring the tables for this project. Data is open source and available upon request from the Healthy Minds

Network. No funding was used to support any of the authors in conjunction with this project.

CRediT authorship contribution statement

Nicholas C. Borgogna: Conceived and designed the study; Analyzed the data; Wrote the paper. **Rachel E. Brenner:** Wrote the paper. **Ryon C. McDermott:** Conceived and designed the study; Wrote the paper.

Declarations of Competing Interest

All authors declare that they have no financial, personal, or professional competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2020.09.069.

References

Alarie, M., Gaudet, S., 2013. "I don't know if she is bisexual or if she just wants to get attention": analyzing the various mechanisms through which emerging adults invisibilize bisexuality. J. Bisex. 13, 191–214. https://doi.org/10.1080/15299716.2013.780004.

Bauerband, L.A., Teti, M., Velicer, W.F., 2019. Measuring minority stress: invariance of a discrimination and vigilance scale across transgender and cisgender LGBQ individuals. Psychol. Sex 10, 17–30. https://doi.org/10.1080/19419899.2018.1520143.

Borgogna, N.C., McDermott, R.C., 2019. Perceived discrimination is disproportionally associated with sexual minority mental health: implications for non-monosexual sexual minorities. J. Gay Lesbian Ment. Health 1–20. https://doi.org/10.1080/19359705.2019.1644571.

Borgogna, N.C., McDermott, R.C., Aita, S.L., Kridel, M.M., 2018. Anxiety and depression across gender and sexual minorities: implications for transgender, gender nonconforming, pansexual, demisexual, asexual, queer, and questioning individuals. Psychol. Sex. Orientat. Gend. Divers. 6, 54–63. https://doi.org/10.1037/sgd0000306.

Burke, P.J., 1991. Identity processes and social stress. Am. Sociol. Rev. 56, 836–849.

Chen, F.F., 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct. Equ. Model. 14, 464–504.

Cheung, G.W., Lau, R.S., 2012. A direct comparison approach for testing measurement invariance. Organ. Res. Methods 15, 167–198. https://doi.org/10.1177/1094428111421987.

Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct. Equ. Model. 9, 233–255. https://doi.org/10.1207/S15328007SEM0902_5.

Cheung, G.W., Rensvold, R.B., 2000. Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. J. Cross. Cult. Psychol. 31, 187–212. https://doi.org/10.1177/0022022100031002003.

Clark, R., Benkert, R.A., Flack, J.M., 2006. Large arterial elasticity varies as a function of gender and racism-related vigilance in black youth. J. Adolesc. Heal. 39, 562–569. https://doi.org/10.1016/j.jadohealth.2006.02.012.

Ghabrial, M.A., Ross, L.E., 2018. Representation and erasure of bisexual people of color: a content analysis of quantitative bisexual mental health research. Psychol. Sex Orientat. Gend. Divers. 5, 132–142. https://doi.org/10.1037/sgd0000286.

Green, M., Bobrowicz, A., Ang, C.S., 2015. The lesbian, gay, bisexual and transgender community online: discussions of bullying and self-disclosure in YouTube videos. Behav. Inf. Technol. 34, 704–712. https://doi.org/10.1080/0144929X.2015.1012649.

Hayes, S.C., Wilson, K.G., Gifford, E.V., Follette, V.M., Strosahl, K., 1996. Experiential avoidance and behavioral disorders: a functional dimensional approach to diagnosis and treatment. J. Consult. Clin. Psychol. 64, 1152–1168. https://doi.org/10.1037/0022-006X.64.6.1152.

Holman, E.G., Oswald, R.F., 2016. A decade of changes: within-group analysis of LGBTQ individuals' perceptions of their community context and the relevance for social service providers. J. Gay Lesbian Soc. Serv. 28, 214–230. https://doi.org/10.1080/10538720.2016.1191406.

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. Am. Psychol. 74, 171–193. https://doi.org/10.1037/amp0000307.

Jablensky, A., 2016. Psychiatric classifications: validity and utility. World Psychiatry 15, 26–31. https://doi.org/10.1002/wps.20284.

Kline, R.B., 2016. Principles and Practice of Structural Equation Modeling, 4th ed. Guilford Press, New York, NY.

Kuper, L.E., Adams, N., Mustanski, B.S., 2018. Exploring cross-sectional predictors of suicide ideation, attempt, and risk in a large online sample of transgender and gender nonconforming youth and young adults. LGBT Heal 5, 391–400. https://doi.org/10.1089/lgbt.2017.0259.

Kyriazos, T.A., 2018. Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. Psychology 9, 2207–2230. https://doi.org/10.4236/psych.2018.98126.

La Roi, C., Meyer, I.H., Frost, D.M., 2019. Differences in sexual identity dimensions between bisexual and other sexual minority individuals: implications for minority stress and mental health. Am. J. Orthopsychiatry 89, 40–51. https://doi.org/10.1037/ort0000369.

Lambe, J., Cerezo, A., O'Shaughnessy, T., 2017. Minority stress, community involvement, and mental health among bisexual women. Psychol. Sex. Orientat. Gend. Divers. 10.1037/sgd0000222.

Lefevor, G.T., Boyd-Rogers, C.C., Sprague, B.M., Janis, R.A., 2019. Health disparities between genderqueer, transgender, and cisgender individuals: an extension of minority stress theory. J. Couns. Psychol. 66, 385–395. https://doi.org/10.1037/cou0000339.

Martin, G., Broadhurst, C., Hoffshire, M., Takewell, W., 2018. "Students at the Margins": student affairs administrators creating inclusive campuses for LGBTQ students in the South. J. Stud. Aff. Res. Pract. 55, 1–13. https://doi.org/10.1080/19496591.2017.1345756.

Matsuno, E., Budge, S.L., 2017. Non-binary/genderqueer identities: a critical review of the literature. Curr. Sex. Heal. Rep. 9, 116–120. https://doi.org/10.1007/s11930-017-0111-8.

McDermott, E., Hughes, E., Rawlings, V., 2018. Norms and normalisation: understanding lesbian, gay, bisexual, transgender and queer youth, suicidality and help-seeking. Cult. Health Sex. 20, 156–172. https://doi.org/10.1080/13691058.2017.1335435.

McLemore, K.A., 2018. A minority stress perspective on transgender individuals' experiences with misgendering. Stigma Heal. 3, 53–64. https://doi.org/10.1037/sah0000070.

Meade, A.W., Johnson, E.C., Braddy, P.W., 2008. Power and sensitivity of alternative fit indices in tests of measurement invariance. J. Appl. Psychol. 93, 568–592. https://doi.org/10.1037/0021-9010.93.3.568.

Meyer, I.H., 2015. Resilience in the study of minority stress and health of sexual and gender minorities. Psychol. Sex. Orientat. Gend. Divers. 2, 209–213. https://doi.org/10.1037/sgd0000132.

Meyer, I.H., 2003. Prejudice, social stress, and mental mealth in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol. Bull. 129, 674–697. https://doi.org/10.1037/0033-2909.129.5.674.

Meyer, I.H., Frost, D.M., 2013. Minority stress and the health of sexual minorities. In: Patterson, C.J., D'Augelli, A.R. (Eds.), Handbook of Psychology and Sexual Orientation. Oxford University Press, New York, NY, pp. 252–266.

Mohr, J.J., Kendra, M.S., 2011. Revision and extension of a multidimensional measure of sexual minority identity: the lesbian, gay, and bisexual identity scale. J. Couns. Psychol. 58, 234–245. https://doi.org/10.1037/a0022858.

Muthén, B.O., Muthén, L.K., 2016. Mplus User's Guide, 7th ed. Muthén & Muthén, Los Angeles, CA.

Niepel, C., Greiff, S., Mohr, J.J., Fischer, J.A., Kranz, D., 2019. The English and German versions of the Lesbian, Gay, and bisexual identity scale: establishing measurement invariance across nationality and gender groups. Psychol. Sex. Orientat. Gend. Divers. 6, 160–174. https://doi.org/10.1037/sgd0000315.

Ong, T.Q., Bandalos, D.L., Swearer, S.M., 2019. Does the spiritual values/religion subscale of the self-description questionnaire III function differentially across heterosexual and non-heterosexual young adults? A measurement invariance study. J. Homosex. 10.1080/00918369.2019.1591785.

Parra, L.A., Benibgui, M., Helm, J.L., Hastings, P.D., 2016. Minority stress predicts depression in lesbian, gay, and bisexual emerging adults via elevated diurnal cortisol. Emerg. Adulthood 4, 365–372. https://doi.org/10.1177/2167696815626822.

Pryor, J.T., Garvey, J.C., Johnson, S., 2017. Pride and progress? 30 years of ACPA and NASPA LGBTQ presentations. J. Stud. Aff. Res. Pract. 54, 123–136. https://doi.org/10.1080/19496591.2016.1206020.

Richards, C., Bouman, W.P., Seal, L., Barker, M.J., Nieder, T.O., Tsjoen, G., 2016. Non-binary or genderqueer genders. Int. Rev. Psychiatry 28, 95–102. https://doi.org/10.3109/09540261.2015.1106446.

Ross, L.E., Salway, T., Tarasoff, L.A., MacKay, J.M., Hawkins, B.W., Fehr, C.P., 2018. Prevalence of depression and anxiety among bisexual people compared to gay, lesbian, and heterosexual individuals: a systematic review and meta-analysis. J. Sex Res. 55, 435–456. https://doi.org/10.1080/00224499.2017.1387755.

Scheer, J.R., Woulfe, J.M., Goodman, L.A., 2019. Psychometric validation of the identity abuse scale among LGBTQ individuals. J. Commun. Psychol. 47, 371–384. https://doi.org/10.1002/jcop.22126.

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Lowe, B., 2006. A brief measure for assessing generalized anxiety disorder. Arch. Intern. Med. 166, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092.

Sutter, M., Perrin, P.B., 2016. Discrimination, mental health, and suicidal ideation among LGBTQ people of color. J. Couns. Psychol. 63, 98–105. https://doi.org/10.1037/cou0000126.

Teymoori, A., Real, R., Gorbunova, A., Haghish, E.F., Andelic, N., Wilson, L., Asendorf, T., Menon, D., von Steinbüchel, N., 2019. Measurement invariance of assessments of depression (PHQ-9) and anxiety (GAD-7) across sex, strata and linguistic backgrounds in a European-wide sample of patients after Traumatic Brain Injury. J. Affect. Disord. https://doi.org/10.1016/j.jad.2019.10.035.

Tucker, R.P., 2019. Suicide in transgender veterans: prevalence, prevention, and implications of current policy. Perspect. Psychol. Sci. 14, 452–468. https://doi.org/10.1177/1745691618812680.

Vandenberg, R.J., Lance, C.E., 2000. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ. Res. Methods 3, 4–70. https://doi.org/10.1177/109442810031002.

Velez, B.L., Watson, L.B., Cox Jr., R., Flores, M.J., 2017. Minority stress and racial or ethnic minority status: a test of the greater risk perspective. Psychol. Sex. Orientat. Gend. Divers. 4, 257–271. https://doi.org/10.1037/sgd0000226.
Wang, J., Wang, X., 2012. Structural Equation Modeling: Applications Using Mplus. Higher Education Press, Hoboken, NJ, USA. 10.1002/9781118356258.
Westen, D., 2012. Prototype diagnosis of psychiatric syndromes. World Psychiatry 10. 1016/j.wpsyc.2012.01.004.
Williams, D.R., Yu, Y., Jackson, J.S., Anderson, N.B., 1997. Racial differences in physical and mental health. Socio-economic status, stress and discrimination. J. Health Psychol. 10.1177/135910539700200305.
Xu, H., Tracey, T.J.G., 2017. Use of multi-group confirmatory factor analysis in examining measurement invariance in counseling psychology research. Eur. J. Couns. Psychol. 6, 75–82. https://doi.org/10.5964/ejcop.v6i1.120.
Yoon, M., Lai, M.H.C., 2018. Testing factorial invariance with unbalanced samples. Struct. Equ. Model. 25, 201–213. https://doi.org/10.1080/10705511.2017.1387859.